Statistics Workshop Notes

SCIENCE & RESEARCH INTERNAL REPORT 187

Department of Conservation
Te Papa Atawhai

Statistics Workshop Notes

SCIENCE & RESEARCH INTERNAL REPORT 187

Jennifer Brown and Bryan F.J. Manly

Published by
Department of Conservation
PO Box 10-420
Wellington, New Zealand

Science & Research Internal reports are written by DOC staff or contracr scientists on matters which are on-going within the Department. They include reports on conferences, workshops and study tours, and also work in progress. Internal reports are not normally subject to peer review.
Publication was approved by the Manager, Southern Regional Office, Department of Conservation, Christchurch.

ISSN 0114-2798
ISBN 0-478-22052-9

Contents

STATISTICS WORKSHOP NOTES

This page intentionaly left blank

Contents

MODULE 1: BACKGROUND TO DATA ANALYSIS

Module 1: Background to Data Analysis

SUMMARY

This module begins by stating the background knowledge of statistics that is needed for fully understanding the material in this and the other modules in this document. Briefly, what is required is a knowledge of what discrete and continuous statistical distributions are, the concept of a sampling distribution, an understanding of the structure of a test of significance, and an understanding of what a confidence interval means. The module then covers a number of general issues related to the design and analysis of studies:

- The difference between observational studies (with passive observation only) and experimental studies (with the manipulation of conditions).
- The difference between true experiments (with randomization, replication and controls), and quasi-experiments (with one or more of these components missing), and how this affects the strength of the conclusions that can be drawn.
- The difference between design-based inference (which draws its validity from random sampling), and model-based inference (which relies on the assumed model being more or less correct).
- The current controversy about the value of tests of significance, and whether using confidence limits instead overcomes the perceived problems.
- The computer-intensive methods of randomization and bootstrapping that are receiving increasing use in all areas of science.
- What pseudoreplication is, and how it can be avoided.
- If and when adjustments for multiple testing should be made when analysing data.
- Meta-analysis methods for combining the results of several studies on the same variable.
- The difference between classical statistical methods and the Bayesian methods that are becoming popular with some data analysts.
- Data quality objective (DQO) procedures for ensuring that when studies are finished they will meet the original objectives.

1.1 The Starting Point

This module and the ones that follow assume that readers are starting with a background knowledge of statistics at the level that is usually expected to be reached or exceeded after taking a typical first year university course. Table 1.1 gives a list of what this should include. What is important is not so much to be familiar with all the details of the items that are listed, but is more the understanding of the concepts involved. For example, with tests of significance it is not necessary to be able to carry out the calculations for a range of tests without looking up the equations in a text book. However, it is important to understand the logic behind these tests, i.e. the idea of setting up a null hypothesis and testing this by comparing the observed value of a test statistic with the distribution of the statistic that will apply if the null hypothesis is correct.

There are many statistics texts available that cover the material in Table 1.1. If you are feeling a little statistically "rusty" then some revision using one of these texts may be useful.

1.2 Drawing Conclusions from Data

Statistics is all about drawing conclusions from data, and in this module we at the basis of some of the methods that are used for drawing conclusions. Quite a variety of topics are considered, including some which are rather important and yet often receive relatively little attention in statistics texts. These include the difference between observational and experimental studies, the difference between inference based on the random sampling design used to collect data and inference based on the assumption of a particular model for the data, criticisms that have been raised about the excessive use of significance tests, the use of the computer-intensive methods of randomization and bootstrapping instead of more conventional methods, the avoidance of pseudoreplication, the use of sampling methods where sample units have different probabilities of selection, the problem of multiple testing, meta-analysis (methods for combining the results from different studies), and the use of Bayesian inference, which is currently receiving a great deal of attention.

TABLE 1.1 STATISTICAL BACKGROUND THAT IS ASSUMED IN THIS AND THE FOLLOWING MODULES

CONCEPT	WHAT SHOULD BE KNOWN
Random variation in data	How observations taken under apparently similar conditions display variation, which can be described by statistical distributions such as the normal distribution for continuous data, and the binomial distribution for discrete (count) data.
Summary statistics	How the mean, standard deviation, etc. are used to summarise a sample or a theoretical distribution.
Distributions for sample statistics	The standard error of the mean, SE($\bar{x}$) = $\sigma/\sqrt{n}$. The use of the t-distribution for inferences about sample means. The uses of the chi-squared distribution with count data.
Tests of significance	The logic behind tests of significance, including the difference between one and two sided tests, the meaning of the significance level, and the role of the null and alternative hypotheses. The use of one and two sample t-tests, chi-squared goodness of fit tests.
Confidence limits	The interpretation of a confidence interval as one within which a population parameter will lie with a stated probability.
Analysis of variance	The partitioning of the total sum of squares about the mean for a set of data into components associated with different factors and their interactions, the summary of this in an analysis of variance table, and F-tests for significant effects, for factorial experiments only (i.e. one factor analysis of variance, two factor analysis of variance, etc.)
Regression	The idea of accounting for the variation in a dependent variable Y in terms of the variation in one or more X variables. The uses of the t-distribution and F-distribution to determine which of the X variables are important.

People often use statistical methods without giving much thought to why these methods lead to valid conclusions - if indeed they do! This module is intended to make you think more critically about these matters.

1.3 Observational and Experimental Studies

When considering the nature of empirical studies there is an important distinction between observational and experimental studies. **With observational studies data are collected by observing populations in a passive manner that as far as possible will not change the processes going on.** For example samples of animals might be collected in order to estimate the proportions in different age classes or the sex ratio. On the other hand, **experimental studies are usually thought of as involving the collection of data with some manipulation of variables that is assumed to affect population parameters**, keeping other variables constant as far as possible. An example of this type would be a study where possums are removed from an area to see whether this leads to improved survival of an endangered plant.

In many cases the same statistical analysis can be used with either observational or experimental data. However the validity of any inferences that result from the analysis depends very much on the type of study. In particular, an effect that is seen consistently in replications of a well designed experiment can only reasonably be explained as being caused by the manipulation of the experimental variables. But with an observational study the same consistency of results might be obtained because all the data are affected in the same way by some unknown and unmeasured variable. Therefore the 'obvious' explanation for an effect that is seen in the results of an observational study may be quite wrong. To put it another way, the conclusions from observational studies are not necessarily wrong. The problem is that there is little assurance that they are right (Hairston, 1989, p. 1).

It is clear that in general it is best to base inferences on experiments rather than observational studies, but this is not always possible. Some experiments cannot be performed either because the variables involved are not controllable, or because the experiment is not feasible. For example, suppose that a researcher wishes to assess the effect of discharges of pollutants from a sewage treatment plant on the organisms in a river. Then systematically changing the levels of pollutants, and in some cases increasing them to higher levels than would normally occur, might either not be possible or be considered to be unethical. Hence in this situation the only study possible might be one involving attempts to relate measurements on the organisms to unplanned variation in pollutant levels, with some allowance for the effects of other factors that may be important.

Having defined two categories of study (observational and experimental), it must be admitted that at times the distinction becomes a little blurred. In particular, suppose that the variables that are thought to determine the state of an ecological system are abruptly changed either by some naturally occurring accident, or are an unintentional result of some human intervention. If the outcome is then studied this appears to be virtually the same as if the changes were made by the observer as part of an experiment. But such 'natural experiments' do not have some of the important characteristics of true experiments. The conclusions that can be drawn might be stronger than those that could be drawn if the system was not subjected to large changes, but they are not as strong as those that could be drawn if the same changes were made as part of a well designed experiment.

Although the broad distinction that has been made between observational and experimental studies is useful, a little thought will show that both of these categories can be further subdivided in meaningful ways. For example, Eberhardt and Thomas (1991) propose a classification of studies into eight different types. However, this elaboration is unnecessary here, where it merely needs to be noted that most of the studies carried out by DOC staff are observational, with all the potential limitations that this implies.

1.4 True Experiments and Quasi-Experiments

At this stage it becomes necessary to better define what is required for a study to be a 'true' experiment. Basically, the three important ingredients are **randomization, replication, and controls.**

Randomization should be used whenever there is an arbitrary choice to be made of which units will be measured out of a larger collection of possible units, or of the units to which different levels of a factor will be assigned. This does not mean that all selections of units and all allocations of factor levels have to be made completely at random. In fact, a large part of the theory of experimental design is concerned with how to restrict randomization and allocation in order to obtain the maximum amount of information from a fixed number of observations. Thus randomization is only required subject to whatever constraints are involved in the experimental design.

Randomization is used in the hope that it will remove any systematic effects of uncontrolled factors of which the experimenter has no knowledge. The effects of these factors will still be in the observations. However, randomization makes these effects part of the experimental errors that are allowed for by statistical theory. Perhaps more to the point, if randomization is not carried out then there is always the possibility of some unseen bias in what seems to be a haphazardous selection or allocation.

Randomization in experiments is not universally recommended. Its critics point to the possibility of obtaining random choices that appear to be unsatisfactory. For example, if different varieties of a crop are to be planted in different plots in a field then a random allocation can result in all of one variety being placed on one side of the field. Any fertility trends in the soil may then appear as a difference between the two varieties, and the randomization has failed to remove potential biases due to positions in the field. Although this is true, common sense suggests that if an experimenter has designed an experiment that takes into account all the obvious sources of variation, such as fertility trends in a field, so that the only choices left are between units that appear to be essentially the same, such as two plots in the same part of a field, then randomization is always worthwhile as one extra safeguard against the effects of unknown factors.

Replication is needed in order to decide how large effects have to be before they become difficult to account for in terms of normal variation. This requires the measurement of normal variation, which can be done by repeating experimental arrangements independently a number of times under conditions that are as similar as possible. Experiments without replication are case studies that may be quite informative and convincing, but it becomes a matter of judgement as to whether the outcome could have occurred in the absence of any manipulation.

Controls provide observations under normal conditions without the manipulation of factor levels. They are included in an experiment to give the standard with which the results under other conditions are compared. In the absence of controls it is usually necessary to assume that if there had been controls then these would have given a particular outcome. For instance, suppose that the yield of a new type of wheat is determined without running control trials with the standard variety. Then in order to decide whether the yield of the new variety is higher than that for the standard it is necessary to make some assumption about what the yield of the standard variety would have been under the experimental conditions. The danger here is obvious: it may be that under the conditions of the study the yield of the standard variety would not be what is expected, so that the new variety is being compared with the wrong standard.

Experiments that lack one or more of the ingredients of randomization, replication and control are sometimes called quasi-experiments. Social scientists realized many years ago that they can often only do quasi-experiments rather that true experiments, and have considered very thoroughly the implications of this in terms of drawing conclusions (Campbell and Stanley, 1963; Manly, 1992, Chapter 5 and 6). It is important point to realize is that the experiments carried out by DOC staff are usually quasi-experiments, so that the social scientists problems are also problems for DOC.

There is no need here to discuss these problems at length. In fact, many of them are fairly obvious with a little thought. Some of the simpler designs that are sometimes used without a proper recognition of potential problems are listed below. Here O_i indicates observations made on a group of experimental units, X denotes an experimental manipulation, and R indicates a random allocation of experimental units to treatment groups. For example R O_1 X O_2 indicates that there is a random allocation to groups, observations are taken on the experimental group, an experimental manipulation is made for this group, and then observations are taken again. The description 'pre-experimental design' is used for the weakest situations, where inferences are only possible by making strong assumptions. 'Quasi-experimental designs' are better, but still not very satisfactory, while 'proper designs' have all the desirable characteristics of true experiments.

Two Pre-Experimental Designs

The one group pretest-posttest design

O_1 X O_2

The two group comparison without randomization

X O_1
O_2

A Quasi-Experimental Design

The comparative change design without randomization

O_1 X O_2
O_3 O_4

Two Proper Designs

The two group comparison with randomization

R X O_1
R O_2

The comparative change design with randomization

R O_1 X O_2
R O_3 O_4

Of these designs, the comparative change ones are of particular interest because they are the same as the before-after-control-impact (BACI) design that is commonly used by DOC scientists. Problems arise in these applications when, as is usually the case, it is not possible to randomly allocate experimental units to the treated and control groups before the treatment is applied. There is then the possibility that somehow the nature of the units is different for the control and treated groups in terms of how they are likely to change with time either with or without any treatment.

1.5 Design-Based and Model-Based Inference

It is often not realized that conclusions from data are reached using two very different philosophies for making scientific inferences. One is design-based, using the randomization used when collecting data, and the other is model-based, using the randomness that is inherent in the assumed model.

All of the classical methods for sampling that are discussed in Module 2 are design-based, because this is how the classical theory for sampling finite populations developed. For example, one important equation used in that module is for the variance of the mean of a random sample of size n from a population of size N with variance σ^2. This equation states that Var($\bar{y}$) = (σ^2/ n)(1 - n/N). What this means is that if the process of drawing a random sample is repeated many times, and the sample means $\bar{y}_1$, $\bar{y}_2$, $\bar{y}_3$, ... are recorded, then the variance of these means will be (σ^2/n)(1 - n/N). Thus this is the variance that is generated by the sampling process. No model is needed for the distribution of the Y values, and in fact the variance applies for any distribution at all.

By way of contrast, consider the testing of the coefficient of X for a simple linear regression equation. In that case the usual situation is that there are n values y_1, y_2, ..., y_n for Y with corresponding values x_1, x_2, ..., x_n for X. The specific model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad (1.1)$$

is then assumed, where β_0 and β_1 are constants to be estimated, and ε_i, is a random value from a normal distribution with a mean of zero and a constant unknown variance σ^2. The values of β_0 and β_1 are then estimated by least-squares as discussed in Module 4. If b_1 is the estimate of β_1, with an estimated standard error $\hat{\mathrm{SE}}(b_1)$, then a test to see whether this is significantly different from zero involves comparing $b_1 / \hat{\mathrm{SE}}(b_1)$ with critical values of the t-distribution with n - 2 degrees of freedom.

In this case, there is no requirement that the units on which X and Y are measured are a random sample from some specific population. In fact, the equation for estimating the standard error of b_1 is based on the assumption that the X values for the data are fixed constants rather than being random, with the difference between b_1 and β_1 being due only to the particular values that are obtained for the errors ε_1, ε_2, ... ε_n in the model. Thus in this case the assessment about whether b_1 is significantly different from zero is not based on any requirement for random sampling of a population. Instead, **it is based on the assumption that the model (1.1) correctly describes the structure of the data,** and that the errors ε_1, ε_2, ... ε_n for the real data are a random sample from a normal distribution with mean zero and constant variance.

What often happens with environmental studies is that design-based and model-based inference are both used at different times. For example, consider evaluation of the impact of an accidental spill of a toxic chemical on some biological variable. There may be:

1. design-based estimates of the exposure to the chemical from random sampling in the field;
2. model-based laboratory assessments of the effects of the chemical based on true experiments with randomization, replication and controls; and
3. model-based simulations of the effects of the chemical exposure using parameters estimated from the random sampling in the field.

An advantage of the design-based approach is that valid inferences are possible which are completely justified by the design of the study and the way that data are collected. The conclusions can then always be defended providing that there is agreement about which variables should be measured, the procedures used to do the measuring, and the design protocol. In this case, any reanalysis of the data by other groups will not be able to declare these original conclusions incorrect. It is possible that a reanalysis using a model-based method may lead to different conclusions, but the original analysis will still retain its validity.

On the other hand, most statistical analysis is model-based, and there can be no question about the fact that this is necessary. The use of models allows much more flexibility in analyses and all of the methods described in the previous chapter are model-based, requiring specific assumptions about the structure of data. The flexibility comes at a price. Sometimes the implicit assumptions of models are hidden below the surface of the analysis. These range from assumptions about the random components of models that may or may not be critical as far as conclusions to assumptions about the mathematical form of the equations that relate different variables which may be absolutely critical particularly if it is necessary to predict the response of some variable when predictor variables are outside of the range observed on the available data. Moreover, whenever conclusions are drawn from a model-based analysis there is always the possibility that someone else will repeat the analysis with another equally reasonable model and reach different conclusions.

A case in point is the use of the lognormal distribution as a model for data. This is frequently assumed for the distribution of the concentration of a chemical in field samples because it has an appropriate shape (Figure 1.1). However, two recent studies have cast doubt on the uncritical use of this model.

Schmoyer *et al.* (1996) simulated data from lognormal, truncated normal and gamma distributions and compared the results from estimation and testing assuming a lognormal distribution with other approaches that do not make this assumption. They found that departures from the lognormal distribution were difficult to detect with the sample sizes that they used, but when they occurred the tests based on the lognormal assumption did not work as well as the alternatives. They concluded that "in the estimation of or tests about a mean, if the assumption of lognormality is at all suspect, then lognormal-based approaches may not be as good as the alternative methods". Because the lognormal distribution will probably seldom hold exactly for real data, this is a serious criticism of the model.

Wiens (1999) example is perhaps more disturbing. The same set of data was analysed two ways. First, the observations (the amount of antibody to hepatitis A virus in serum samples) were analysed assuming a generalized linear model (as discussed in Module 4) with a possible mean difference between two groups, and lognormal errors. The difference in the two groups was approaching significance ($p = 0.10$) with the second group estimated to have a higher mean. Next, the data were analysed assuming gamma distributed errors, where the gamma distribution is one that often has the same type of shape as the lognormal. In this case the difference between the two groups was nowhere near significant ($p = 0.71$), but the first group was estimated to have a higher mean. Hence, the modelling assumptions made when analysing the data are rather crucial. Wiens notes that with this particular example a non-parametric test can be used to compare the two groups, which is better than the model-based approach. However, he also points out that with more complicated data sets a model-based approach may be preferred because of the flexibility that this permits in the analysis. He therefore proposes the ad-hoc solution of analysing data like this using both the lognormal and gamma models and investigating further if the results do not agree. This, of course, raises the possibility that both models are wrong, with similar misleading outcomes that go undetected.

The moral from all this is that although a strict adherence to design-based analyses is not possible for all DOC studies, it is a good idea to rely on design-based analyses as much as possible. The value of at least a few indisputable design-based statistical inferences may, for example, be of great value for defending a study in a court case.

1.6 Tests of Significance and Confidence Intervals

Tests of significance are very commonly used for drawing conclusions from data. However, these tests have certain limitations which have led over the years to a number of authors questioning their use, or at least the extent to which they are used.

The two basic problems can be illustrated in terms of the comparison of the mean of a variable at a site which was once contaminated and is now supposed to be cleaned up with the mean at a reference site which was never contaminated. The first problem is that the two sites cannot be expected to have exactly the same mean even if the cleaning operation has been very effective. Therefore, if large samples are taken from each site there will be a high probability of finding a significant difference between the two sample means, irrespective of the effectiveness of the cleanup. The second problem is that if the difference between the sample means is not significant then it does not mean that no difference exists. An alternative explanation is that the sample sizes used are not large enough to detect the existing differences.

These well-known problems have been discussed many times by social scientists (e.g. Oakes, 1986), medical statisticians (e.g. Gardner and Altman, 1986), environmental scientists (e.g. McBride *et al.* 1993), wildlife scientists (e.g. Cherry, 1998 and Johnson, 1999), statisticians (e.g. Nelder, 1999), and no doubt by those working in other areas as well. A common theme is that too often hypothesis tests are used when it is obvious in advance that the null hypothesis is not true, and that as a result scientific papers are becoming cluttered up with unnecessary p-values.

In truth, **there is not much point in testing hypotheses that are known to be false.** Under such circumstances it makes more sense to estimate the magnitude of the effect of interest, with some indication of the likely accuracy of results. However, there are situations where the truth of a null hypothesis really is in question, and then carrying out a significance test is an entirely reasonable thing to do. Once evidence for the existence of an effect is found, it is then reasonable to start measuring its magnitude.

Testing the effect of poisoned 1080 baits on invertebrates numbers is a case in point. It may be entirely plausible initially that a drop of poisoned baits in an area has no perceptible effect on invertebrate numbers one week after a drop. It is then sensible to ask whether an observed mean change in the invertebrate density before and after experimental drops is statistically significant.

No doubt arguments about the value of tests of significance will continue. The point of view adopted here is that it does often happen that the existence of an effect is in doubt, in which case testing the null hypothesis that the effect does not exist is sensible. However, in other cases it is more or less certain that an effect exists and the main question of interest is the size of the effect. In that case a confidence interval may provide the necessary information. Thus both tests of significance and confidence intervals are important tools for data analysis, but under different circumstances.

1.7 Randomization Tests

Randomization is a computer-intensive method that is receiving more use as time goes by for the analysis of biological data (Manly, 1997), although it has a long history, going back about 65 years to the work of Sir Ronald Fisher, one of developers of many of the statistical methods used today (Fisher, 1935, 1936). What a randomization test does is to see whether a pattern in a set of data is likely to have occurred by chance if there is actually no effect for the factor being studied.

The simplest situation for understanding what is mean by a randomization test is the two group comparison, as proposed by Fisher (1936). In this situation there is one sample of values x_1, x_2, ..., x_m, with mean $\bar{x}$, and a second sample of values y_1, y_2, ..., y_n, with mean $\bar{y}$. The question of interest is whether the two samples come from the same distribution or, more precisely, whether the absolute mean difference $|\bar{x} - \bar{y}|$ is small enough for this to be plausible. The test proceeds as follows:

1. The observed absolute mean difference is labelled d_1.
2. It is argued that if the null hypothesis is true (the two samples come from the same distribution) then any one of the observed values x_1, x_2, ..., x_m and y_1, y_2, ..., y_n could equally well have occurred in either of the samples. On this basis, a new sample 1 is chosen by randomly selecting m out of the full set of $n + m$ values, with the remaining values providing

the new sample 2. The absolute mean difference $d_2 = | \bar{x} - \bar{y} |$ is then calculated for this randomized set of data.

3. Step (2) is repeated a large number R - 1 of times to give a total of R differences $d_1, d_2, ..., d_R$.
4. The R differences are put in order from the smallest to largest.
5. If the null hypothesis is true then d_1 should look like a typical value from the set of R differences, and is equally likely to appear anywhere in the list. On the other hand, if the two original samples come from distributions with different means then d_1 will tend to be near the top of the list. On this basis, d_1 is said to be significantly large at the 100"% level if it is among the top 100α% of values in the list. If 100α% is small (say 5% or less) then this is regarded as evidence against the null hypothesis.

It is an interesting fact that this test is exact in a certain sense even when R is quite small. For example, suppose that R = 99. Then if the null hypothesis is true and there are no tied values in the differences $d_1, d_2, ..., d_{100}$, the probability of d_1 being one of the largest 5% of values (i.e. one of the largest 5) is exactly 0.05. This is precisely what is required for a test at the 5% level: the probability of a significant result when the null hypothesis is true is equal to 0.05.

The test just described is two-sided. A one-sided version is easily constructed by using the signed difference $\bar{x} - \bar{y}$ as the test statistic, and seeing whether this is significantly high (assuming that the alternative to the null hypothesis of interest is that the values in the first sample come from a distribution with a higher mean than that for the second sample).

An advantage that the randomization approach has over a conventional parametric test on the sample mean difference is that it is not necessary to assume any particular type of distribution for the data, such as normal distributions for the two samples for a *t*-test. The randomization approach also has an advantage over a non-parametric test like the Mann-Whitney U-test because it allows the original data to be used rather than just the ranks of the data. Indeed, the Mann-Whitney U-test is really just a type of randomization test for which the test statistic only depends on the ordering of the data values in the two samples being compared.

Example: The Effect of 1080 Poison Pellets on Invertebrates

A DOC study of the effect of 1080 (sodium monofluoroacetate) pellets on invertebrates was carried out at Ohakune in the North Island of New Zealand. Trials were run for either 9 or 18 days, using neighbouring grids of pellets, with one grid as a control receiving placebo pellets, and the other grid receiving 1080 poison pellets. Pellets were replaced every day, and the data obtained were counts of the invertebrates found on the pellets at a certain time each day. For the 9 day trials pellets were put out as shown in Table 1.2 (a). For the 18 day trials the same design was used but was repeated for another 9 days, as shown in Table 1.2 (b). More details about the experimental arrangements are given by Sherley and Wakelin (1998). This is a type of before-after-control-impact (BACI) design.

TABLE 1.2 EXPERIMENTAL DESIGN FOR THE 9 AND 18 DAY TRIALS TO COMPARE INVERTEBRATE NUMBERS ON PLACEBO AND 1080 PELLETS.

(A) 9 DAY TRIALS			
Grid	Days 1 - 3	Days 4 - 6	Days 7 - 9
Control	Placebo pellets	Placebo pellets	Placebo pellets
Treated	Placebo pellets	1080 pellets	Placebo pellets
(B) 18 DAY TRIALS (AS ABOVE PLUS)			
Grid	Days 10 - 12	Days 13 - 15	Days 16 - 18
Control	Placebo pellets	Placebo pellets	Placebo pellets
Treated	Placebo pellets	1080 pellets	Placebo pellets

There are many different approaches that might be used to analyse the data from this experiment including, for example, attempting to model the counts of invertebrates on individual pellets. However, a fairly straightforward randomization test for the effect of 1080 is also possible. It is important to realize in this respect that this was an initial experiment to determine whether there is any effect at all.

To carry out a randomization test it is first necessary to measure the effect of poisoning, if any. A simple way to do this is illustrated in Table 1.3, based on the results for days 1 to 9 of the first trial. Two measures are calculated. The first is E_1, which is the change in the treated - control difference between the initial 3 days of the experiment (when no poison was laid) and the next 3 days (when 1080 pellets were laid on the treated grid). The treated - control difference was initially 0.35, but changed to -0.39, to give E_1 = -0.39 - 0.35 = - 0.74. Thus the apparent effect of the 1080 pellets was to reduce the invertebrate count by an average of 0.74 per pellet. This is a measure of the immediate effect of using 1080 pellets.

The second measure is E_2, which is the change in the treated - control difference between the initial 3 days and day 9. The initial difference of 0.35 changed to -1.13, giving E_2 = -1.13 - 0.35 = -1.47. This is a measure of the residual effect of the treatment after it has been discontinued for 3 days.

TABLE 1.3 CALCULATION OF E_1, THE IMMEDIATE EFFECT OF POISON PELLETS, AND E_2, THE RESIDUAL EFFECT OF POISON PELLETS FOR THE FIRST 9 DAYS OF TRIAL 1.

	AVERAGE INVERTEBRATES PER BAIT		
	BEFORE	DURING	AFTER
Control	0.66	0.78	1.89
Treated	1.01	0.39	0.76
Difference	0.35	-0.39	-1.13
Effect		E_1 = -0.74	E_2 = -1.47

When the trial was repeated on days 10 to 18 it is possible to calculate another two statistics E_3 and E_4. Here E_3 is the same as E_1 but calculated for the second 9 days instead of the first 9 days. It measures the effect of the treatment when it is repeated. Similarly, E_4 is the same as E_3 but calculated for the last 9 days. This measures the residual effect for a second treatment.

TABLE 1.4 FULL RESULTS FOR THE POISON PELLET TRIALS IN TERMS OF THE STATISTICS E_1 (THE IMMEDIATE EFFECT), E_2 (THE RESIDUAL EFFECT), E_3 (THE IMMEDIATE EFFECT FROM A REPEATED APPLICATION), AND E_4 (THE RESIDUAL EFFECT FROM A REPEATED APPLICATION).

	Immediate Effects		**Repeated Effects**	
Trial	**E_1**	**E_2**	**E_3**	**E_4**
1	-0.74	-1.47	-1.85	-0.52
2	-0.38	-0.98	-1.79	-0.68
3	-0.19	-0.10	-0.90	-0.59
4	-0.11	-0.01	-0.66	-0.18
5	0.09	-0.59		
6	-0.51	0.68		
7	-0.12	0.46		
8	-0.15	0.05		
9	-0.20	0.01		
10	-2.13	0.30	-0.03	-0.53
11	-1.98	-0.51	0.12	-0.89
12	-0.66	0.10		
13	-0.79	0.02		
Mean	-0.60	-0.16	-0.85	-0.56

Table 1.4 shows the full experimental results for E_1 to E_4 for the 13 trials carried out. The mean values of these statistics are shown at the foot of the table, and the question to be considered is whether these mean values indicate a significant impact of the poison pellets.

To run a randomization test it is only necessary to note that switching round the control and treated data for one trial just has the effect of changing the signs to E_1 to E_4. It can therefore be argued that on the null hypothesis of no treatment effect the signs in any row of Table 1.4 were equally likely to be as they are shown, or reversed. The test therefore involves seeing whether the mean values shown in the table are significantly low in comparison with the distributions found by randomly reversing the signs of E_1 to E_4 for the individual trials with probability 0.5. A one sided-test is called for because it is hard to imagine how the use of 1080 pellets could increase invertebrate numbers.

The @RISK software (Pallisade, 1995) was used to generate 10,000 randomized sets of data. The proportion (p) of times that a mean value as low as that shown in Table 1.4 was then determined, for each of the mean values for E_1 to E_4. The mean of E_1 was very significantly low (p = 0.0001), giving very strong evidence of an immediate reduction in invertebrates from the poison pellets. The mean of E_2 was not significantly low (p = 0.19), so there is no real evidence of a residual effect from one treatment. The mean of E_3 was fairly significantly low (p = 0.047), giving some evidence of a reduction in invertebrate numbers following a second treatment. The mean of E_4 was quite significantly low (p = 0.015), giving some evidence of a residual effect of the second treatment.

This example is really a complicated application of Fisher's (1935) randomization analysis for the method of paired comparisons. There is more that could be said about it, plus some questions such as why the experimental design was chosen. It seems a bit ad-hoc. Also, of course many statisticians would prefer a model-based analysis using the raw data rather than the summary statistics E_1 to E_4. I like the analysis given here because it is easy to understand, and quite convincing.

1.8 Bootstrapping

Bootstrapping as a general tool for analysing data was first proposed by Efron (1979). Initially the main interest was in using this method to construct confidence intervals for population parameters using the minimum of assumptions, but more recently there has been interest in bootstrap tests of hypotheses (Hall and Wilson, 1991; Manly, 1997).

The basic idea behind bootstrapping is that when only sample data are available, and no assumptions can be made about the distribution that the data are from, **then the best guide to what might happen by taking more samples from the distribution is provided by resampling the sample. This is a very general idea, and the way that it might be applied is illustrated by the following example.**

Example: Finding a 95% Confidence Interval for the Mean Chlorophyll-a in Lakes

A random sample of 25 lakes is taken from the large number in a certain area, and the values for chlorophyll-a are determined. The sample mean is 50.30 and the sample standard deviation is 50.02. There is interest in calculating a 95% confidence interval for the mean of chlorophyll-a in all the lakes in the area.

If the chlorophyll-a values are approximately normally distributed then the confidence interval can be calculated using the t-distribution. The interval is then

$$\bar{x} - 2.06s/\sqrt{25} < \mu < \bar{x} + 2.06s/\sqrt{25}, \qquad (1.2)$$

where $\bar{x}$ is the sample mean, s is the sample standard deviation, and 2.06 is the value that is exceeded with probability 0.025 for the t-distribution with 24 degrees of freedom. For the data in question, $\bar{x}$ = 50.30 and s = 50.02, so the interval is

$$29.66 < \mu < 70.95.$$

Unfortunately, the values of chrorophyll-a are very far from being normally distributed, as is clear from Figure 1.1. There is therefore a question about whether this t-distribution method for determining the interval really gives the required level of 95% confidence.

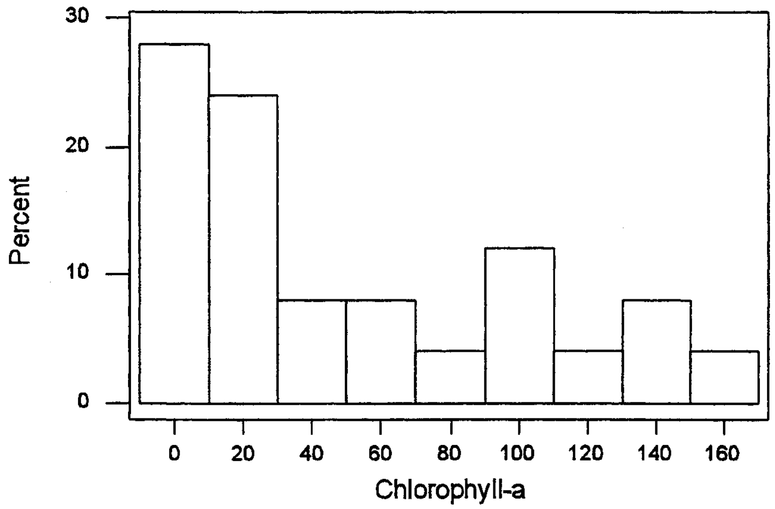


Figure 1.1 The distribution of chlorophyll-a for 25 lakes in a region, with the height of the histogram bars reflecting the percentage of the distribution in different ranges.

Bootstrapping offers a possible method for obtaining an improved confidence interval, with the method that will now be described being called bootstrap-*t*. This works by using the bootstrap to approximate the distribution of

$$t = (\bar{x} - \mu)/(s/\sqrt{25})$$

instead of assuming that this follows a *t*-distribution with 24 degrees of freedom, which it would for a sample from a normal distribution.

An algorithm to do this is as follows, where this was easily carried out in a spreadsheet program:

(a) The 25 sample observations of chlorophyll-a is set up as the bootstrap population to be sampled. This population has the known mean of μ_B = 50.30.

(b) A bootstrap sample of size 25 is selected from the population by making each value in the sample equally likely to be any of the 25 population values. This is sampling with replacement, so that a population value may occur 0, 1, 2, 3 or more times.

(c) The t-statistic $t_1 = (\bar{x} - \mu_\beta)/(s/\sqrt{25})$ is calculated from the bootstrap sample.

(d) Steps (b) and (c) are repeated 5000 times to produce 5000 t-values t_1, t_2, ..., t_{5000} to approximate the distribution of the t-statistic for samples from bootstrap population.

(e) Using the bootstrap distribution obtained, two critical values t_{low} and t_{high} are estimated such that

$$\text{Prob}\quad [(\bar{x} - \mu_\beta)/(s/\sqrt{25}) < t_{\text{low}}] = 0.025,$$

and

$$\text{Prob}\quad [(\bar{x} - \mu_\beta)/(s/\sqrt{25}) < t_{\text{high}}] = 0.025.$$

At this point it is assumed that the critical values that are obtained at step (e) also apply for random samples of size 25 from the distribution of chlorophyll-a from which the original set of data was drawn. Thus it is asserted that

$$\text{Prob} \quad [t_{\text{low}} < (\bar{x} - \mu)/(s/\sqrt{25}) < t_{\text{high}}] = 0.95,$$

where $\bar{x}$ and s are now the values calculated from the original sample, and μ is the mean chlorophyll-a value for all lakes in the region of interest. Rearranging the inequalities then leads to the statement that

$$\text{Prob} \quad [\bar{x} - t_{\text{high}} s/\sqrt{25} < \mu < \bar{x} - t_{\text{low}} s/\sqrt{25}] = 0.95,$$

so that the required 95% confidence interval is

$$\bar{x} - t_{\text{high}} s/\sqrt{25} < \mu < \bar{x} - t_{\text{low}} s/\sqrt{25}]. \qquad (1.3)$$

The intervals (1.2) and (1.3) differ to the extent that t_{low} and t_{high} vary from 2.064. When the process was carried out it was found that the bootstrap distribution of $t = (\bar{x} - \mu_\beta)/(s/\sqrt{25})$ is quite close to the t-distribution with 24 degrees of freedom, as shown by Figure 1.2, but with t_{low} = -2.6 and t_{high} = 2.0. Using the sample mean and standard deviation, the bootstrap-*t* interval therefore becomes

$$50.30 - 2.0(50.02/5) < \mu < 50.30 + 2.6(50.02/5),$$

or

$$30.24 < \mu < 76.51.$$

These compare with the limits of 29.66 to 70.95 obtained using the t-distribution. Thus the bootstrap-*t* method gives a rather higher upper limit, presumably because this takes better account of the type of distribution being sampled.

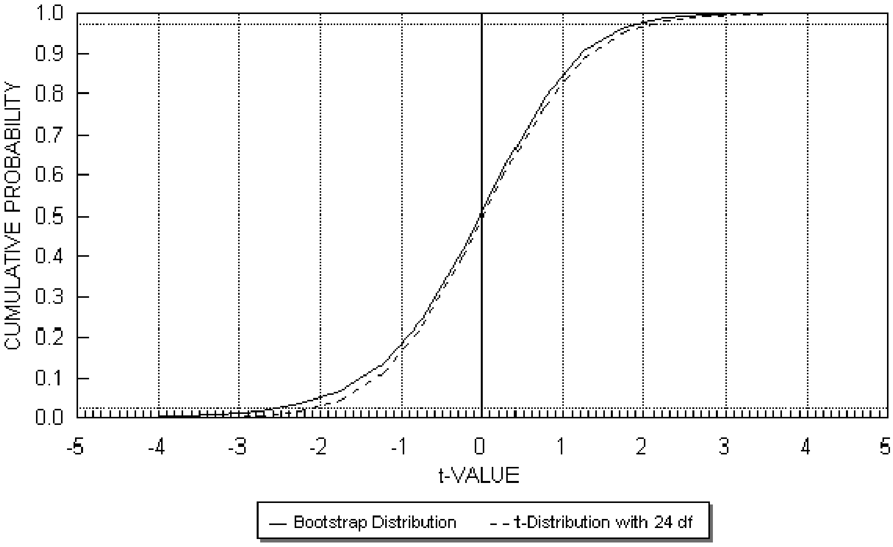


Figure 1.2 Comparison between the bootstrap distribution of ($\bar{x}$ - μ_β)/(s/√25) and the *t*-distribution with 24 degrees of freedom. According to the bootstrap distribution, the probability of a value less than t_{low} = -2.6 is approximately 0.025, and the probability of a value higher than t_{high} = 2.0 is approximately 0.025.

1.9 Pseudoreplication

The term "pseudoreplication" causes some concern, particularly among field ecologists, with the clear implication that when an investigator believes that replicated observations have been taken, this may not really the case at all. Consequently, there is some fear that the conclusions from studies will not be valid because of unrecognized pseudoreplication.

The concept of pseudoreplication was introduced by Hurlbert (1984) with the definition **"the use of inferential statistics to test for treatment effects with data from experiments where either treatments are not replicated, or replicates are not statistically independent"**. Two examples of pseudoreplication are:

- A sample of metre square quadrats randomly located within a 1 ha study region randomly located in a larger burned area is treated as a random sample from the entire burned area.
- Repeated observations on the location of a radio-tagged animal are treated as a simple random sample of the habitat points used by the animal, although in fact successive observations tend to be close together in space.

In both of these examples it is the application of inferential statistics to dependent replicates as if they were true replicates from the population of interest that causes the pseudoreplication. However, it is important to understand that using a single observation per treatment or replicates that are not independent data is not necessarily wrong. Indeed it may be unavoidable in some field studies. What is wrong is to ignore this in the analysis of the data.

There are two common aspects of pseudoreplication. One of these is the **extension of a statistical inference observational study beyond the specific population studied to other unstudied populations.** This is the problem with the first example above on sampling of burned areas. The other aspect is the analysis of dependent data as if they are independent data. This is the problem with the example on radio-tagged animals.

When dependent data are analysed as if they are independent, the sample size used is larger than the effective number of independent observations. This often results in too many significant results being obtained from tests of significance, and confidence intervals being narrower than they should be. To avoid this, a good rule to follow is that statistical inferences should be based on only one value from each independently sampled unit, unless the dependence in the data is properly handled in the analysis. For example, if five quadrats are randomly located in a study area, then statistical inferences about the area should be based on five values, regardless of the number of plants, animals, soil samples, etc., that are counted or measured in each quadrat. Similarly, if a study uses data from 5 radio-tagged animals then statistical inferences about the population of animals should be based on a sample of size 5, regardless of the number of times each animal is relocated.

When data are dependent because they are collected close together in time or space there are a very large number of analyses available to allow for this.

Some of these methods are discussed in later modules. For now it is just noted that unless it is clearly possible to identify independent observations from the study design used then a method of analysis that allows for dependent data needs to be used.

1.10 Multiple Testing

Suppose that an experimenter is planning to run a number of trials to determine whether a chemical at a very low concentration in the environment has adverse effects. A number of variables will be measured (survival times of fish, growth rate of plants, etc.) with comparisons between control and treated situations, and the experimenter will end up doing 20 tests of significance at the 5% level each. She decides that if any of these tests give a significant result then there is evidence of adverse effects. This experimenter has a multiple testing problem.

To see this, suppose that the chemical has no perceptible affects at the level tested so that the probability of a significant effect on any one of his 20 tests is 0.05. Suppose also that the tests are on independent data. Then the probability of none of the tests being significant is $0.95^{20} = 0.36$, so that the probability of obtaining at least one significant result is 1 - 0.36 = 0.64. Hence the likely outcome of the experimenter's work is to conclude that the chemical has an adverse effect even when it is harmless.

Many solutions to the multiple testing problem have been proposed. The best known of these relate to the specific problem of comparing the mean values at different levels of a factor in conjunction with analysis of variance. These are discussed at length in general statistics texts (e.g. Steel and Torrie, 1980, Chapter 8; Underwood, 1997, Section 8.6), and also in the more specialised texts of Hochberg and Tamhane (1987) and Westfall and Young (1993). These multiple comparison procedures are also available in standard statistical computer packages, although they are not accepted as necessarily being useful by all statisticians (e.g. Saville, 1986, 1990; Mead, 1988, p. 311; Nelder, 1999).

There are also some procedures that can be applied more generally when several tests are to be conducted at the same time. Of these, the **Bonferroni procedure** is the simplest. This is based on the fact that if m tests are carried out at the same time using the significance level $(100\alpha\%)/m$, and all of the null hypotheses are true, then the probability of getting any result significant is less than α. Thus the experimenter with 20 tests to carry out can use the significance level (5%)/20 = 0.25% for each test, and this ensures that the probability of getting any significant results is less than 0.05 when no effects exist.

An argument against using the Bonferroni procedure is that it requires very conservative significance levels when there are many tests to carry out. This has led to the development of a number of improvements that are designed to result in more power to detect effects when they do really exist. **Of these, the method of Holm's (1979) appears to be the one which is easiest to apply** (Peres-Neto, 1999). This does not however take into account the correlation between the results of different tests. If some correlation does exist because the different test statistics are based partially on the same data then in principle methods which allow for this should be better, such as the randomization procedure described by Manly (1997, Section 6.8) which can be applied in a wide variety of different situations (e.g. Holyoak and Crowley, 1993), or several approaches that are described by Troendle and Legler (1998).

Holm's method works using the following algorithm:

1. Decide on the overall level of significance α to be used (the probability of declaring anything significant when the null hypotheses are all true).
2. Calculate the *p*-value for the m tests being carried out.
3. Sort the p-values into the ascending order, to give p_1, p_2, ..., p_m, with any tied values being put in a random order.
4. See if $p_1 \leq \alpha/k$, and if so declare the corresponding test to give a significant result, otherwise stop. Next see if $p_2 \leq \alpha/(k\text{-}1)$, and if so declare the corresponding test to give a significant result, otherwise stop. Next see if $p_3 \leq \alpha/(k\text{-}2)$, and if so declare the corresponding test to give a significant result, otherwise stop. Continue this process until an insignificant result is obtained, or until it is seen whether $p_\mathrm{k} \leq \alpha$, in which case the corresponding test is declared to give a significant result. Once an insignificant result is obtained, all the remaining tests are also insignificant, because their *p*-values are at least as large as the insignificant one.

Example: Multiple Tests on Correlations Between Characters for Brazilian Fish

This example used to illustrate the Holm (1979) procedure is also the one used by Peres-Neto (1999). The situation is that five morphological characters have been measured for 47 species of Brazilian fish and there is interest in which pairs of characters show significant correlation. Table 1.5 shows the ten pairwise correlations obtained with their probability values based on the assumptions that the 47 species of fish are a random sample from some population, and that the characters being measured have normal distributions for this population. (For the purposes of this example the validity of these assumptions will not be questioned.) The calculations for Holm's procedure, using an overall significance level of 5% (α = 0.05) are shown in Table 1.6. It is found that just two of the correlations are significant after allowing for multiple testing.

TABLE 1.5 CORRELATIONS (R) BETWEEN CHARACTERS FOR 47 SPECIES OF BRAZILIAN FISH, WITH CORRESPONDING *p*-VALUES. FOR EXAMPLE, THE CORRELATION BETWEEN CHARACTERS 1 AND 2 IS 0.110, WITH $p = 0.460$.

		CHARACTER			
CHARACTER		1	2	3	4
2	*r*	0.110			
	p-value	0.460			
3	*r*	0.325	0.345		
	p-value	0.026	0.018		
4	*r*	0.266	0.130	0.142	
	p-value	0.070	0.385	0.340	
5	*r*	0.446	0.192	0.294	0.439
	p-value	0.002	0.196	0.045	0.002

1.11 Meta-Analysis: Methods for Combining Results from Several Studies

The term ‘meta-analysis’ is used to describe methods for combining the results from several studies to reach an overall conclusion. This can be done in a number of different ways, with the emphasis either on determining whether there is overall evidence of the effect of some factor, or of producing the best estimate of an overall effect.

TABLE 1.6 CALCULATIONS AND RESULTS FROM THE HOLM (1979) METHOD FOR MULTIPLE TESTING USING THE CORRELATIONS AND *p*-VALUES FROM TABLE 5.1, AND $\alpha = 0.05$.

i	*r*	*p*-VALUE	0.05/(m+i-1)	SIGNIFICANCE
1	0.439	0.002	0.005	yes
2	0.446	0.002	0.006	yes
3	0.345	0.018	0.006	no
4	0.325	0.026	0.007	no
5	0.294	0.045	0.008	no
6	0.266	0.070	0.010	no
7	0.192	0.196	0.013	no
8	0.142	0.340	0.017	no
9	0.130	0.385	0.025	no
10	0.110	0.460	0.050	no

A simple approach to combining the results of several tests of significance was proposed by Fisher (1970). This is based on three well-known results: (a) if the null hypothesis is true for a test of significance then the *p*-value from the test has a uniform distribution between 0 and 1 (i.e. any value in this range is equally likely to occur); (b) If *p* has a uniform distribution, then $-2\log_e(p)$ has a chi-squared distribution with 2 degrees of freedom; and (c) if $X_1, X_2, ..., X_n$ all have independent chi-squared distributions then their sum, $S = \Sigma X_i$ also has a chi-squared distribution, with the number of degrees of freedom being the sum of the degrees of freedom for the components. It follows from these results that if n tests are carried out on the same null hypothesis using independent data and yield p-values of $p_1, p_2, ..., p_n$, then a sensible way to combine the test results involves calculating

$$S_1 = -2\Sigma\log_e(p_i), \tag{1.4}$$

where this will have a chi-squared distribution with $2n$ degrees of freedom if the null hypothesis is true for all of the tests. A significantly large value of S_1 is evidence that the null hypothesis is not true for at least one of the tests, where this will occur if one or more of the individual *p*-values is very small, or if most of the *p*-values are fairly small.

There are a number of alternative methods that have been proposed for combining *p*-values, but Fisher's method seems generally to be about the best providing that the interest is in whether the null hypothesis false for any of the sets of data being compared (Folks, 1984). However, Rice (1990) argued that sometimes this is not quite what is needed. Instead, the question is whether a set of tests of a null hypothesis are in good agreement about whether there is evidence against the null hypothesis. Then a consensus *p*-value is needed to indicate whether, on balance, the null hypothesis is supported or not. For this purpose Rice suggests using the Stouffer method described by Folks (1984).

The Stouffer method proceeds as follows. First the *p*-value from each test is converted to an equivalent *z*-score, i.e. the *p*-value p_i for the ith test is used to find the value z_i such that

$$\text{Prob}(Z < z_i) = p_i, \tag{1.5}$$

where Z is a random value from the standard normal distribution with a mean of zero and a standard deviation of one. If the null hypothesis is true for all of the tests then all of the z_i values will be random values from the standard normal distribution, and it can be shown that their mean $\bar{z}$ will be normally distributed with a mean of zero and a variance of $1/\sqrt{n}$. The mean *z*-value can therefore be tested for significance by seeing whether

$$S_2 = \bar{z}/(1/\sqrt{n}), \tag{1.6}$$

is significantly less than zero.

There is a variation on the Stouffer method that is appropriate when for some reason it is desirable to weight the results from different studies differently. This is called the Liptak-Stouffer method by Folks 1984). In this case, let w_i be the weight for the *i*th study, and define the test statistic

$$S_3 = (w_1z_1 + w_2z_2 + ...w_nz_n)/\sqrt{(w_1^2 + w_2^2 + ...w_n^2)}. \tag{1.7}$$

If the null hypothesis is true for all studies then this will follow a standard normal distribution. If it is significantly low in comparison with the standard normal distribution then this is evidence that the null hypothesis is not always true.

Meta-analysis as generally understood involves more than just combining the p-values from several sets of data. In fact, the usual approach is to take a series of studies and for each one calculate an estimated effect size, which is often just the mean difference between the treated and control groups in units of the estimated standard deviation of individual observations. Questions of interest are then:

- How large is the effect overall?
- Is it generally a positive or a negative effect, and is it usually different from zero?
- Are the effect sizes similar for all studies?
- If there is variation between studies, can this be related to the different types of study involved?

There is a large literature on this type of meta-analysis. Comprehensive sources for more information are the books by Hedges and Olkin (1985, 1999), while the introductions to the topic provided by Gurevitch and Hedges (1993, 1999) and Fernandez-Duque (1997) will be useful for beginners in this area. Also, a recent Special Feature in the journal *Ecology* gives an up-to-date review of applications of meta-analysis in this area (Osenberg *et al.* 1999).

1.12 Bayesian Inference

So far the methods discussed in this book have all been based on a traditional view of statistics, with tests of significance and confidence intervals being the main tools for inference, with these being justified either by the study design (for design-based inference) or an assumed model (with model-based inference). There is, however, **another fundamentally different approach to inference that is being used increasingly in recent times because certain computational difficulties that used to occur have now been overcome.**

This alternative approach is called Bayesian inference because it is based on a standard result in probability theory called Bayes' theorem. To see what this theorem says, consider the situation where it is possible to state that a certain parameter θ must take one, and only one of a set of n specific values denoted by θ_1, θ_2, ... θ_n, and where before any data are collected it is known that the prior probability of θ_i (i.e. the probability of this being the correct value for θ) is $P(\theta_i)$, with $P(\theta_1) + P(\theta_2) + ... + P(\theta_n) = 1$. Some data that are related to θ are then collected. Under these conditions, Bayes' theorem states that the posterior probability that θ_i (i.e. the probability that this is the correct value of θ, given the evidence in the data), is

$$P(\theta_i | \text{data}) = P(\text{data} | \theta_i)P(\theta_i)/ \Sigma P(\text{data} | \theta_k)P(\theta_k), \qquad (1.8)$$

where the summation is for $k = 1$ to n, and $P(\text{data} | \theta_i)$ is the probability of observing the data if $\theta = \theta_i$.

This result offers a means of being able to calculate the probability that a particular value is the correct one for θ on the basis of the data, which is the best that can be hoped for in terms of inferences about θ. The stumbling block is that in order to do this it is necessary to know the prior probabilities before the data are collected.

There are two approaches used to determine prior probabilities when, as is usually the case, these are not really known. The first approach uses the investigator's subjective probabilities, based on general knowledge about the situation. The obvious disadvantage of this is that another investigator will likely not have the same subjective probabilities so that the conclusions from the data will depend to some extent at least on who does the analysis. It is also very important that the prior probabilities are not determined after the data have been examined because equation (1.8) does not apply if the prior probabilities depend on the data. Thus inferences based on equation (1.8) with the prior probabilities depending partly on the data are not Bayesian inferences. In fact, they are not justified at all.

The second approach is based on choosing prior probabilities that are uninformative, so that they do not have much effect on the posterior probabilities. For example, the n possible values of θ can all be given the prior probability 1/n. One argument for this approach is that it expresses initial ignorance about the parameter in a reasonable way, and that providing there is enough data the posterior probabilities do not depend very much on whatever is assumed for the prior probabilities.

Equation (1.8) generalizes in a straightforward way to situations where there are several or many parameters of interest, and where the prior distributions for these parameters are discrete or continuous. For many purposes, all that needs to be known is that

$$\text{P(parameters} \mid \text{data)} \propto \text{P(data} \mid \text{parameters)P(parameters)},$$

i.e. the probability of a set of parameters values given the data is proportional to the probability of the data given the parameter values, multiplied by the prior probability of the set of parameter values. This result can be used to generate posterior probability distributions using possibly very complicated models when the calculations are done using a special technique called Markov chain Monte Carlo.

There is one particular aspect of Bayesian inference that should be appreciated. **It is very much model-based in the sense discussed in Section 5.** This means that it is desirable with any serious study that the conclusions from an analysis should not be quite robust to both the assumptions made about prior distributions, and the assumptions made about the other components in the model. Unfortunately, these types of assessments are often either not done, or not done very thoroughly.

A brief introduction to Bayesian methods is included here because it seems likely that there will be increasing use of these methods in the future as means of drawing conclusions from data. More information about them with the emphasis on Markov Chain Monte Carlo methods is provided in the books by Manly (1997) and Gilks *et al.* (1996). For more on Bayesian data analysis in general see the book by Gelman *et al.* (1995), and for arguments why these approaches should be viewed with caution see Dennis (1996).

1.13 Data Quality Objectives (DQO) Process

The Data Quality Objectives (DQO) process was developed by the United States Environmental Protection Agency to ensure that when a data collection endeavour has been completed it will have accomplished two goals:

- provided sufficient data to make required decisions within a reasonable certainty
- collected only the minimum amount of necessary data

The idea is to have the least expensive data collection scheme, but not at the price of providing answers that have too much uncertainty (United States Office of Environmental Management, 1997).

At the heart of the use of the process is the assumption that there will always be two problems with environmental decision making: (1) the resources available to address the question being considered are not infinite, and (2) there will never be a 100% guarantee that the right decision has been reached. Generally, more resources can be expected to reduce uncertainty. The DQO process therefore attempts to get the right balance between resource use and uncertainty. There are two main activities involved:

- the questions to be answered are stated very specifically for the problem being considered
- the amount of uncertainty that can be tolerated is stated very specifically

The DQO process then provides a complete and defensible justification for the data collection methods used, covering:

- the questions that are important
- whether the data will answer the questions
- what quality of data is needed
- how much data are needed
- how the data will actually be used in decision making

This is all done before the data are collected, and preferably agreed to by all the stakeholders involved.

There are seven steps to the DQO process:

1. State the problem: describe the problem, review prior work, understand the important factors.
2. Identify the decision: find what questions need to be answered and the actions that might be taken, depending on the answers.
3. Identify inputs to the decision: determine the data needed to answer the important questions.
4. Define the study boundaries: specify the time periods and spatial areas to which decisions will apply, determine when and where to gather data.
5. Develop a decision rule: define the parameter of interest, specify action limits, integrate the previous DQO outputs into a single statement that describes the logical basis for choosing among alternative possible actions.

6. Specify limits on decision errors: specify tolerable decision error probabilities (probabilities of making the wrong decisions) based on the consequences of incorrect decisions.

7. Optimize the design for obtaining data: generate alternative sampling designs, choose the one that meets all the DQOs with the minimum use of resources.

The output from each step influences the choices made later but it is important to realize that the process is iterative and the carrying out of one step may make it necessary to reconsider one or more earlier steps. Steps 1-6 should produce the Data Quality Objectives that are needed to develop the sampling design at step 7.

When used by the US EPA, a DQO planning team usually consists of technical experts, senior managers, a statistical expert and a quality assurance/quality control (QA/QC) advisor. The final product is a data collection design that meets the qualitative and quantitative needs of the study, and much of the information generated during the process is used for the development of Quality Assurance Project Plans (QAPPs) and the implementation of the Data Quality Assessment (DQA) Process. These are all part of the US EPA's system for maintaining quality in their operations.

More information about the DQO process with reference documents can be obtained from the world-wide web (United States Office of Environmental Management, 1997). A good start is to read the EPA's guidance document (United States Environmental Protection Agency, 1994).

1.14 Key Points in This Module

- The prior knowledge of statistics needed to get the maximum benefit from this and the other modules is summarised
- In drawing conclusions from data, an important distinction is between observational and experimental studies. In general, observational studies are more likely to be affected by uncontrolled factors, leading to incorrect conclusions.
- Experimental studies can be either true experiments, or quasi-experiments. True experiments incorporate randomization, replication and controls, while quasi-experiments lack one of these components. Many studies in environmental science are really quasi-experiments, so it is important to realize the limitations that this imposes on inferences.
- There are two quite distinct philosophies that are used for drawing conclusions with conventional statistical methods. One is design-based, drawing its justification from the randomization used in sampling, or in the random allocation of experimental units to different treatments. The other is model-based, drawing its justification from the random variation inherent in a model assumed to describe the nature of observations. In general it is recommended that where possible inferences should be design-based because this requires fewer assumptions and is always valid providing that randomizations are properly carried out.
- There are limitations with tests of significance which has led to their use being criticised, at least with some applications. Two particular problems are: (1) often tests are carried out when the null hypothesis is known to probably be untrue so that a significant result is very likely if enough data are collected, and (2) a non-significant result does not mean that the null hypothesis is false because the sample size may just not be large enough to detect the existing effect.
- It is argued that null hypotheses are relevant in situations where there really is doubt about whether a null hypothesis is true or not. If this is not the case then it is more reasonable to estimate the magnitude of effects with some measure of how accurate the estimation is, using a confidence interval, for example.
- Randomization tests have been used quite often with environmental data. The idea is to compare the value of an observed test statistic with the distribution obtained by randomly reallocating the data to different samples in some sense. These tests have the advantage of requiring fewer assumptions than more conventional tests. They are, however, computer-intensive and may need special computer software.
- Bootstrapping is another computer-intensive method. It is based on the idea that in the absence of any knowledge about a distribution other than the values in a random sample from the distribution, the best guide to what would happen by resampling the population is to resample the sample. In principle, bootstrapping can be used to conduct tests of significance and to construct confidence intervals for population parameters.

- Pseudoreplication occurs when standard statistical methods are used to test treatment effects where either treatments are not replicated, or replicates are not statistically independent. Two errors can be made in this respect: (1) inferences can be extended outside the population actually sampled, and (2) observations that are correlated because they are close in space or time are not analysed taking this into account.

- When several related hypothesis tests are carried out at the same time and all the null hypotheses are true, the probability of at least one significant result increases with the number of tests. This is a well-known problem which has led to the development of a range of procedures to take into account the multiple testing. Multiple comparison methods to compare means at different factor levels after analysis of variance are procedures of this type that are widely available in statistical software, but not favoured by all statisticians. More general approaches are also used, including using a Bonferroni adjustment for the significance level used with individual tests, and Holm's stepwise method for adjusting these levels.

- Meta-analysis is concerned with the problem of combining the results from a number of different studies on the same subject. This can be done by combining the p-values obtained from different studies using Fisher's method, or the Stouffer method. The Stouffer method also has a variation called the Liptak-Stouffer method which allows the results of different studies to receive different weights. Alternatively, rather than using p-values, an effect size is estimated for each study and these effect sizes are examined in terms of the overall effect, the extent to which the effect varies from study to study, and whether the variation between studies can be explained by differences between the type of studies used.

- Bayesian inference is different from conventional statistical inference, and is becoming more widely used. With this approach a prior distribution assumed for a parameter of interest is changed using Bayes' theorem into a posterior distribution, given the information from some new data. Modern computing methods, particularly Markov chain Monte Carlo, make the Bayesian approach much easier to use than was the case in the past. However, it is cautioned that Bayesian inference is very much model-based, with all the potential problems that this implies.

- The United States Data Quality Objective (DQO) process is described. This is a formal seven step mechanism for ensuring that sufficient data are collected to make required decisions with a reasonable probability that these are correct, and that only the minimum amount of necessary data are collected.

1.15 Questions About This Module

After completing this module you should be able to give reasonable answers to the following questions:

1. For the following examples, decide whether it is an observational or experimental study:

   (a) the trials using 1080 pellets in the example of a randomization test used in Section 1.7;

   (b) monitoring of *Dactylanthus taylorii* near the summit of Mount Pirongia, 1997-99 (Data Set 3 in the Appendix), where various measures such as the number of chewed buds on plants were compared for plants that were uncaged or caged;

   (c) collection of data on blue cod lengths from Paterson Inlet, 1994-98 (Data Set 5 in the Appendix), where some sampling sites are in a proposed marine reserve, and other sites are not;

   (d) collection of data on the stomach contents of galaxid and trout in Otago streams (Data Set 8 in the Appendix).

2. What is the difference between a 'true' experiment and a quasi-experiment?

3. Are the before-after-control-impact (BACI) studies carried out by DOC usually true experiments or quasi-experiments?

4. What is the difference between design-based inference and model-based inference, and which is usually easiest to defend?

5. What is it about tests of significance which has led to their use being attacked by a number of authors in recent years?

6. Given a sample of adult animals, and their lengths, how would you use a randomization test to decide whether there is evidence that males tend to be larger than females?

7. Consider the bootstrap example on finding a 95% confidence interval for the mean chlorophyll-a in the lakes in a region used in Section1.8. How could bootstrapping be used to decide on the number of lakes that need to be sampled in order to estimate the population mean with a 95% confidence interval of ± 10.0.

8. What are the two common types of pseudoreplication?

9. Under what circumstances should an adjustment for multiple testing be used when carrying out a series of statistical tests, and what is the likely outcome if no adjustment is made?

10. What is mean by a 'meta-analysis', and why might you want to combine the p-values from several studies as part of a study of this type?

11. In order to use Bayesian inference it is necessary to specify prior distributions for the parameters of interest (e.g. the prior distribution for the mean amount of possum browse on the canopy of the trees in an area). What exactly does this prior distribution represent?

12. What are the main goals of the US EPA's Data Quality Objectives (DQO) process, and to what extent is this relevant with New Zealand conditions?

REFERENCES

Campbell, D.T. and Stanley, J.C. (1963). *Experimental and Quasi-Experimental Designs for Research.* Houghton Mifflin, Boston.

Cherry, S. (1998). Statistical tests in publications of the Wildlife Society. *Wildlife Society Bulletin* 26: 947-53.

Dennis, B. (1996). Discussion: should ecologists become Bayesians? *Ecological Applications* 6: 1095-103.

Eberhardt, L.L. and Thomas, J.M. (1991). Designing environmental field studies. *Ecological Monographs* 61: 53-73.

Efron, B. (1979). Bootstrap methods - another look at the jackknife. *Annals of Statistics* 7: 1-26.

Fisher, R.A. (1935). *The Design of Experiments.* Oliver and Boyd, Edinburgh.

Fisher, R.A. (1936). The coefficient of racial likeness and the future of craniometry. *Journal of the Royal Anthropological Institute* 66: 57-63.

Fisher, R.A. (1970). *Statistical Methods for Research Workers*, 14th Edit. Oliver and Boyd, Edinburgh.

Folks, J.L. (1984). Combination of independent tests. In *Handbook of Statistics 4, Nonparametric Methods* (eds. P.R. Krishnaiah and P.K. Sen), pp. 113-21. North-Holland, Amsterdam.

Gardner, M.J. and Altman, D.G. (1986). Confidence intervals rather than p-values: estimation rather than hypothesis testing. *British Medical Journal* 292: 746-50.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.R. (1995). *Bayesian Data Analysis.* Chapman and Hall, London.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J., eds. (1996). *Markov Chain Monte Carlo in Practice.* Chapman and Hall, London.

Gurevitch, J. and Hedges, L.V. (1993). Meta-analysis: combining the results of independent studies in experimental ecology. In *The Design and Analysis of Ecological Experiments* (eds. S. Scheiner and J. Gurevitch), pp. 378-98. Chapman and Hall, New York.

Gurevitch, J. and Hedges, L.V. (1999). Statistical issues in ecological meta-analysis. *Ecology* 80: 1142-49.

Hairston, N.G. (1989). *Ecological Experiments: Purpose, Design, and Execution.* Cambridge University Press, Cambridge.

Hall, P. and Wilson, S. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics* 47: 757-62.

Hedges, L.V. and Olkin, I. (1985). *Statistical Methods for Meta-Analysis.* Academic Press, New York.

Hedges, L.V. and Olkin, I. (1999). *Statistical Methods for Meta-Analysis in the Medical and Social Sciences.* Academic Press, New York.

Hochberg, Y. and Tamhane, A.C. (1987). *Multiple Comparison Procedures.* Wiley, New York.

Holm, S. (1979). A simple sequential rejective multiple test procedure. *Scandinavian Journal of Statistics* 6: 65-70.

Holyoak, M. and Crowley, P.H. (1993). Avoiding erroneously high levels of detection in combinations of semi-independent tests. Oecologia 95: 103-14.

Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54:187-211.

Johnson, D.H. (1999). The insignificance of significance testing. *Journal of Wildlife Management* 63: 763-72.

Manly, B.F.J. (1992). *The Design and Analysis of Research Studies.* Cambridge University Press, Cambridge.

Manly, B.F.J. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd Edit. Chapman and Hall, London.

McBride, G.B., Loftis, J.C. and Adkins, N.C. (1993). What do significance tests really tell us about the environment? *Environmental Management* 17: 423-32.

Mead, R. (1988). *The Design of Experiments: Statistical Principles for Practical Applications.* Cambridge University Press, Cambridge.

Nelder, J.A. (1999). Statistics for the millenium: from statistics to statistical science. *Statistician* 48: 257-69.

Oakes, M. (1986). *Statistical Significance: a Commentary for the Social and Behavioural Sciences.* Wiley, New York.

Osenberg, C.W., Sarnelle, O. and Goldberg, D.E. (1999). Special feature on meta-analysis on ecology: concepts, statistics, and applications. *Ecology* 80: 1103-4.

Pallisade (1995). *@RISK Risk Analysis and Simulation Add-In for Microsoft Excel or Lotus 123.* Palisade Corporation, 31 Decker Road, Newfield, NY 14867.

Peres-Neto, P.R. (1999). How many tests are too many? The problem of conducting multiple ecological inferences revisited. *Marine Ecology Progress Series* 176: 303-6.

Rice, W.R. (1990). A consensus combined p-value test and the family-wide significance of component tests. *Biometrics* 46: 303-8.

Saville, D.J. (1986). The inconsistency of multiple comparison procedures. In *Pacific Statistical Congress* (eds. I.S. Francis, B.F.J. Manly and F.C. Lam), pp. 286-9. Elsevier Science Publishers B.V., Amsterdam.

Saville, D.J. (1990). Multiple comparison procedures: the practical solution. *American Statistician* 44: 174-80.

Schmoyer, R.L., Beauchamp, J.J., Brandt, C.C. and Hoffman, F.O. (1996). Difficulties with the lognormal model in mean estimation and testing. *Environmental and Ecological Statistics* 3: 81-97.

Steel, R.G.D. and Torrie, J.H. (1980). *Principles and Procedures of Statistics: a Biometrical Approach.* McGraw-Hill, New York.

Sherley, G. and Wakelin, M. (1998). Impact of monofluroacetate ("1080") on forest invertebrates at Ohakune, North Island, New Zealand. Submitted for publication.

Troendle, J.F. and Legler, J.M. (1998). A comparison of one-sided methods to identify significant individual outcomes in a multiple outcome setting: stepwise tests or global tests with closed testing. *Statistics in Medicine* 17: 1245-60.

Underwood, A.J. (1997). *Experiments in Ecology: Their Logical Design and Interpretation Using Analysis of Variance.* Cambridge University Press, Cambridge.

United States Environmental Protection Agency (1994). *Guidance for the Data Quality Objectives Process.* Report EPA/600/R-96/055, Office of Research and Development, Washington DC 20460.

United States Office of Environmental Management (1997). Data Quality Objectives: why use the DQO process? Web page http://etd.pnl.gov:2080/DQO/why.html.

Westfall, P.H. and Young, S.S. (1993). *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment.* Wiley, New York.

Wiens, B.L. (1999). When lognormal and gamma models give different results: a case study. *American Statistician* 53: 89-93.

Contents

MODULE 2: SAMPLE SURVEY DESIGNS

Module 2: Sample Survey Designs

SUMMARY

When data are collected to assess environmental quality, or change in quality, they are usually only a sample from the population. It is important therefore to understand at least the basic ideas of the theory of statistics concerning sampling designs and the estimation of population parameters. This module introduces the idea of simple random, stratified and systematic sampling, as well as some more complex designs, and discusses how they are applied to environmental management.

2.1 Population Parameters and Sample Statistics

Environmental data is usually only a sample of the population of interest. For example, a data set of ground water quality measurements is only a subset of what could have been collected if there was an infinitely large budget. The subset is a sample from a population which, in this example, may be defined spatially (e.g. the Waimakariri Basin) and temporally (e.g. the calendar year of 1998). The goal of sampling is to summarise the characteristics of the population, e.g. the quality of the water. A summary of these characteristics might then consist of estimates of averages and variability. These estimates are called “estimators” of the population. As another example, if we are interested in breeding rate of the rodents a sample of the rodents may be collected and the weight and sex of each animal recorded. The mean and standard deviation of the weight of animals and the proportion of females are used to characterise the population.

In the terminology of statistics, the population is defined to be the collection of all items that are of interest in an investigation. This definition should include the spatial and temporal location. The water quality dataset is used to estimate water quality in the Waimakariri Basin, and in the period of 1998. The items in a population may be individual animals or plants, or small plots of land, pieces of rock, or groups of animal, or units of water and air. The crucial thing as far as statistical theory is concerned is that the items that make up the population are sampled using an appropriate procedure. For this reason the items are often called the sample units.

Sometimes population sizes are small enough to allow every item to be examined. This then provides a census. However, the populations of interest are usually too large to make a census practical. In other situations information may be available on all items but it is not possible to process all the data in the time allowed or, in the time it takes to collect the information on all the items the population would have changed. For example, it would be possible, given enough people, to survey the entire West Coast Region and measure the coverage of Old Man’s Beard. Given the growth rate of the vine, if the survey were conducted in summer the coverage would probably have increased between the start of the survey and the end!

The measures that are used to summarise a population are called population parameters and the corresponding sample values are called statistics. For example, the population mean (a parameter) may be estimated by a sample mean (a statistic). Similarly, a population sex ratio of females (a parameter) may be estimated by a sample ratio (a statistic). Remember - Population – Parameter Sample - Statistics. Some of the commonly used notation for populations and samples are:

Sample mean - $\bar{y}$	Population mean - μ
Sample variance - s^2	Population variance - σ^2

2.2 Variability in Environmental Data

One of the typical features of environmental data is high variability. Variability means that the measurements taken from the population unit are not the same. We expect to have variability in environmental data and often the amount of variation is what we are specifically interested in. For example, the abundance of a pest population is expected to increase following a control operation. The goal in monitoring may be to track this change, or variation, over time to trigger a follow-up control operation once the population has reached a target threshold level. Temporal variation can be confounded by other sources of variation, e.g. seasonal changes and variation due to the fact we measure only a sample rather than the entire population (Link *et al.* 1994). What is important in sampling is to design a survey that minimises these confounding sources of variation, or allows for them to be quantified.

2.3 Simple Random Sampling

Whenever inferences are to be made about population parameters on the basis of the sample result the sample design should have some element of random selection. This is because of the concept of statistical inference, i.e. inferring about the population from the sample data is based on the laws of probability. Random sampling is not subjective sampling, where units in the population are chosen because they “seem to be representative”, or haphazard sampling where the units in the sample are those that are “convenient” to select. Random sampling involves the selection of units using a well defined and carefully carried out randomisation procedure that (in its simplest application) ensures that all possible samples of the required size are equally likely to be chosen. When subjective or haphazard sampling is conducted results should not be used to infer about the whole population.

All possible samples from a population have equal chance of being selected with random sampling and this method may produce exactly the same units as a subjectively or haphazardly drawn sample. It is important therefore to appreciate that the key to the value of random sampling is the properties of this procedure rather than the specific units that are obtained. In fact it is not uncommon to feel uncomfortable with the result of random sampling because it does not look “representative” enough on close inspection. This will not be a valid objection to the sample providing that the procedure used to select it was defined and carried out in an appropriate manner.

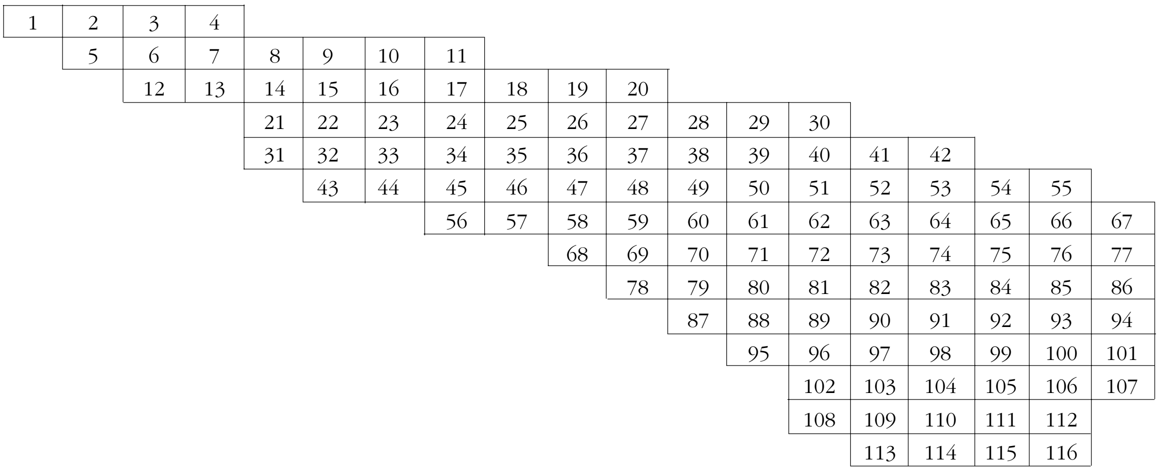
Simple random sampling involves giving each sample unit the same probability of being selected. This can be with replacement, in which case every selected unit is chosen from the full population, irrespective of which units have already been included in the sample, or without replacement, in which case a sample unit can occur at most once in the sample. As a general principle, sampling without replacement is preferable to sampling with replacement because it gives slightly more accurate estimation of population parameters. However, the difference between the two methods of sampling is not great when the population size is much larger than the sample size.

Example: Sampling Weeds in a Large Study Area

Suppose we need to estimate the density of a weed species in a large study area. One approach would be to set up a grid over the entire area. Figure 1 indicates the type of result that might then be obtained, where in this case there are 116 quadrats covering an irregular shaped area. The quadrats are the sampling units within the population of interest. The list of these units is sometimes called the sampling frame. The variable that will be measured in each sample unit, in this example, is the number of plants in the quadrat.

The next step would be to decide on a sample size *n*, i.e. how many quadrats should be randomly sampled to estimate the population mean number of plants per quadrat with an acceptable level of accuracy. Methods for choosing sample sizes are discussed later in this module. For this example it will be assumed that a sample size of 10 is needed. Note that we have had to decide on what quadrat size to use in order to determine the desired number of quadrats.

There are various ways to determine the random sample. One approach involves labelling the quadrats in the population from 1 to 116 (as in Figure 2.1) and then selecting the ones to sample by generating on a computer 10 random integers from 1 to 116. Because sampling should be without replacement the same quadrat would not be allowed to occur more than once. Any repeated selections would therefore be ignored and the process of selecting random integers continued until 10 different quadrats have been chosen.



1	2	3	4															
	5	6	7	8	9	10	11											
		12	13	14	15	16	17	18	19	20								
				21	22	23	24	25	26	27	28	29	30					
				31	32	33	34	35	36	37	38	39	40	41	42			
					43	44	45	46	47	48	49	50	51	52	53	54	55	
							56	57	58	59	60	61	62	63	64	65	66	67
									68	69	70	71	72	73	74	75	76	77
										78	79	80	81	82	83	84	85	86
											87	88	89	90	91	92	93	94
												95	96	97	98	99	100	101
													102	103	104	105	106	107
													108	109	110	111	112	
														113	114	115	116	

Figure 2.1 A study area divided into 116 square quadrats to be used as sample units.

Computer programs are available to generate random integers within the specified range of 1 to 116. If one of these is not available then a table of random numbers such as the one provided here in Table 2.1 can be used. It is definitely not allowable to just think up the numbers. Human beings do not have random number generators in their heads!

TABLE 2.1 A RANDOM NUMBER TABLE WITH EACH DIGIT CHOSEN SUCH THAT 0, 1, ..., 9 WERE EQUALLY LIKELY TO OCCUR. THE GROUPING INTO GROUPS OF FOUR DIGITS IS ARBITRARY SO THAT, FOR EXAMPLE, TO SELECT NUMBERS FROM 0 TO 99999 THE DIGITS CAN BE CONSIDERED FIVE AT A TIME.

1252	9045	1286	2235	6289	5542	2965	1219	7088	1533
9135	3824	8483	1617	0990	4547	9454	9266	9223	9662
8377	5968	0088	9813	4019	1597	2294	8177	5720	8526
3789	9509	1107	7492	7178	7485	6866	0353	8133	7247
6988	4191	0083	1273	1061	6058	8433	3782	4627	9535
7458	7394	0804	6410	7771	9514	1689	2248	7654	1608
2136	8184	0033	1742	9116	6480	4081	6121	9399	2601
5693	3627	8980	2877	6078	0993	6817	7790	4589	8833
1813	0018	9270	2802	2245	8313	7113	2074	1510	1802
9787	7735	0752	3671	2519	1063	5471	7114	3477	7203
7379	6355	4738	8695	6987	9312	5261	3915	4060	5020
8763	8141	4588	0345	6854	4575	5940	1427	8757	5221
6605	3563	6829	2171	8121	5723	3901	0456	8691	9649
8154	6617	3825	2320	0476	4355	7690	9987	2757	3871
5855	0345	0029	6323	0493	8556	6810	7981	8007	3433
7172	6273	6400	7392	4880	2917	9748	6690	0147	6744
7780	3051	6052	6389	0957	7744	5265	7623	5189	0917
7289	8817	9973	7058	2621	7637	1791	1904	8467	0318
9133	5493	2280	9064	6427	2426	9685	3109	8222	0136
1035	4738	9748	6313	1589	0097	7292	6264	7563	2146
5482	8213	2366	1834	9971	2467	5843	1570	5818	4827
7947	2968	3840	9873	0330	1909	4348	4157	6470	5028
6426	2413	9559	2008	7485	0321	5106	0967	6471	5151
8382	7446	9142	2006	4643	8984	6677	8596	7477	3682
1948	6713	2204	9931	8202	9055	0820	6296	6570	0438
3250	5110	7397	3638	1794	2059	2771	4461	2018	4981
8445	1259	5679	4109	4010	2484	1495	3704	8936	1270
1933	6213	9774	1158	1659	6400	8525	6531	4712	6738
7368	9021	1251	3162	0646	2380	1446	2573	5018	1051
9772	1664	6687	4493	1932	6164	5882	0672	8492	1277
0868	9041	0735	1319	9096	6458	1659	1224	2968	9657
3658	6429	1186	0768	0484	1996	0338	4044	8415	1906
3117	6575	1925	6232	3495	4706	3533	7630	5570	9400
7572	1054	6902	2256	0003	2189	1569	1272	2592	0912
3526	1092	4235	0755	3173	1446	6311	3243	7053	7094
2597	8181	8560	6492	1451	1325	7247	1535	8773	0009
4666	0581	2433	9756	6818	1746	1273	1105	1919	0986
5905	5680	2503	0569	1642	3789	8234	4337	2705	6416
3890	0286	9414	9485	6629	4167	2517	9717	2582	8480
3891	5768	9601	3765	9627	6064	7097	2654	2456	3028

To use Table 2.1 to choose the sample, first start at an arbitrary place in the table such as the beginning of row five. The first three digits in each block of four digits can then be considered, to give the series 698, 419, 008, 127, 106, 605, 843, 378, 462, 953, 745, and so on. The first ten different numbers between 1 and 116 then gives a simple random sample of quadrats: 8, 106, and so on.

Once the sample quadrats are chosen the quadrats are surveyed to find the number of plants that each contains. The average for the 10 sample quadrats then gives an estimate of the mean number of plants in the area of one quadrat over the entire study region. The likely level of error can be determined by methods to be discussed below. If necessary, the estimate and its error can then be converted to be in terms of plants per square metre, or the total number of plants in the area, or any other measure of density.

Another way to select a random sample is to select random sample unit locations from a map. Using horizontal and vertical co-ordinates from the map (e.g. the eastings and northings) and random numbers, pairs of random co-ordinates can be selected. For example in a particular study area the horizontal map co-ordinates extended from 890 to 990 (or for 100 units) and the vertical co-ordinates extended from 700 to 800 (100 units). The RND function on a calculator was used and the first two random numbers were 0.403 and 0.414. The first sample location is at: 890 + (100 * 0.403) = 930.3 = 930 east and .700 + 100 * 0.414 = 741.4 = 741 north. This sequence can be repeated for as many sample points are needed. If any sample point falls outside the study area it is discarded and a new set of random numbers generated. There are many other ways to select random co-ordinates for a map. The important point is that the co-ordinates should be randomly, not subjectively, selected.

2.4 Estimation of Mean Values

Assume that a simple random sample of size *n* is taken from a population of N units, and that the variable of interest Y has values $y_1, y_2, \ldots, y_n$, for the sampled units. Then sample statistics that are commonly computed are:

the sample mean
$$\bar{y} = \frac{(y_1 + y_2 + \ldots y_n)}{n} = \frac{\sum_{i=1}^{n} y_i}{n} \quad (1)$$

the sample variance
$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1} \quad (2)$$

the sample standard deviation $s = \sqrt{s^2}$

the estimated coefficient of variation, $\hat{cv} = \frac{s}{\bar{y}}$

Note that the use of a caret (^) is to indicate an estimate rather than the actual true value, e.g. $\bar{y} = \hat{\mu}$. The difference between the sample mean $\bar{y}$ and the true population μ mean, is the sampling error. This difference will vary from sample to sample if the sampling process is repeated. It can be shown theoretically that if the random sampling process is repeated many times then the sampling error will average out to zero. Therefore the sample mean is an "unbiased" estimator of the population mean.

If there were many samples taken from the same population then the variance among the sample means is a measure of the precision of the sampling procedure. If the sample means were all very similar then the sampling has high precision. If the sample means are quite different than the sampling would have low precision. In real life we will usually only take one sample from a population but the precision of the sample can still be measured although this measure is only an estimate. The estimate of the precision of a sample mean is:

$$\hat{\text{var}}(\bar{y}) = \frac{s^2}{n}\left(1 - \frac{n}{N}\right) \qquad (3)$$

The factor (1-*n/N*) is called the finite population correction factor (fpc). This fpc adjusts the estimate of precision according to what proportion of the population has been sampled. For example if only 10% of the population were sampled, (say N = 1000 and n = 100) then the fpc will be 0.9. But if 40% of the population were sampled then the fpc = (1 - 0.4) = 0.6. Multiplying the term s^2/n by 0.6 has a large effect in shrinking the estimate of variance. Note that the fpc is only used where you have a finite population, that is, where you know how many units there are in the population. In the above example there were 116 quadrats, and N = 116. In other examples the population size may be unknown. If possums were being sampled in a design where individual animals were the sample unit we don't know how many possums there are in NZ and could not use a fpc. A population may be truly infinite e.g. where sampling occurs at regular intervals through time and may continue forever. In this situation a fpc would not be used either. In other situations the fraction of the population that is sampled is so small the fpc has little real effect.

The square root of $\hat{\text{var}}(\bar{y})$ is the estimated standard error of the sample mean. It is usually denoted by SE, or $\text{SE}(\bar{y})$. Strictly speaking it should be $\hat{\text{SE}}(\bar{y})$ because it is the "estimated" standard error. The estimated coefficient of variation of the mean is $\hat{cv}(\bar{y}) = \frac{\text{SE}}{\bar{y}}$. This is written simply as *cv* although this makes it difficult to distinguish between $cv(y)$ and $cv(\bar{y})$.

The terms "standard error of the mean" and "standard deviation" are often confused when encountered for the first time. What must be remembered is that the standard error of the mean is just the standard deviation of the mean rather than the standard deviation of individual observations. More generally, the term "standard error" is used to describe the standard deviation of any sample statistic that is used to estimate a population parameter.

The coefficient of variation of the mean is an index that reflects the precision of estimation relative to the magnitude of the mean. This can be used to compare the results of several studies, to see which have relatively better precision than others.

Example: Surveys of Blue Cod from Paterson Inlet, Stewart Island (Appendix Data Set 5)

As an example of the calculation of the statistics that have just been defined, consider the data from six sites that were randomly selected from inside and six from outside the proposed Marine Reserve area at Paterson Inlet. Data collected from each site included: the average length of cod, the number of cod, the standard deviation of the cod lengths and the coefficient of variation of the cod lengths, *cv(y)*. The sample unit is the survey sites, and not the individual fish that were measured. This is discussed further in module 3 but briefly, the survey sites were randomly selected rather than the individual fish. Therefore the appropriate unit is the site. The size of the sample is n = 6 for the survey inside the Reserve area and n = 6 for the survey outside the Reserve area. The sample statistics from the data collected from the Reserve area in 1998 in April were calculated in SPSS and some of the output is in Table 2.2.

TABLE 2.2 SUMMARY STATISTICS FOR SURVEYS INSIDE (RESERVE = IN) AND OUTSIDE (RESERVE = OUT) THE PROPOSED MARINE RESERVE AT PATERSON INLET IN APRIL 1998. THE VARIABLES ARE: THE AVERAGE COD LENGTH (AVERAGE), THE NUMBER OF COD (NUMBER), THE STANDARD DEVIATION OF THE COD LENGTHS (SD) AND THE COEFFICIENT OF VARIATION OF THE COD LENGTHS (CV).

RESERVE		N	MEAN	STD. ERROR	STD. DEVIATION	VARIANCE
In	Average	6	295.1884	9.3289	22.8512	522.175
	Number	6	53.8333	11.5684	28.3367	802.967
	sd	6	60.0110	7.3736	18.0616	326.223
	cv	6	.2031	2.450E-02	6.001E-02	3.602E-03
Out	Average	6	296.0682	7.6332	18.6974	349.592
	Number	6	64.0000	10.1259	24.8032	615.200
	sd	6	50.4157	5.5507	13.5964	184.862
	cv	6	.1706	1.845E-02	4.520E-02	2.043E-03

The mean of the average cod length among survey sites within the proposed Reserve was $\bar{y}$ = 295.19, $\mathrm{SE}(\bar{y})$ = 9.33, s = 22.85 and s^2 = 522.18. The $\hat{cv}(\bar{y})$ = 295.19/22.85 = 12.92. These calculations have been carried through to two decimal places. As a general rule it is reasonable for statistics to be quoted with at least one more decimal place than the original data. SPSS does not include the fpc in calculating the standard error. In this example the fpc is not needed because there are an infinite number of sites that could have been sampled. If the fpc were to be used the standard error produced by SPSS would need to be adjusted.

The summary statistic for the variable "Number" can also be read from the SPSS output. The meaning of summary statistics for the other variables is a little more complicated. The variable *cv* was used as a measure of the variation in cod length within each survey site. The mean of the *cv*, 0.20, is a measure of the average amount of this variation of the within-survey site cod lengths. The other three statistics, $\mathrm{SE}(\bar{y})$, s and s^2 are measures of how variable, or how different, the within-site variation is, i.e. whether some sites have more or less variation in cod length than other sites.

The summary statistics for the survey sites outside the Reserve can be compared to the statistics from the sites inside the Reserve. For example, the average cod length in sites in the Reserve area ($\bar{y}_{in}$ = 295.19) is similar to the average length of cod in sites outside the Reserve area ($\bar{y}_{out}$ = 296.07). The next step would be to test for statistical significance, e.g. by a t-test.

The accuracy of a sample mean for estimating the population mean is often represented by a 100(1-α)% confidence interval of the form

$$\bar{y} \pm z_{\alpha/2,}\mathrm{SE} \tag{4}$$

where $z_{\alpha/2}$ refers to the value that is exceeded with probability α/2 for the standard normal distribution. Common values used for the confidence level with the corresponding values of z are: $z_{0.25} = 0.68$ for 50% confidence, $z_{0.16} = 1.00$ for 68% confidence, $z_{0.05} = 1.64$ for 90% confidence, $z_{0.025} = 1.96$ for 95% confidence, and $z_{0.005} = 2.58$ for 99% confidence. The meaning of "confidence interval" is best understood by an example. For instance the interval $\bar{y} \pm 1.64(\mathrm{SE})$ will contain the true population mean with a probability of approximately 0.90. In other words, if there were many samples taken from a population, and for each a confidence interval were calculated, approximately 90% of the intervals would contain the true population mean.

The interval (4) is only valid for large samples. For small samples (say with $n < 20$) it is better to use

$$\bar{y} \pm t_{\alpha/2,n-1}\mathrm{SE}$$

where $t_{\alpha/2,n-1}$ is the value that is exceeded with probability α/2 for the t-distribution with n-1 degrees of freedom. This requires the assumption that the variable being measured is approximately normally distributed in the population being sampled. If this is not the case then no simple method exists for calculating an exact confidence interval. For the cod data example above the 90% confidence interval for the average cod length in sites inside the proposed Reserve in April 1998 is:

$$295.19 \pm 2.02(9.33)$$

$$295.19 \pm 18.80$$

The value 2.02 is $t_{0.025,\,5.}$ This can be found in statistical tables, or as a function in EXCEL using the function TINV(0.1,5), or in SPSS with the transformation IDF.T(0.05,5). Note, EXCEL gives the two-tailed distribution (which is what we want here) and SPSS the one-tailed distribution. We can say with 90% confidence that the true average cod length for the proposed Marine Reserve area is between 295.19 - 18.80 = 276.39 and 295.19 + 18.80 = 313.99. For the area outside the Reserve the 90% confidence interval is 296.07 ± 2.02(7.63), or between 280.69 and 311.45. These confidence intervals overlap and there is no evidence that there are differences in average cod length between the area inside and outside the proposed Reserve.

2.5 Sample Sizes for Estimation of Means

One of the key considerations in designing a study is choosing what sample size is use. The sample size should be large enough to give an adequate level of precision, but should not be unnecessarily large. The sample size is determined by the properties desired for the estimates and the available resources. In most studies the resources available have the greatest influence.

One property for estimates that plays an important role in sample size determination is the precision. This refers to how close the estimate is to the quantity being estimated and usually we think in terms of the maximum allowable precision. The precision may be defined to be:

(i) absolute so that, for example, a mean is to be estimated to within 10 of the true figure;

(ii) relative so that, for example, an estimate is to be within 5% of the true figure; or

(iii) in terms of the coefficient of variation of the mean, $cv(\bar{y})$ so that, for example $cv(\bar{y})$ should be no more than 0.02.

Usually precision is defined in terms of absolute or relative precision. However, some researchers prefer the use of the coefficient of variation.

The other property of estimates that is important is the reliability, which refers to the probability that the estimate will be as close to the parameter being estimated as is specified by the measure of precision. For example, a researcher may wish to ensure with probability 0.9 (or with 90% confidence) that a sample mean is within 30 units of the population mean.

One further definition is needed in discussing the determination of sample sizes. This relates to the precision of a confidence interval. As discussed earlier, the end points of a 100(1-α)% confidence interval are often approximated by the values

(Estimate) ± $z_{\alpha/2}$SE(Estimate).

The precision of the confidence interval is then d=±$z_{\alpha/2}$SE(Estimate) and the relative precision of the interval is r=±$z_{\alpha/2}$SE(Estimate)/Estimate. A typical question in this respect is then: how large a sample is needed to ensure that d is less than some chosen value d_0 or r is less than some chosen r_0?

This type of question can be answered approximately by setting d or r equal to the desired value and solving the resulting equation for the sample size n. Thus in order to obtain a 100(1-α)% confidence interval for the mean of $\bar{y}\pm d_0$ it is required from equation (4) that

$$z_{\alpha/2}\mathrm{SE}(\bar{y}) = d_0,$$

so that

$$z_{\alpha/2}(s/\sqrt{\ }n)\sqrt{\ }(1-n/N) = d_0\cdot$$

Solving this equation for n yields

$$n = z_{\alpha/2}{}^2 s^2/(d_0{}^2 + z_{\alpha/2}{}^2 s^2/N)\cdot \tag{5}$$

Because the standard deviation is not known in advance it is necessary to guess what this might be, for example using the value from an earlier sample. Notice that if the population size N is large then

$$n \approx (z_{\alpha/2}s/d_0)^2\cdot \tag{6}$$

In fact this is a conservative equation in the sense that, for all population sizes N, equation (6) gives a larger value of n than equation (5).

If it is a relative precision of r_0 that is needed for a confidence interval for the population mean then this requires that

$$z_{\alpha/2}\mathrm{SE}/\bar{y} = r_0,$$

so that

$$z_{\alpha/2}(s/\sqrt{}\,n)\sqrt{}\,(1-n/N)/\bar{y} = r_0.$$

Solving for n then gives

$$n = z_{\alpha/2}^{\;2}(s/\bar{y})^2/\left\{r_0^{\;2} + z_{\alpha/2}^{\;2}(s/\bar{y})^2/N\right\},$$

or

$$n = z_{\alpha/2}^{\;2}\hat{cv}(y)^2/\left\{r_0^{\;2} + z_{\alpha/2}^{\;2}\hat{cv}(y)^2/N\right\}. \qquad (7)$$

This equation can be applied with an estimated, or guessed, value for the coefficient of variation in the population in place of the unknown $\hat{cv}(y)$. For a large population size N it reduces to

$$n = \{z_{a/2}^{\;2}\hat{cv}(y)/r_0\}^2, \qquad (8)$$

which always gives a larger value than equation (7).

Suppose the survey in Paterson Inlet proposed Marine Reserve were a pilot study. The April 1998 data from within the Reserve site were used to estimate how large a sample should be taken to estimate the mean of the average cod length to within 10% of the true mean using a 90% confidence interval. This is often referred to as a margin of error of 0.1. Using equation 8 and note that the $\hat{cv}(y)$ = 22.85/295.19 = 0.08 the approximate sample size should be

n = {1.64^2 (0.8/0.10^2) = 20.82 or 21 sites.

Equation (5) to (7) should be regarded as giving no more than a rough indication of adequate sample sizes when they are used with guessed values for standard deviations or coefficients of variation. Nevertheless, as a general principle it should be remembered that any effort spent in determining appropriate sample sizes is better than no effort at all.

All of the discussion so far about sample sizes has been in terms of just one variable of interest in a study but in most studies several different variables have to be considered at the same time. If all variables require sample sizes of about the same magnitude then the size used can be the maximum required for any variable. However if this is not possible within the available resources then some lower size may need to be used, and some variables may have to be estimated with less precision than was originally desired.

2.6 Estimating Totals

In many situations estimates of the total of all values in a population, rather than the mean per population unit is required. For example, the total weight of new growth of all the plants in a region might be more important than the mean growth on individual plants. Similarly, the total amount of a pollutant may be more important than the average amount per sample unit.

The estimation of a population total is straightforward if the population size N is known and an estimate of the mean is available. The general equation is the mean per unit, multiplied by the number of units,

$$\hat{T} = N\bar{y} \qquad (9)$$

The sampling variance of an estimated total, $\hat{\text{var}}(\hat{T})$, is

$$\hat{\text{var}}(\hat{T}) = N^2\,\hat{\text{var}}(\bar{y}) \qquad (10)$$

and its standard error, $\mathrm{SE}(\hat{T})$ is

$$\mathrm{SE}(\hat{T}) = (N)(\mathrm{SE}(\bar{y})) \tag{11}$$

2.7 Stratified Sampling

A valid criticism of simple random sampling is that it leaves too much to chance. For example, consider a survey to estimate species diversity in Buller Gorge Scenic Reserve where the locations of 100 plots were randomly assigned over the study area. By chance all the plots could be located within the valleys and none were on the upper slopes. This survey may not be considered to be "representative" of the population. One way to overcome this problem while still keeping the advantages of random sampling is to use stratified random sampling. This involves dividing the units in the population into non-overlapping strata, and selecting an independent simple random sample from each of these strata. For the Buller Gorge survey the strata may be altitude bands, e.g. the valley floors, mid slopes and upper slopes could be considered strata.

In general there is nothing to loose by using stratified sampling over simple random sampling and there are some potential gains. Firstly, if the individuals within strata are more similar than individuals in general then the estimate of the overall population mean will have a smaller standard error than can be obtained with the same simple random sample size. This is because the estimate of the sampling variance for stratified sampling is the weighted-sum of the sampling variance within each stratum. If the units within a stratum are similar the sampling variance within the stratum will be small and the weighted sum will be at least as small as the variance obtained by pooling the sample data together. Secondly, there may be value in having separate estimates of population parameters for the different strata. Often interest is in not only the overall population mean, but in differences among the strata. Using the example of the Buller Gorge survey the overall species diversity can be estimated. With a stratified design the species diversity in each stratum, and the differences among strata, can be estimated. A third advantage is that stratification makes it possible to sample different parts of a population in different ways, which may make some cost savings possible. For practical reasons dividing the survey into separate strata can make logistical planning easier.

Example: Deer Pellet Survey in the Murchison Mountains Takahe Special Area (Appendix Data Set 9)

A survey to count the number of deer pellet groups in 2.5m radius plots was conducted in 1998. The survey design used 20 long lines along which there were circular plots at regular intervals. The lines ran at right angles to the river and crossed between two strata - the valley floor and the valley sides. The average number of pellet groups in plots along the 20 lines can be summarised by strata by using SPSS. The average number of pellet groups in plots was greater on the valley sides (mean = 0.0425) compared with the valley floor (mean = 0.0300).

TABLE 2.3 SUMMARY STATISTICS FOR THE AVERAGE NUMBER OF DEER PELLET GROUPS IN PLOTS IN THE VALLEY FLOOR AND VALLEY SIDES OF THE TAKAHE SPECIAL AREA, MURCHISON MOUNTAINS, 1998. THERE WERE 20 LINES WITH PLOTS SPACED AT REGULAR INTERVALS ALONG THE LINE.

STRATUM	N	MEAN	STD. ERROR
valley-floor	20	0.03000	0.01469
valley-sides	20	0.04252	0.01238

Generally, the types of stratification that should be considered are those based on: spatial location, i.e. areas within which the population is expected to be uniform, and, the size of sampling units. For example, to sample an animal population over a large area a map could be used to partition the area into a few apparently homogeneous strata based on factors such as altitude and vegetation type. In sampling insects on trees the population could be stratified on the basis of small, medium and large tree diameters. In sampling households a town could be divided into regions within which the age and class characteristics are relatively uniform. Usually the choice of how to stratify is just a question of common sense.

The question of how many strata should also be considered. The aim in stratified sampling is for the units within each stratum to be as similar as possible. At first it would seem to be sensible to create many small strata to ensure that the within-stratum variance is low. This is true to an extent, but if too many strata are created then each may have only a small sample size. The problem with small sample sizes is that the SE can be quite large when n is small (remember that $\text{SE} = s / \sqrt{n}$. Another problem with having too many strata and small within-stratum sample sizes is that there maybe too few sample units to allow for optimal allocation of effort among the strata. Allocation of sample effort among strata is discussed below but if stratum-sample sizes are too small it may be impossible to have sufficient “flexibility” in the allocation of sample units to achieve an optimal design. As a general rule of thumb the minimum number of sample units in a stratum is five. Be warned that this rule is a guide only for the minimum and should not be used as a rule to decide on the total sample size for the entire study area by multiplying five by the number of strata!

Another useful rule is that the more information you have about the population the more strata you can create, or more realistically, the less information, the fewer the strata. If you have very little information you should consider simple random sampling (i.e. one stratum). Dividing the population into strata should be based on sound environmental reasoning, and not on the will to make a survey “appear” to be sophisticated. Having said that a useful rule of thumb is five units per stratum there are examples of well designed surveys with only two units per stratum, e.g. the US EPA drinking well survey in 1990. In this example there was a lot of information on spatial variation in water quality to create a design with many strata. A sample size of two within a stratum is definitely the minimum stratum-sample size because with any fewer it is not possible to estimate stratum variances.

Assume that K strata have been chosen, with the i^{th} of these having size N_i and the total population size being $\Sigma N_i = N$. If a random sample with size n_i is taken from the i^{th} stratum the sample mean $\bar{y}_i$ will be an unbiased estimate of the true stratum mean μ_i with estimated variance

$$\hat{\text{var}}(\bar{y}_1) = \frac{s_i^2}{n_i}\left(1 - \frac{n_i}{N_i}\right)$$

where s_i is the sample standard deviation for the stratum. These results follow by simply applying the results discussed above for simple random sampling to the i^{th} stratum only.

An estimate of the overall population mean is the weighted average of the stratum sample means,

$$\bar{y}_s = \sum_{i=1}^{K} W_i \bar{y}_i$$

where $W_i = N_i / N$. The estimated variance is

$$\hat{\text{var}}(\bar{y}_s) = \sum_{i=1}^{K} W_i^2 \, \hat{\text{var}}(\bar{y}_i) \qquad (13)$$

The estimated standard error, $\text{SE}(\bar{y}_s)$ is the square root of the estimated variance, and an approximate confidence interval for the population mean is given by

$$y_s \pm t_{\alpha/2,n-1}\text{SE}(\bar{y}_s) \qquad (14)$$

The estimate of a total with stratified sampling is $\hat{T}_s = N\bar{y}_s$. Estimates of the variance and standard error for a total when stratified sampling is used are $\hat{\text{var}}(\hat{T}_s) = N^2 \, \hat{\text{var}}(y_s), \text{SE}(\hat{T}_s) = (N)(\text{SE}(\bar{y}_s))$.

2.8 Allocation of Sample Units to Strata

The simplest way to allocate survey effort among strata is to make the sample sizes for the different strata to be proportional to the strata sizes:

$$n_i / n = N_i / N. \qquad (15)$$

This is called stratification with proportional allocation. The samples are self-weighting in that the chance of an individual unit being selected is exactly the same as in simple random sampling. However, the chance of getting a “bad sample” is less. For example, consider a population with 2400 male fish and 1600 female fish that was stratified on gender. If $n = 400$ then with stratified sampling with proportional allocation 240 males and 160 females will be selected. The disadvantage of simple random sampling is that by chance, all 400 sample units could be male fish.

Although proportional allocation is often used because it is convenient, it is not necessarily the most efficient use of resources. Another method to allocate effort among strata is to consider the costs of sampling. Sampling costs are usually considered to have two components - fixed and variable costs. Fixed costs include things like travel time to the site, set up costs, equipment use etc. Variable costs are the costs of sampling units within each stratum and include travel time within the site to get to the stratum, the cost of measuring each sample unit etc. A general cost equation that can be used is,

$$\text{Total Cost } = F + \Sigma c_i n_i ,$$

where c_i is the variable cost of sampling one unit from stratum *i.*

If the costs of sampling a unit within a stratum vary and there are differences in how variable the strata are, optimal allocation of survey effort will be when:

$$\frac{n_i}{n} \equiv \frac{W_i s_i / \sqrt{c_i}}{\Sigma(W_i s_i / \sqrt{c_i})} \qquad (16)$$

In general stratum sample sizes will be larger if:

(i) the stratum is large

(ii) it is more variable

(iii) sampling is cheaper.

If the cost/sample unit is the same for all strata than it is known as “Neyman Allocation” and $\frac{n_i}{n} \equiv \frac{N_i s_i}{\Sigma(N_i s_i)}$.

The Neyman Allocation is a special kind of optimal allocation.

2.9 Post-Stratification

With some populations that are suitable for stratification it is difficult to know the stratum to which units belong until a survey has been conducted, although the strata sizes are known accurately for the population. For example, the stratification of sample quadrats on the basis of habitat type may not be possible until the quadrats have been visited. Information collected during the sample can be used to post-stratify where a simple random sample of *n* is taken from the population and the sampled units are classified into *K* strata. Then the usual stratified sampling estimator can be used to estimate the population mean. Even if the stratum sizes are not known the final post-stratified sample should be like stratified sampling with proportional allocation if the sample is large enough.

Post-stratification should not be used as a technique to use to reduce sample variance by creating artificial strata. The strata should be based on sound environmental reasons and not formed just to group similar sample units together.

2.10 Systematic Sampling

Systematic sampling is a very useful design for environmental assessment. Systematic sampling can be carried out whenever a population can be listed in order or it covers a well-defined spatial area. In the former case, every k^{th} item in the list can be sampled, starting at an item chosen at random from the first k. In the second case sampling points can be set out on a grid at equally spaced intervals.

There are two reasons why systematic sampling is sometimes used in preference to random sampling. First, systematic sampling is often easier to carry out than random sampling. Second, it seems likely that a systematic sample will be more 'representative' than a random sample, and hence more precise, because it gives uniform coverage of the whole of the population of interest.

Systematic sampling suffers from the disadvantage of not allowing any simple determination of the level of sampling errors unless it is assumed that the items in the population are in a more or less random order. If that is the case then a systematic sample can be treated as being effectively a simple random sample and the various results given earlier for this type of sampling can all be used.

A common criticism of systematic sampling is that if the population has some periodic trend that matches the spacing between sample units in the systematic sample then using the formulae for simple random sampling will underestimate the sampling variance, i.e. the sample will appear more precise that it really it. What critics often fail to mention is that the opposite is true when the units within the systematic sample are highly variable and more variable than the population. A systematic sample may be more precise than it appears, in other words, the estimate of precision is conservative. This will occur when there is a gradient or trend, or large areas with population units with similar values (EPA 1989). As an example if there were a sequential trend up a mountain slope of reducing tree heights a systematic sample up the slope at every 100m in altitude would produce a sample with highly variable sample units. The sample variance from using the simple random sampling formulae would be an overestimate. This concept is easiest understood by thinking about what the sample means would be if there were many systematic samples taken (using 100m spacing). While each sample would have a lot of variation among units within the sample and a large standard deviation, the sample means would all be quite similar. When all the potential sample means are similar the survey has high precision (or low sampling variance). In this example the true sampling variance would be low, but would appear to be high because the estimator is based on the sample standard deviation.

If there are concerns with estimating the variance from systematic sampling there are special analysis methods that can be used. One of these methods is to aggregate the adjacent sample points into groups and then treat each group as a stratum in a stratified design (Yates 1961). The groups should be created on geographic boundaries not on the basis of their sample value. Another method is to use estimate precision from the pairs of adjacent sample points, i.e. the standard error is calculated from the difference between the 1st and 2nd sample points, the 2nd and 3rd sample points and so on. This approach of treating the sample points as a linear sequence is referred to as a serpentine pattern (EPA 1989).

There are many modifications to systematic sampling that can be used. One is called the interpenetrating systematic sample. Despite its name it is quite a simple design and is an effective way of dealing with a periodic trend although estimation for this survey design is slightly more complex. In this design rather than taking say, a 1 in 20 survey where every 20th item is selected, five 1 in 100 surveys are taken. Five random start points are chosen between 1 and 100 and every 100th item selected beyond each start point.

Another design, two-dimensional systematic sampling, has one systematic sampling at intervals along one axis and another at intervals along the other axis that is at right-angles to the first. With both these designs data are best analysed as a cluster sample where each of the five systematic surveys is considered a cluster (see some of the texts listed below for details on cluster sampling).

It is worth revisiting the concept of the systematic interval matching an underlying trend or spatial pattern because in environmental studies this trend or pattern may be one of the variables of interest. For example, in wildlife studies gaining information on spatial pattern, such as size of home-ranges, can be one of the purposes of the survey. A systematic survey can be an appropriate design with regular and close spacing between sample units. One way to use survey data would be to compare the correlation between sample units at a range of spatial scales. The change in correlation among spatial scales can provide information on spatial patchiness. However, if the sole purpose of the survey is to estimate population means the systematic interval needs to be carefully chosen so not to match environmental patterns.

2.11 Composite Sampling

Composite sampling is a technique where multiple samples collected in the field are combined into a new sample. The new sample is then mixed and either all, or part, is analysed. This approach has application in soil sampling where the cost of collecting samples in the field is low relative to the cost of analysing the samples in the laboratory. The composite sample will give estimates of the overall population average but because the individual samples are "lost" will not give information on the smaller scale variation among the original sample points. The estimation for composite samples can be complex. Procedures are reviewed by Gilbert (1987) and Patil (1995a).

2.12 Rank-set Sampling

Rank-set sampling is not that commonly used in environmental science which is surprising given the potential gains in precision. In rank set sampling the units are grouped into sets of m units based on e.g. their spatial location. Typically sets are small, around three or four units. Within the first set the units are ranked by a quick estimate and the unit highest value is measured. Within the second set the unit with the second highest value is measured and so on up to the m^{th} set. The unit with the lowest value is measured in the m^{th} set. The sequence is then repeated r times to give a total sample size of $n = rm$.

The idea behind rank-set sampling is to use expert knowledge about the environment. As an example consider a survey for possum browse on trees in a forest. A sample of 20 trees is needed. Trees are grouped into sets of four. In the first step four sets of four trees are randomly selected. The four trees in the first set are quickly inspected and the one with the highest level of browse is selected and more detailed browse measurements are recorded. Then, within the next set of four trees the tree with the second highest quick-assessment score is measured, and so on up to the fourth set of four trees where the tree with the lowest quick-assessment score is measured. In the next step another four sets of four trees are randomly selected and the 16 trees (four groups of four trees) are surveyed by the same method. This process is repeated in total five times to give a sample size, n = 20. This design will give a more precise estimate than if 20 trees were randomly selected. If the ordering of trees within each group were correct then efficiencies up to 300% improvement over simple random sampling can be expected. If the ordering were entirely wrong, i.e. the quick-assessment score were nothing more than a random guess than the sample precision will be what would be expected from a random sample (Barnett and Moore 1997). A rank set sample can be analysed as a simple random sample but there are also more complicated estimators that can be used (Patil 1995b).

2.13 What Sample Unit to Use

One of the decisions in planning a survey is choosing the sample unit. What is the appropriate sample unit is usually defined by the population, that is the physical characteristics of the habitat and the type of organisms (Resh 1979). For example, in surveys of low growing weeds a square plot (a quadrat) is often used. In freshwater fisheries surveys where electric fishing is used the sample unit is a site, measured in metres. Air samples may be units that are a volume of air collected over a time period. In marine fisheries surveys trawls may be the sample unit.

The actual sampling device often determines the size of the sample unit, for example, a trawl net is a fixed size and although the length of the tow may vary, smaller, or larger nets may not be feasible. However, plots, are not a fixed physical unit and can be e.g. 1m^2, or 0.25m^2, or 10m^2 etc. depending on the population of interest. One rule of thumb for plot size is that the plot should be 20 times the size of the individual in the population (Green 1979).

In general, the larger the sample size the better. For surveys that use plots many, small plots rather than a few large, plots is recommended. However there must be a balance between the size of the sample and the size of the unit within the sample. For example, if the plot is so small that it is the same size as the individuals in the population the sample will be highly variable because it will consist of either plots of zero counts and plots with counts of one. At the other extreme, if the plots are very large the variability among sample units will be low, but there will be few plots.

Often the best solution to what size sample unit to use is to conduct a pilot survey with various sample unit sizes to give information on the precision and total sample cost of each (Green 1979). Other methods are to use a nested design in the pilot survey where many small units are used. Estimates of precision and cost are calculated using the smallest sample unit size. Then, adjacent units are combined to give an effective sample unit size that is twice as large and new estimates of precision and cost are calculated. These units can be combined again, and so on, giving a range of successively larger sample units and estimates of precision and cost can be compared between the various sizes.

Another general rule for sample unit selection is that the size of the unit should not match the scale of any patchiness in the environment. For example, if the plots used to sample the low growing weed were 1m^2 and the weeds occurred in patches that were about 1m^2 in size then some of the plots would have very high counts (when the plot was located entirely within a patch), and others would have very low counts (when the plot was located entirely outside a patch). The sample would have high variance, and low precision. The plot should either be very much larger than the scale of patchiness, or very much smaller.

Example: Deer Pellet Survey in the Murchison Mountains Takahe Special Area (Appendix Data Set 9)

Consider the survey in the Special Takahe Area in the Murchison Mountains in 1998 described above. Data on the presence or absence of groups of deer pellets were also collected from within the Chester Burn catchment from 493 plots. Two plot sizes, 1.14m and 2.5m radius were set up from the same plot centre. In total there were 493 plot centres. Of these 12 of the smaller 1.14m radius plots had groups of deer pellets present but 30 of the 2.5 radius plots had groups of pellets present. Although the larger plots were 4.8 times the size of the smaller plots there were only 2.5 times as many large plots with pellets. This difference may be due to patchiness in pellet distribution. If groups of pellet tended to occur in patches, or aggregates, then when there was one group of pellets it was likely that there would be a second group near by. It appears from the data that the scale of this clustering of pellet groups was at around 2 to 4m. A cluster of pellet groups could contain only one large plot but there could be many small plots.

One other consideration in using plots is their shape. A long, thin rectangular plot is an efficient shape because the plot "spreads" across more of the study area. This has the effect of minimising correlation between individuals within the plot and therefore the plot is more informative. A circular plot has less spatial "spread" and higher correlation within the plot. The advantage of a circular plot is that it will have less edge than a long, thin rectangular plot. The problem with a plot shape with a lot of edge is there is more chance of mistakenly recording an individual is "in" the plot when in fact it is "out" and vice versa.

2.14 Errors in Sample Surveys

In general there are four sources of error or variation in scientific studies (Cochran, 1977):

a) There are sampling errors due to the variability between experimental units and the random selection of units included in a sample. Different random samples will generally produce different estimates of population parameters. This variation reflects the sampling errors.

b) There may be measurement errors due to the lack of uniformity in the manner in which a study is conducted. The measurement procedure may be biased, imprecise or both biased and imprecise. This type of error results solely from the manner in which the observations are made. For example, fisherman may report incorrect lengths and weights of fish caught, human subjects may lie about their age or weight, etc.

c) There may be missing data due to the failure to measure some units in the sample.

d) Gross errors may be introduced in coding, tabulating, typing and editing data.

An understanding of sampling errors and their effects is the basis of statistical inference procedures. The control of sampling errors is therefore primarily the responsibility of the statistician. Random measurement errors can be modelled but their control and reduction must come from careful experimental design. In fact, in many fields of study the presence of measurement error is barely recognised and its influence is played down. Many statisticians follow the rule of thumb that the measurement error should be small relative to the sampling error, especially in utilising statistical procedures such as regression and correlation analysis. Certainly for many studies conducted in ecology measurement errors cannot be ignored and standard analysis procedures such as regression analysis may not be applicable until this source of error is under control.

2.15 Key Points in This Module

- Environmental data is typically only a sample from the population of interest.
- The goal of sampling is to summarise the characteristics of the entire population.
- The measures that are used to summarise a population are called population parameters and the corresponding sample values are called statistics.
- Whenever inferences are to made about population parameters on the basis of the sample result the sample design must have some element of random selection if statistical sampling theory is to be used.
- One of the typical features of environmental data is high variability.
- Estimates of population parameters should always be quoted with their associated level of precision, and usually this is a confidence interval.
- An understanding of sampling errors and their effects is the basis of statistical inference procedures.

2.16 Questions About This Module

After completing this module you should be able to give reasonable answers to the following questions.

1. In a study in Otago of the fish diet the content of fish guts were analysed. Data were collected on the prey species: species name, number of each individual prey species and the head-length of each prey item. At least ten fish from seven streams were sampled. What is the population of interest? What is the sample population? What population parameters can be estimated?
2. Using the example above for the surveys of cod in the potential Patterson Inlet Marine Reserve, what is your estimate of the sample size if estimate of the mean of the average cod length was required to be within 20% of the true mean using a 90% confidence interval? What is your estimate of the

sample size if the *cv* was twice as large, i.e. *cv* = 0.16? What is the affect of variation and desired precision on the required sample size? Which has more influence?

3. A possum control operation was estimated to achieve a residual trap catch (*rtc*) of 10% with a 90% confidence interval from 2% and 18%. The target *rtc* is 5%. The survey for estimating the *rtc* used data on the proportion of traps that caught a possum after three nights of trapping. The sample size was five, i.e. there were five lines of traps. One explanation for a *rtc* of 10% is that the control operation failed to reduce the possum population to the desired low level. What are some other explanations - think about the sources of error.

4. An estimate is required of the number of a pingao on a section of foreshore. The beach foreshore area is some 8km long and defined as being 1km wide. Twenty randomly located plots, 50mx50m (= 0.0025km^2) in size were used. The pingao habitat is variable and there are sections where the plant numbers will be low and other sections where the plant numbers will be high. The foreshore is stratified into three strata on the basis of habitat types: north, middle, and south. The number of sections surveyed within each stratum is proportional to the size of the stratum. The northern stratum is 2km long, and the middle is 3.5km and the southern 2.5km long. Use proportional allocation to decide on how to allocate the 20 sample plots among the three strata.

   In a pilot study of the pingao area the survey results for the three strata were:

   $\bar{y}_1$ = 17.2 plants, s_1 = 1.2

   $\bar{y}_2$ = 20.5 plants, s_2 = 6.3

   $\bar{y}_3$ = 12.1 plants, s_3 = 2.4

   where $\bar{y}_1$ and s_1 is the sample mean and standard deviation for the north stratum.

   How would you allocate the 20 sample plots using optimal allocation (assume the costs of sampling each stratum are equal).

5. What are the advantages of using stratified sampling over simple random sampling?

REFERENCES

Cochran, W.G. (1977) *Sampling Techniques,* 3rd edition. Wiley

Environmental Protection Agency (1989) Methods for Evaluating the Attainment of Cleanup Standard, vol. 1. Soils and Soil Media. Statistical Policy Branch (PM-223), Office of Policy, Planning and Evaluation, US Environmental Protection Agency, 401 M Street, S.W., Washington, DC 20460.

Gilbert, R.O. (1987) *Statistical Methods for Environmental Pollution Monitoring.* van Nostrand Reinhold Company, New York.

Green, R.H. (1979) *Sampling Design and Statistical Methods for Environmental Biologists,* Wiley.

Link, W.A., Barker, R.J., and Sauer, J.R. (1994) Within-site variability in surveys of wildlife populations. Ecology 75:1097-1108.

Patil, G.P. (1995a) Editorial: composite sampling. *Environmental and Ecological Statistics* 2:169-79.

Patil, G.P. (1995b) Editorial: ranked set sampling. *Environmental and Ecological Statistics* 2:271-85.

Resh, V.H. (1979) Sampling variability, life history features, and the experimental design of aquatic insect studies. *Journal of the Fisheries Research Board of Canada* 36:290-311.

Yates, F. (1981) *Sampling Methods for Censuses and Surveys,* 4th ed., Oxford University Press.

Some other useful references are:

Schaeffer, R.L., Mendenhall, W., and Ott, L. (1986) *Elementary Survey Sampling,* 3 rd ed., Duxbury.

Lohr, S.L. (1999) *Sampling: Design and Analysis,* Duxbury.

Thompson, S.K. (1992) *Sampling,* Wiley.

Contents

MODULE 3: DESIGNS FOR MONITORING SCHEMES

Module 3: Designs for Monitoring Schemes

SUMMARY

The objective of many monitoring schemes is to detect the effect of human impacts. What sites to monitor is usually a question of balancing the use of the same sites over and over (which is good for detecting trends) or bringing in new sites (which is better for estimating averages). Augmented designs are a way of combining these two approaches. One of the key considerations in designing a monitoring scheme is whether there will be sufficient statistical power to detect trends in the population parameters of interest. Several factors affect power: sample size, variability of the samples, and magnitude of the difference or trend to be detected. Strategies to improve power include reducing variation among and within sites, and matching treatment sites to control sites. Pseudoreplication is an example of confounding where inferences to the wider population are made from the results of a study where there is either no replication, or replication is at the wrong scale.

3.1 Introduction

The purpose of many monitoring schemes is to detect the effect of a human impact on natural populations (Underwood 1992). The appropriate analysis therefore is to measure the change in an associated environmental variable. However, natural populations display considerable temporal and spatial variation so the sampling design and analysis must be capable of distinguishing between what is normal variation and variation that may be attributed to the human impact (Skalski and McKenzie 1982, Underwood 1994).

3.2 Choice of Monitoring Sites

There are a number of different decisions to make in designing a monitoring scheme including what sites to select for the surveys and to analyse the data. The scheme chosen should relate directly to the specific objectives of monitoring. For example, in designing a monitoring scheme for Karner blue butterflies in Wisconsin, USA, the two objectives were to monitor: i) the overall effectiveness of the conservation efforts by assessing regional trends; and ii) the effectiveness of individual conservation strategies to allow comparison among strategies (Brown and Boyce 1996).

The first objective required data collected over time from a sample of sites that were representative of the Wisconsin State. The purpose of statistical analysis was to separate natural variation from variation and trends resulting from the conservation efforts. The second objective required data to be collected from the areas managed under the conservation strategy of interest and for this to be compared with data from control sites that are governed by natural ecological processes (Eberhardt 1976, Green 1979 p.29).

The special requirements of environmental monitoring schemes has led to interest recently in more complicated designs that include aspects of random sampling, good spatial cover, and the gradual replacement of sampling sites over time (Skalski, 1990; Stevens and Olsen, 1991; Overton *et al.* 1991, Urquhart *et al.* 1993). Monitoring designs that are optimum in some sense have also attracted interest in recent years (Fedorov and Mueller, 1989; Caselton *et al.* 1992).

3.3 Temporal Replication of Sites - Monitoring over Time

Environmental monitoring typically requires a number of years of sampling to be able to detect real biological trend (Barker and Sauer 1992). There are four general sources of pattern in population data:

i) trend resulting from a population change, i.e. the population trend that we are wanting to detect in monitoring;

ii) irregular environmental perturbations e.g. unusual weather events;

iii) autocorrelation due to population processes, i.e. the population size in one year is expected to be related to the population size in the previous year; and

iv) stochasticity associated with sampling.

If too few years of data are collected it can be difficult to separate population trends from these other sources of underlying environmental stochasticity.

There are many analysis techniques for trend detection, some of which are discussed in later modules. The appropriate analysis technique should be chosen prior to sampling to ensure that sufficient data is collected, and that the sampling design meets the assumptions of the analysis technique. For example, many techniques assume the samples are independent. Other techniques assume the sites are independent but require repeated samples to be taken from the same site over time.

3.4 Spatial Replication of Sites - Purposely Chosen or Randomly Chosen Monitoring Sites

For practical reasons often long-term monitoring sites are not randomly chosen. For example, the nine sites of the United Kingdom Environmental Change Network (ECN) were chosen on the basis of:

i) good geographical distribution covering a wide range of environmental conditions and the principal natural and managed ecosystems

ii) some guarantee of long-term physical and financial security

iii) a known history of consistent management reliable and accessible records of past data, preferably for ten or more years; and,

iv) sufficient size to allow the opportunity for further experiments and observations.

The interest in the ECN is in monitoring the change in these sites and therefore it does not matter that the sites were not initially all similar in their status. The ECN is attempting to relate the differences in the change in sites to measured meteorological and geographical differences.

The alternative design to purposely choosing sites is to randomly select sites. The potential problem with purposely chosen sites is that they may not be as "representative" of the population as thought. In some situations it may not be possible to purposely choose sites because there is insufficient knowledge about which sites to chose. A random selection of sites ensures there is no bias in the estimation of population parameters. This attribute of random selection was discussed in a previous module.

Even with randomly selected sites there is still the question of what to monitor over time - do you measure the same sites at each time period, or randomly reselect sites at time period? The answer depends on the sampling objective.

- If the objective is to estimate the mean value following the most recent survey, e.g. environmental status, then it is best to reselect a fresh sample (i.e. new sample locations).
- If the objective is to estimate the change in population means, i.e. trends in environmental status it is best to use the same sites for each survey (Skalski 1990).

With the former case, by reselecting the sites each year the population parameter will not be consistently over- or under-estimated. With the later case, resampling the same site each year will eliminate random variation among sites that could confound the survey results.

3.5 Some Special Designs for Choosing Monitoring Sites - Augmented Rotating Panel Design

Monitoring can often have both objectives described above - to detect status as well as to detect trends. Skalski (1990) suggested a rotating panel design with augmentation for long-term monitoring. This design combines both ideas where some sites are sampled every year and others are rotated.

The design takes the form shown in Table 3.1 if there are eight sites that are visited every year and four sets of ten sites that are rotated. Site set 7, for example, consists of ten sites that are visited in years 4 to 6 of the study. The number of sites in different sets is arbitrary. Preferably, the sites will be randomly chosen from an appropriate population of sites.

This design has some appealing properties: the sites that are always measured can be used to detect long-term trends but the rotation of blocks of ten sites ensures that the study is not too dependent on an initial choice of sites that may be unusual in some respects. However, Urquart *et al.* (1993) have provided evidence that the serially alternating design that is discussed next is more efficient because more sites are measured in the first few years of the study.

The practical reasoning behind the rotating panel design is very sensible. When there is limited information on an environmental impact it is difficult to know where and when to monitor. Without knowing the extent of the area effected by e.g. a coastal sewer out-fall the monitoring design should include many sites that are situated from the out-fall in all directions (land and sea) and for a good distance. Such large spatial coverage is often beyond the scope of most budgets. This design provides a method to “add in” new sites and allows for the dynamic nature of populations (human and non-human). For example, monitoring of an out-fall in Akaroa Harbour may focus on the areas where people live. Over time the new housing developments may mean that new monitoring sites should be added into the survey design.

TABLE 3.1 ROTATING PANEL DESIGN WITH AUGMENTATION. IN THIS EXAMPLE, EVERY YEAR 48 SITES ARE VISITED. OF THESE, 8 ARE ALWAYS THE SAME AND THE OTHER 40 SITES ARE IN FOUR BLOCK OF SIZE TEN, SUCH THAT EACH BLOCK OF TEN REMAINS IN THE SAMPLE FOR FOUR YEARS AFTER THE INITIAL START UP PERIOD.

SITE SET	NUMBER					YEARS							
	OF SITES	1	2	3	4	5	6	7	8	9	10	11	12
repeated	8	x	x	x	x	x	x	x	x	x	x	x	x
1	10	x											
2	10	x	x										
3	10	x	x	x									
4	10	x	x	x	x								
5	10		x	x	x	x							
6	10			x	x	x	x						
7	10												
..													
..													
14	10											x	x
15	10												x

3.6 Some Special Designs for Choosing Monitoring Sites - Serially Alternating, Augmented Rotating Panel Design

The name of this design says it all! This design is similar to the design discussed above but, using the example in Table 3.1, rather than surveying 30 out of 40 sites in the rotating panel the next year, each year a rotating selection of 40 sites are surveyed (Table 3.2). The US Environmental Protection Agency Environmental Monitoring and Assessment Program (EMAP) was based on this design.

TABLE 3.2 SERIALLY ALTERNATING ROTATING PANEL DESIGN WITH AUGMENTATION. IN THIS EXAMPLE, EVERY YEAR 48 SITES ARE VISITED. OF THESE, EIGHT ARE ALWAYS THE SAME AND THE OTHER 40 SITES ARE FROM A ROTATING PANEL

SITE SET	NUMBER					YEARS							
	OF SITES	1	2	3	4	5	6	7	8	9	10	11	12
repeated	8	x	x	x	x	x	x	x	x	x	x	x	x
1	40	x				x				x			
2	40		x				x				x		
3	40			x				x				x	
4	40				x				x				x

The advantage of this serially alternating design is that more sites are visited. Compare Table 3.1 and 3.2. At the end of the first year 48 sites will have been surveyed. By the end of year 2 with the serially alternating design 88 sites will have been visited, but with the design shown in Table 3.1 only 58 sites will have been visited. By the end of year 3, 128 will have been visited with the serially alternating design and only 68 with the other design, and so on. The point is that with the serially alternating design more sites are visited, and more information is collected from among the sites. This is discussed further in the next section on power, but generally it is better to collect information from as many sites as possible. With both designs eventually all 168 sites will be surveyed. With the serially alternating design this will take 4 years, but with the other design it will take 13 years.

3.7 Some Special Designs for Choosing Monitoring Sites - Optimal Designs

The concept of an optimal design has developed largely from the field of spatial statistics. The definition of optimal is based on the idea of minimising variances of prediction models. One example of a prediction model is the prediction of SO_2 levels in the eastern USA. There are 34 long-term monitoring sites in the area extending from Minnesota though to the east coast. Data from the weekly SO_2 measurements is used to predict pollution over the entire study area. The optimal design is one that does "best" on average, i.e. minimises the prediction error. Another criterion is the optimal design is one that has the smallest worst prediction error, i.e. it minimises the worst case. A third criterion is based on the concept of entropy (Cox *et al* 1997).

Designs for geostatistical analyses are usually most efficient if they are a regular grid of sampling sites, when efficiency is measured by the average, or maximum prediction variance (Cressie 1991). Regular, or geometric designs of triangular, square of hexagonal grids are often used. The EMAP design was based on a triangular grid, which was randomly located over the USA, landmass. The distance between grid points was 27km.

3.8 Statistical Power

One of the major considerations in designing a monitoring scheme is whether you will be able to detect a true change, or trend, in the population parameter of interest (Taylor and Gerrodette, 1993, Fairweather 1991). This ability to detect a trend, e.g. the effect of human-induced change, which occurs over and above the amount of variation that natural populations exhibit is referred to as power.

When planning a monitoring study the number of samples, the likely effect that can be detected, and the number of years required to be able to detect a trend are all considerations that relate to power. Calculating the power of a trend survey can be difficult because it requires estimates of variance and until monitoring is undertaken there may not be any estimates of the variances (Gerrodette 1987, Steidl et al. 1997). However, approximations of likely power can be made by using data from other studies and from pilot studies.

In statistical terms the power of a test is the ability to reject the null hypothesis when it is false, that is, it is the probability of correctly rejecting the null hypothesis. In trend detection power refers to the ability to detect a true increasing or decreasing trend. Several factors affect power, such as sample size, variability of the samples, and magnitude of the difference or trend to be detected. Designs with small sample sizes and high variability will have low power. If the size of the difference or trend is small compared with the natural population variability it will be difficult to detect any effect.

Example: What is the Power to Detect an Increase in the Blue Cod from Paterson Inlet, Stewart Island? (Appendix Data Set 5)

Using the example discussed in module 2 the average head length of blue cod inside the proposed Marine Reserve at Paterson Inlet was estimated to be 295.1884 with a standard deviation of 22.8512 from a sample of size 6 in 1998. Remember the sample units were the survey sites. What is the power to detect an increase in average head length of 10mm to 305.1884 using a 5% significance level and with a sample size of 6? The calculations are as follows.

1. The null and alternative hypotheses are

   H_o: μ = 295.1884

   H_a: μ > 295.1884

2. H_o is rejected if the t statistic exceeds the upper 5% of t_5 = 2.015 (there are 5 degrees of freedom). The t statistic is:

$$t = \frac{\bar{y} - \mu}{s / \sqrt{n}}$$

$$= \frac{\bar{y} - 295.1884}{s / \sqrt{6}}$$

   Therefore, when $t = \dfrac{\bar{y} - 295.1884}{s / \sqrt{6}} \geq 2.105$ H_o will be rejected.

3. The choice of the s can be difficult when there is no data collected yet on the future population. The best estimate we have is the estimate from the 1998 survey, s = 22.8512. Using this value and rearranging the above equation we can calculate what size the sample mean needs to be for H_o to be rejected,

$$\bar{y} \geq \left[2.015 \left(\frac{22.8512}{\sqrt{6}} \right) \right] + 295.1884$$

$$\geq 313.9863$$

4. The power is the probability that $\bar{y} \geq 313.9863$ when μ = 305.1884. Using σ = 22.8512,

$$P\left(\bar{y} \geq 313.9863 \,\middle|\, \mu = 305.1884 = P\left[z \geq \frac{313.9863 - 305.1884}{22.8512 / \sqrt{6}} \right]\right.$$

This probability can be calculated by solving the right hand-side of the equation,

$P(\geq 0.9431)$
=1-0.8264
=0.1736

In this example the power is very low, there is only 17.36% chance that if the population increased the average head length to 305.1884 mm that this would be detected in a sample of six sites.

If the sample size were increased to 20 sites then the power would improve. Reworking the above equations the power is now 0.5902. In fact these equations can be easily done on a spreadsheet such as EXCEL and the values changed to explore the effect on power of larger effects (here the effect is 305.1884), smaller standard deviations, and larger sample sizes (Figure 3.1). Power quickly improves by sampling more sites. Power can also be improved by reducing the sample standard deviation. And finally, larger effects, that is, larger increases in head length can be more easily detected. This is discussed more in the next section.

This example shows one way to estimate power. For more complicated designs and analysis, e.g. estimating the power of a multi-factor experiment there are other techniques that can be used. For example, one approach is to simulate likely data sets from a statistical distribution on a computer. Such techniques are beyond the scope of this workbook. Special software is available for estimating power but all require a good understanding of the principles of power analysis to be used properly.

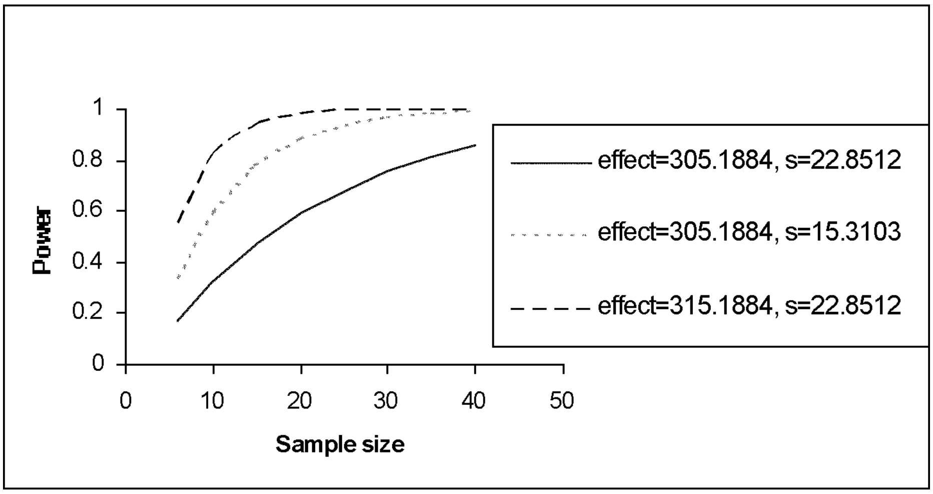


Figure 3.1 Estimates of power with varying sample sizes for detecting an increase in the average head length of cod from the 1998 estimate of 295.1884cm to 305.1884 and 315.1884cm (effect = size of increased head length). Also shown is the effect on power of the standard deviation, s = 22.8512 and s = 15.3103. The sample unit is the survey site.

3.9 How to Improve Power

Variation in data collected from monitoring studies is due to:

i) within-site variation which reflects the inexactness of the data collection,

ii) the variation among sites due to the environmental heterogeneity, and

iii) temporal variation.

(Millard and Lettenmaier 1986, Gerrodette 1987, Link *et al.* 1994). The power of monitoring, for example, the ability to detect if there is a true difference between a treatment and a non-treatment site, or to detect a regional-population trend, will improve if these sources of variation are reduced.

For trend detection sampling more units is generally preferable to increasing sampling effort within a unit (Millard and Lettenmaier 1986, Link *et al.* 1994, Brown and Miller 1998). Millard and Lettenmaier (1986) found that in their study to maximise power the optimal design was a spatially extensive one with many sampling units. With a design with many sample sites the among-site variation is reduced.

Strategies to reduce within-site variation (or measurement error) are to have strict guidelines of when and how sampling should be undertaken and to sample a site more than once. Modelling the environmental factors that effect the observed sample values can also reduce within-site variation. For example, consider the example described above for monitoring Karner blue butterflies in Wisconsin. Butterflies are less mobile and less detectable on cool days compared with warm days. Counts of butterflies seen during surveys (i.e. the observed sample data) on cool days can be inflated to adjust for differences in daily temperature. In the above example for blue cod sampling more fish within each survey site can decrease the within-site variation.

So far we have been discussing designs for long-term monitoring to detect trends in population status. Monitoring is also undertaken to detect a possible change following some specific management action, e.g. to detect if a site has been cleaned up after a remedial action, or to detect whether a rat population has been reduced after a rat-poisoning operation. In these situations one way to improve power is to have “treatment” and “non-treatment” sites. This is addressed in more detail in later modules but the discussion of power is relevant here.

With “treatment” and “non-treatment” sites the differences between the sites can be compared over time. The power to detect the difference between trends in treatment and non-treatment sites will generally be higher than the power to detect the individual trend at either site. If the variation among time intervals for the treatment sites was identical to the variation for the control sites, by using the differences between the two, this source of variation would be eliminated (Stewart-Oaten *et al.* 1986). Even if the correlation is not perfect, Stewart-Oaten argue that the variation in the differences over time would be small compared to other sources of variation, particularly from sampling error.

3.10 Pseudoreplication

Pseudoreplication is defined as:

"*the use of inferential statistics to test for treatment effects with data from experiments where either treatments are not replicated or replicates are not statistically independent.*" (Hurlbert 1984).

This concept was introduced in module 1. Continuing with the example from module 1, to assess diversity in grassland following burning a 1ha study area was randomly located within a burnt and unburnt field. Within each of the 1ha areas 15 - 1m^2 quadrats were randomly located. Does this design give 15 independent replicate samples from the burnt and unburnt area? The answer is no if the question is about the general effect of burning. The problem is that the "replicate" quadrats are not replicated at the correct scale. The idea of replication is to provide a measure of the intrinsic variability of the area that has nothing to do with the treatment (in this example, the treatment is burning) (Underwood 1997). The observed differences in the burnt and unburnt quadrats could be due site differences and not the effect of burning. The observed differences are confounded by other differences. Underwood, in fact, uses the term confounding in preference to the term pseudoreplication because it draws attention to what is needed;

"*It is not replication as such that is the problem. The difficulty is to separate out the differences among treatments that are due to the experimental factor from any differences due to other factors. The logic of the experiment requires this, so that any differences found can be unambiguously attributed to the process proposed in the model. Other differences confound such conclusions, making logical interpretation impossible.*" (Underwood 1997).

As another example, in a study of the diurnal pattern of Hector's dolphin in Akaroa Harbour an observational design was used. Each morning and afternoon the direction the dolphins were swimming in when they were at the harbour entrance was observed. The basic sampling unit was the pods of dolphins, not the individual dolphins. The use of the individual dolphins as the data points would be pseudoreplication because all dolphins within a pod will be swimming in the same direction. The dolphin pods were the level at which the sampling units are independent. Using the individual dolphins as the sample unit artificially inflates the sample size.

3.11 Two Levels of Pseudoreplication

Extension of the statistical (inductive) conclusions from an observational design beyond the specific study areas/populations to other unstudied areas/ populations is one of the common forms of pseudoreplication (Hurlbert 1984, Stewart-Oaten *et al.* 1986). Consider an example of the accidental spill of oil into Lyttelton Harbour. Deductive inferences concerning general conclusions of cause-and-effects of the oil (that extend beyond the specific study areas/ populations) may be possible if enough independent studies of different discharges of the oil are observed to produce similar effects. However, statistical (inductive) inferences beyond the study areas/populations are not possible using a simple observational study within the one harbour.

Results of observational studies such as evaluation of environmental impacts of accidental spills of oil or chemicals are often referred to a pseudoreplication in the biological literature. This use of the word is misleading unless it is qualified by adding the comment that the subsampling of treatment and non-treatment areas/populations is pseudoreplication if statistical conclusions are extrapolated beyond the treatment and non-treatment areas/populations. In other words, random sampling within an observational study is not pseudoreplication if the statistical inferences are limited to the specific areas or populations studies. It is the actual application of inferential statistics to unreplicated treatments or dependent replicates that causes "pseudoreplication." Single replicates per treatment or dependent data are not necessarily bad, or avoidable in field studies. However, it is dangerous to extrapolate inductive conclusions from such data using inferential statistics.

A second level of pseudoreplication occurs if dependent data from the basic sampling units of observational studies are analysed as if they are independent. Essentially the "sample size" is artificially inflated by analysing more than one datum per basic sampling unit. The importance of identifying and maintaining the integrity of data from the basic sampling units cannot be overemphasised. Things can get complicated. A good rule to follow is that statistical inferences should be based on only one value from each sample unit (unless the dependent data are properly handled in the analysis). For example, if 5 quadrats are randomly located in a study area, then design/data-based statistical inferences to the area should be based on 5 values; regardless of the number of plants, organisms, split-samples, etc, which many be present and measured or counted. The sample units for the analysis discussed above for surveys of cod in Paterson Inlet are the six sites and not the individual fish. If the purpose of the analysis were to compare variation within survey sites then the sample units should be the fish. Each site is then essentially a replicate experiment.

As another example, heights of individual plants were recorded for all plants in randomly located quadrats within a study area. The variation from plant to plant within a quadrat is an inappropriate measure of variation for statistical comparisons of a pair of treatment and non-treatment sites. A researcher would be guilty of pseudoreplication if the within quadrat variance is used in the statistical tests to compare mean height of plants in a particular pair of assessment and control sites.

3.12 Identifying Pseudoreplication

Problems associated with incorrect identification of data from the sampling units can give rise to incorrect statistical precision of estimated end-points. A simple example of pseudoreplication occurs if a single collection of material (sediment, plant tissue, etc.) might be taken at one point in an area, and then split several times in the laboratory. Analyses of each subset of the material might be conducted. Variation among such "replicates" is proper for study of the accuracy and precision of the laboratory measurement procedures, but does not represent spatial or temporal variation in the area and/or variation due to the field sampling. Variation among such replicates is not the correct measure of variation for comparison of assessment and control areas by statistical procedures. This example of pseudoreplication is presented because it is relatively easy to understand and has occurred in practice. The problem is usually easy to fix. In this example, simply average the results of the toxicity analyses on the subsets (split samples) to produce one number for each location of collection in the field. In general, use one datum per sampling unit from the field in the statistical procedures.

Although Hurlbert (1984) states that pseudoreplication widely occurs in both observational studies and manipulative experiments, he focuses the majority of his review on manipulative experiments. He also defines temporal pseudoreplication to occur when multiple samples are taken sequentially over time on the same sampling units (i.e. time series data), but the data are analysed as if they are independent. The experimenter should assume that the potential for false treatment effects is high because successive samples from a single unit taken over time are likely correlated. For example, in the study of growth of a weed, 30 successive measurements of the size of one patch is not equivalent to relocations of 30 randomly sampled patches from the population. Hurlbert did not discuss the use of "repeated measurement experimental designs" (the analysis of repeated measurements on the same experimental units over time). This theory has potential for solving many of the temporal pseudoreplication problems in manipulative experiments (e.g. Milliken and Johnson 1984).

Example: Sampling Vegetation Cover at Pupu Springs (Appendix Data Set 14)

Data from three transects on the proportion of bare substrate has been collected at Pupu Springs since 1991. The data is recorded in 5m sections, that is the proportion of bare substrate in the first 5m section of transect is recorded, then the proportion in the next 5m section, and so on. Two of the transects are 35m in length and the third is 50m in length. In total there are 24 - 5m sections. Some of the data is displayed in Table 3.3.

TABLE 3.3 PROPORTION OF BARE SUBSTRATE IN THREE TRANSECTS AT PUPU SPRINGS RECORDED IN 5M SECTIONS IN 1999.

		PROPORTION BARE SUBSTRATE - 1999			
	SECTION	TRANSECT 1	TRANSECT 2	TRANSECT 3	OVERALL
	0	0.3		0.1	
	5	0.5	0.1	0.85	
	10	0.35	0.5	0.5	
	15	0.35	0.4	0.25	
	20	0.55	0.2	0.3	
	25	0.6	0.05	0.8	
	30	0.4	0.1	0.15	
	35		0.02	0.3	
	40			0.2	
	45			0.3	
Transect average		0.435714	0.195714	0.375	0.335476
Transect standard deviation		0.114434	0.184739	0.260608	0.222674
Transect *cv*		0.262636	0.943923	0.694955	0.633838

To answer the question "Is there a change in the proportion of bare substrate over time", the appropriate level of analysis is the transect averages or totals (Figure 3.2). The proportion of bare substrate from each 5m section of transect is averaged over the number of sections within each transect. The sample size is therefore three since there are three transects. There is no obvious trend in increasing, or decreasing amounts of bare substrate over time. One test to see if the amounts change significantly among years would be to conduct a repeated measures analysis. An example of this using SPSS is in the appendix.

This analysis used the transect averages and ignores information in the variation of the bare substrate within transects. For example, consider two transects which both have on average 0.3 proportion of the surface area as bare substrate. On one transect the first seven 5m sections have no bare substrate and then the last three sections are all bare. On the other transect all ten sections have 0.3 proportion bare substrate. These two transects have quite different coverage and the management implications differ. In the first transect perhaps there has been some disturbance in the end sections of the transect, while in the second transect there may be the same level of disturbance over all the transect. To compare within-transect variation the standard deviation and *cv* of the proportion of bare substrate *within* each transect is calculated. The average cv over the three transects is also shown in Figure 3.2.

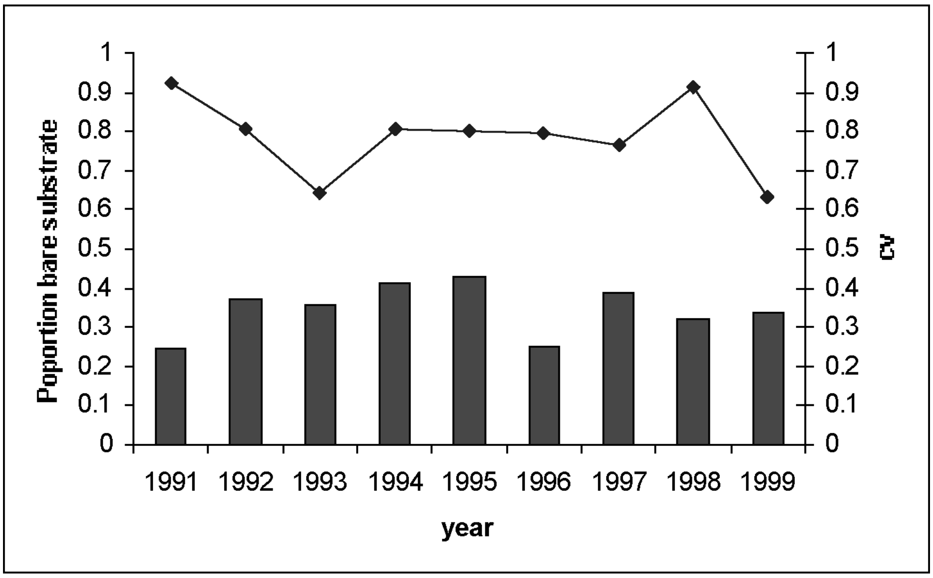


Figure 3.2 Average amount of the proportion of bare substrate in three transects at Pupu Springs recorded in 5m sections from 1991 and 1999. Also shown is the average amount of within - transect variation in the proportion of bare substrate. This is measured by the *cv* of the 5m sections within each transect.

3.13 Key Points in This Module

- The design of a monitoring scheme should relate directly to the specific objectives of monitoring.
- Environmental data typically is highly variable. Monitoring programmes should be designed to separate population trends from underlying environmental stochasticity.
- If the monitoring objective is to estimate the mean value following the most recent survey, e.g. environmental status, then it is best to reselect a fresh sample (i.e. new sample locations). If the objective is to estimate the change in population means, i.e. trends in environmental status it is best to use the same sites for each survey. Augmented designs are a way of combining these two approaches.
- Power is one of the crucial factors that should be considered in designing a monitoring programme.
- Power can be improved by reducing variation among and within sites, and matching treatment sites to control sites.
- Pseudoreplication is defined as the use of inferential statistics to test for treatment effects with data from experiments where either treatments are not replicated or replicates are not statistically independent.

3.14 Questions About This Module

After completing this module you should be able to give reasonable answers to the following questions.

1 A study is being planned to monitor the long-term effects of a herbicide on soil invertebrates within a special reserve. The herbicide is only used in the one reserve but there are many similar reserves in the region. What sites would you select for the monitoring? First define what the monitoring objectives are.

2 To survey of birds timed bird counts are often used where the number of birds heard singing during a time period are counted. In designing such a survey there are many levels that need to be considered in allocating survey effort: how long should the count be conducted for (5 or 10 minutes), how many locations within a site should be surveyed, how many sites should be surveyed, how many times should the survey be repeated within a day, within a season, within a year and among years? What factors should be considered in making these decisions? What are some likely survey objectives and for each, what measure of variation should be used in the analysis?

3 Using the above example, a survey is designed where 5 minute counts are taken at 10 locations within 5 sites in Canterbury. The surveys are repeated twice a day and over 5 days in summer for 2 years. Therefore in total there were 10 x 5 x 2 x 5 x 2 = 1000 surveys conducted. Is the sample size 1000? What is the correct sample size? Note the sample size depends on the objective of the survey so chose a sensible definition first.

REFERENCES

Barker, R.J. and Sauer, J.R. (1992). Modelling population change from time series data. In: *Wildlife 2001: Populations*. eds. McCullogh, D.R. and Barrett, R.H. Elsevier Applied Science, London.

Brown, J.A. and Miller, C.M. (1998). Monitoring stoat *Mustela erminea* control operations: power analysis and design. Science for Conservation: 96, Department of Conservation, Wellington

Brown, J.A. and Boyce, M.S. (1996). Monitoring of Karner blue butterflies (*Lycaeides melissa samuelis*) for the proposed habitat conservation plan, Wisconsin. Report to the US National Fish and Wildlife Foundation.

Caselton, W.F., Kan, L. and Zidek, J.V. (1992). Quality data networks that minimize entropy. In *Statistics in the Environmental and Earth Sciences* (eds. A.T. Walden and P. Guttorp), pp. 10-38. Edward Arnold, London.

Cox, D.D, Cox, L.H. and Ensor, K.B (1997). Spatial sampling and the environment: some issues and directions. *Environmental and Ecological Statistics* 4: 219-233.

Cressie, N. (1991). *Statistics for Spatial Statistics.* Wiley, New York.

Eberhardt, L.L. (1976). Quantitative ecology and impact assessment. *Journal of Environmental Management* 4: 27-70.

Fairweather. P.G. (1991). Statistical power and design requirements for environmental monitoring. *Australian Journal of Marine and Freshwater Research* 42: 555-567.

Fedorov, V. and Mueller, W. (1989). Comparison of two approaches in the optimal design of an observation network. *Statistics* 20: 339-51.

Gerrodette, T. (1987). A power analysis for detecting trends. *Ecology* 68: 1364-1372.

Green, R.H. (1979). *Sampling Design and Statistical Methods for Environmental Biologists.* Wiley, New York.

Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54(2):187-211.

Millard, S.P and Lettenmaier, D.P. 1986. Optimal design of biological sampling programs using analysis of variance. *Estuarine, Coastal and Shelf Science* 22:637-656.

Link, W.A., Barker, R.J., Sauer, J.R. and Droege, S. (1994). Within-site variability in surveys of wildlife populations. *Ecology* 75:1097-1108.

Milliken, G. A., and D. E. Johnson. (1984). *Analysis of Messy Data*, Volume I: Designed experiments. Lifetime Learning Publications, Belmont, California.

Overton, W.S., White, D. and Stevens, D.L.. (1991). *Design Report for EMAP, the Environmental Monitoring and Assessment Program.* U.S. Environmental Protection Agency report EPA/ 600/3-91/053, Washington, D.C.

Skalski, J.R. (1990). A design for long term status and trends monitoring. *Journal of Environmental Management* 30:139-144.

Skalski, J.R. and McKenzie, D.H. (1982). A design for aquatic monitoring programs. *Journal of Environmental Management* 14:237-251.

Steidl, R.J., Hayes, J.P., Schauber, E. 1997. Statistical power analysis in wildlife research. *Journal of Wildlife Management* 61:270-279.

Stevens, D.L. and Olsen, A.R. (1991). Statistical issues in environmental monitoring and assessment. In: *Proceedings of the Section on Statistics and the Environment*, American Statistical Association, Alexandria, Virginia.

Stewart-Oaten, A., Murdoch, W.W. and Parker, K.R. (1986). Environmental impact assessment: “Psuedoreplication” in time? *Ecology* 67:929-940.

Taylor, B.L., Gerrodette, T. (1993). The uses of statistical power in conservation biology: the vaquita and northern spotted owl. *Conservation Biology* 7:489-500.

Underwood, A.J. (1992). Beyond BACI: The detection of environmental impacts in the real, but variable, world. *Journal of Experimental and Marine Biological Ecology* 161:145-178.

Underwood, A.J. (1994). On beyond BACI: Sampling designs that might reliably detect environmental disturbances. *Ecological Applications* 4:3-15.

Underwood, A.J. (1997). *Experiments in Ecology.* Cambridge University Press, Cambridge.

Urquhart, N.S., Overton, W.S. and Birkes, D.S. (1993). Comparing sampling designs for monitoring ecological status and trends: impact of temporal patterns. In: *Statistics for the Environment*, eds. Barnett, V. and Turkman, K.F. Wiley and Sons.

Contents

MODULE 4: MODELS FOR ANALYSIS

Module 4: Models for Analysis

SUMMARY

An essential part of analysing many sets of data is choosing a model. This module covers a large number of topics related to this operation. First, some standard discrete and continuous statistical distributions are defined. Then one of the most popular models, the linear regression model, is described. This leads on to the related models for one, two and three factor analysis of variance. Finally, a very flexible class of models called generalized linear models, which are being used increasingly, is described.

4.1 Introduction

Many statistical analyses are based on a specific model for a set of data, where this consists of one or more equations that describe the observations in terms of parameters of distributions and random variables. For example a simple model for the measurement *X* made by an instrument might be

$$X = \theta + \in,$$

where θ is true value of what is being measured, and ε is a measurement error which is equally likely to be anywhere in the range from $-\frac{1}{2}$ to $+\frac{1}{2}$.

In situations where a model is used, an important task for the data analyst is to select a plausible model and to check, as far as possible, that the data are in agreement with this model. This includes both examining the form of the equation assumed, and the distribution or distributions that are assumed for the random variables.

To aid in this type of modelling process there are many standard distributions available, the most important of which are mentioned here. In addition, there are some standard types of model that have been found to be useful in practice for many sets of data. These are also reviewed briefly in this module. More on discrete and continuous distributions will be found in many introductory statistics texts. A useful summary of over the properties of over 30 distributions is in the @Risk Manual (Palisade, 1997, Appendix A).

4.2 Discrete Probability Distributions

A discrete distribution is one for which the random variable being considered can only take on certain specific values, rather than any value within some range. By far the most common situation in this respect is where the random variable is a count and the possible values are 0, 1, 2, 3, etc., possibly with no upper limit.

It is usual to denote a random variable by a capital *X* and a particular observed value by a lower case *x*. A discrete distribution is then defined by a list of the possible values x_1, x_2, x_3, ..., for X, and the probabilities $P(x_1)$, $P(x_2)$, $P(x_3)$, ... for these values. Of necessity,

$$P(x_1) + P(x_2) + P(x_3) + ... = 1,$$

i.e. the probabilities must add to 1. Also of necessity $P(x_i) \geq 0$ for all i, with $P(x_i) = 0$ meaning that the value x_i can never occur. Often there is a specific equation for the probabilities defined by

$$P(x) = \text{Prob}(X = x),$$

where $P(x)$ is some function of x.

The following discrete distributions are the ones which occur most often in environmental and other applications of statistics.

The Hypergeometric Distribution

The hypergeometric distribution arises when a random sample of size n is taken from a population of N units. If the population contains R units with a certain characteristic then the probability that the sample will contain exactly x units with the characteristic is

$$P(x) = {}^{R}C_{x}\,{}^{N-R}C_{n-x}/{}^{N}C_{n}, \text{ for } x = 0, 1, ..., \text{Min}(n,R),$$

where ${}^{A}C_{B}$ denotes the number of combinations of A things taken B at a time. The mean is $\mu = nR/N$ and the variance is $\sigma^2 = R(N - R)n/N^2$.

As an example of a situation where this distribution applies, suppose that a grid is set up over a study area and the intersection of the horizontal and vertical grid lines defines N possible sample locations. Let R of these locations have values in excess of a constant C. If a simple random sample of n from the N locations is taken then $P(x)$ gives the probability that exactly x out of the n sampled locations will have a value exceeding C. Figure 4.1(a) shows some examples of probabilities for hypergeometric distributions.

The Binomial Distribution

Suppose that it is possible to carry out a certain type of trial and that when this is done the probability of observing a positive result is always p, irrespective of the outcome of any other trial. Then if n trials are carried out the probability of observing exactly x successes is given by

$$P(x) = {}^{n}C_{x}p^{x}(1 - p)^{n-x}, \text{ for } x = 0, 1, 2, ..., n.$$

This is the binomial distribution. The mean is $\mu = np$ and the variance is $\sigma^2 = np(1 - p)$.

An example of this distribution occurs with the use of mark-recapture methods to estimate survival rates of fish in rivers. In that setting, if n fish are tagged and released into a river and there is a probability p of being recorded while passing a detection station downstream for each of the fish, then the probability of recording a total of exactly x fish downstream is given by the binomial distribution.

Figure 4.1(b) shows some examples of probabilities calculated for some particular binomial distributions.

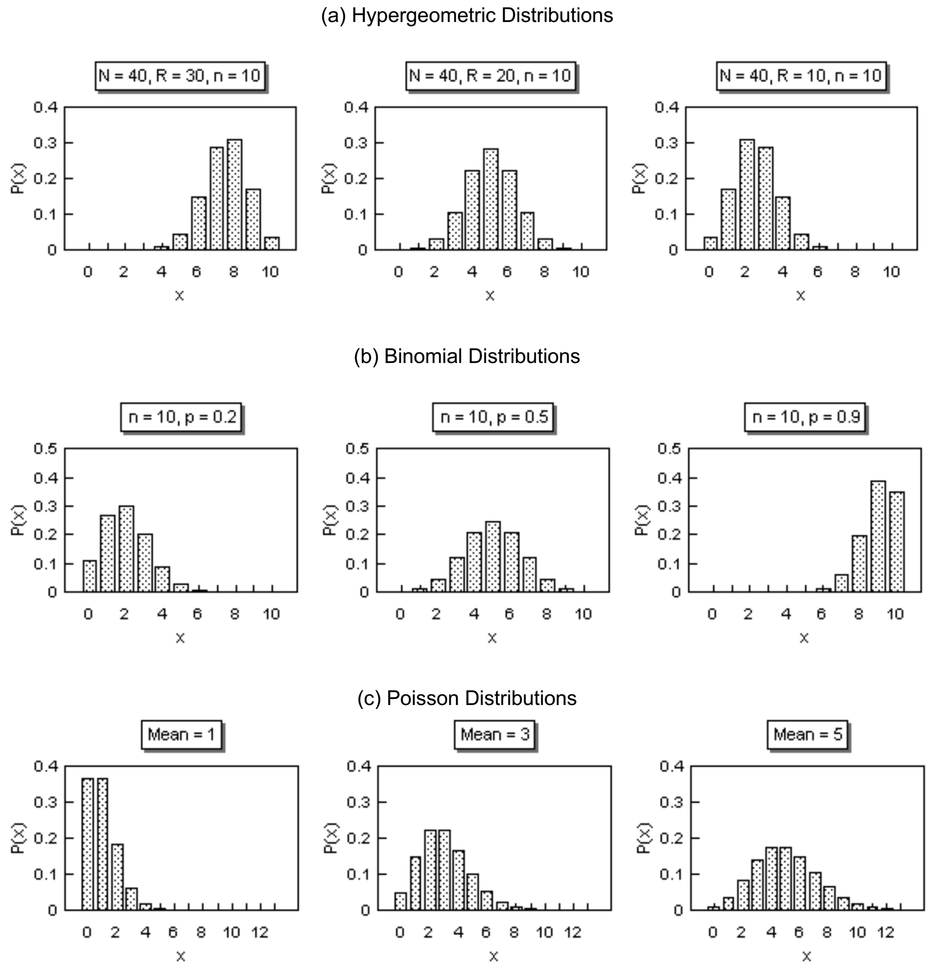


Figure 4.1 Examples of discrete probability distributions.

The Poisson Distribution

One derivation of the Poisson distribution is as the limiting form of the binomial distribution as μ tends to infinity and p tends to zero, with the mean $\mu = np$ remaining constant. More generally, however, it is possible to derive it as the distribution of the number of events in a given interval of time or a given area of space when the events occur at random, independently of each other at a constant mean rate. The probability function is

$$P(x) = \exp(-\mu)\mu^x/x!, \text{ for } x = 0, 1, 2, \ldots,$$

where the mean and variance of the distribution are both equal to μ.

In terms of events occurring in time, the type of situation where a Poisson distribution might occur is for counts of the number of occurrences of minor oil leakages in a region per month, or the number of cases per year of a rare disease in the same region. For events occurring in space a Poisson distribution might occur for the number of rare plants found in randomly selected metre square quadrats taken from a large area. In reality, though, counts of these types often display more variation than is expected for the Poisson distribution because of some clustering of the events. Indeed, the ratio of the variance of sample counts to the mean of the same counts, which should be close to one for a Poisson distribution, is sometimes used as an index of the extent to which events do not occur independently of each other.

Figure 1(c) shows examples of probabilities calculated for some particular Poisson distributions.

4.3 Continuous Statistical Distributions

Continuous distributions are described by a probability density function, $f(x)$, in such a way that the area under this function between two limits a and b gives the probability of an observation within this range, as shown in Figure 4.2. The following continuous distributions are ones that often occur in environmental and other applications of statistics.

The Exponential Distribution

The probability density function for the exponential distribution with mean μ is

$$f(x) = (1/\mu)\exp(-x/\mu), \text{ for } x \geq 0,$$

which has the form shown in Figure 4.3. The standard deviation is equal to the mean, μ. The main applications for this distribution is as a model for the time until a certain event occurs, such as the failure time of an item being tested, the time between the reporting of cases of a rare disease, etc.

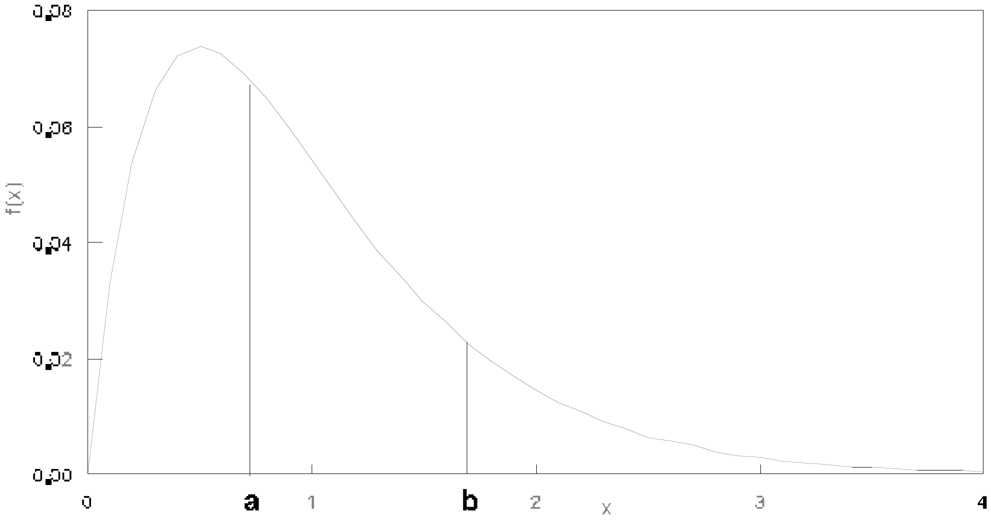


Figure 4.2 The probability density function $f(x)$ for a continuous distribution. The probability of a value between a and b is the area under the curve between these values, i.e. the area between the two vertical lines at $x = a$ and $x = b$.

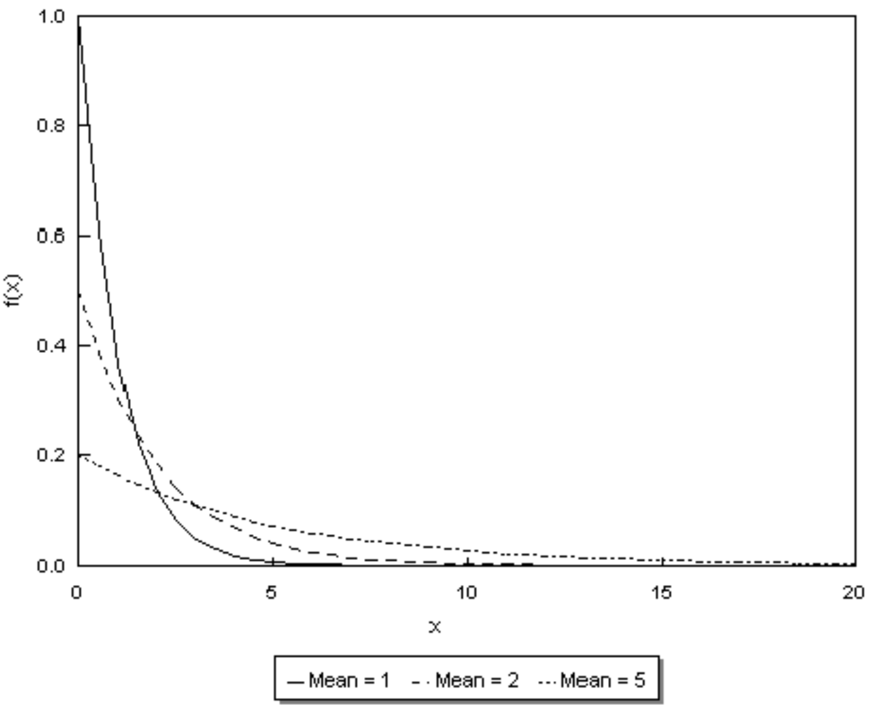


Figure 4.3 Examples of probability density functions for exponential distributions.

The Normal or Gaussian Distribution

The normal or Gaussian distribution with a mean of μ and a standard deviation of F has the probability density function

$$f(x) = \{1/\sqrt{(2\pi\sigma^2)}\}\ \exp\{-(x - \mu)^2/(2\sigma^2)\},\ \text{for } -\infty < x < +\infty,$$

as illustrated by Figure 4.4. Most of the distribution is within the range of the mean plus and minus two standard deviations, and virtually all within the range of the mean plus and minus three standard deviations. To be more precise, 68.3% of the distribution is within the interval $\mu \pm \sigma$, 95.4% of the distribution is within the interval $\mu \pm 2\sigma$, and 99.7% of the distribution is within the interval $\mu \pm 3\sigma$.

This is the 'default' that is often assumed for a distribution that is known to have a symmetric bell-shaped form, at least roughly. It is often observed for biological measurements such as the height of humans, and it can be shown theoretically (through something called the central limit theorem) that the normal distribution will tend to result whenever the variable being considered consists of a sum of contributions from a number of other distributions. In particular, mean values, totals, and proportions from simple random samples will often be approximately normally distributed.

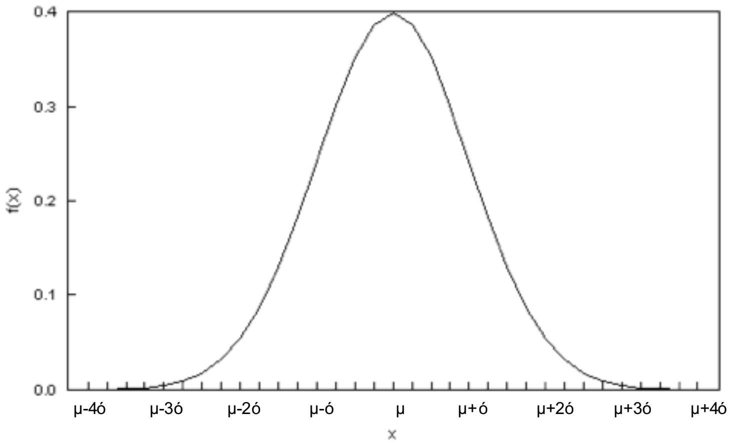


Figure 4.4 The probability density function for the normal distribution with a mean of μ and a standard deviation of σ.

The Lognormal Distribution

It is a characteristic of the distribution of many environmental variables that they are not symmetric like the normal distribution. Instead, there are many fairly small values and occasional extremely large values. Distribution of this type are said to be skewed to the right.

Often with data like this only positive values can occur and it turns out that the logarithm of the measurements has a normal distribution, at least approximately. In that case the distribution of the original measurements can be assumed to be a lognormal distribution, with probability density function

$$f(x) = [1/\{x\sqrt{(2\pi\sigma^2)}\}]\exp[-\{\log_e(x) - \mu\}^2/\{2\sigma^2\}], \text{ for } x > 0.$$

Here μ and F are the mean and standard deviation of the natural logarithm of the original measurement. The mean and variance of the original measurement itself are $\exp(\mu - \sigma^2)$ and $\exp(2\mu - \sigma^2)\{\exp(\sigma^2) - 1\}$, respectively. Figure 4.5 shows some examples of probability density functions for some lognormal distributions.

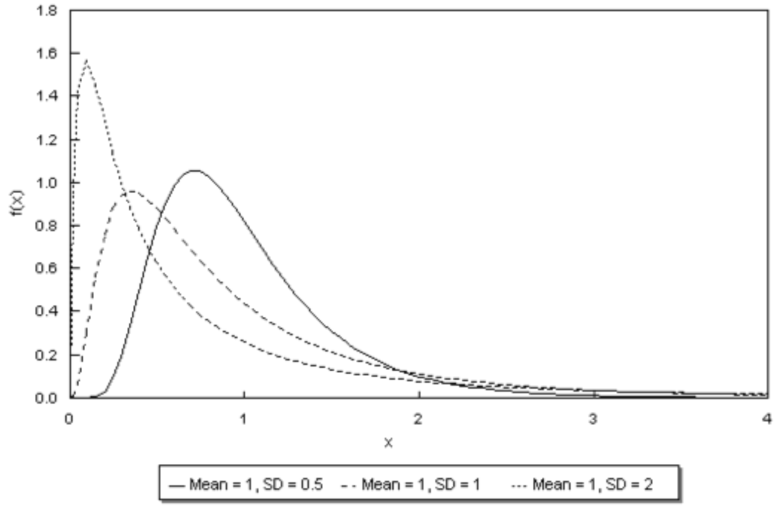


Figure 4.5 Examples of lognormal distributions with a mean of 1.0. The standard deviations are 0.5, 1.0 and 2.0.

4.4 Linear Regression

Linear regression is one of the most frequently used statistical tools. The purpose is to relate a single observed variable (*Y*) to one or more other variables (X_1, X_2, ..., X_p), in an attempt to account for the variation in the first variable as a result of variation in the other variables. With only one other variable this is often referred to as simple linear regression.

The usual situation is that the data available consist of *n* observations y_1, y_2, ..., y_n for the dependent variable *Y*, with corresponding values for the *X* variables. The model that is assumed is

$$y = \text{ß}_0 + \text{ß}_1 x_1 + \text{ß}_2 x_2 + \ldots + \text{ß}_p x_p + \varepsilon,$$

where ε is a random error with a mean of zero and a constant standard deviation σ. The model is estimated by finding the coefficients of the *X* values that make the error sum of squares (SSE) as small as possible. In other words, if the estimated equation is

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p,$$

then the b values are chosen so as to minimise

$$\text{SSE} = \Sigma(y_i - \hat{y}_i)^2,$$

where the $\hat{y}_i$ is the value given by the fitted equation that corresponds to the data value y_i, and the sum is over the *n* data values. Statistical packages or spreadsheets are readily available to do these calculations.

There are various ways that the usefulness of a fitted regression equation can be assessed. One involves partitioning the variation observed in the *Y* values into parts that can be accounted for by the *X* values, and a part (SSE, above) which cannot be accounted for. To this end, the total variation in the *Y* values is measured by the total sum of squares

$$\text{SST} = \Sigma(y_i - \bar{y})^2.$$

This is partitioned into the error sum of squares SSE and the sum of squares accounted for by the regression, so that SST = SSR + SSE. The proportion of the variation in *Y* accounted for by the regression equation is then the coefficient of multiple determination,

$$\text{R}^2 = \text{SSR/SST} = 1 - \text{SSE/SST},$$

which is a good indication of the effectiveness of the regression.

There are a variety of inference procedures that can be applied in the multiple regression situation when the regression errors ε, are assumed to be independent random variables from a normal distribution with a mean of zero and constant variance σ^2. For example, a test for whether the fitted equation accounts for a significant proportion of the total variation in *Y* can be based on a technique called the analysis of variance, and the estimated regression coefficients can also be tested individually to see whether they are significantly different from zero using a t-test. There is also sometimes value in considering the variation in *Y* that is accounted for by a variable X_j when this is included in the regression after some of the other variables are already in, again using the analysis of variance.

This has been an extremely brief introduction to the uses of multiple regression, which is a widely used tool. For a more detailed discussion see Manly (1992, Chapter 4). Some further aspects of the use of this method are also considered in the following example. All of the calculations can be done either in Excel or SPSS.

Example: Caging and the Flowering of Woodrose (Appendix Data Set 3)

As an example of the use of regression, consider Data Set 3 (Monitoring of *Dactylanthus taylorii* near the summit of Mount Pirongia) from the Appendix to these notes. In brief, this data set gives information on various characteristics of *D. taylorii* at eight locations from 1997 to 1999, where some of the plants were caged and others uncaged. Records are not available at all locations in all years. Here we consider the number of flowers per plant as the dependent variable, which means that data can only be used from six locations.

Table 4.1 gives a summary of the data used for the analysis. There are 19 sample units, with each unit consisting of the plants at one location that were either caged or not. There are eight *X* variables used in the analysis, labelled *X1* to *X8* in the table. These require some explanation because they have been set up to allow for differences between locations, years and the cage status. It is not valid to just use the code numbers for these variables in a regression analysis because, for example, location 6 does not have six times as much 'location' as location 1.

Variables LOC1 to LOC5 are indicator variables for the location. Thus LOC1 is 1 for an observation at location 1 or is otherwise 0, LOC2 is 1 for an observation at location 2 or is otherwise 0, and so on up to LOC5. Actually, there are six locations. However, location 6 is not assigned an indicator variable. This means that location 6 becomes the 'standard' location, which applies if LOC1 to LOC5 are all equal to 0. Including LOC1 to LOC5 in a regression equation has the effect of allowing the mean value of *Y* (the number of flowers per plant) to vary from location to location. This setting up of indicator variables is done automatically by some regression programs for a variable like location that is a code to indicate different categories of something. Unfortunately, this is not the case with the regression options in Excel or SPSS.

TABLE 4.1 REGRESSION DATA FROM A STUDY OF THE NUMBER OF FLOWERS ON CAGED AND UNCAGED WOODROSE PLANTS IN SIX LOCATIONS, IN 1997 TO 1999.

				INDICATOR VARIABLES FOR EFFECTS[1]										
				LOCATIONS					YEARS		CAGE			
CASE	LOCATION[2]	YEAR[3]	CAGED[4]	LOC1	LOC2	LOC3	LOC4	LOC5	YEAR1	YEAR2	CAGE	PLANTS	FLOWERS	*Y*[5]
1	1	2	2	1	0	0	0	0	0	1	1	3	14	4.67
2	1	3	1	1	0	0	0	0	0	0	0	3	21	7.00
3	1	3	2	1	0	0	0	0	0	0	1	2	22	11.00
4	2	1	1	0	1	0	0	0	1	0	0	8	18	2.25
5	2	2	1	0	1	0	0	0	0	1	0	1	16	16.00
6	2	2	2	0	1	0	0	0	0	1	1	8	46	5.75
7	2	3	1	0	1	0	0	0	0	0	0	2	3	1.50
8	2	3	2	0	1	0	0	0	0	0	1	7	32	4.57
9	3	1	1	0	0	1	0	0	1	0	0	9	3	0.33
10	3	2	1	0	0	1	0	0	0	1	0	10	0	0.00
11	3	3	1	0	0	1	0	0	0	0	0	6	16	2.67
12	3	3	2	0	0	1	0	0	0	0	1	8	101	12.63
13	4	3	1	0	0	0	1	0	0	0	0	3	0	0.00
14	4	3	2	0	0	0	1	0	0	0	1	5	55	11.00
15	5	3	2	0	0	0	0	1	0	0	1	2	10	5.00
16	6	1	1	0	0	0	0	0	1	0	0	6	6	1.00
17	6	2	1	0	0	0	0	0	0	1	0	6	1	0.17
18	6	3	1	0	0	0	0	0	0	0	0	2	8	4.00
19	6	3	2	0	0	0	0	0	0	0	1	4	42	10.50

Notes:

[1] See text for an explanation of the indicator variables.

[2] Locations are: 1, By hut; 2, Goblin Wood; 3, Hihikiwi, 4, South Hihikiwi; 5, Middle Bell track; and 6, North of Hihikiwi.

[3] Years are: 1, 1997; 2, 1998; and 3, 1999.

[4] 1 = uncaged, 2 = caged.

[5] *Y* = Flowers per Plant.

Variables YEAR1 and YEAR2 are indicator variables for the year, while CAGE is an indicator variable for the cage status. These are defined similarly to the indicator variables for location: YEAR1 is 1 for 1997 or is otherwise 0, YEAR2 is 1 for 1998 or is otherwise 0, and CAGE is 1 for caged plants or is otherwise 0. Hence 1999 becomes the 'standard' year (YEAR1 = YEAR2 = 0), and uncaged plants become the standard type (CAGE= 0).

To examine the effects if any of caging on the number of flowers per plant, consider the regression model including all of the *X* variables,

$$Y = \beta_0 + \beta_1\text{LOC1} + \ldots + \beta_5\text{LOC5} + \beta_6\text{YEAR1} + \beta_7\text{YEAR2} + \beta_8\text{CAGE} + \varepsilon,$$

which claims that the mean number of flowers per plant may vary from location to location (allowed for by including LOC1 to LOC5), the year (allowed for by including YEAR1 and YEAR2), and whether or not plants are caged or not (allowed for by including CAGE).

There is one further complication with this example. The number of plants for an observation varies from 1 to 10, and it can be expected that the random error in a Y value will become less variable as the number of plants increases.

In fact, since *Y* is a mean value it can be expected that its variance will be proportional to 1/(Number of Plants). This can be allowed for in the SPSS regression module by using the number of plants as a regression weight. There is an option for this on the main regression menu.

Part of the output from the SPSS regression module is shown in Table 4.2, with some comments in italics. The location variables were entered in a block, as were the year variables and the cage variable (i.e. there were three blocks of variables defined). The remove option was then used in SPSS to allow these blocks of variables to be removed if there effect is not significant. For example, this option allows the location variables LOC1 to LOC5 to be removed if their combined relationship with the dependent variable is not significant at the 10% level (the SPSS default). As the output shows, the actual operation of SPSS seems to be a bit different. The cage variable ends up being removed although it is highly significant (p = 0.003).

TABLE 4.2 SPSS REGRESSION ANALYSIS FOR THE DATA IN TABLE 4.1, WITH COMMENTS IN ITALICS.

Variables Entered/Removed

MODEL	VARIABLES ENTERED	VARIABLES REMOVED	METHOD
1	CAGE2, YEAR2, LOC1, LOC5, LOC4, LOC3, YEAR1, LOC2		Enter
2		LOC2, LOC4, LOC3, LOC5, LOC1	Remove
3		CAGE2	Remove

a All requested variables entered.
b All requested variables removed.
c Dependent Variable: RATIO
d Weighted Least Squares Regression - Weighted by PLANTS

In the last step the cage variable has been removed! Only the year variables are left in.

Model Summary

MODEL	R	R SQUARE	ADJUSTED R SQUARE	STD. ERROR OF THE ESTIMATE
1	.818	.669	.405	7.7470
2	.797	.635	.562	6.6457
3	.586	.344	.261	8.6308

a Predictors: (Constant), CAGE2, YEAR2, LOC1, LOC5, LOC4, LOC3, YEAR1, LOC2
b Predictors: (Constant), CAGE2, YEAR2, YEAR1
c Predictors: (Constant), YEAR2, YEAR1
d Dependent Variable: RATIO
e Weighted Least Squares Regression - Weighted by PLANTS

ANOVA

MODEL		SUM OF SQUARES	df	MEAN SQUARE	F	SIG.
1	Regression	1215.425	8	151.928	2.531	.085
	Residual	600.154	10	60.015		
	Total	1815.579	18			
2	Regression	1153.095	3	384.365	8.703	.001
	Residual	662.484	15	44.166		
	Total	1815.579	18			
3	Regression	623.732	2	311.866	4.187	.034
	Residual	1191.847	16	74.490		
	Total	1815.579	18			

a Predictors: (Constant), CAGE2, YEAR2, LOC1, LOC5, LOC4, LOC3, YEAR1, LOC2
b Predictors: (Constant), CAGE2, YEAR2, YEAR1
c Predictors: (Constant), YEAR2, YEAR1
d Dependent Variable: RATIO
e Weighted Least Squares Regression - Weighted by PLANTS

Coefficients

MODEL		UNSTANDARDIZED COEFFICIENTS *B*	STD. ERROR	STANDARDIZED COEFFICIENTS BETA	*t*	SIG.
1	(Constant)	3.482	2.401		1.450	.178
	LOC1	.939	3.389	.060	.277	.787
	LOC2	-.929	2.477	-.095	-.375	.715
	LOC3	.132	2.277	.014	.058	.955
	LOC4	-.431	3.500	-.027	-.123	.904
	LOC5	-4.600	5.977	-.151	-.770	.459
	YEAR1	-2.038	2.486	-.200	-.820	.431
	YEAR2	-2.984	2.043	-.311	-1.460	.175
	CAGE2	6.118	2.049	.688	2.986	.014
2	(Constant)	3.481	1.437		2.423	.029
	YEAR1	-2.308	1.996	-.226	-1.156	.266
	YEAR2	-2.931	1.654	-.306	-1.772	.097
	CAGE2	5.603	1.618	.631	3.462	.003
3	(Constant)	7.047	1.301		5.416	.000
	YEAR1	-5.874	2.221	-.576	-2.645	.018
	YEAR2	-4.296	2.086	-.448	-2.059	.056

a Dependent Variable: RATIO
b Weighted Least Squares Regression - Weighted by PLANTS

It is a mystery why CAGE2 was removed for model 3 because it is the most significant variable - highly significant, in fact. The best model seems to be Model 2, including year and cage effect.

Excluded Variables

MODEL		BETA IN	*t*	SIG.	PARTIAL CORRELATION	COLLINEARITY STATISTICS TOLERANCE
2	LOC1	.089	.549	.592	.145	.967
	LOC2	-.091	-.531	.603	-.141	.875
	LOC3	.057	.340	.739	.090	.932
	LOC4	-.008	-.049	.962	-.013	.893
	LOC5	-.143	-.891	.388	-.232	.959
3	LOC1	.121	.575	.574	.147	.971
	LOC2	.096	.458	.654	.117	.979
	LOC3	-.085	-.406	.690	-.104	.996
	LOC4	-.013	-.060	.953	-.016	.893
	LOC5	-.070	-.334	.743	-.086	.975
	CAGE2	.631	3.462	.003	.666	.733

a Predictors in the Model: (Constant), CAGE2, YEAR2, YEAR1
b Predictors in the Model: (Constant), YEAR2, YEAR1
c Dependent Variable: RATIO
d Weighted Least Squares Regression - Weighted by PLANTS

Residuals Statistics

	MINIMUM	MAXIMUM	MEAN	STD. DEVIATION	*N*
Predicted Value	1.1726	7.0466	4.9887	2.5310	19
Residual	-7.0466	13.2489	.2765	4.6929	19
Std. Predicted Value	-	-	-	-	0
Std. Residual	-	-	-	-	0

a Not computed for Weighted Least Squares regression.
b Dependent Variable: RATIO
c Weighted Least Squares Regression - Weighted by PLANTS

The overall conclusion from this analysis is that a model including a year effect and a cage effect seems appropriate for the data. The effect of caging seems to be to increase the number of flowers by about 5.6 per plant, with the standard error of this estimate being 1.6.

A regression analysis is not complete without plots of the data and residuals to ensure that the model being considered is reasonable. With a weighted regression SPSS does not provide these plots automatically for residuals, but the output tells you what to do to get the graphs. The ones in Figure 4.6 were produced in a spreadsheet. Apart from the fact that the residuals are all close to zero in location 6 and year 1, quite likely due to small sample sizes, there seems nothing unusual in these plots.

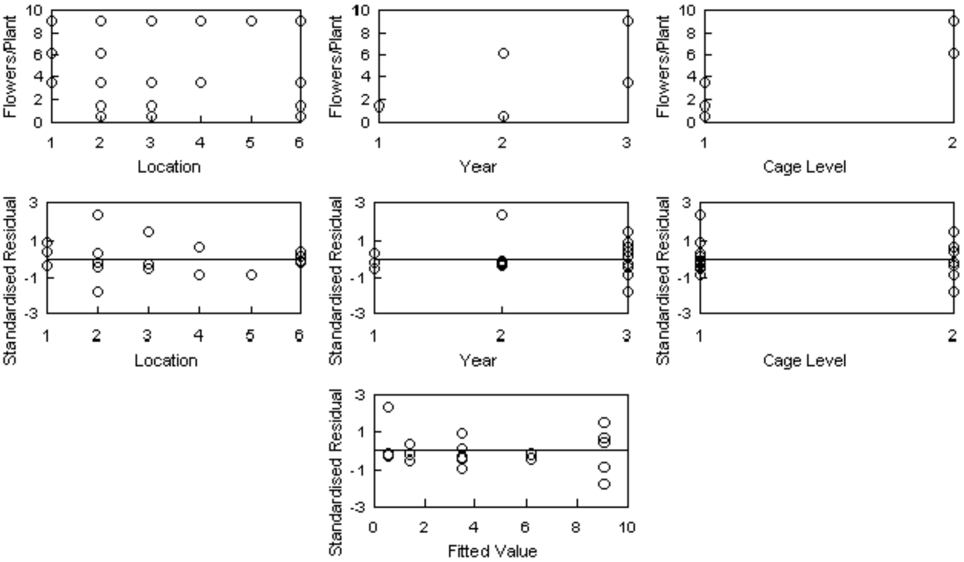


Figure 4.6 Plots of the number of flowers per plant and standardised residuals against locations, years and the cage status, plus a plot of standardized residuals against the expected number of flowers.
NB: Standardised residuals equals the difference between the observed number of flowers per plant and the expected number from the fitted regression equation, divided by the estimated standard error of the observed value. For a good model the standardized residuals will almost all be in the range -2 to +2, with no relationship with what they are plotted against.

4.5 Analysis of Variance

An important distinction in statistical modelling is that between variables and factors. A variable is something like the phosphorus concentration or nitrogen concentration in lakes. A factor, on the other hand, has a number of levels and in terms of a regression model it might be thought plausible that the response variable being considered has a mean level that changes with these levels. The location in the example just considered is therefore a factor.

Thus if an experiment is carried out to assess the effect of 1080 poison pellets on invertebrate densities, then the density of invertebrates might be related by a regression model to the level of 1080 in the pellets, perhaps at four concentrations. The 1080 level, would then be treated as a variable. If the experiment was carried out at three different locations, then the location would be a factor, which could not just be entered as a variable. The locations could be labelled 1 to 3, and what would be required in the regression equation is an allowance for the invertebrate density to vary with the location, just like the number of flowers per plant was allowed to vary in the example just considered.

The type of regression model that could then be appropriate would be

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon,$$

where Y is the density of invertebrates, X_i for $i = 1$ to 3 are indicator variables such that $X_i = 1$ if the location is i, or is otherwise 0, and X_4 is the concentration of 1080. The effect of this formulation is that for location 1 the expected invertebrate density with a 1080 concentration of X_4 is $\beta_1 + \beta_4 X_4$, for location 2 the expected invertebrate density with this concentration is $\beta_2 + \beta_4 X_4$, and for location 3 the expected invertebrate density with this concentration is $\beta_3 + \beta_4 X_4$. Hence in this situation the location factor at three levels can be allowed for by introducing three 0-1 variables into the regression equation and omitting the constant term β_0. This is a slight modification of the approach used in the previous example where a constant term was left in the regression equation and the number of 0-1 indicators was one less than the number of locations.

The equation above allows for a factor effect, but only on the expected invertebrate density. If the effect of the concentration of 1080 may also vary with the location then the model can be extended to allow for this, by adding products of the 0-1 variables for the location with the concentration variable to give

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_4 + \beta_5 X_2 X_4 + \beta_6 X_3 X_4 + \varepsilon.$$

For locations 1 to 3 the expected invertebrate densities are then $\beta_1 + \beta_4 X_4$, $\beta_2 + \beta_5 X_4$, and $\beta_3 + \beta_6 X_4$, respectively. The effect is then that there is a linear relationship between the invertebrate density and the concentration of 1080, which may differ for the three locations.

When there is only one factor to be considered in a model it can be handled reasonably easily by using dummy indicator variables as just described. However, with more than one factor this gets cumbersome and it is more usual to approach modelling from the point of view of what is called the analysis of variance. This is based on a number of standard models and the theory can get quite complicated. Nevertheless, the use of analysis of variance in practice can be quite straightforward using SPSS to do the calculations. An introduction to experimental designs and their corresponding analyses of variance is given by Manly (1992, Chapter 7). Here only four simple situations will be considered. A number of texts have been written covering mainly just the topics of linear regression and analysis of variance. Three are those of Neter *et al.* (1983), Younger (1985) and Mead *et al.* (1993).

One factor Analysis of Variance

With a single factor the analysis of variance model is just a model for comparing the means of I samples, where I is two or more. This model can be written as

$$x_{ij} = \mu + a_i + \varepsilon_{ij},$$

where x_{ij} is the jth observed value of the variable of interest at the *i*th factor level (i.e. in the *i*th sample), μ is an overall mean level, a_i is the deviation from μ for the *i*th factor level with $a_1 + a_2 + \ldots a_I = 0$, and ε_{ij} is the random component of x_{ij}, which is assumed to be independent of all other terms in the model, with a mean of zero and a constant variance.

To test for an effect of the factor an analysis of variance table is set up, where this takes the form shown in Table 4.3. Here the sum of squares for the factor is just the sum of squares accounted for by allowing the mean level to change with the factor level in a regression model, although it is usually computed somewhat differently. The F-test requires the assumption that the random components ε_{ij} in the model have a normal distribution.

Two Factor Analysis of Variance

With a two factor situation there are I levels for one factor (*A*) and *J* levels for the other factor (*B*). It is simplest if m observations are taken for each combination of levels, which is what will be assumed here. The model can be written

$$x_{ijk} = \mu + a_i + b_j + (ab)_{ij} + \varepsilon_{ijk},$$

where x_{ijk} denotes the *k*th observation at the *i*th level for factor *A* and the jth level for factor *B*, μ denotes an overall mean level, a_i denotes an effect associated with the *i*th level of factor *A*, b_j denotes an effect associated with the jth level of factor *B*, $(ab)_{ij}$ denotes an interaction effect so that the mean level at a factor combination does not have to be just the sum of the effects of the two individual factors, and ε_{ijk} is the random part of the observation x_{ijk}, which is assumed to be independent of all other terms in the model, with a mean of zero and a constant variance.

TABLE 4.3 FORM OF THE ANALYSIS OF VARIANCE TABLE FOR A ONE FACTOR MODEL, WITH I LEVELS OF THE FACTOR AND N OBSERVATIONS IN TOTAL.

SOURCE OF VARIATION	SUM OF SQUARES[1]	DEGREES OF FREEDOM	MEAN SQUARE[2]	F[3]
Factor	SSF	I - 1	MSF = SSF/(I - 1)	MSF/MSE
Error	SSE	n - 1	MSE = SSE/(n - 1)	
Total	SST = $\sum\sum(x_{ij} - \bar{x})^2$	n - 1		

[1] SSF = sum of squares between factor levels, SSE = sum of squares for error (variation within factor levels), and SST = total sum of squares for which the summation is over all observations at all factor levels.

[2] MSF= mean square between factor levels, and MSE = mean square error.

[3] The *F*-value is tested for significance by comparison with critical values for the *F*-distribution with I - 1 and n - 1 degrees of freedom.

Moving from one to two factors introduces the complication of deciding whether the factors have what are called fixed or random effects. With a fixed effects factor the levels of the factor for which data are collected are regarded as all the levels of interest. The effects associated with that factor are then defined to add to zero. Thus if A has fixed effects then $a_1 + a_2 + \ldots + a_I = 0$ and $(ab)_{1j} + (ab)_{2j} + \ldots + (ab)_{Ij} = 0$, for all j. If, on the contrary, A has random effects then the values a_1 to a_I are assumed to be random values from a distribution with mean zero and variance σ^2_A, while $(ab)_{1j}$ to $(ab)_{Ij}$ are assumed to be random values from a distribution with mean zero and variance σ^2_{AB}.

An example of a fixed effect is when an experiment is run with low, medium and high levels for the amount of a chemical because in such a case the levels can hardly be thought of as a random choice from a population of possible levels. An example of a random effect is when one of the factors in an experiment is the brood of animals tested, where these broods are randomly chosen from a large population of possible broods. In this case the brood effects observed in the data will be random values from the distribution of brood effects that are possible.

The distinction between fixed and random effects is important because the way that the significance of factor effects is determined depends on what is assumed about these effects. Some statistical packages allow the user to choose which effects are fixed and which are random, and carries out tests based on this choice. The 'default' is usually fixed effects for all factors, in which case the analysis of variance table is as shown in Table 4.4.

If there is only m = 1 observation for each factor combination then the error sum of squares shown in Table 4.4 cannot be calculated. In that case it is usual to assume that there is no interaction between the two factors, and the interaction sum of squares becomes the error sum of squares, and the factor effects are tested using F-ratios that are the factor mean squares divided by this error sum of squares.

Three Factor Analysis of Variance

With three factors with levels *I, J*, and *K*, and m observations for each factor combination, the analysis of variance model becomes

$$x_{ijku} = \mu + a_i + b_j + c_k + (ab)_{ij} + (ac)_{ik} + (bc)_{jk} + (abc)_{ijk} + \varepsilon_{ijku},$$

where x_{ijku} is the uth observation for level *i* of factor *A*, level *j* of factor *B*, and level *k* of factor *C*, a_i, b_j and c_k are the main effects of the three factors, $(ab)_{ij}$, $(ac)_{ik}$ and $(bc)_{jk}$ are terms that allow for first order interactions between pairs of factors, $(abc)_{ijk}$ allows for a three factor interaction (where the mean for a factor combination is not just the sum of the factor and first order interaction effects), and ε_{ijku} is a random component of the observation, independent of all other terms in the model with a mean of zero and a constant variance.

TABLE 4.4 FORM OF THE ANALYSIS OF VARIANCE TABLE FOR A TWO FACTOR MODEL WITH FIXED EFFECTS, AND WITH I LEVELS FOR FACTOR A, *j* LEVELS FOR FACTOR B, *m* OBSERVATIONS FOR EACH COMBINATION OF FACTOR LEVELS, AND *n* = *IJm* OBSERVATIONS IN TOTAL.

SOURCE OF VARIATION	SUM OF SQUARES[1]	DEGREES OF FREEDOM	MEAN SQUARE	F[2]
Factor A	SSA	I - 1	MSA = SSA/(I - 1)	MSA/MSE
Factor B	SSB	J - 1	MSB = SSB/(J - 1)	MSB/MSE
Interaction	SSAB	(I - 1)(J - 1)	MSAB = SSAB/{(I - 1)(J - 1)}	MSAB/MSE
Error	SSE	I J(*m* - 1)	MSE = SSE/{I J(*m* - 1)}	
Total	SST = $\Sigma\Sigma\Sigma(x_{ijk} - \bar{x})^2$	*N* - 1		

[1] The sum for SST is over all levels for *i, j* and *k*, i.e. over all *n* observations.
[2] The F-ratios for the factors are for fixed effects only.

The analysis of variance table generalises in an obvious way in moving from two to three factors. There are now sums of squares, mean squares and F-ratios for each of the factors, the two factor interactions, the three factor interaction, and the error term. The interpretation of the analysis of variance table depends on what assumptions are made about fixed and random effects.

Example: Poisoning and Bird Counts at Hurunui Mainland Island (Appendix Data Set 12)

This is an example of a three factor analysis of variance done in SPSS. Five minute bird counts were conducted in the summers of 1995/96, 1996/97, 1997/98 and 1998/99 in two areas on Hurunui Mainland Island in October or November, and then again in February (except in the 1998 summer, when the second sampling was done in late November). Between the first and second sampling times in each summer there was a stoat poisoning operation in area 1 only, and there is interest in whether this affected bird numbers.

The original data are in Data Set 12 in the Appendix. For the purposes of this example, only a summary will be used, as shown in Table 4.5. This table gives the mean counts of all bird species in two replicate sets of 40 5-minute counts (Count 1 and Count 2), for the two study areas (control and treated), for the two sample times (before and after poisoning), for the four summers (1995/96 to 1998/99).

TABLE 4.5 MEANS FROM 40 5-MINUTE BIRD COUNTS (ALL SPECIES) FROM TWO REPLICATE COUNTS IN EACH OF THE TREATED AND CONTROL AREAS, BEFORE AND AFTER STOAT POISONING IN FOUR SUMMERS.

		1995/96		1996/97		1997/98		1998/99	
AREA	COUNT	BEFORE	AFTER	BEFORE	AFTER	BEFORE	AFTER	BEFORE	AFTER
Treated	1	12.50	17.78	7.93	7.28	12.38	8.70	11.38	10.08
	2	11.70	14.30	5.45	4.08	11.15	8.10	5.98	7.93
Control	1	12.70	14.88	9.58	7.83	16.40	8.43	13.03	10.90
	2	9.50	13.65	5.68	4.08	12.10	6.63	9.60	6.90

For the analysis of variance the factors are the area with two levels, the summer with four levels, and the sample time in the summer with two levels. This gives 2x4x2 = 16 factor combinations, with two counts that are treated as replicates for each of these combinations, giving 32 observations altogether.

Table 4.6 shows part of the output from the analysis in SPSS, with comments in italics.

TABLE 4.6 OUTPUT FROM SPSS FOR THE THREE FACTOR ANALYSIS OF VARIANCE USING THE GENERAL LINEAR MODEL OPTION FOR FACTORIAL DATA.

Between-Subjects Factors

		N
AREA	1	16
	2	16
SUMMER	1	8
	2	8
	3	8
	4	8
TIME	1	16
	2	16

Descriptive Statistics

Dependent Variable: M_COUNT

AREA	SUMMER	TIME	MEAN	STD. DEVIATION	N
1	1	1	12.1000	.5657	2
		2	16.0400	2.4607	2
		Total	14.0700	2.7018	4
	2	1	6.6900	1.7536	2
		2	5.6800	2.2627	2
		Total	6.1850	1.7526	4
	3	1	11.7650	.8697	2
		2	8.4000	.4243	2
		Total	10.0825	2.0215	4
	4	1	8.6800	3.8184	2
		2	9.0050	1.5203	2
		Total	8.8425	2.3803	4
	Total	1	9.8088	2.9004	8
		2	9.7812	4.3206	8
		Total	9.7950	3.5549	16
2	1	1	11.1000	2.2627	2
		2	14.2650	.8697	2
		Total	12.6825	2.3017	4
	2	1	7.6300	2.7577	2
		2	5.9550	2.6517	2
		Total	6.7925	2.4112	4
	3	1	14.2500	3.0406	2
		2	7.5300	1.2728	2
	Total	10.8900	4.3214	4	
	4	1	11.3150	2.4254	2
		2	8.9000	2.8284	2
		Total	10.1075	2.5635	4
	Total	1	11.0738	3.2042	8
		2	9.1625	3.6941	8
		Total	10.1181	3.4833	16
Total	1	1	11.6000	1.4652	4
		2	15.1525	1.8223	4
		Total	13.3763	2.4391	8
	2	1	7.1600	1.9633	4
		2	5.8175	2.0188	4
		Total	6.4887	1.9783	8
	3	1	13.0075	2.3221	4
		2	7.9650	.9232	4
		Total	10.4862	3.1529	8
	4	1	9.9975	3.0225	4
		2	8.9525	1.8549	4
		Total	9.4750	2.3878	8
	Total	1	10.4412	3.0238	16
		2	9.4719	3.8964	16
		Total	9.9566	3.4659	32

Tests of Between-Subjects Effects

Dependent Variable: M_COUNT

SOURCE	TYPE III SUM OF SQUARES	df	MEAN SQUARE	F	SIG.
Corrected Model	294.739	15	19.649	4.048	.004
Intercept	3172.260	1	3172.260	653.584	.000
AREA	.835	1	.835	.172	.684
SUMMER	193.860	3	64.620	13.314	.000
TIME	7.518	1	7.518	1.549	.231
AREA * SUMMER	8.258	3	2.753	.567	.645
AREA * TIME	7.097	1	7.097	1.462	.244
SUMMER * TIME	74.365	3	24.788	5.107	.011
AREA * SUMMER * TIME	2.806	3	.935	.193	.900
Error	77.658	16	4.854		
Total	3544.657	32			
Corrected Total	372.397	31			

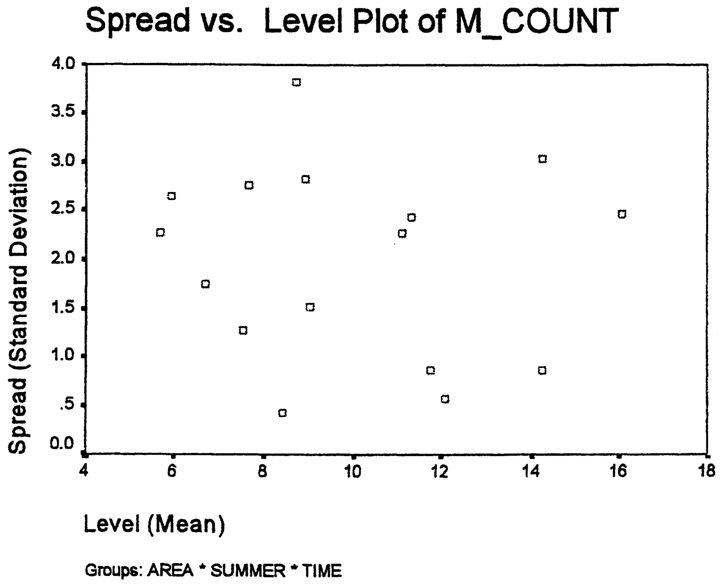
a R Squared = .791 (Adjusted R Squared = .596)

There are significant differences between summers, and a significant summer by time interaction, i.e. differences between the sample times (before and after poisoning) seem to vary from summer to summer. The type III sums of squares are the usual ones used for a balanced analysis like this. They represent the variation accounted for by the factor after adjusting for any effects that do not contain the effect in question.

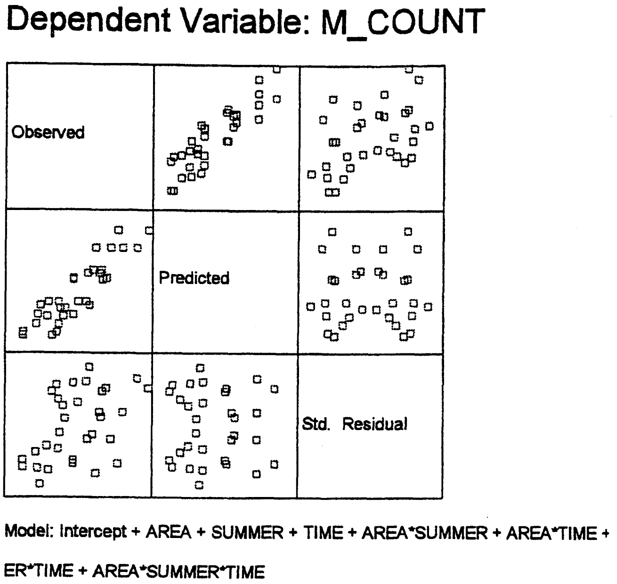
AREA * SUMMER * TIME

Dependent Variable: M_COUNT

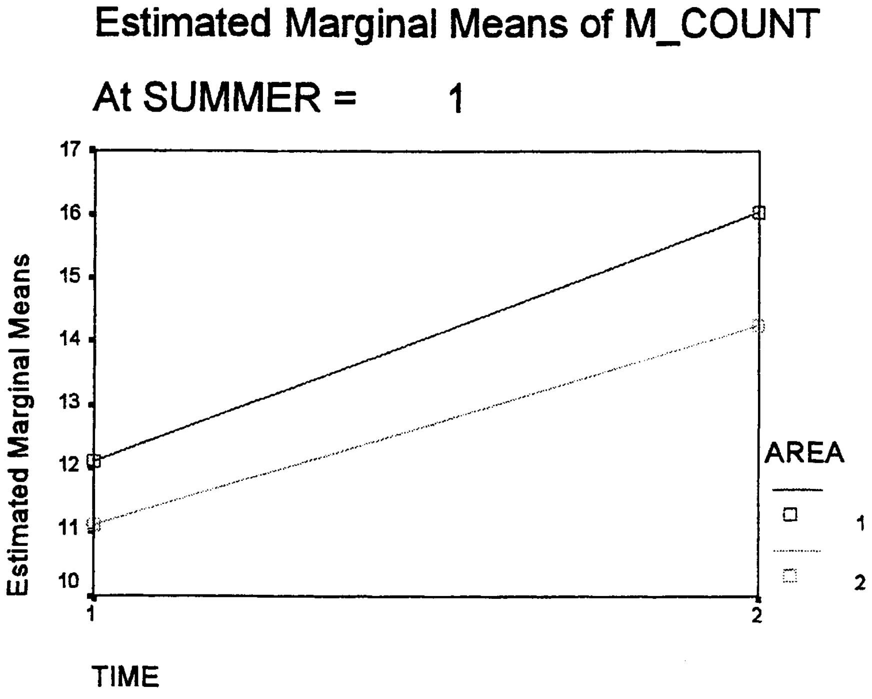
AREA	SUMMER	TIME	MEAN	STD. ERROR	95% CONFIDENCE INTERVAL	
					LOWER BOUND	UPPER BOUND
1	1	1	12.100	1.558	8.798	15.402
		2	16.040	1.558	12.738	19.342
	2	1	6.690	1.558	3.388	9.992
		2	5.680	1.558	2.378	8.982
	3	1	11.765	1.558	8.463	15.067
		2	8.400	1.558	5.098	11.702
	4	1	8.680	1.558	5.378	11.982
		2	9.005	1.558	5.703	12.307
2	1	1	11.100	1.558	7.798	14.402
		2	14.265	1.558	10.963	17.567
	2	1	7.630	1.558	4.328	10.932
		2	5.955	1.558	2.653	9.257
	3	1	14.250	1.558	10.948	17.552
		2	7.530	1.558	4.228	10.832
	4	1	11.315	1.558	8.013	14.617
		2	8.900	1.558	5.598	12.202



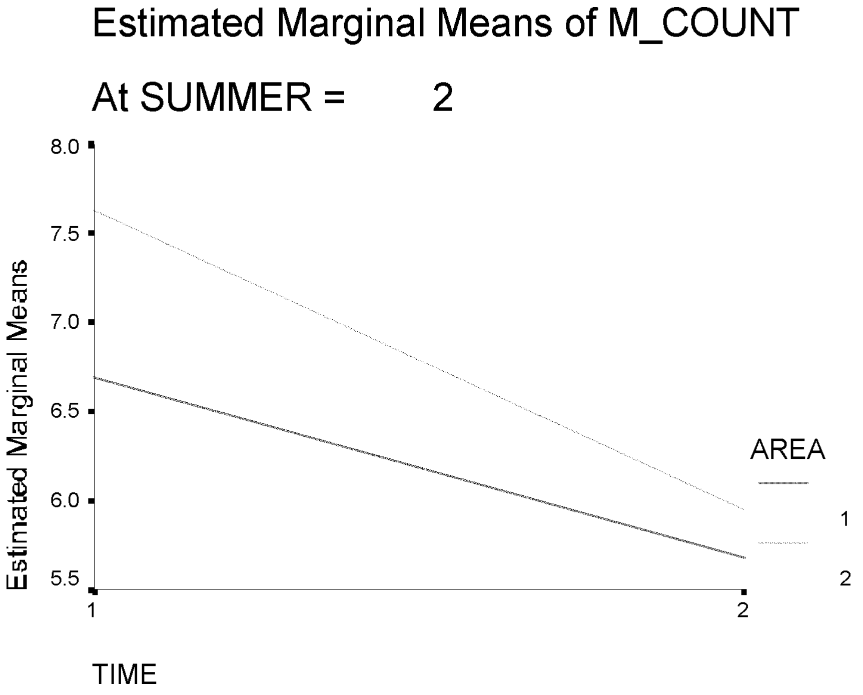
The variation in replicate counts seems fairly constant.



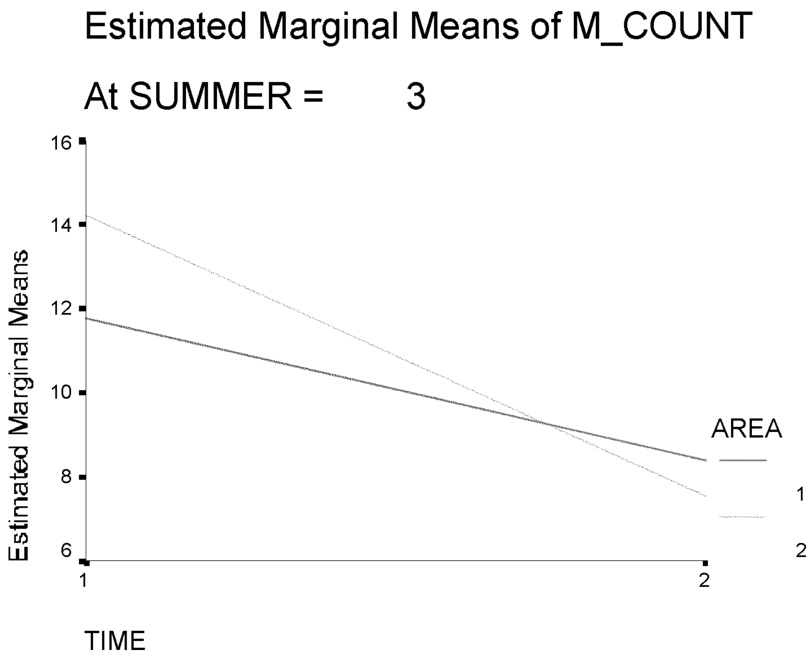
The key plot here is the one for the standardised residual against the predicted values (the middle plot on the bottom row). The assumption of a constant variance seems reasonable.



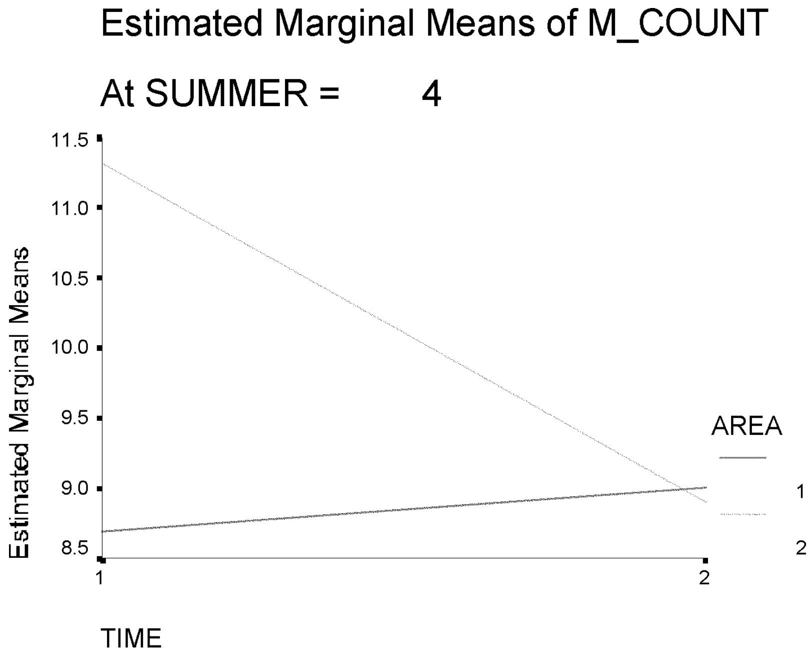
This is 1995/96. The top line above is area 1 (treated) and the bottom line is area 2 (control). Bird numbers increased pretty much the same from before and after poisoning in both areas.



This is 1996/1997. The top line is area 2 (control) and the bottom line is area 1 (treated). These are colour coded in the SPSS output. There has been a reduction in numbers in both areas from before to after poisoning, but more of a reduction in the treated area.



Here the top line at time 1 is area 2 (control) and the bottom line is area 1 (treated). The bird numbers went down after poisoning in both areas, but more in the control area.



Here the top line is area 2 (control) and the bottom line is area 1 (treated). The mean bird count decreased considerably in the control area after poisoning, but increased slightly in the treated area.

It is the differences between the the last four graphs that have resulted in the significant summer by time interaction. They certainly do not indicate a reduction in bird numbers due to poisoning in the treated area.

Repeated Measures Designs

Many data sets that are collected by DOC scientists have a repeated measures type of design. An example would be vegetation monitoring where there are three areas, one with no possum control, one with some possum control, and one with intensive possum control. Within each area five randomly placed plots might be set up, and then the percentage foliage cover measured for six years on those plots. This would then result in data of the form shown in Table 4.7.

In this example the area is a between plot factor at three levels, and the year is a within plot factor. There is a special option in SPSS to analyse data of this type, which can have more than one between plot factor, and more than one within plot factor. This should not be analysed as a factorial design with three factors (Area, Year and Plot), because that assumes that the plots in different areas match up, e.g. plot 1 in areas 1, 2 and 3 have something similar about them. Generally, this will not be true. On the other hand, the repeated measurements on one plot in one area are assumed to possibly have some similarity.

Multiple Comparisons and Contrasts

Many statistical packages for analysis of variance (including SPSS) allow the user to make comparisons of the mean level of the dependent variable for different factor combinations, with the multiple testing being allowed for in various ways. These tests are then intended to help users to understand how mean levels differ with factor levels. There are 18 different approaches that can be used in SPSS, and the help information should be read carefully before deciding which, if any, of these to use. Use of a Bonferroni correction is one possibility that is straightforward, although this may not have the power of other methods. What these multiple comparison methods do is to produce confidence intervals for the difference between the means for different factor levels. If one such interval does not include zero, then there is evidence that the population mean is not the same for the different factor levels.

Be warned that some statisticians do not like multiple comparison methods. To quote one leading expert on the design and analysis of experiments (Mead, 1988, p. 310):

> A*lthough each of these methods for multiple comparisons was developed for a particular, usually very limited, situation, in practice these methods are used very widely with no apparent thought as to their appropriateness. For many experimenters, and even editors of journals, they have become automatic in the less desirable sense of being used as a substitute for thought ... I recommend strongly that multiple comparison methods be avoided unless, after some thought and identifying the situation for which the test you are considering was proposed, you decide that the method is exactly appropriate.*

He goes on to suggest that simple graphs of means against factor levels will often be much more informative than multiple comparison tests.

On the other hand, Mead (1988) does make use of contrasts for interpreting experimental results, where these are linear combinations of mean values that reflect some aspect of the data that is of particular interest. For example, one contrast might be the mean value in year 1 compared to the mean value for all other years combined. Alternatively, a set of contrasts might be based on comparing each of the other years with year 1.

The package SPSS allows these types of comparisons to be made easily. As for the multiple comparison methods, a good starting point for using contrasts involves looking at the package's help facility.

4.6 Generalized Linear Models

The regression and analysis of variance models described in the previous two sections can be considered to be special cases of a general class of what are called generalized linear models. These were first defined by Nelder and Wedderburn (1972), and used to develop GLIM, a computer program for fitting these models to data (Francis *et al* 1993). They include many of the regression types of model that are likely to be of most use for analysing data. A very thorough description of the models and the theory behind them is provided by McCullagh and Nelder (1989).

The characteristic of generalized linear models is that there is a dependent variable Y, which is related to some other variables X_1, X_2, ..., X_p by an equation of the form

$$Y = f(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p) + \varepsilon,$$

where f(x) is one of a number of allowed functions, and ε is a random value with a mean of zero from one of a number of allowed distributions. For example, setting $f(x) = x$ and assuming a normal distribution for ε, just gives the usual multiple regression model discussed in Section 4.4.

Setting $f(x) = \exp(x)$ makes the expected value of Y equal to

$$\mathrm{E}(Y) = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p).$$

Assuming that Y has a Poisson distribution then gives what is called a log-linear model, which is a popular assumption for analysing count data (Manly, 1992, Section 8.5). The description 'log-linear' comes about because the logarithm of the expected value of Y is a linear combination of the X variables.

TABLE 4.7 THE FORM OF DATA FROM A REPEATED MEASURES EXPERIMENT WITH FIVE PLOTS IN EACH OF THREE DIFFERENT TREATMENT AREAS MEASURED FOR SIX YEARS. A MEASUREMENT OF PERCENTAGE FOLIAGE COVER IS INDICATED BY X.

AREA	PLOT	YEAR 1	YEAR 2	YEAR 3	YEAR 4	YEAR 5	YEAR 6
No possum control	1	X	X	X	X	X	X
	2	X	X	X	X	X	X
	3	X	X	X	X	X	X
	4	X	X	X	X	X	X
	5	X	X	X	X	X	X
Low possum control	1	X	X	X	X	X	X
	2	X	X	X	X	X	X
	3	X	X	X	X	X	X
	4	X	X	X	X	X	X
	5	X	X	X	X	X	X
High possum control	1	X	X	X	X	X	X
	2	X	X	X	X	X	X
	3	X	X	X	X	X	X
	4	X	X	X	X	X	X
	5	X	X	X	X	X	X

Alternatively, setting $f(x) = \exp(x)/\{1 + \exp(x)\}$ makes the expected value of Y equal to

$$\mathrm{E}(Y) = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p)/\{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{\mathrm{p}} X_{\mathrm{p}})\}.$$

This is the logistic model for a random variable Y that takes the value 0 (indicating the absence of a event) or 1 (indicating that an event occurs), where the probability of $Y = 1$ is given as a function of the X variables on the right-hand side of the equation.

There are many other possibilities for modelling within this framework, although SPSS only includes the most commonly used options, which are ordinary regression, logistic or probit regression for proportion data, log-linear models for count data, and the Cox regression model for survival data. Even then, the option within SPSS are limited and cannot handle the analysis of many common types of data.

Generalized linear models are usually fitted to data using the principle of maximum likelihood, i.e. the unknown parameter values are estimated as those values that make the probability of the observed data as large as possible. The goodness of fit of a model is then measured by the deviance, which is minus twice the maximized log-likelihood, with associated degrees of freedom equal to the number of observations minus the number of estimated parameters.

With models for count data with Poisson errors the deviance gives a direct measure of the absolute goodness of fit. If the deviance is significantly large in comparison with critical values from the chi-squared distribution then the model is a poor fit to the data. Conversely, a deviance that is not significantly large shows that the model is a reasonable fit. Similarly, with data consisting of proportions with binomial errors the deviance is an absolute measure of the goodness of fit when compared to the chi-squared distribution providing that the numbers of trials that the proportions relate to are not too small, say generally more than five.

With data from distributions other than the Poisson or binomial, or for binomial data with small numbers of trials the deviance can only be used as a relative measure of goodness of fit. The key result then is that if one model has a deviance of D_1 with υ_1 degrees of freedom and another model has a deviance of D_2 with υ_2 degrees of freedom, and the first model contains all of the parameters in the second model plus some others, then the first model gives a significantly better fit than the second model if the difference $D_2 - D_1$ is significantly large in comparison with the chi-squared distribution with $\upsilon_2 - \upsilon_1$ degrees of freedom. Comparing several models in this way is called an analysis of deviance by analogy to the analysis of variance. These tests using deviances are approximate but they should give reasonable results, except perhaps with rather small sets of data.

The individual estimates in a generalized linear model can also be tested to see whether they are significantly different from zero. This just involves comparing the estimate divided by its standard error, $z = \hat{\beta}/\mathrm{SE}(\hat{\beta})$, with critical values for the standard normal distribution. Thus if the absolute value of z exceeds 1.96 then the estimate is significantly different from zero at about the 5% level.

More about the theory and practice of generalized linear models can be found in the books by Healy (1988), McCullagh and Nelder (1989) and Lindsey (1989), as well as in the very comprehensive manual for the program GLIM (Francis *et al* 1993).

Example: Dolphin Bycatch in Trawl Fisheries

This example concerns the accidental bycatch of the common dolphin (*Delphinus delphis*) and the bottlenose dolphin (*Tursiops truncatus*) in the Taranaki Bight trawl fishery for jack mackerel (*Trachurus declivis*, *T. novae zealeandiae*, and *T. murphyi*) off the west coast of New Zealand.

The New Zealand Ministry of Fisheries puts official observers on about 10% of fishing vessels to monitor dolphin bycatch, and Table 4.8 shows a summary of the data collected by these observers for the six fishing seasons 1989/90 to 1994/95, as originally published by Baird (1996, Table 3). The table shows the number of observed trawls and the number of dolphins accidentally killed categorised by eight conditions for each fishing year: the fishing area (the northern or southern Taranaki Bight), the gear type (bottom or midwater), and the time (day or night). Excluding five cases where there were no observed trawls, this gives 43 observations on the bycatch under different conditions, in different years. Some results from fitting a generalized linear model to these data are also shown in the last two columns of the table.

Because the dependent variable (the number of dolphins killed) is a count, it is reasonable to try fitting the data using a log-linear model with Poisson errors. A simple model of that type for the ith count is

$$Y_i = T_i \exp\{\alpha(f_i) + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}\} + \varepsilon_i,$$

where T_i is the number of tows involved; $\alpha(f_i)$ depends on the fishing year f_i when the observation was collected in such a way that $\alpha(f_i) = \alpha(1)$ for observations in 1989/90, $\alpha(f_i) = \alpha(2)$ for observations in 1990/91, and so on up to $\alpha(f_i) = \alpha(6)$ for observations in 1994/95; X_{i1} is 0 for North Taranaki and 1 for South Taranaki; X_{i2} is 0 for bottom trawls and 1 for mid-water trawls; and X_{i3} is 0 for day and 1 for night. The fishing year is then being treated as a factor at six levels while the three X variables indicate the absence and presence of different particular conditions. The number of trawls is included as a multiplying factor in the equation for the model because, other things being equal, the amount of bycatch is expected to be proportional to the number of trawls made.

The model was fitted using GLIM4 (Francis *et al.* 1993) to produce the estimates that are shown in Table 4.9. There is a log-linear model option in SPSS but it is designed for a very restricted type of data consisting of counts in contingency tables, with one line of data for every time that a certain event occurs.

The estimates for the effects of different years that are shown in Table 4.8 are not easy to interpret because their estimated standard errors are quite large. Nevertheless there are significant differences between years from an analysis of deviance. The other coefficients are very significantly different from zero.

TABLE 4.8 BYCATCH OF DOLPHINS IN THE TARANAKI BIGHT TRAWL FISHERY FOR JACK MACKEREL.

				OBSERVED		DOLPHINS KILLED	
SEASON	AREA	GEAR TYPE	TIME	TOWS	OBSERVED	FITTED	RATE[1]
1989-90	North	Bottom	Day	40	0	0	0.1
	North	Bottom	Night	6	0	0	0.6
	North	Mid-water	Night	1	0	0	3.9
	South	Bottom	Day	139	0	0.6	0.4
	South	Mid-water	Day	6	0	0.2	2.8
	South	Bottom	Night	6	0	0.2	3.6
	South	Mid-water	Night	90	23	21.9	24.4
1990-91	North	Bottom	Day	2	0	0	0
	South	Bottom	Day	47	0	0	0
	South	Mid-water	Day	110	0	0	0
	South	Bottom	Night	12	0	0	0
	South	Mid-water	Night	73	0	0	0
1991-92	North	Bottom	Day	101	0	0.4	0.4
	North	Mid-water	Day	4	0	0.1	2.8
	North	Bottom	Night	36	2	1.3	3.6
	North	Mid-water	Night	3	5	0.7	24.3
	South	Bottom	Day	74	1	1.9	2.5
	South	Mid-water	Day	3	0	0.5	17.1
	South	Bottom	Night	7	5	1.5	22.1
	South	Mid-water	Night	15	16	22.6	150.4

				OBSERVED		DOLPHINS KILLED	
SEASON	AREA	GEAR TYPE	TIME	TOWS	OBSERVED	FITTED	RATE[1]
1992-93	North	Bottom	Day	135	0	0.1	0.1
	North	Mid-water	Day	3	0	0	0.5
	North	Bottom	Night	22	0	0.1	0.6
	North	Mid-water	Night	16	0	0.7	4.2
	South	Bottom	Day	112	0	0.5	0.4
	South	Bottom	Night	6	0	0.2	3.9
	South	Mid-water	Night	28	9	7.4	26.3
1993-94	North	Bottom	Day	78	0	0	0
	North	Mid-water	Day	19	0	0	0.2
	North	Bottom	Night	13	0	0	0.2
	North	Mid-water	Night	28	0	0.4	1.6
	South	Bottom	Day	155	0	0.2	0.2
	South	Mid-water	Day	20	0	0.2	1.1
	South	Bottom	Night	14	0	0.2	1.4
	South	Mid-water	Night	71	8	6.8	9.6
1994-95	North	Bottom	Day	17	0	0	0.1
	North	Mid-water	Day	80	0	0.3	0.4
	North	Bottom	Night	9	0	0	0.5
	North	Mid-water	Night	74	0	2.5	3.4
	South	Bottom	Day	41	0	0.1	0.4
	South	Mid-water	Day	73	6	1.8	2.4
	South	Bottom	Night	13	0	0.4	3.1
	South	Mid-water	Night	74	15	15.8	21.3

[1] Dolphins expected to be captured per 100 tows according to the fitted model.

Table 4.10 shows the analysis of deviance table obtained by adding effects into the model one at a time. All effects are highly significant in terms of the reduction in the deviance that is obtained by adding them in. The final model gives a reasonable fit to the data (chi-squared = 42.08 with 34 degrees of freedom, p = 0.161). Finally, the last two columns of Table 4.7 show the expected counts of dolphin deaths to compare with the observed counts, and the expected number of deaths per 100 tows. The expected number of deaths per 100 tows is usually fairly low but has the very large value of 150.4 for mid-water tows, at night, in South Taranaki, in 1991/92. In summary, it seems clear that bycatch rates seem to have varied greatly with all of the factors considered in this example

TABLE 4.9 ESTIMATES FROM FITTING A LOG-LINEAR MODEL TO THE DATA IN TABLE 4.8.

PARAMETER	ESTIMATE	STANDARD ERROR
“(1), year effect 1989/90	-7.328	0.59
“(2), year effect 1990/91	-17.52	21.38
“(3), year effect 1991/92	-5.509	0.537
“(4), year effect 1992/93	-7.254	0.612
“(5), year effect 1993/94	-8.26	0.636
“(6), year effect 1994/95	-7.463	0.551
Area effect (south v north)	1.822	0.411
Gear effect (mid-water v bottom)	1.918	0.443
Time effect (night v day)	2.177	0.451

TABLE 4.10 ANALYSIS OF DEVIANCE FOR A LOG-LINEAR MODEL FITTED TO THE DATA IN TABLE 4.8.

				CHANGE
		DEGREES OF		DEGREES OF
EFFECT	DEVIANCE	FREEDOM	DEVIANCE	FREEDOM
No effects	334.331	42		
			58.482	5
+ Year	275.851	37		
			60.712	1
+ Area	215.141	36		
			139.162	1
+ Gear type	75.981	35		
			33.912	1
+ Time	42.07	34		

[1] Significantly large at the 0.1% level, indicating that the model gives a poor fit to the data.

[2] Significantly large at the 0.1% level, indicating that bycatch is very strongly related to the effect added to the model.

4.7 Data Transformations

Data transformations are usually made for one or more of the following reasons: (a) to make the variance the same for all values of the mean, (b) to make the data more normally distributed, and (c) to ensure that the effects of different factors or variables are additive. With luck, all of these objectives will be achieved at the same time, at least to a reasonable approximation.

Sometimes the choice of a transformation to achieve the first two objectives (normality and constant variance) is straightforward, because there are some standard rules that can be applied. For example:

- With count data the square root transformation gives a more normal distribution with approximately constant variance if the original counts have Poisson distributions. In this case, replace the data values X by $\sqrt{X}$ before running an analysis.
- With data consisting of proportions (X successes out of n trials), the arc sine transformation should give a more normal type of distribution with an approximately constant variance. In this case, replace the observed proportion $p = X/n$ by $\arcsin(\sqrt{p})$, before running an analysis.
- If the original data are positively skewed, with the standard deviation being proportional to the mean, then a logarithmic transformation may produce a more normally distributed variable with a more constant variance. In this case replace the data value X with $\log(X)$ using logarithms to base 10 or base e.

There are variations on these standard transformations, for example using $\sqrt{(X + 0.5)}$ instead of $\sqrt{X}$, that are discussed in some statistics texts. There is also the possibility of choosing what transformation to use from a whole range of possible transformations. For example, a Box-Cox transformation is one for which a data value is replaced by $(X^{\tau} - 1)/\tau$ if $\tau > 0$, or $\log(X)$ if $\tau = 0$, with τ chosen to make the data as normally distributed as possible (Box and Cox, 1964; Madansky, 1988, p. 158).

Having said all this, there are good arguments for avoiding transformations as much as possible. These days log-linear models can be used to analyse count data, and logistic or probit regression can be used to analyse proportions. Approximate methods for handling these types of data are therefore now usually unnecessary.

The logarithmic transformation is however a special case. It often happens that the effects of a factor can be expected to operate multiplicatively rather than additively, in which case it is very appropriate to use logarithms in place of the original values in an analysis. In such a case it may well be found, for example, that an analysis of variance on logarithms produces a simpler regression or analysis of variance model, with residuals that have better properties than the residuals from an analysis on the original data.

4.8 Power

An important question that should be considered before any data are collected is whether the sample sizes planned will be large enough to detect effects that are of interest. With complicated data analysis methods such as an analysis of variance with several factors, or log-linear modelling it may not be easy to answer this question, in which case it may be best to collect some data and see whether this seems to be enough to detect reasonable effects. If not, then obviously more data need to be collected.

An alternative approach, which can always be used, involves producing simulated data sets and then running the proposed analysis on each of these. The simulated data might then be based on previous studies, possibly with bootstrapping to produce the new data sets (Manly, 1992, p. 329). If no past data are available then it is necessary to imagine what type of data might occur (e.g. a log-normal distribution with the same standard deviation for each factor combination).

The basic principle behind this type of empirical power study is to make up data sets that are as similar as possible to what might occur in reality and see how often an effect of interest is detected for a range of possible sample sizes. The percentage of sets of data for which an effect is significant is an estimate of the power for the proposed analysis with the particular sample size being used.

4.9 Key Points in This Module

- Statistical models describe observations on variables in terms of parameters of distributions and the nature of the random variation involved.
- The properties of discrete random variables are briefly described, and the hypergeometric, binomial, and Poisson distributions are defined.
- The properties of continuous random variables are briefly described, and the exponential, normal, and lognormal distributions are defined.
- The theory of linear regression is summarised for relating the values of a variable Y to the corresponding values of some other variables X_1, X_2, ..., X_p.
- An example on effectiveness of caging to protect woodrose flowers from browsing by possums is used to illustrate the use of multiple regression.
- The difference between factors and variables is described. The models for one, two and three factor analysis of variance are defined.
- What is meant by a repeated measures design is explained.
- A three factor example on the effect of stoat poisoning on bird counts is used to illustrate analysis of variance, where the three factors are the summer (1995/96 to 1998/99), two types of area (control and treated), and the time of counting (before and after poisoning).
- The use of multiple comparison methods and contrasts in conjunction with analysis of variance is discussed.
- The structure of generalized linear models is defined.
- The use of a generalized linear model is illustrated by an example where the number of dolphins accidentally killed during commercial fishing operations is related to the year of fishing, the type of fishing gear used, and the time of day of fishing. A log-linear model with the number of dolphins killed assumed to have a Poisson distribution is found to give a good fit to the data.
- Reasons for transforming data before an analysis, and some of the transformations that might be used, are discussed.
- A method for determining the power of a data analysis for different levels of sample sizes using simulated data is briefly covered.

4.10 Questions About This Module

After completing this module you should be able to give reasonable answers to the following questions:

1. What is the difference between a discrete and a continuous random variable?
2. Under what circumstances might each of the following distributions be used: hypergeometric, binomial, Poisson, exponential, normal, and lognormal?
3. Describe a set of data that you are familiar with where it is interesting to relate the values of a variable *Y* to other variables X_1 to X_p using multiple regression. Why exactly would the results be interesting to you?
4. Again for a situation that you are familiar with, explain how you would set up an experimental or observational study with either a two factor or a three factor factorial design. How would you ensure random sampling of replicate values within factor combinations? What would the analysis of variance table look like for your experimental results?
5. What is the difference between a factorial design and a repeated measures design?
6. Describe situations that you are familiar with where (i) a log-linear model, and (ii) logistic regression would give a useful data analysis.
7. When would you consider using each of the following transformations: square root, arcsine, and logarithmic?
8. Suppose that you have to design a study in which logistic regression will be used to analyse the results. You are required to demonstrate in advance that your proposed sample size will give a reasonable power to detect an effect of interest if it exists. How would you do this?

REFERENCES

Baird, S.J. (1996). Nonfish Species and Fisheries Interactions Working Group Report, May 1996. New Zealand Fisheries Assessment Working Group Report 96/1. Ministry of Fisheries, Wellington, New Zealand.

Box, G.E.P. and Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society* B26: 211-52.

Francis, B., Green, M. and Payne, C. (1993). *The GLIM System, Release 4 Manual.* Clarendon Press, Oxford.

Healy, M.J.R. (1988). *GLIM: An Introduction.* Clarenden Press, Oxford.

Madansky, A. (1988). *Prescriptions for Working Statisticians.* Springer-Verlag, New York.

Manly, B.F.J. (1992). *The Design and Analysis of Research Studies.* Cambridge University Press, Cambridge.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models.* Chapman and Hall, London.

Mead, R. (1988). *The Design of Experiments: Statistical Principles for Practical Application.* Cambridge University Press, Cambridge.

Mead, R., Curnow, R.N. and A.M. Hasted (1993). *Statistical Methods in Agriculture and Experimental Biology,* 2nd Edit. Chapman and Hall, London.

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society* A135: 370-84.

Neter, J., Wasserman, W. and Kutner, M.H. (1983). *Applied Linear Regression Models.* Irwin, Homewood, Illinois.

Palisade (1997). *@Risk, Risk Analysis and Simulation Add-In for Microsoft Excel or Lotus 123, Windows Version, July 1997.* Palisade Corporation, Newfield, New York 14867, USA.

Younger, M.S. (1985). *A First Course in Linear Regression.* Duxbury Press, Boston.

Contents

MODULE 5: DETECTION OF TRENDS AND CHANGE POINTS

Module 5: Detection of Trends and Change Points

SUMMARY

This module is concerned with various aspects of the detection of changes in environmental variables. Two situations are considered:

- Measurements are taken on an environmental variable at one location at various points in time, so that a single time series is available for analysis. There is interest in whether the mean level changes abruptly at any time, or whether there are gradual changes in the mean of the series.
- Measurements on an environmental variable are taken at a number of fixed sites in a spatial region at various points in time, so that several time series are available. There is interest in whether there are changes with time in the distribution of the variable (the mean, the standard deviation, etc.) over the sampled area.

A range of statistical methods are presented for handling these situations, ranging from the setting up and fitting of multiple linear regression models to non-parametric tests that require the minimum of assumptions for a valid analysis.

5.1 Introduction

The primary reason for many monitoring schemes is the detection of abrupt changes and gradual trends in important variables. There are many statistical tools available for this purpose, of which only a few will be mentioned here. To begin with, it will be assumed that values of a variable are measured at equally-spaced points in time at one location to form a time series, and that it is the analysis of this single time series that is of interest. The situation with several time series from different monitoring sites is considered later in the module.

Serial correlation is always a possibility with time series data. When present, this is usually positive so that values in the series tend to be similar when they are close in time. Serial correlation complicates analyses. At moderate to high levels it should not be ignored because if it is then the result tends to be an excessive number of significant results on statistical tests. With some of the methods described here references for modifications to allow for serial correlation are provided. If there is no mention of serial correlation then it is assumed to be absent.

5.2 The Change-Point Problem

Suppose that a variable is observed at a number of points of time, to give a time series x_1, x_2, ..., x_n. The change point problem is then to detect a change in the mean of the series if this has occurred at an unknown time between two of the observations. The problem is much easier if the point where a change might have occurred is known, which requires what is sometimes called an intervention analysis.

A formal test for the existence of a change point seems to have first been proposed by Page (1955) in the context of industrial process control. Since that time a number of other approaches have been developed, as reviewed by Jandhyala and MacNeill (1986), and methods for detecting a change in the mean of an industrial process through control charts and related techniques have been considerably developed (Montgomery, 1991). Bayesian methods have also been investigated (Carlin *et al.*, 1992), and Sullivan and Woodall (1996) suggest a useful approach for examining data for a change in the mean and/or the variance at an unknown time.

It is not valid to look at the time series, decide where a change point may have occurred, and then test for a significant difference between the means for the observations before and after the change. This is because the maximum mean difference between two parts of the time series may be quite large by chance alone and is liable to be statistically significant if it is tested ignoring the way that it was selected.

A Randomization Test for a Change-Point

One way to overcome the problem of not knowing where a change may have occurred before looking at the data involves using a randomization test (Manly, 1997, Chapter 1). Suppose therefore that there is interest in testing for a change in the mean between two unspecified observations against the null hypothesis that the series consists of independent observations from the same distribution. This test is not valid when there is serial correlation in the series being considered.

Consider the first i observations in the series and suppose that these have the mean $\bar{x}_{1i}$ and variance s_{1i}^2, taking $s_{1i}^2 = 0$ if $i = 1$. Similarly, consider the last n - i observations in the series and let these have mean and variance $\bar{x}_{2i}$ and s_{2i}^2. An appropriate test statistic is then $t_{\max}$, the maximum of the t-statistics t_1, t_2, ..., $t_{n\text{-}1}$ where

$$t_1 = \left|\bar{x}_{1i} - \bar{x}_{2i}\ \right| / \{s_i \sqrt{\{1/i + 1(n-i)\}} \,,$$

is the usual t-statistic for comparing the means of the first i and last n-i observations, with

$$s_i^2 = \{(i-1)s_{1i}^2 + (n-i-1)s_{2i}^2\}/(n-2)$$

being the usual pooled estimate of the variance. The test then involves comparing $t_{\max}$ with the distribution that is generated by computing the same statistic using the series values in a random order.

To be precise, suppose that the observed value of $t_{\max}$ is $t_{\max,1}$, and R-1 random orderings of the series yield values of $t_{\max,2}$, to $t_{\max,\mathrm{R}}$. Then if $t_{\max,1}$ is included among the largest 100α% of the values $t_{\max,1}$ to $t_{\max,\mathrm{R}}$ it can be declared to be significantly large at the 100α% level. The justification for this is that if the null hypothesis is true then the observed ordering of values in the series is just another random order so that the probability of it yielding one of the largest 100α% of the set of values is exactly α (assuming that there are no ties).

The reason for using t_i to measure the difference between the first *i* values in the series and the rest is that this measure will have approximately the same distribution for all *i*. This is ensured by dividing the mean difference between $\bar{x}_{1i}$ and $\bar{x}_i$ by the estimated standard error of this difference. Using $\bar{x}_{1i}$ - $\bar{x}_{2i}$ directly, for example, would mean that large differences would tend to be at one or other end of the series and these differences would often dominate $t_{\max}$.

Example: Flow Variation in a South Island River

As an example of the randomization test just described, consider the data shown in Table 5.1, and displayed graphically in Figure 5.1. These data were collected at one point along a South Island river in order to determine whether flow rates have become more variable in recent years. This question is important to users of the river for irrigation as fast changes in the flow rate are different are averages for the first, second, third and fourth quarters of the year 1987 to 1996, and the first two quarters of 1997. With these quarterly observations there is little indication of seasonal effects or serial correlation.

TABLE 5.1 MEAN VALUES FOR THE VARIABILITY OF FLOW RATES, MEASURED AS ABSOLUTE PERCENTAGE CHANGES IN FLOW RATES IN A SOUTH ISLAND RIVER FROM 1987 TO 1997.

QUARTER	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	MEAN
1	5.80	5.74	3.84	3.98	3.12	3.41	3.20	4.18	3.51	4.48	8.25	4.50
2	4.66	6.98	4.92	4.22	3.54	4.48	4.01	5.18	4.68	5.02	9.68	5.22
3	4.15	6.20	3.65	3.65	4.00	1.30	3.13	3.41	5.52	7.08		4.21
4	5.37	3.48	5.56	3.89	4.27	4.08	4.45	4.70	5.02	10.38		5.12

The calculations needed can be carried out using a computer program called CHANGEPT.EXE, available from BFJM (Brian F.J.Manly). The input and output are shown in Table 5.2. It is found that the largest absolute t value for a difference in means between two stretches of the series is 7.96 for the difference between the periods before and after the middle of 1996. This is therefore $t_{\max,1}$. When 9999 random permutations of the series were made to generate $t_{\max,2}$ to $t_{\max,10000}$ it was found that $t_{\max,1}$ was never equalled or exceeded with the randomized sets of data. Therefore, the test statistic for the observed series is significantly different from zero at the 1 in 10000, or 0.01% level. There is very strong evidence that the mean was not constant from 1987 to 1997.

The test just carried out allows for multiple testing. If the individual t-statistics are tested then it is found that there is a significant difference at the 5% level if the time series is split between any of the last 18 observations.

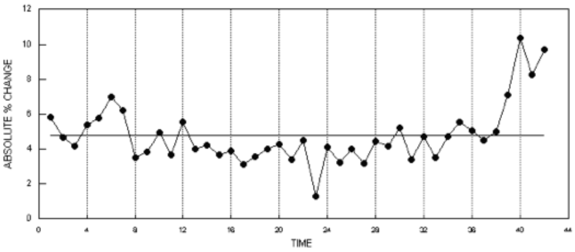


Figure 5.1 Mean values for the variability of flow rates, measured as absolute percentage changes in flow rates in a South Island river from 1987 to 1997. Time is 1 for the first quarter in 1987, increasing to 42 for the second quarter in 1997. The horizontal line is at the overall mean.

TABLE 5.2 INPUT AND OUTPUT FOR THE PROGRAM FOR A RANDOMIZATION TEST FOR A CHANGEPOINT IN THE AMOUNT OF VARIATION IN FLOW RATES IN A SOUTH ISLAND RIVER.

Input to CHANGEPT.EXE

```
South Island River Variability
42
5.80  4.66  4.15  5.37  5.74  6.98  6.20  3.48  3.84  4.92
3.65  5.56  3.98  4.22  3.65  3.89  3.12  3.54  4.00  4.27
3.41  4.48  1.30  4.08  3.20  4.01  3.13  4.45  4.18  5.18
3.41  4.70  3.51  4.68  5.52  5.02  4.48  5.02  7.08 10.38
8.25  9.68
```

Output from the Program

```
############################################################
#                        CHANGEPT                          #
# Program to test for a change point in a time series by   #
# testing the maximum t-statistic for a difference between #
# values in two parts of the series.                       #
#                                                          #
#                Version 1.0 (August, 1997)                #
############################################################

Date:  2/10/1999  Time: 16:43

Data: South Island River Variability (Year, Quarter, Time, %
Change)

Length of series =   42

Observations
5.8    4.7    4.2    5.4    5.7    7.0    6.2    3.5    3.8    4.9
3.7    5.6    4.0    4.2    3.7    3.9    3.1    3.5    4.0    4.3
3.4    4.5    1.3    4.1    3.2    4.0    3.1    4.4    4.2    5.2
3.4    4.7    3.5    4.7    5.5    5.0    4.5    5.0    7.1   10.4
8.2    9.7

Random number seed =    666

For the observed series maximum t =   7.96

Values for changes after times 1, 2, 3, etc.
0.61   0.39   0.11   0.28    0.52   1.06   1.35   0.98  0.73   0.73
0.48   0.63   0.46   0.35    0.13  -0.03  -0.33  -0.55 -0.68  -0.77
-1.02 -1.07  -1.75  -1.90   -2.24  -2.44  -2.87  -3.00 -3.23  -3.21
-3.74 -3.93  -4.63  -5.00   -5.06  -5.49  -6.57  -7.96 -7.53  -4.22
-3.22
```

```
Significance level for maximum t-value (% of randomizations
giving absolute values as large or larger than the observed
value) =    0.01

Significance levels for individual times of change
42.36  66.57  91.52  79.02  61.88  29.03  17.32  33.70  48.37  48.36
64.28  54.25  65.26  73.94  90.06  97.48  74.37  59.98  51.26  46.21
32.62  29.89   8.86   6.17   2.88   1.82   0.52   0.42   0.28   0.29
0.04   0.05   0.01   0.01   0.02   0.01   0.01   0.01   0.01   0.32
4.62
```

5.3 Trend Detection

A common problem in terms of examining changes in an environmental time series is the detection of a monotonic trend (Taylor and Loftis, 1989; Zetterqvist 1991; Harcum *et al*, 1992). Complications include seasonality and serial correlation in the observations.

When considering the evidence for a trend in a time series it is most important to define the time scale that is of interest. As pointed out by Loftis *et al.* (1991), in most analyses that have been conducted in the past there has been an implicit assumption that what is of interest is a trend over the time period for which data happen to be available. For example, if 20 yearly results are known, then a 20 year trend has implicitly been of interest. This then means that an increase in the first ten years followed by a decrease in the last ten years to the original level has been considered to give no overall trend, with the intermediate changes possibly being thought of as due to serial correlation. This is clearly not appropriate if systematic changes over a five year period (say) are thought of by managers as being ‘trend’.

Regression Methods

When serial correlation is negligible, regression analysis provides a very convenient framework for testing for trend. In simple cases, a regression of the measured variable against time will suffice, with a test to see whether the coefficient of time is significantly different from 0. In more complicated cases there may be a need to allow for seasonal effects and the influence of one or more exogenous variables. Thus, for example, if the dependent variable is measured monthly then the type of model that might be investigated is

$$Y_{ij} = ß_1M_{1j} + ß_2M_{2j} + \ldots + ß_{12}M_{12j} + \alpha X_{ij} + \theta t_{ij} + \varepsilon_{ij}$$

where Y_{ij} is the observation in month j of year i, M_{kj} is a month indicator that is 1 when $j = k$ or otherwise 0, X_{ij} is a relevant covariate in month j in year i, t_{ij} is the time in months from the start of the series, and ε_{ij} is random noise. Then the parameters $ß_1$ to $ß_{12}$ allow for differences in Y values related to months of the year, the parameter α allows for an effect of the covariate, and θ is the change in Y per month after adjusting for any seasonal effects and effects due to differences in X from month to month. If the estimate of θ obtained by fitting the regression equation is significant then this provides the evidence for a trend.

A small change can be made to the model in order to test for the existence of seasonal effects. One of the month indicators (say the first or last) can be omitted from the model and a constant term introduced. A comparison between the fit of the model with just a constant in and the model with the constant and month indicators then shows whether the mean value appears to vary from month to month.

If a regression equation such as the one above is fitted to data then a check for serial correlation in the error variable ε_{ij} should always be made using the Durbin-Watson (1951) test (Manly, 1992, p. 108). If serial correlation may be present then the model can still be used. However, it should be fitted using a method that is more appropriate than ordinary least-squares. Edwards and Coull (1987), Judge *et al.* (1988, pp. 388-93 and 532-8), Neter *et al.* (1983, Chapter 13) and Zetterqvist (1991) all describe how this can be done. There is more than one method available and the calculations can be done in a standard package like SPSS with some manipulation of the data.

The Mann-Kendall Test

Researchers in the area of environmental monitoring have tended to favour non-parametric tests for monotonic trends in recent years because of the need to analyse large numbers of series with a minimum amount of time devoted to considering the special needs of each series. Thus transformations to normality, choosing the order of autoregressive models etc. are to be avoided if possible. The non-parametric methods that currently appear to be most useful are the Mann-Kendall test, the seasonal Kendall test, and the seasonal Kendall test with a correction for serial correlation (Taylor and Loftis, 1989; Harcum *et al.* 1992). Unfortunately, these tests are generally missing from standard statistical packages, including SPSS, although they are part of packages designed specifically for analysing water quality data.

The Mann-Kendall test is appropriate for data that do not display seasonal variation, or for seasonally corrected data. For a series $X_1, X_2, \ldots, X_n$ the test statistic is the sum of the signs of the differences between all pairwise observations,

$$S = \sum_{i=2}^{n} \sum_{j=1}^{i-1} sign\ (x_i - x_j)$$

where sign(z) is -1 for $z < 0$, 0 for $z = 0$, and +1 for $z > 0$. For a series of values in a random order the expected value of S is zero and the variance is

$$\mathrm{Var(S)} = n(n\text{-}1)(2n\text{+}5)/18.$$

To test whether S is significantly different from zero it is best to use a special table if n is ten or less and S is not close to zero in comparison with its standard deviation (Helsel and Hirsch, 1992, p. 469). For larger values of n $Z_S = S/\sqrt{\mathrm{Var(S)}}$ can be compared with critical values for the standard normal distribution.

To accommodate seasonality in the series being studied, Hirsch *et al.* (1982) introduced the seasonal Kendall test. This involves calculating the statistic S separately for each of the seasons of the year (weeks, months, etc.) and uses the sum for an overall test. Thus if S_j is the value of S for season j, then on the null hypothesis of no trend $S_T = \Sigma S_j$ has an expected value of 0 and a variance of $\mathrm{Var}(S_T) = \Sigma \mathrm{Var}(S_j)$. The statistic $Z_T = S_T/\sqrt{\mathrm{Var}(S_T)}$ can therefore be used for an overall test of trend by comparing it with the standard normal distribution.

Apparently the normal approximation is good providing that the total series length is 25 or more.

An assumption with the seasonal Kendall test is that the statistics for the different seasons are independent. When this is not the case an adjustment for serial correlation can be made when calculating Var(ΣS_T) (Hirsch and Slack, 1984; Zetterqvist, 1991). An allowance for missing data can also be made in this case.

Example: CPUE at Lake Taupo (Appendix Data Set 10)

As an example of testing for trend using the Mann-Kendall statistic consider the data in Table 5.3 that were obtained from surveys of anglers at Lake Taupo for the summers of 1991/92 to 1998/99 (Data Set 10 from Appendix 3). The catch per unit effort (CPUE) for trout is the total number of fish kept for all surveyed anglers, divided by the total fishing time of these anglers. The Mann-Kendall test can be used to test for trend in the CPUE series.

TABLE 5.3 TOTAL HOURS FISHED, TROUT KEPT, AND CATCH PER UNIT EFFORT (CPUE) AT LAKE TAUPO.

SUMMER	FISHING HOURS	FISH KEPT	CPUE
1991/92	321.7	47	0.146
1992/93	260.3	63	0.242
1993/94	1493.5	366	0.245
1994/95	1840.6	410	0.223
1995/96	3185.9	481	0.151
1996/97	3358.8	633	0.188
1997/98	2387.4	425	0.178
1998/99	3087.7	516	0.167

The calculation of S is shown in Table 5.4. The observed value is S = -6, suggesting a downward trend. With a series of length n = 8 a special table sometimes has to be used to decide whether or not the result is significant. However, in the present case the standard error associated with S is 8.1. Clearly, therefore, the observed value of -6 is not significantly different from zero. It follows that the Mann-Kendall test gives no evidence of a trend.

TABLE 5.4 CALCULATION OF THE MANN-KENDALL TEST STATISTIC FOR TESTING FOR TREND IN THE CPUE.

		SUMMER							
		1	2	3	4	5	6	7	8
SUMMER	CPUE	0.146	0.242	0.245	0.223	0.151	0.188	0.178	0.167
1	0.146								
2	0.242	1							
3	0.245	1	1						
4	0.223	1	-1	-1					
5	0.151	1	-1	-1	-1				
6	0.188	1	-1	-1	-1	1			
7	0.178	1	-1	-1	-1	1	-1		
8	0.167	1	-1	-1	-1	1	-1	-1	
								S =	-6
								Var(S) =	65.3
								SE(S) =	8.1

NB. The body of the table shows values for sign(x_i - x_j) for rows *i* and columns *j*. The sum is S = -6, with standard error 8.1.

Decomposition of Seasonal Series

Many statistical packages, including SPSS, have an option to decompose a seasonal time series into a trend component, seasonal components, and a random, 'unexplained' part. Either an additive or a multiplicative model is used, where the additive model assumes that

$$X_i = T_i + S_i + e_i$$

and the multiplicative model assumes that

$$X_i = T_i (S_i + e_i)$$

where, in both cases, X_i is the ith observation in the series, which is expressed in terms of the trend T_i, the seasonal component S_i, and the unexplained part e_i of the series. The trend is estimated by a moving average of the appropriate length. For example, a moving average of 12 is used for a monthly series because in this way all average values include all months of the year. Hence seasonal effects are assumed to be eliminated with a moving average of this length.

A disadvantage of this way of analysing a time series is that there is no guide to the accuracy or the significance of the different components in the model. Nevertheless, this type of analysis may be interesting for an initial look at data.

Example: Seasonal Decomposition of Flow Variation Series

The seasonal decomposition option in SPSS was used with the variability of flow rate data in Table 5.1 The output obtained is shown in Table 5.5 and plots of the components of the series are shown in Figure 5.2. According to this analysis, the mean of the series changed from 1987 to 1997, first decreasing, and then increasing (see the trend plot in Figure 5.2). The seasonal pattern is estimated to be quite distinct, with the series increasing from quarter 1 to quarter 2, then decreasing in quarter 3, and then increasing again (see the seasonal factors plot in Figure 5.2).

TABLE 5.5 SEASONAL DECOMPOSITION OF THE FLOW VARIABILITY DATA IN TABLE 5.1 USING THE SPSS OPTION FOR THIS PURPOSE, ASSUMING AN ADDITIVE MODEL. THE VARIABLE IS CALLED 'CHANGE'.

```
Results of SEASON procedure for variable CHANGE.
Additive Model.  Equal weighted MA method.  Period = 4.

                                            Seasonally Smoothed
                     Moving           Seasonal adjusted   trend- Irregular
       DATE_   CHANGE averages   Ratios factors   series   cycle component

Q1   1987   5.800        .        .  -.421   6.221   5.300      .921
Q2   1987   4.660        .        .   .364   4.296   5.021     -.725
Q3   1987   4.150    4.995    -.845  -.396   4.546   4.938     -.392
Q4   1987   5.370    4.980     .390   .453   4.917   5.231     -.314
Q1   1988   5.740    5.560     .180  -.421   6.161   5.855      .306
Q2   1988   6.980    6.073     .908   .364   6.616   5.923      .693
Q3   1988   6.200    5.600     .600  -.396   6.596   5.500     1.097
Q4   1988   3.480    5.125   -1.645   .453   3.027   4.663    -1.636
Q1   1989   3.840    4.610    -.770  -.421   4.261   4.288     -.027
Q2   1989   4.920    3.973     .947   .364   4.556   4.268      .288
Q3   1989   3.650    4.493    -.842  -.396   4.046   4.458     -.412
Q4   1989   5.560    4.528    1.032   .453   5.107   4.514      .592
Q1   1990   3.980    4.353    -.373  -.421   4.401   4.358      .043
Q2   1990   4.220    4.352    -.132   .364   3.856   4.112     -.256
Q3   1990   3.650    3.935    -.285  -.396   4.046   3.852      .194
Q4   1990   3.890    3.720     .170   .453   3.437   3.613     -.176
Q1   1991   3.120    3.550    -.430  -.421   3.541   3.588     -.047
Q2   1991   3.540    3.638    -.098   .364   3.176   3.628     -.452
Q3   1991   4.000    3.733     .268  -.396   4.396   3.838      .558
Q4   1991   4.270    3.805     .465   .453   3.817   3.911     -.094
Q1   1992   3.410    4.040    -.630  -.421   3.831   3.717      .114
Q2   1992   4.480    3.365    1.115   .364   4.116   3.427      .689
Q3   1992   1.300    3.318   -2.018  -.396   1.696   3.114    -1.418
Q4   1992   4.080    3.265     .815   .453   3.627   3.253      .374
Q1   1993   3.200    3.148     .053  -.421   3.621   3.403      .218
Q2   1993   4.010    3.605     .405   .364   3.646   3.651     -.004
Q3   1993   3.130    3.698    -.567  -.396   3.526   3.787     -.261
Q4   1993   4.450    3.943     .508   .453   3.997   4.079     -.082
Q1   1994   4.180    4.235    -.055  -.421   4.601   4.307      .294
Q2   1994   5.180    4.305     .875   .364   4.816   4.390      .427
Q3   1994   3.410    4.368    -.957  -.396   3.806   4.231     -.424
Q4   1994   4.700    4.200     .500   .453   4.247   4.150      .097
Q1   1995   3.510    4.075    -.565  -.421   3.931   4.293     -.362
Q2   1995   4.680    4.603     .077   .364   4.316   4.606     -.290
Q3   1995   5.520    4.682     .838  -.396   5.916   4.927      .989
Q4   1995   5.020    4.925     .095   .453   4.567   4.923     -.356
Q1   1996   4.480    5.010    -.530  -.421   4.901   5.171     -.270
Q2   1996   5.020    5.400    -.380   .364   4.656   5.913    -1.257
Q3   1996   7.080    6.740     .340  -.396   7.476   7.241      .236
Q4   1996  10.380    7.683    2.698   .453   9.927   8.450     1.477
Q1   1997   8.250    8.848    -.598  -.421   8.671   9.305     -.634
Q2   1997   9.680        .        .   .364   9.316   9.421     -.105
```

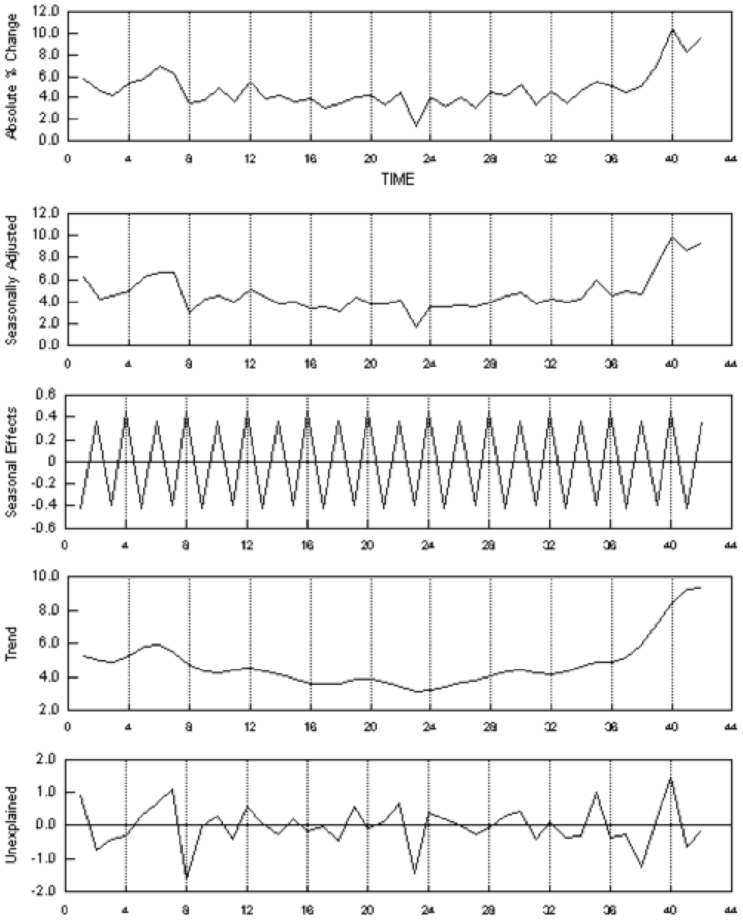


Figure 5.2 Components of the change time series as estimated from the seasonal decomposition option in SPSS.
In order from the top to the bottom plot these are (a) the original series, (b) the seasonally adjusted series, which is the original series minus the estimated seasonal factors, (c) the estimated seasonal factors, (d) the estimated trend in the series, which is the seasonally adjusted values after smoothing, and (e) the part of the original series which is not accounted for by the trend or the seasonal effects.

The problem with the analysis just done is that we have no idea how accurate the results are. For example, there is no way of knowing whether the seasonal variation that is apparently present could actually be what is estimated for a series with no seasonal variation at all. Regression analysis is distinctly better in this respect. Therefore the data will now be reanalysed using a regression model.

Regression Model for Flow Variation Series

The model used is

$$\text{Change} = \beta_0 + \beta_1\text{Qtr1} + \beta_2\text{Qtr2} + \beta_3\text{Qtr3} + \beta_4\text{T1} + \beta_5\text{T2} + \beta_6\text{T3} + \varepsilon$$

where Change is the flow variation, Qtr1 to Qtr3 are quarter indicators such that Qtr_i is 1 for an observation in quarter i or is otherwise 0, and T1 to T3 are the time, time squared, and time cubed, where time is 1 for the first observation and 42 for the last observation.

The output from a regression analysis done using SPSS is shown in Table 5.6. It is apparent that T1 to T3 are all needed in the model but there is no evidence of the existence of seasonal effects. It seems, therefore, that there was a non-linear trend in the mean level of the series, but little if any variation associated with the time of the year.

This example would benefit from more diagnostic tests of assumptions. However, these tests will not be done here. One thing that it is worth noting, however, is that the Durbin-Watson statistic (which measures the degree of autocorrelation displayed by the regression residuals) has the value 2.012. With no autocorrelation at all the expected value of this statistic is approximately 2.0. Therefore, there is certainly no evidence for autocorrelation with this data set.

It is values of the Durbin-Watson statistic less than 2.0 that indicate positive autocorrelation (a tendency for observations that are close in time to have similar values for the unexplained part of observations). Details of how to see whether an observed value is significant are provided by Manly (1992, Section 4.9). Methods for dealing with autocorrelation if this is necessary have been referenced above.

TABLE 5.6 OUTPUT FROM A REGRESSION ANALYSIS ON THE FLOW VARIATION DATA CARRIED OUT USING SPSS.

Variables Entered/Removed

MODEL	VARIABLES ENTERED	VARIABLES REMOVED	METHOD
1	T1	.	Enter
2	T2	.	Enter
3	T3	.	Enter

a All requested variables entered.
b Dependent Variable: CHANGE

Model Summary

MODEL	R	R SQUARE	ADJUSTED R SQUARE	STD. ERROR OF THE ESTIMATE	CHANGE STATISTICS					DURBIN-WATSON
					R SQUARE CHANGE	F CHANGE	df1	df2	SIG. F CHANGE	
1	.290	.084	.061	1.6592	.084	3.662	1	40	.063	
2	.768	.589	.568	1.1253	.505	47.957	1	39	.000	
3	.830	.689	.665	.9913	.100	12.254	1	38	.001	2.012

a Predictors: (Constant), T1
b Predictors: (Constant), T1, T2
c Predictors: (Constant), T1, T2, T3
d Dependent Variable: CHANGE

ANOVA

MODEL		SUM OF SQUARES	df	MEAN SQUARE	F	SIG.
1	Regression	10.082	1	10.082	3.662	.063
	Residual	110.118	40	2.753		
	Total	120.200	41			
2	Regression	70.813	2	35.406	27.959	.000
	Residual	49.388	39	1.266		
	Total	120.200	41			
3	Regression	82.855	3	27.618	28.103	.000
	Residual	37.345	38	.983		
	Total	120.200	41			

a Predictors: (Constant), T1
b Predictors: (Constant), T1, T2
c Predictors: (Constant), T1, T2, T3
d Dependent Variable: CHANGE

Coefficients

		UNSTANDARDIZED		STANDARDIZED	t	SIG.	95% CONFIDENCE INTERVAL FOR B	
		COEFFICIENTS		COEFFICIENTS				
MODEL		B	STD. ERROR	BETA			LOWER BOUND	UPPER BOUND
1	(Constant)	3.897	.521		7.475	.000	2.843	4.951
	T1	4.042E-02	.021	.290	1.914	.063	-.002	.083
2	(Constant)	6.785	.547		12.409	.000	5.679	7.891
	T1	-.353	.059	-2.532	-6.026	.000	-.472	-.235
	T2	9.159E-03	.001	2.910	6.925	.000	.006	.012
3	(Constant)	5.150	.671		7.677	.000	3.792	6.508
	T1	7.760E-02	.134	.556	.581	.565	-.193	.348
	T2	-1.561E-02	.007	-4.959	-2.177	.036	-.030	-.001
	T3	3.840E-04	.000	4.942	3.501	.001	.000	.001

a Dependent Variable: CHANGE

Excluded Variables

		BETA IN	t	SIG.	PARTIAL CORRELATION	COLLINEARITY STATISTICS
MODEL						TOLERANCE
1	T2	2.910	6.925	.000	.743	5.967E-02
	T3	1.910	7.953	.000	.786	.155
	QTR1	.092	.604	.549	.096	.997
	QTR2	-.137	-.906	.370	-.144	1.000
	QTR3	.046	.297	.768	.048	.997
2	T3	4.942	3.501	.001	.494	4.102E-03
	QTR1	.091	.883	.383	.142	.997
	QTR2	-.135	-1.327	.192	-.210	1.000
	QTR3	.044	.426	.672	.069	.997
3	QTR1	.120	1.336	.190	.215	.989
	QTR2	-.135	-1.517	.138	-.242	1.000
	QTR3	.016	.175	.862	.029	.989

a Predictors in the Model: (Constant), T1
b Predictors in the Model: (Constant), T1, T2
c Predictors in the Model: (Constant), T1, T2, T3
d Dependent Variable: CHANGE

Casewise Diagnostics

CASE NUMBER	STD. RESIDUAL	CHANGE	PREDICTED VALUE	RESIDUAL
1	.592	5.80	5.2127	.5873
2	-.591	4.66	5.2462	-.5862
3	-1.113	4.15	5.2531	-1.1031
4	.136	5.37	5.2356	.1344
5	.549	5.74	5.1962	.5438
6	1.859	6.98	5.1371	1.8429
7	1.149	6.20	5.0605	1.1395
8	-1.502	3.48	4.9689	-1.4889
9	-1.033	3.84	4.8645	-1.0245
10	.172	4.92	4.7496	.1704
11	-.985	3.65	4.6266	-.9766
12	1.072	5.56	4.4976	1.0624
13	-.389	3.98	4.3651	-.3851
14	-.011	4.22	4.2314	-1.1385E-02
15	-.453	3.65	4.0987	-.4487
16	-.080	3.89	3.9693	-7.9282E-02
17	-.732	3.12	3.8455	-.7255
18	-.191	3.54	3.7298	-.1898
19	.379	4.00	3.6242	.3758
20	.745	4.27	3.5313	.7387
21	-.044	3.41	3.4532	-4.3155E-02
22	1.097	4.48	3.3922	1.0878
23	-2.069	1.30	3.3507	-2.0507
24	.756	4.08	3.3310	.7490
25	-.137	3.20	3.3354	-.1354
26	.649	4.01	3.3662	.6438
27	-.298	3.13	3.4256	-.2956

CASE NUMBER	STD. RESIDUAL	CHANGE	PREDICTED VALUE	RESIDUAL
28	.942	4.45	3.5160	.9340
29	.545	4.18	3.6398	.5402
30	1.393	5.18	3.7991	1.3809
31	-.591	3.41	3.9963	-.5863
32	.470	4.70	4.2337	.4663
33	-1.012	3.51	4.5136	-1.0036
34	-.160	4.68	4.8384	-.1584
35	.312	5.52	5.2102	.3098
36	-.617	5.02	5.6315	-.6115
37	-1.639	4.48	6.1045	-1.6245
38	-1.626	5.02	6.6315	-1.6115
39	-.136	7.08	7.2148	-.1348
40	2.545	10.38	7.8568	2.5232
41	-.312	8.25	8.5597	-.3097
42	.357	9.68	9.3259	.3541

a Dependent Variable: CHANGE

Residuals Statistics

	MINIMUM	MAXIMUM	MEAN	STD. DEVIATION	N
Predicted Value	3.3310	9.3259	4.7660	1.4216	42
Residual	-2.0507	2.5232	-4.5466E-16	.9544	42
Std. Predicted Value	-1.009	3.208	.000	1.000	42
Std. Residual	-2.069	2.545	.000	.963	42

a Dependent Variable: CHANGE

5.4 Use of Control Charts

Control charts are used to monitor industrial processes (Montgomery, 1991) and they can be used equally well with environmental data. An advantage of this approach is that the graphs are relatively easy to interpret, and wide experience has demonstrated their effectiveness. The program SPSS includes an option for producing Shewhart control charts (the 'usual' type), but it is not very flexible. It may therefore be better to set up control charts in a spreadsheet rather than use this option.

Example: Ring Widths of Andean Alders

As an example, consider the data Table 5.7 (kindly provided by Dr Alfredo Grau of Invermay Agricultural Centre, Mosgiel, New Zealand). These are yearly ring widths for samples of six Andean alders (*Alnus acuminanta*) taken from Taficillo Ridge at about 1700 m altitude, in Tucuman, Argentina, every year from 1970 to 1989. A question here is whether there is evidence of changes in the mean width from year to year, which can be addressed by constructing control charts for the sample means and ranges (maximum values - minimum values), as shown in Figure 5.3.

TABLE 5.7 RING WIDTHS FOR SAMPLES OF ANDEAN ALDERS TAKEN FROM TAFICILLO RIDGE, TUCUMAN, ARGENTINA AT AN ALTITUDE OF ABOUT 1700 M.

	Ring Widths							
Year	1	2	3	4	5	6	Mean	Range
70	1.8	7.3	1.1	3.3	5.0	5.5	4.00	6.2
71	2.7	2.6	4.0	1.4	1.0	3.0	2.45	3.0
72	4.0	4.0	3.6	3.0	2.0	4.4	3.50	2.4
73	5.0	3.5	2.0	5.4	6.0	1.3	3.87	4.7
74	5.7	0.5	4.6	4.7	3.5	2.8	3.63	5.2
75	5.1	3.4	8.4	3.5	4.5	4.0	4.82	5.0
76	2.6	3.3	7.4	4.2	5.8	6.3	4.93	4.8
77	7.5	4.7	1.0	7.4	2.6	8.8	5.33	7.8
78	8.5	5.5	4.3	3.3	4.7	2.4	4.78	6.1
79	3.0	2.0	4.7	3.0	4.1	3.6	3.40	2.7
80	4.0	3.9	2.5	3.0	2.6	0.7	2.78	3.3
81	3.8	2.9	4.8	3.9	2.4	2.0	3.30	2.8
82	2.6	4.0	2.0	2.9	2.7	1.1	2.55	2.9
83	4.9	1.2	3.5	2.4	2.2	0.9	2.52	4.0
84	6.0	6.6	3.1	1.8	6.0	3.1	4.43	4.8
85	0.2	0.7	1.3	1.5	3.6	0.5	1.30	3.4
86	2.5	2.9	0.5	1.3	0.8	0.3	1.38	2.6
87	4.1	4.1	2.6	0.9	0.5	2.4	2.43	3.6
88	0.6	3.4	2.5	0.3	2.3	0.6	1.62	3.1
89	1.7	0.5	0.4	0.8	0.9	0.8	0.85	1.3
						Mean	3.19	4.0

The method for constructing the control chart for sample means, which is often referred to as an $\bar{x}$ chart, involves the following stages:

(a) Assuming that the mean did not change over the 20 year period, this is estimated by the overall mean of 3.19. Similarly, the standard deviation of ring widths is assumed to have remained constant and this is estimated on the basis of a known relationship between the mean range and the standard deviation for samples of size six from a normal distribution. From Table 5.8, this relationship is $\sigma = 0.395W$, where W is the sample mean range, giving in the present case $\bar{x} \approx 0.395x4.0 = 1.58$. A similar result holds for other sample sizes but with the sample mean range multiplied by a different factor.

(b) The standard error of the mean for samples of size 6 is estimated to be

$$\hat{SE}(\bar{x}) = 1.58/\sqrt{6} = 0.65.$$

(c) Warning limits are set at the mean plus and minus 1.96 standard errors, i.e. at 3.19 ± 1.96x0.65, or 1.92 and 4.46. If the mean and standard deviation were constant over time then only about one in 20 (5%) of sample means should be outside one of these limits. Control limits are set at the mean plus and minus 3.09 standard errors, i.e. 3.19 ± 3.09x0.65, or 1.18 and 5.20. Only about one in 500 (0.2%) of sample means should plot outside these limits.

The rationale behind constructing the $\bar{x}$ chart in this way is that it shows the changes in the sample means with time, and, furthermore, the warning and control limits indicate whether these changes are too large to be due to sampling errors if the mean ring width remained constant over the 20 year period. In fact, the chart indicates very clearly that the mean ring width was not constant. It seems that the mean increased from 1971 until 1977, at which point the upper control limit was crossed. The mean then decreased until 1989, at which point the lower control limit was crossed.

With control charts it is conventional to measure process variability using sample ranges on the grounds of simplicity, although standard deviations or variances could be used instead. Like $\bar{x}$ charts, range charts can have warning limits placed so that the probability of crossing one of these is 0.05 (5%), assuming that the level of variation is stable. Similarly, control limits can be placed so that the probability of crossing one of them is 0.002 (0.2%) when the level of variation is stable. The setting of these limits requires the use of tabulated values that are provided and explained in Table 5.8. The limits from this table are not symmetric about the mean range, and are more accurate than the symmetric limits produced by the SPSS control chart option.

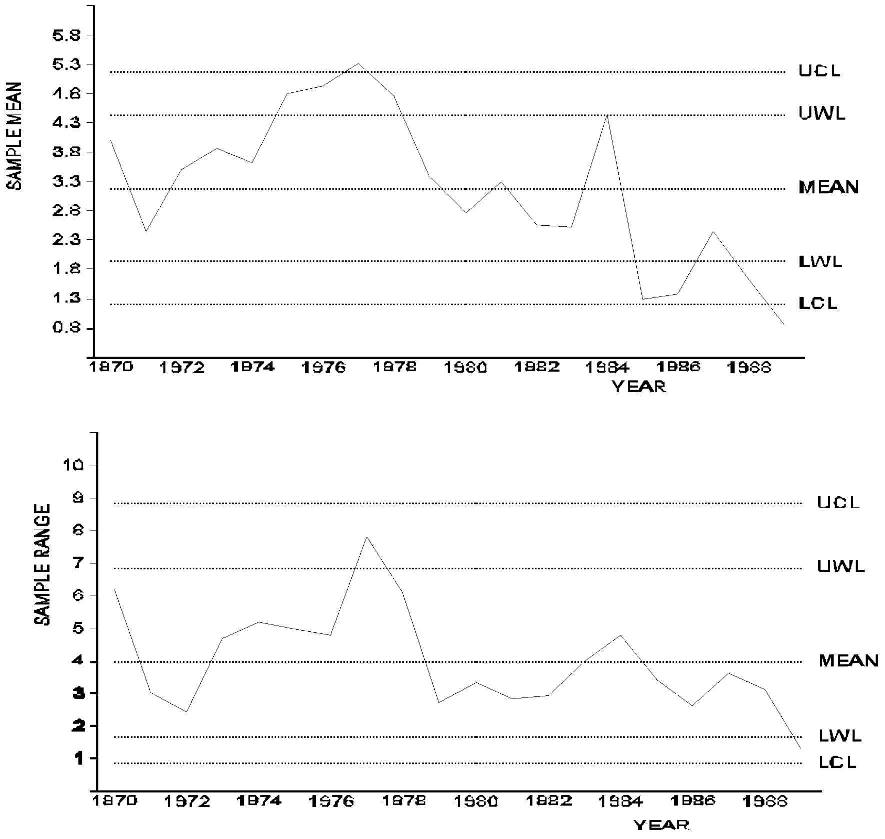


Figure 5.3 Control charts for means and ranges for ring widths for samples of Andean alders. NB. Abbreviations for the lines shown on the charts are: UCL, upper control limit (0.2%); UWL, upper warning limit (2.5%); LWL, lower warning limit (2.5%); and LCL, lower control limit (0.1%). The percentages in parenthesis here are the probabilities of crossing the lines if the distribution of annual ring widths was constant over time.

The range chart is included in Figure 5.3. From this chart it can be seen that the variability in annual ring widths seems to some extent to have mirrored the changes in the mean. In particular, the variability appears to have been lower at the end of the 20 year period than it was at the beginning, although none of the plotted points is outside a control limit. As is the case here, it is often the pattern in the points that are plotted on a control chart that is meaningful rather than the number of points outside the control limits.

TABLE 5.8 CONTROL CHART LIMITS FOR SAMPLE RANGES, ASSUMING SAMPLES FROM NORMAL DISTRIBUTIONS.

NB. TO FIND THE LIMITS ON THE RANGE CHART, MULTIPLY THE MEAN RANGE BY THE TABULATED VALUE. FOR EXAMPLE, FOR SAMPLES OF SIZE N = 5 THE LOWER ACTION LIMIT IS 0.16W, WHERE W IS THE MEAN RANGE. WITH A STABLE DISTRIBUTION A WARNING LIMIT IS CROSSED WITH PROBABILITY 0.05 (5%) AND A CONTROL LIMIT WITH PROBABILITY 0.002 (0.2%). THE LAST COLUMN IS THE FACTOR THAT THE MEAN RANGE MUST BE MULTIPLIED BY TO OBTAIN THE STANDARD DEVIATION. FOR EXAMPLE, FOR SAMPLES OF SIZE 3 THE STANDARD DEVIATION IS 0.591W. SOURCE: TABLES G1 AND G2 OF DAVIES AND GOLDSMITH (1972).

Sample	Lower Limits		Upper Limits		SD
Size	Control	Warning	Warning	Control	Factor
2	0.00	0.04	2.81	4.12	0.887
3	0.04	0.18	2.17	2.99	0.591
4	0.10	0.29	1.93	2.58	0.486
5	0.16	0.37	1.81	2.36	0.430
6	0.21	0.42	1.72	2.22	0.395
7	0.26	0.46	1.66	2.12	0.370
8	0.29	0.50	1.62	2.04	0.351
9	0.32	0.52	1.58	1.99	0.337
10	0.35	0.54	1.56	1.94	0.325

5.5 Cumulative Sum (CUSUM) Methods

Cumulative sum (CUSUM) methods were introduced for industrial quality control with the idea that the graphs used are better for displaying changes than ordinary control charts. Here two different types of CUSUM method are considered. The first is the ‘standard’ method that is described in many textbooks and available in some computer programs. This applies when there is a single time series being considered. The second method is not so widely known. It can be used when there are observations at a number of sampling stations for several different times.

The Standard CUSUM Method

As an example of the usual CUSUM method, suppose that 50 monthly samples are available for the phosphorus concentration (μgL^{-1}) in a lake. Suppose that it can be assumed that level was stable for the first half of the period. Are there changes in the underlying mean of X in the second half of the series? For the present the potential complications of seasonal variation, autocorrelation in time, etc. can be ignored. Figure 5.4(a) shows a comparison between a plot of the original observations and a CUSUM plot. The latter is obtained by plotting the cumulative sum of deviations from the mean

$$\text{CUSUM}(i) = \{X(1) - \text{Mean}\} + \{X(2) - \text{Mean}\} + \ldots + \{X(i) - \text{Mean}\}$$

against the observation number, i. The mean that is used is that for the first 25 observations, where the true mean is supposed to have been constant.

It can be seen that high observations (30-40) have given a positive slope on the CUSUM. In general, an increasing CUSUM plot indicates that observations tend to be above the target, and a decreasing plot indicates that observations tend to be below the target. The idea is that the CUSUM emphasises this more than a plot of the original values.

Figure 5.4(b) shows plots using logarithms of the original observations, which might be preferred because of the skewed nature of the distribution of the variable being considered. The CUSUM plot looks quite similar for either the original or the transformed data.

A useful general reference is MacNally and Hart (1997). They discuss the use of CUSUM methods for monitoring water storage facilities. Programs for CUSUM methods include AARDVARK for water quality analyses (Van Dijk and Ellis, 1995), WQSTAT also for water quality analysis (Intelligent Decision Technologies, 1998), and are included in some general purpose statistical packages like MINITAB (Minitab, Inc., 1994).

(a) Original Observations Plotted

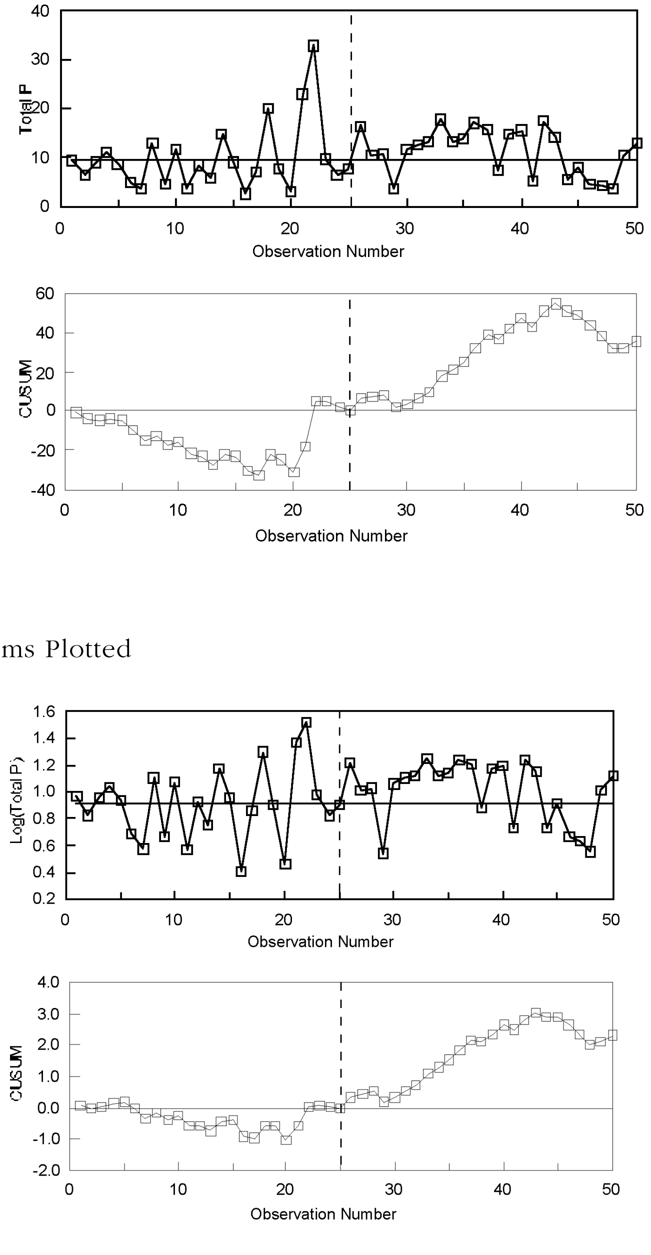


Figure 5.4 Plot of 50 monthly observations of phosphorus and the derived CUSUM plot using the mean of the first 25 observations as the target.

Another CUSUM Type Method

With the second CUSUM method to be considered, the situation is where there are a number of sampling sites in a region (perhaps 5-100), and observations on a variable on some or all of the sites at a number of sample times (perhaps 3-50). Interest is in whether there are any overall consistent patterns of change at the sampled sites. This is examined by making a CUSUM plot for each sample time, which indicates how the distribution of the variable compares with the distribution at other times, and carrying out some tests, which show whether the differences between sample times could be due to chance. The approach was originally developed for monitoring examination marks of students for different subjects, but was then extended to use with environmental monitoring (Manly, 1988, 1994).

Suppose that there are n sample units measured at m different times. Let x_{ij} be the measurement on sample unit i at time t_j, and let $\bar{x}_i$ be the mean of all the measurements on the unit. Assume that the units are numbered in order of their values for $\bar{x}_i$, so that $\bar{x}_1$ is the smallest mean and $\bar{x}_n$ is the largest mean. Then it is possible to construct m cumulative sum (CUSUM) charts (one for each sample time) by calculating

$$S_{ij} = (x_{1j} - \bar{x}_1) + (x_{2j} - \bar{x}_2) + \ldots + (x_{ij} - \bar{x}_i),$$

for j from 1 to m and i from 1 to n, and plotting the S_{ij} values against i. Missing values are easily handled.

The CUSUM chart for time t_j indicates the manner in which the observations made on units at that time differ from the average values for all sample times, with a positive slope for the CUSUM showing that the values for time period j are higher than the average values for all periods - a CUSUM slope of D over a series of observations (the rise divided by the number of observations) indicates that those observations are on average D higher than the corresponding means for the sample units. Thus a constant difference between the values on a sample unit at one time and the mean for all times is indicated by a constant slope of the CUSUM going either up or down from left to right. On the other hand, a positive slope on the left-hand side of the graph followed by a negative slope on the right-hand side indicates that the values at the time being considered were high for units with a low mean but low for units with a high mean.

Randomization methods can be used to decide whether the CUSUM plot for time t_j indicates systematic differences between the data for this year and the average for all years. Three approaches based on the null hypothesis that the values for each sample unit are in a random order are:

(a) A large number of randomized CUSUM plots can be constructed, where for each one of these the observations on each sample unit are randomly permuted. Then for each value of i the maximum and minimum values obtained for S_{ij} can be plotted on the CUSUM chart. This gives an envelope within which any CUSUM plot for real data can be expected to lie: if the real data plot goes outside the envelope then there is clear evidence that the null hypothesis is not true.

(b) Using the randomizations it is possible to determine whether S_{ij} is significantly different from 0 for any particular value of i.

(c) A statistic (Z_{max}) that measures the maximum extent to which a CUSUM plot differs from what is expected on the basis of the null hypothesis can be calculated and the randomizations can be used to see whether the value obtained is likely to occur by chance.

See Manly (1994) for more details about how these tests are carried out.

This CUSUM method can be modified to allow for autocorrelation in the observations taken at one location (Manly and MacKenzie, 1999). A Windows program CAT (CUSUM Analysis Tool) is available from BFJM to do the calculations.

Example: Foliage Cover in Whareorino Forest (Appendix Data Set 4)

The data shown in Table 5.9 were extracted from Data Set 4 in Appendix 3. They are the percentage foliage cover browse on 99 plots sampled in Whareorino forest from 1995 to 1999. For the purpose of this example the plots will be considered as 99 sites, and it will be assumed that spatial correlation is negligible, although this has not been checked. In fact, the 99 plots are on nine lines and the plots along one line may not give truly independent data.

TABLE 5.9 FOLIAGE COVER FOR 99 PLOTS IN THE WHAREORINO FOREST, 1995 TO 1999. VALUES OF -1 INDICATE THAT NO VALUE IS AVAILABLE FOR A PARTICULAR PLOT FOR THE YEAR IN QUESTION.

			YEAR		
PLOT	95	96	97	98	99
1	63	63	68	58	53
2	50	55	60	50	50
3	12	22	42	48	45
4	35	40	45	35	45
5	21	33	43	29	41
6	65	75	65	65	65
7	35	35	45	25	30
8	58	58	58	48	45
9	35	75	85	75	85
10	32	48	55	45	42
11	15	35	35	25	15
12	52	68	65	52	35
13	55	65	65	53	55
14	55	55	60	60	65
15	65	70	75	60	65
16	65	65	55	45	55
17	69	69	65	55	70
18	45	60	65	60	70
19	35	45	55	65	65
20	60	60	65	60	60
21	35	35	55	35	15
22	48	62	62	45	65
23	75	75	75	55	65
24	46	62	68	62	67
25	62	62	68	55	58
26	67	63	65	61	49
27	25	20	15	10	10

			YEAR		
PLOT	95	96	97	98	99
28	62	62	65	55	55
29	65	65	65	65	55
30	75	75	75	65	75
31	55	55	70	65	75
32	65	65	75	65	65
33	72	68	75	65	72
34	50	30	45	45	65
35	80	70	85	75	80
36	5	15	35	35	55
37	35	35	35	35	35
38	43	54	60	54	57
39	33	35	50	48	50
40	45	48	65	63	58
41	18	22	45	52	58
42	82	82	85	75	68
43	75	75	70	65	65
44	53	45	55	45	60
45	70	65	60	63	65
46	62	63	67	68	68
47	22	24	44	36	42
48	35	27	50	53	65
49	30	15	45	40	35
50	50	50	65	50	60
51	50	45	70	65	70
52	57	55	61	47	55
53	68	58	65	58	60
54	75	65	65	55	55
55	80	75	65	65	65
56	67	67	69	59	55
57	15	15	31	15	18
58	58	55	68	62	65
59	25	32	45	48	45
60	-1	20	40	40	45
61	-1	20	40	40	50
62	-1	72	75	65	68
63	-1	45	50	50	50
64	-1	32	45	42	35
65	42	48	52	48	42
66	62	65	68	58	68
67	75	75	75	65	65
68	35	40	35	35	45
69	55	55	55	45	55
70	75	70	-1	65	60
71	55	65	75	65	70
72	15	15	25	15	25
73	65	65	65	63	63
74	-1	-1	68	62	68
75	-1	-1	70	65	68
76	-1	-1	65	58	62
77	-1	-1	48	42	42
78	-1	-1	58	48	55
79	-1	-1	60	58	63
80	-1	-1	43	43	55
81	-1	27	49	49	57
82	-1	23	43	43	53
83	-1	40	68	67	73
84	60	65	65	50	60
85	75	65	65	58	62
86	70	60	55	50	45
87	75	65	65	55	60

	YEAR				
PLOT	95	96	97	98	99
88	40	45	45	35	40
89	55	55	-1	65	65
90	70	65	75	55	60
91	45	45	50	35	35
92	65	65	65	60	55
93	-1	-1	36	38	29
94	-1	-1	68	62	65
95	-1	-1	62	62	72
96	-1	-1	68	70	75
97	-1	-1	51	40	49
98	-1	-1	43	45	45
99	-1	-1	35	40	55

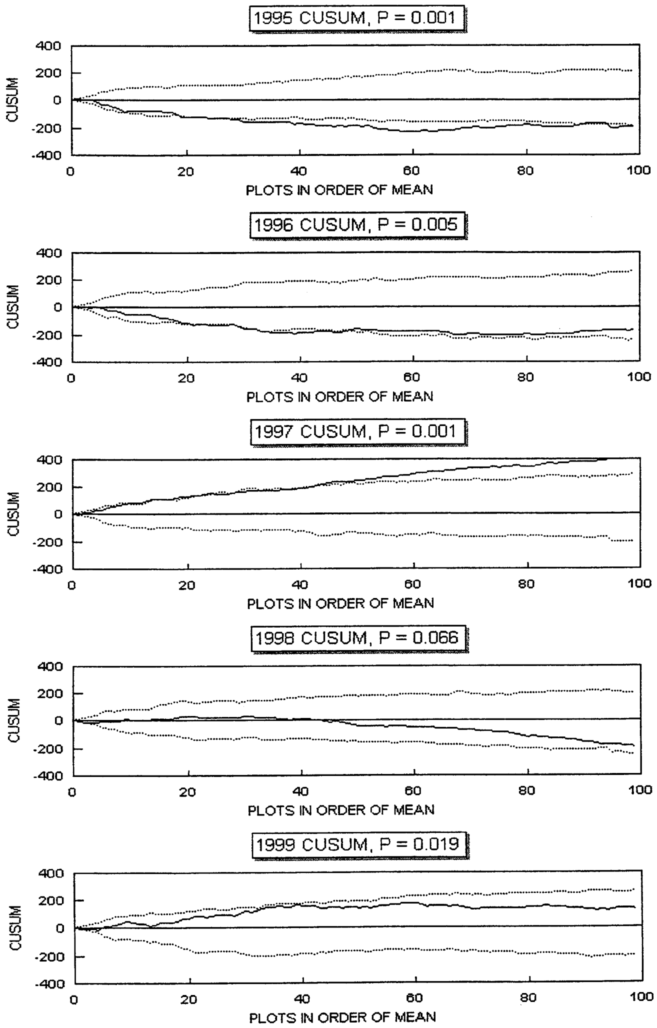


Figure 5.5. CUSUM plots of foliage cover for the years 1995 to 1996, each compared to the mean for all years.

Figure 5.5 shows the CUSUM plots for the years 1995 to 1999, each compared to the mean for all years. There is a great deal of evidence that the distribution of foliage cover varied from year to year, with 1995 and 1996 being years with relatively low cover, and 1997 and 1999 being years with relatively high cover. For 1998 the CUSUM plot is not very unusual in comparison with randomized plots, but if anything the foliage was low for plots where the mean for all years was high. The overall test for differences between years also gives a highly significant result (p = 0.001).

5.6 Another Method for Detecting Change in a Distribution

Stehman and Overton (1994) describe a set of tests for a change in a distribution, which they suggest will be useful as a screening device. These tests can be used whenever observations are available on a random sample of units at two times. If the first observation in a pair is *x* and the second one is *y*, then *y* is plotted against *x* and three chi-squared calculations are made, as shown on Figure 5.6.

The first test compares the number of points above a 45 degree line with the number below, as indicated in Figure 5.6(a). A significant difference indicates an overall shift in the distribution either upwards (most points above the 45 degree line) or downwards (most points below the 45 degree line). In Figure 6(a) there are 30 points above the line and 10 points below. The expected counts are both 20 if *x* and *y* are from the same distribution. Hence there is a chi-squared statistic of (30-20)2/20 + (10-20)2/20 = 10.00 with 1 degree of freedom (df). This is significantly large at the 1% level, giving clear evidence of a shift in the general level of observations. Because most plotted points are above the 45 degree line the observations tend to be higher at the second sample time.

The second test uses a shifted line at 45 degrees such that equal numbers of points are above and below it and counts the number of points in four quadrats, as shown in Figure 5.6(b). The counts then form a 2x2 contingency table for which a significant result indicates a change in shape of the distribution from one time period to the next. In Figure 5.6(b) the observed counts are as shown in Table 5.10. These are exactly equal to the expected counts on the assumption that the probability of a point plotting above the line is the same for high and low observations, leading to a chi-squared value of zero with 1 df.. There is therefore no evidence of a change in shape of the distribution.

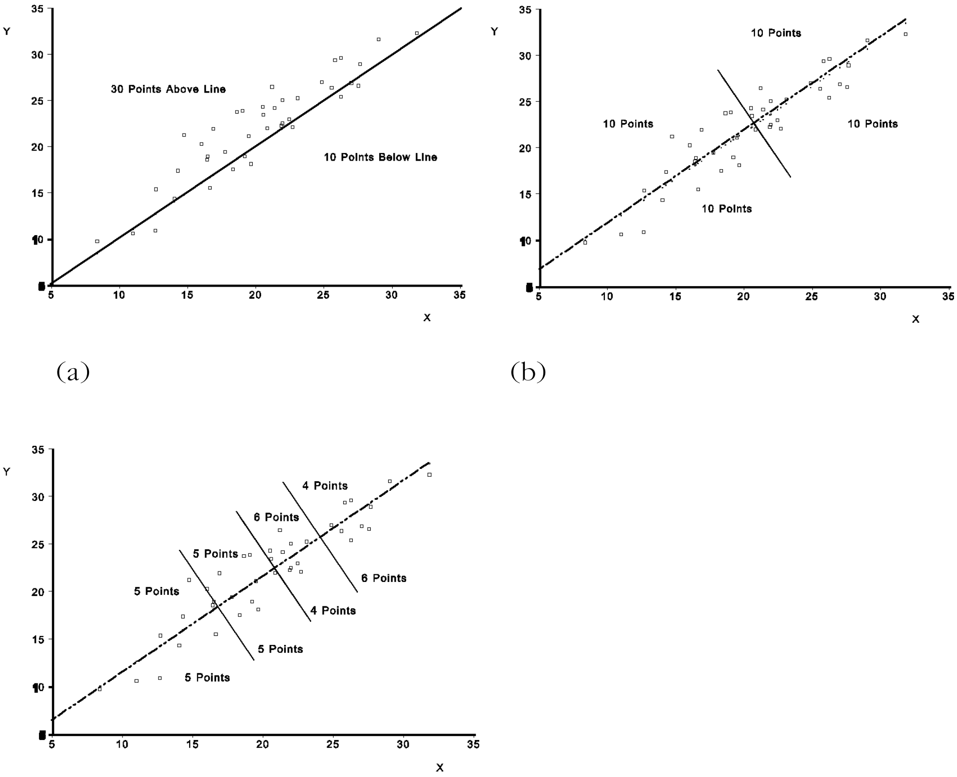
TABLE 5.10 COUNTS ABOVE AND BELOW THE 45 DEGREE LINE IN FIGURE 5.6(B).

	LEFT	RIGHT	TOTAL
Above line	10	10	20
Below line	10	10	20
	20	20	40

Finally, the third test involves dividing the points into quartiles, as shown in Figure 5.6(c). The counts in the different parts of the plot now make a 4x2 contingency table for which a significant chi-squared statistic indicates a change in distribution. The contingency table from Figure 5.6 (c) is shown in Table 5.11. The chi-squared statistic is 0.80 with 3 df, again giving no evidence of a change in shape of the distribution.

TABLE 5.11 COUNTS IN QUARTILES ABOVE AND BELOW THE 45 DEGREE LINE IN FIGURE 5.6(C).

	QUARTILE				
	1	2	3	4	TOTAL
Above line	5	5	6	4	20
Below line	5	5	4	6	20
	10	10	10	10	40



(a)

(b)

(c)

Figure 5.6 Stehman and Overton's screening tests for a change in the distribution of a monitored variable: (a) test for a shift in the distribution; (b) test for a change in shape of the distribution; (c) extension of test (b) with more division of points.

5.7 Use of Analysis of Variance

This module has not discussed analysis of variance types of approach for detecting trends and change points. In practice these may be the easiest approach to use, particularly if missing values are not a problem. Anyone confronted with analysing monitoring data should therefore consider whether a straightforward analysis of variance should be used before considering alternatives that may be more difficult to carry out.

5.8 Key Points in This Module

- The change-point problem concerns deciding whether the mean of a time series has changed and, if so, where that change occurs. A randomization test for a change-point based on t-tests is described.
- Two methods for detecting a trend in a series are a regression test and the Mann-Kendall test. The Mann Kendall test is popular because it makes few assumptions. The seasonal Mann-Kendall test also makes an allowance for serial correlation.
- A useful initial analysis of a time series involves decomposing it into trend, seasonal components, and unexplained variation.
- Control charts as used for industrial processes are also useful with environmental series.
- The standard cumulative sum (CUSUM) method is an alternative to the normal control charts that may show changes more clearly.
- A different type of CUSUM method is suggested for situations where observations are taken on a variable at a number of different locations on several occasions and the problem is to detect systematic differences between different sample times.
- A chi-squared testing procedure is described for comparing the distribution of a variable at two different times or two different places.

5.9 Questions About This Module

1. How does the change point problem differ from the problem of trend detection?
2. Why is there a multiple testing problem with identifying a change point in a time series?
3. Supposing you have a monthly time series which may include trend and seasonal variation. What type of regression model would you consider for describing it?
4. Why is the Mann-Kendall test popular for detecting trend among some environmental scientists?
5. Under what circumstances would you consider using Shewhart control charts or a CUSUM chart for monitoring a process?
6. What advantages are there in using the non-standard CUSUM method described in this module, with the associated randomization tests, instead of regression for example?
7. Under what circumstances would you consider using Stehman and Overton’s test for a change in a distribution between two sample times?

REFERENCES

Carlin, B.P., Gelfand, A.E. and Smith, A.F.M. (1992). Hierarchical Bayesian analysis of changepoint problems. *Applied Statistics* 41: 389-405.

Davies, O.L. and Goldsmith, P.L., eds. (1972). *Statistical Methods in Research and Production.* Oliver and Boyd, Edinburgh.

Edwards, D. and Coull, B.C. (1987). Autoregressive trend analysis: an example using long-term ecological data. *Oikos* 50: 95-102.

Harcum, J.B., Loftis, J.C. and Ward, R.C. (1992). Selecting trend tests for water quality series with serial correlation and missing values. *Water Resources Bulletin* 28: 469-78.

Helsel, D.R. and Hirsch, R.M. (1992). *Statistical Methods in Water Resources.* Elsevier, Amsterdam.

Hirsch, R.M. and Slack, J.R. (1984). A nonparametric trend test for seasonal data with serial dependence. *Water Resources Research* 20: 727-32.

Hirsch, R.M., Slack, J.R. and Smith, R.A. (1982). Techniques of trend analysis for monthly water quality data. *Water Resources Research* 18: 107-21.

Intelligent Decision Technologies (1998). *WQSTAT Plus User's Guide.* Intelligent Decision Technologies, Longmont, Colorado 80501, USA.

Jandhyala, V.K. and MacNeill, I.B. (1986). The change point problem: a review of applications. In *Statistical Aspects of Water Quality Monitoring* (eds. A.H. El-Shaarawi and R.E. Kwiatkowski), pp. 381-7. Elsevier, Amsterdam.

Judge, G.G., Hill, R.C., Griffiths, W.E., Lutkepohl, H. and Lee, T. (1988). *Introduction to the Theory and Practice of Econometrics.* Wiley, New York.

Loftis, J.C., McBride, G.B. and Ellis, J.C. (1991). Considerations of scale in water quality monitoring and data analysis. *Water Resources Bulletin* 27: 255-64.

MacNally and Hart (1997), Use of CUSUM methods for water-quality monitoring in storages, *Environmental Science and Technology* 31: 2114-9.

Manly, B.F.J. (1988). The comparison and scaling of student assessment marks in several subjects. *Applied Statistics* 37: 385-95.

Manly, B.F.J. (1992). *The Design and Analysis of Research Studies.* Cambridge University Press.

Manly, B.F.J. (1994). CUSUM methods for detecting changes in environmental variables. In *Statistics in Ecology and Environmental Monitoring* (D.J. Fletcher and B.F.J. Manly, eds.), pp. 225-38. Otago University Press, Dunedin.

Manly, B.F.J. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd Edit. Chapman and Hall, London.

Manly, B.F.J. and MacKenzie, D. (1999). A cumulative sum type of method for environmental monitoring. *Australian and New Zealand Journal of Statistics* (in press).

Minitab Inc. (1994). *Minitab Reference Manual, Release 10 for Windows.* Minitab Inc., State College, Pennsylvania 16801-3008, USA.

Montgomery, D.C. (1991). *Introduction to Statistical Quality Control*, 2nd Edit. Wiley, New York.

Neter, J., Wasserman, W. and Kutner, M.H. (1983). *Applied Linear Regression Analysis.* Irwin, Homewood, Illinois.

Page, E.S. (1955). A test for a change in a parameter occurring at an unknown time point. *Biometrika* 42: 523-6.

Stehman, S.V. and Overton, W.S. (1994). Environmental sampling and monitoring. In *Handbook of Statistics 12: Environmental Statistics* (eds. G.P. Patil and C.R. Rao), pp. 263-306. Elsevier, Amsterdam.

Taylor, C.H. and Loftis, J.C. (1989). Testing for trend in lake and ground water quality time series. *Water Resources Bulletin* 25: 715-26.

Van Dijk, P.A.H. and Ellis, J.C. (1995). *User Guide to AARDVARK for Windows, Release 2.01.* WRC Plc., Marlow, Buckinghamshire SL7 2HD, UK.

Zetterqvist, L. (1991). Statistical estimation and interpretation of trends in water quality time series. *Water Resources Research* 27: 1637-48.

This page intentionaly left blank

Contents

MODULE 6: BACI DESIGNS

Module 6: BACI Designs

SUMMARY

Before-After/Control-Impact (BACI) designs are used in environmental monitoring to compare the environmental status in treatment sites and control sites in the period prior to and the period after an "impact". Environmental parameters, such as species abundance, vary naturally through time and among spatial locations. A BACI design can be used to separate these sources of natural variation from variation due to the impact of an activity. The idea behind BACI can still be used when data is not available on the "before-impact" period, or from control sites.

6.1 Before-After/Control-Impact (BACI)

A design advocated by Green (1979) for environmental monitoring is to compare the environmental status in treatment sites and control sites in the period prior to and the period after an "impact". Examples of an "impact" activities are the beginning of a new management strategy; a new property development; or change in the protective status of some land. The idea is that an impact can be assessed as the change in ratio between pre- and post-impact. These designs are commonly known as BACI designs and there is a large body of relevant literature (e.g. Skalski and McKenzie 1982, Bernstein and Zalinski 1983, Stewart-Oaten *et al.* 1986, Underwood 1991, 1992, 1994, Skalski and Robson 1992).

The BACI design was developed in response to two major problems of study design (Marine Review Committee 1991). First, environmental effects, such as the abundance of plant species, or nutrient loading in a water body, vary naturally through time. Any change observed in an assessment area between the pre- and post-incident periods could conceivably be unrelated to the activity. Large natural changes are expected during any extended study period.

Second, there are always differences in the environment between any two areas. Simply observing a difference between assessment and control areas following the impact does not necessarily mean that the activity was the cause of the difference.

The BACI design may overcome these difficulties. By collecting data at both control and assessment areas using exactly the same protocol during both pre-impact and post-impact periods questions can be asked such as: Did the average difference in abundance between the control areas and the area of the oil spill change after the incident? The name BACI refers to the common use of the terms Before After Control Impact. The term Impact can be misleading as it has a negative meaning. An "impact" can be an event that has positive effects on the environment. The term Control is also confusing especially in association with pest control. The Control site is the "non-treatment" site, or the reference site. Although the use of these terms can be confusing in this module Before After Control Impact are used to be consistent with literature on this subject and for ease of notation.

Figure 6.1 contains a simplified sketch of point estimates of a BACI design with two periods of data collection before and following an incident. Significant impact due to an incident is judged to have occurred if there is a statistically significant change in the ratio (or difference) of the assessment area to the control area. If there were no significant environmental impact the response curves for the assessment and references area would be approximately parallel. The more data collection periods before and after an incident the better the scientific confidence in assessment of impacts.

Some principles of design apply to all BACI experiments. The study is essentially an observational study and the application of the findings to other populations should be done cautiously. Usually though the impact being considered is specific to the activity and location and the only population of interest is the one being studied. Another principle is that it is preferable to take equally spaced observations through time to minimise serial correlation and to maximise the amount of useful information (e.g. sample when the species is most abundant) (Stewart-Oaten *et al.* 1986). Thirdly, it is always a good idea to initially plot the data on graphs and charts. Such "exploratory data analysis" is no substitute for formal inference because the underlying variation in the data can not be separated from the variation due to the impact, and conclusions can not be supported by estimates of confidence.

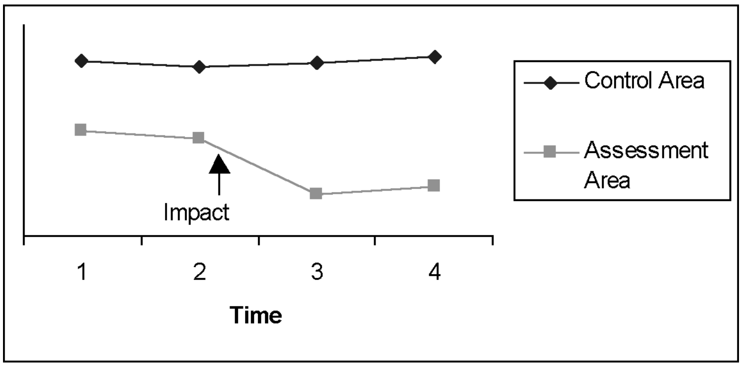


Figure 6.1 Survey data from control and impact areas collected over 4 sample dates. The impact occurred immediately after time 2.

6.2 Replication - temporal and spatial

One of the key features of a BACI design is replication of reference areas over time (Underwood 1994) for "...better guidance for detection of human disturbances." Compare Figure 6.2 with Figure 6.1. In Figure 6.2 point estimates of an indicator variable indicate recovery from injury by the fifth time period following the incident. Confidence would be gained in the assessment if the responses curves were approximately parallel for multiple years after the incident.

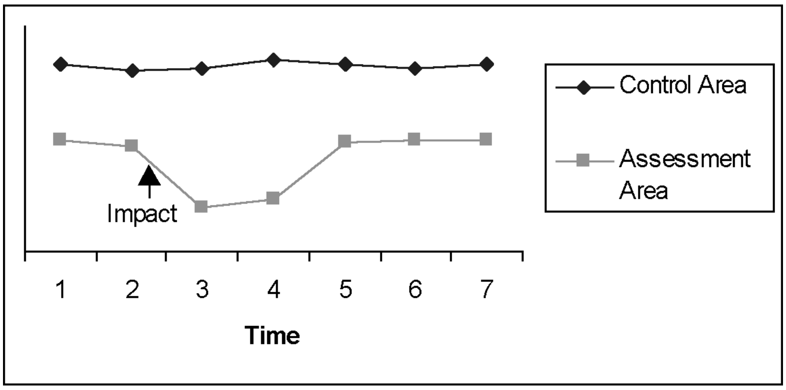


Figure 6.2 Survey data from control and impact areas collected over 7 sample dates. The impact occurred immediately prior to time 2.

It is also desirable to have spatial replication, i.e. more than one control area to compare to the impact area (Underwood 1994). It can be difficult to find a single site that is suitable to act as control sites for comparison with impact areas. The idea of having multiple control sites can help overcome this.

Replicate control sites should be chosen randomly. Underwood (1994) argues it is not necessary for them to mimic the impact area perfectly. Rather, sites should be a sample from a representative range of habitats similar to the habitat in the impact area.

One problem with choosing spatial replicates is deciding on the scale for the replication. The effect of the impact event is usually not known prior to the event and therefore it is difficult to decide where the control sites should be. For example, consider an estuary with a new sewage out-fall. Should the control sites be selected within the same estuary but say 100m away or, is the scale of the impact likely to be larger and control sites should be in estuaries say 10km away? Underwood (1994) recommends, for this example, sampling at two scales - both at the scale of within the estuary and among estuaries.

6.3 Differences between Control and Impact Sites

An early criticism of the BACI design was that repeated samples of the same site over time would be correlated (pseudoreplication) if the samples are analysed as independent replicates (Hurlbert 1984). If the designs have control and impact sites that are sampled at the same time then the difference between the control and impact can be used in the analysis (Stewart-Oaten *et al.* 1986). The differences between the sites may be uncorrelated, even if the successive samples are not.

Statistical analysis of these designs depends on the sampling procedures used for selection of sites and the amount of information collected on site-specific variables. For example, the differences in the average of the impact and control sites in each time period can be used. The average of these average-differences before the impact can be compared with those after the impact:

$$\text{Effect size} = \Delta_{B.} - \Delta_{A.} \qquad (1)$$

where $\Delta_{B.}$ is the average of Δ_{Bi} over i sampling dates. The Δ_{Bi} is the difference between the control site and the impact site at the i^{th} sampling date in the

Before period. Similarly, $\Delta_{A.}$ is the average of Δ_{Ai} over i sampling dates in the After period. In a matched design the control and impact sites are paired. The effect size is an estimate of the magnitude of the environmental impact (Osenburg *et al.* 1994).

With multiple control or impact sites the average of the control sites and the average of the impact sites are first calculated, for each sampling date. The Δ_{Bi} is then the difference between the average of the control sites and the average of the impact sites at the i^{th} sampling date in the Before period. Multiple control or impact sites are spatial replicates that can be used to estimate variation among locations. This variance is not the primary variance of interest. What is of primary interest in assessing an impact is the temporal variation in the control-impact differences. The variance of the effect size (1) is estimated from the differences among the Δ_{Bi} and the differences among the Δ_{Ai}. Therefore the variance of the effect size incorporates within-site sampling error, but not among site variation. Note that the variation among replicate control, or impact, sites at any one time period may still be of interest especially if there are large differences between the control site spatial variation and the impact site spatial variation.

The matching of control and impact sites is called a Control-Treatment Paired (CTP) design (Skalski and Robson 1992: Chapter 6) or a Before/After-Control/Impact-Pairs (BACIP) design (Stewart-Oaten *et al.* 1986).

Example: Invertebrate Feeding on Pellets in Rangataua Forest (Appendix Data Set 1)

Data on the proportion of pellets fed on by inveterbrates in a Control and Impact areas, Before and After a simulated pest control operation was collected in a study in Rangataua Forest in 1997. There were four survey times before the impact and two survey times after the impact. The “survey time” was in fact three nights, i.e. the data on the proportion of baits fed on was collected over three nights at each survey time. Only data where the control and impact areas were surveyed on the same nights is shown (Table 6.1). Note, for this analysis there was a single control and single impact site.

TABLE 6.1 PROPORTION OF PELLETS FED ON BY INVETERBRATES IN A CONTROL AND IMPACT AREA, BEFORE AND AFTER A SIMULATED PEST CONTROL OPERATION WAS COLLECTED IN A STUDY IN RANGATAUA FOREST IN 1997

BEFORE	CONTROL	SURVEY		PELLETS		BEFORE	CONTROL	SURVEY		PELLETS	
/AFTER	/IMPACT	TIME	LEFT	FED ON	PROPORTION	/AFTER	/IMPACT	TIME	LEFT	FED ON	PROPORTION
b	c	1	98	33	0.34	b	i	1	98	24	0.24
b	c	1	94	54	0.57	b	i	1	26	13	0.5
b	c	1	87	24	0.28	b	i	1	0	0	.
b	c	2	100	33	0.33	b	i	2	95	18	0.19
b	c	2	83	35	0.42	b	i	2	33	5	0.15
b	c	2	75	27	0.36	b	i	2	13	1	0.08
b	c	3	99	49	0.49	b	i	3	97	34	0.35
b	c	3	90	61	0.68	b	i	3	41	24	0.59
b	c	3	84	43	0.51	b	i	3	11	3	0.27
b	c	4	98	65	0.66	b	i	4	84	38	0.45
b	c	4	97	79	0.81	b	i	4	24	13	0.54
b	c	4	87	36	0.41	b	i	4	0	0	.
a	c	1	100	37	0.37	a	i	1	100	18	0.18
a	c	1	90	29	0.32	a	i	1	100	26	0.26
a	c	1	88	26	0.30	a	i	1	100	33	0.33
a	c	2	100	46	0.46	a	i	2	100	16	0.16
a	c	2	89	38	0.43	a	i	2	100	29	0.29
a	c	2	86	40	0.47	a	i	2	100	22	0.22

The data for the three nights within each survey time was not considered to be independent. The average of the proportions of pellets fed on within each 3 night survey time was calculated (Table 6.2)

TABLE 6.2 AVERAGE, OVER THREE NIGHTS, OF THE PROPORTION OF BAITS FED ON IN THE CONTROL AND IMPACT AREA, BEFORE AND AFTER A SIMULATED PEST CONTROL OPERATION IN RANGATAUA FOREST.

NB. THE DIFFERENCE BETWEEN THESE AVERAGED PROPORTIONS IN CONTROL AND IMPACT AREAS IN THE BEFORE AND AFTER PERIODS IS IN THE TWO RIGHT-HAND COLUMNS.

BEFORE		AFTER		DIFFERENCE	
CONTROL	IMPACT	CONTROL	IMPACT	BEFORE	AFTER
0.395688	0.372449	0.329226	0.256667	0.023239	0.072559
0.370562	0.139304	0.450694	0.223333	0.231258	0.227361
0.561544	0.40287			0.158674	
0.630497	0.497024			0.133473	

A *t*-test can be used to compare the before and after-differences. In this example there is no evidence of a “difference in the differences” between the control and impact sites prior to the pest operation compared with after the pest operation, *t* = -0.16585, P = 0.876 (the *t*-test was done in SPSS). In Figure 6.3 the control and impact lines are roughly the same distance apart in both the before and after impact period. Notice that despite the considerable variation in the average proportion of bait fed on, the difference between the two lines, shown as the dashed line in Figure 6.3, has little variation.

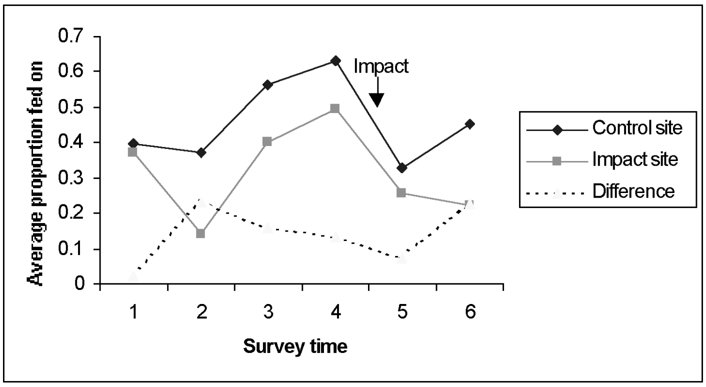


Figure 6.3 Survey data from control and impact areas collected over 7 sample dates. The difference between the control and impact area is shown as a dashed line.

6.4 Impact-Control Designs

Often an environmental impact can occur but there is no baseline "before" data for either the impact area or the control area. Because the BACI design cannot be used, such studies are called "after-only" or Impact-Control designs. In these designs data collected following the incident is compared between the impact and control areas. The obvious problem with such designs is the possibility of the confounding effects of natural factors on any observed difference between impact and control areas. Such designs typically have low power (Osenberg *et al.* 1994).

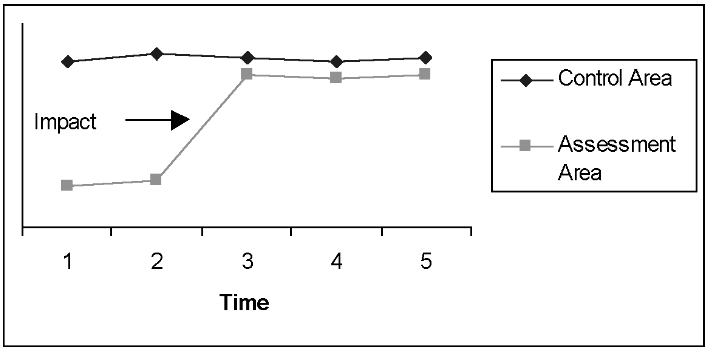


Figure 6.4 Survey data from control and impact areas collected over 5 sample dates. The impact occurred immediately prior to the first survey.

With designs that have no "before impact" data the analysis focuses on finding evidence of an area-time interaction. An area-time interaction means that the differences between the areas (impact and control) are different among the sampling dates. For example, in Figure 6.4 there are large differences following an impact but over time these differences decrease and become more or less constant when the response curves are parallel.

6.5 Before-After Designs

In this design data is available from the impact area prior to the incident. Examples of this situation are where long term monitoring has been occurring and an accidental impact occurs. With Before-After designs where there are no control sites there is a risk that any observed difference is due to some natural environmental fluctuation. Just as with the Impact-Control design, confounding from natural factors can occur. The observed change in mean abundance of a species after an impact maybe due to some other factor and occurred, by chance, at the same time of the disturbance event. In the example in Figure 6.5 the magnitude and abruptness of the perturbation can be evidence of an impact.

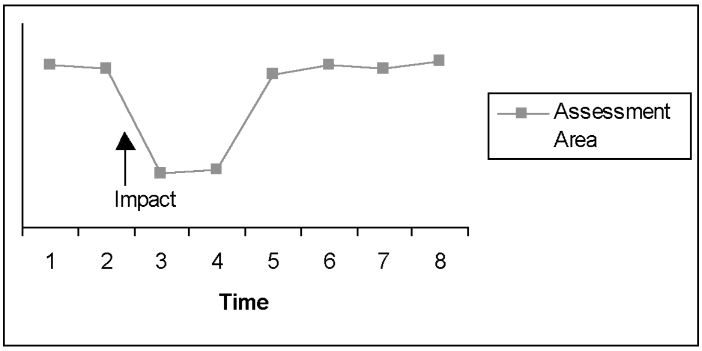


Figure 6.5 Survey data from an impact area collected over 8 sample dates. The impact occurred immediately prior to time 3.

6.6 Analysis of BACI Designs

Analysis of BACI designs can become quite complex. The simplest approach is the one suggested by Stewart-Oaten *et al.* (1986) where *t*-tests of differences are conducted. This method was used in the example above. However, there are some limitations with this approach. One is that the test assumes effects are "additive" when in fact, for many biological systems, effects are multiplicative - the difference between control and impact sites tends to be greater when species are abundant than when they are sparse (Stewart-Oaten *et al.* 1986). One way to overcome this problem of non-additive effects is to transform the data, e.g. with a log transformation. Another problem is that the analysis assumes that in the difference between the control and impact sites there is no obvious trend in the period before the onset of the impact event, i.e. the impact and control areas should not be changing relative to each other prior to the impact.

Underwood's 1991, 1992 and 1994 papers describe various alternative analysis methods based on ANOVA models. He discusses further some limitations with the Stewart-Oaten *et al.* approach. The analyses presented by Underwood can be quite complex and often use nested (or hierarchical) ANOVA designs. The simplest of Underwood's methods are described here. The range of papers should be reviewed for more complex designs.

- *Before-After*

  For a design where there is only Before sample data collected at *t* times and After sample data collected at *t* times at a single site the appropriate ANOVA would be, using the notation of Underwood (1991):

SOURCE OF VARIATION	df	F
Before vs After B	1	$MS_B / MS_{T(B)}$
Time(Before vs After) T(B)	$2(t-1)$	
Residual	$2t(n-1)$	
Total	$2tn - 1$	

The factor Time(Before vs After) means Time is *nested* in the factor Before vs After. This factor is measuring whether the average abundance of the organism being measured differed over time within the Before period and differed over time within the After period. The number of replicate samples at the site at each time is *n*.

- *Before-After-Control-Impact - One Control site and One Impact site - One survey Before and One survey After*

  When a single Control and a single Impact site are surveyed once Before and once After (i.e. there are only two sites and two survey times) Underwood (1991) recommends a simple two factor ANOVA:

SOURCE OF VARIATION	df	F
Before vs After B	1	
Control vs Impact L	1	
Interaction B x L	1	MS $_{\text{B x L}}$ / MS $_{\text{Residual}}$
Residual	4(*n* -1)	
Total	4*n* - 1	

  In this design the interest is in the interaction B x L which is the interaction between the Before and After times (factor B) and the Control vs Impact sites (factor L). The interpretation of a significant interaction is that the difference between the averages of the Control and Impact survey data are not the same in the Before period as in the After period.

- *Before-After-Control-Impact - One Control site and One Impact site - Repeat surveys Before and Repeat surveys After*

  The next design is where a single Control and single Impact site are sampled more than once, i.e. there is more than one survey before and more than one after the impact. Again a nested design is used. Note that in this analysis the surveys of the Control and Impact sites are assumed to occur at the same time.

SOURCE OF VARIATION	df	F
Before vs After B	1	
Control vs Impact L	1	
Interaction B x L	1	MS $_{\text{B x L}}$ / MS $_{\text{L x T(B)}}$
Time(Before vs After) T(B)	2(*t* - 1)	
Interaction L x T(B)	2(*t* - 1)	MS $_{\text{L x T(B)}}$ / MS $_{\text{Residual}}$
L x T(B)-Before	*t* - 1	MS $_{\text{L x T(B) Before}}$ / MS $_{\text{Residual}}$
L x T(B)-After	*t* - 1	MS $_{\text{L x T(B) After}}$ / MS $_{\text{Residual}}$
Residual	4*t*(*n* -1)	
Total	4*tn* - 1	

The partitioning of the interaction of location and time into the Before and After components is to allow testing of whether the differences in the Impact and Control sites varied over time Before the impact. Similarly it allows testing of whether the differences in the Impact and Control sites varied over time After the impact and for comparison of these two effects. If there were variation over time in the difference between the Control and Impact sites this would confound the detection of difference between before and after the impact.

More complex analysis, e.g. using use asymmetrical variances for designs with multiple control sites are discussed in Underwood (1992).

6.7 Comparison of Variances in a BACI Design

The BACI analysis described above relies on comparison of averages - long term running averages between impact and control sites. In some situations the focus on "averages" is not appropriate and the interest is more in the variation, e.g. after disturbance does the fish community fluctuate in time more or less than what it was doing previous to the disturbance? Questions may be asked about the rate and magnitude of fluctuations.

The ability to detect temporal fluctuations depends on the temporal scale of the surveys. For example, if surveys are conducted once a year, unusual fluctuations within the year will not be detected (Figure 6.6).

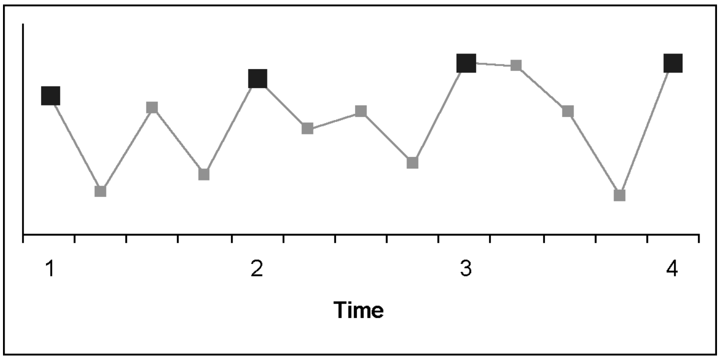


Figure 6.6 Survey data from an impact area collected over years.
NB The four annual sample dates are indicated by the ■, show little variation. However the fluctuations between each annual sample date vary among the four sample dates. If sampling had occurred more frequently this level of variation would have been detected.

To detect if there is similar variation over time after an impact event compared with the period before the event it may be necessary to sample at different temporal scales. The idea is similar to having spatial replication at varying scales. For example, control and impact sites are surveyed in three time periods before and after the impact event, e.g. a survey is done in March over 3 years before and 3 years after the impact event. Within each survey period there are three survey times, e.g. at weekly intervals. The appropriate analysis would be using nested design where Before vs After is the major source of variation, survey periods are nested within the factor Before vs After and times nested within survey periods (Underwood 1991).

6.8 Key Points in This Module

- Before-After/Control-Impact (BACI) designs are a very useful design in environmental monitoring to assess the effect of an activity.
- Environmental status in treatment sites and control sites are compared in the period prior to and the period after an "impact".
- Environmental effects, such as species abundance vary naturally through time and among spatial locations. A BACI design can be used to separate these sources of natural variation from variation due to the impact of an activity.
- When an environmental impact occurs and there is no baseline "before" data for either the impact area or the control area an "after-only" or Impact-Control design can be used. Data collected following the incident is compared between the impact and control areas.
- When data is available from the impact area prior to the incident a Before-After designs can be used.
- With both the Impact-Control and Before-After designs there is a risk that any observed difference is confounded by natural environmental fluctuation.

6.9 Questions About This Module

After completing this module you should be able to give reasonable answers to the following questions.

1. The long-term impact on forest bird populations of a new predator programme is to be assessed. The programme is to begin next month. How would you design the monitoring of such a programme, e.g. how would you choose control sites and impact sites? What spatial and temporal replication would you have? What variables would you measure?
2. In the above scenario what is the crucial feature of the design that is lacking to make it a true BACI design?
3. To assess the impact of a new marine reserve data were collected from a control site and the "impact" site both before and after the establishment of the marine reserve at three survey times (i.e. there were three survey times before and three after the reserve was established). The analysis used the factor L where L = control or impact, the factor B where B = before or after and the factor T where T = survey time 1, 2 or 3. There was a significant interaction between B and L. What is the interpretation of this result?
4. There was also a significant interaction for the marine reserve example, between the factors L and $T(B)$ in the Before period. Would you be confident about using the changes in the differences between the Control and Impact sites Before and After impact to assess the effect of the marine reserve?

REFERENCES

Bernstein, B.B. and Zalinski, J. (1983) An optimum sampling design and power tests for environmental biologists. *Journal of Environment Management* 16:35 – 43.

Green, R.H. (1979). *Sampling design and statistical methods for environmental biologists.* John Wiley and Sons, NY.

Marine Review Committee. (1991). Marine review committee response to comments by Southern California Edison on the MRC's Final Report to the California Coastal Commission. W.W. Murdoch (Chairman), Second Draft: Report to the California Coastal Commission, University of California.

Osenberg, C.W., Schmitt, R.J., Holbrook, S.J., Abu-Saba, K.E. and Flegal, A.R. (1994). Detection of environmental impacts: natural variability, effect size, and power analysis. *Ecological Applications* 4: 16-30.

Skalski, J.R., and McKenzie, D.H. (1982). A design for aquatic monitoring programs. *Journal of Environmental Management* 14:237-251.

Skalski, J.R., and D.S. Robson. (1992). *Techniques for wildlife investigations: Design and Analysis of capture data.* Academic Press, Inc., San Diego, CA.

Stewart-Oaten, A., W.W. Murdoch, and K.R. Parker. 1986. Environmental impact assessment: "Pseudoreplication" in time? *Ecology* 67:929-940.

Underwood, A.J. 1991. Beyond BACI: experimental designs for detecting human environmental impacts on temporal variations in natural populations. *Australian Journal and Marine Freshwater Research* 42:569-587.

Underwood, A.J. 1992. Beyond BACI: the detection of environmental impacts on populations in the real, but variable, world. *Journal of Experimental Marine Biological Ecology* 161:145-178.

Underwood, A.J. 1994. On beyond BACI: sampling designs that might reliably detect environmental disturbances. *Ecological Applications* 4:3-15.

Appendix 1: Worksheets for Practice on SPSS by Workshop Participants

Appendix 2: Examples of Analysis in SPSS

OUTPUT FROM SPSS LOGISTIC REGRESSION

```
      Total number of cases:      1000 (Unweighted)
      Number of selected cases:   1000
      Number of unselected cases: 0

      Number of selected cases:                 1000
      Number rejected because of missing data:  0
      Number of cases included in the analysis: 1000

Dependent Variable Encoding:

Original       Internal
Value          Value
       0       0
       1       1

                               Parameter
                 Value   Freq  Coding
                                (1)    (2)    (3)    (4)    (5)    (6)    (7)
NATION
                     2     36  1.000   .000   .000   .000   .000   .000   .000
                     3     13   .000  1.000   .000   .000   .000   .000   .000
                     6     56   .000   .000  1.000   .000   .000   .000   .000
                     7     28   .000   .000   .000  1.000   .000   .000   .000
                     8     38   .000   .000   .000   .000  1.000   .000   .000
                     9    311   .000   .000   .000   .000   .000  1.000   .000
                    10     44   .000   .000   .000   .000   .000   .000  1.000
                    11    420   .000   .000   .000   .000   .000   .000   .000
                    12     38   .000   .000   .000   .000   .000   .000   .000
                    13     16   .000   .000   .000   .000   .000   .000   .000

                          (8)    (9)
NATION
                     2   .000   .000
                     3   .000   .000
                     6   .000   .000
                     7   .000   .000
                     8   .000   .000
                     9   .000   .000
                    10   .000   .000
                    11  1.000   .000
                    12   .000  1.000
                    13   .000   .000

                               Parameter
                 Value   Freq  Coding
                                (1)    (2)    (3)    (4)    (5)    (6)    (7)
TARGET
                    1     98  1.000   .000   .000   .000   .000   .000   .000
                    2     13   .000  1.000   .000   .000   .000   .000   .000
                    3    599   .000   .000  1.000   .000   .000   .000   .000
                    4    276   .000   .000   .000  1.000   .000   .000   .000
                    5      2   .000   .000   .000   .000  1.000   .000   .000
                    6      6   .000   .000   .000   .000   .000  1.000   .000
                    7      1   .000   .000   .000   .000   .000   .000  1.000
                    8      1   .000   .000   .000   .000   .000   .000   .000
                    9      4   .000   .000   .000   .000   .000   .000   .000

                          (8)
TARGET
                    1   .000
                    2   .000
                    3   .000
                    4   .000
                    5   .000
                    6   .000
                    7   .000
                    8  1.000
                    9   .000
```

```
                               Parameter
                 Value   Freq  Coding
                                 (1)    (2)    (3)    (4)
FYEAR
                     2    177  1.000   .000   .000   .000
                     3    179   .000  1.000   .000   .000
                     4    282   .000   .000  1.000   .000
                     5    132   .000   .000   .000  1.000
                     6    230   .000   .000   .000   .000
TDAY
                     1    144  1.000   .000   .000
                     2    350   .000  1.000   .000
                     3    324   .000   .000  1.000
                     4    182   .000   .000   .000
SEASON
                     1    225  1.000   .000   .000
                     2    637   .000  1.000   .000
                     3     12   .000   .000  1.000
                     4    126   .000   .000   .000
GEAR
                     1    469  1.000
                     2    531   .000

Dependent Variable.   SLION

Beginning Block Number  0.  Initial Log Likelihood Function

-2 Log Likelihood   164.06795

* Constant is included in the model.

Estimation terminated at iteration number 6 because
Log Likelihood decreased by less than .01 percent.

Classification Table for SLION
The Cut Value is .50
                 Predicted
                  0       1      Percent Correct
                  0  I    1
Observed     +-------+-------+
   0     0   I  984  I    0  I  100.00%
             +-------+-------+
   1     1   I   16  I    0  I     .00%
             +-------+-------+
                        Overall  98.40%

----------- Variables in the Equation ------------

Variable           B      S.E.     Wald    df      Sig        R    Exp(B)

Constant     -4.1190     .2520 267.1209     1    .0000

Beginning Block Number  1.  Method: Backward Stepwise (LR)

Variable(s) Entered on Step Number
1.        NATION
          FYEAR
          SEASON
          TARGET
          GEAR
          TDAY
          LOGDUR
          LOGWT

Estimation terminated at iteration number 11 because
Log Likelihood decreased by less than .01 percent.

 -2 Log Likelihood      137.165
 Goodness of Fit        546.123
 Cox & Snell - R^2         .027
 Nagelkerke - R^2          .175

                     Chi-Square    df Significance
 Model                  26.903    30        .6283
 Block                  26.903    30        .6283

 Step                   26.903    30        .6283

Classification Table for SLION
The Cut Value is .50
                 Predicted
                  0       1      Percent Correct
                  0   I    1
Observed     +-------+-------+
   0     0   I  984  I    0  I  100.00%
             +-------+-------+
   1     1   I   16  I    0  I     .00%
             +-------+-------+
                        Overall  98.40%

----------- Variables in the Equation ------------

Variable            B      S.E.     Wald    df      Sig       R    Exp(B)

NATION                            2.8404     9    .9703   .0000
 NATION(1)     1.7187  130.1639    .0002     1    .9895   .0000    5.5773
 NATION(2)     -.7126  163.0693    .0000     1    .9965   .0000     .4904
 NATION(3)     2.8515  130.1622    .0005     1    .9825   .0000   17.3132
 NATION(4)     4.3086  130.1666    .0011     1    .9736   .0000   74.3360
 NATION(5)      .2544  137.3739    .0000     1    .9985   .0000    1.2897
 NATION(6)    -2.8388  151.2774    .0004     1    .9850   .0000     .0585
 NATION(7)     1.5056  130.1580    .0001     1    .9908   .0000    4.5069
 NATION(8)     1.5253  130.1533    .0001     1    .9906   .0000    4.5967
 NATION(9)     1.4325  130.1482    .0001     1    .9912   .0000    4.1891
FYEAR                             3.7931     4    .4347   .0000
 FYEAR(1)      -.1283    1.0043    .0163     1    .8984   .0000     .8796
 FYEAR(2)       .2121    1.0923    .0377     1    .8461   .0000    1.2362
 FYEAR(3)     -1.4848    1.1092   1.7921     1    .1807   .0000     .2265
 FYEAR(4)       .6962     .9089    .5868     1    .4437   .0000    2.0062
SEASON                             .5231     3    .9138   .0000
 SEASON(1)    -9.2818   28.3422    .1072     1    .7433   .0000     .0001
 SEASON(2)    -1.0949    1.6646    .4326     1    .5107   .0000     .3346
 SEASON(3)     -.0406  148.2897    .0000     1   1.0000   .0000     .9602
TARGET                            5.8606     8    .6628   .0000
 TARGET(1)   -11.2995   68.4538    .0272     1    .8689   .0000     .0000
 TARGET(2)    -5.5347  141.9855    .0015     1    .9689   .0000     .0039
 TARGET(3)    -3.9655    1.6413   5.8376     1    .0157  -.1529     .0190
 TARGET(4)      .8315   77.8882    .0001     1    .9915   .0000    2.2968
 TARGET(5)   -11.7876  315.0652    .0014     1    .9702   .0000     .0000
 TARGET(6)   -15.1836  169.4920    .0080     1    .9286   .0000     .0000
 TARGET(7)   -12.2337  446.5065    .0008     1    .9781   .0000     .0000
 TARGET(8)   -10.2162  446.5098    .0005     1    .9817   .0000     .0000
GEAR(1)         .8943    1.5218    .3453     1    .5568   .0000    2.4455
TDAY                               .1643     3    .9831   .0000
 TDAY(1)       -.4074    1.0231    .1586     1    .6905   .0000     .6654
 TDAY(2)       -.1912     .7936    .0581     1    .8096   .0000     .8260
 TDAY(3)       -.1485     .7934    .0350     1    .8516   .0000     .8620
LOGDUR          .8999     .7030   1.6385     1    .2005   .0000    2.4593
LOGWT           .1535     .2117    .5260     1    .4683   .0000    1.1659
Constant      -2.1904  130.1786    .0003     1    .9866

-------- Model if Term Removed ---------

Term         Log                              Significance
Removed      Likelihood    -2 Log LR    df    of Log LR

NATION          -69.972        2.780     9          .9724
FYEAR           -70.850        4.535     4          .3385
SEASON          -72.251        7.337     3          .0619
TARGET          -72.203        7.241     8          .5109
GEAR            -68.772         .379     1          .5384
TDAY            -68.666         .167     3          .9827
LOGDUR          -69.508        1.851     1          .1737
LOGWT           -68.861         .557     1          .4555
```

```
Variable(s) Removed on Step Number
2.       TDAY

Estimation terminated at iteration number 11 because
Log Likelihood decreased by less than .01 percent.

 -2 Log Likelihood      137.332
 Goodness of Fit        533.282
 Cox & Snell - R^2         .026
 Nagelkerke - R^2          .174

                     Chi-Square    df Significance

 Model                   26.736    29        .5859
 Block                   26.736    27        .4781
 Step                     -.167     3        .9827

Note:  A negative Chi-Square value indicates that the Chi-Square
       value has decreased from the previous step.

Classification Table for SLION
The Cut Value is .50
                 Predicted
                  0       1      Percent Correct
                  0  I    1
Observed     +-------+-------+
   0     0   I  984  I    0  I  100.00%
             +-------+-------+
   1     1   I   16  I    0  I     .00%
             +-------+-------+
                        Overall  98.40%

----------- Variables in the Equation ------------
Variable            B     S.E.      Wald   df       Sig       R    Exp(B)

NATION                              3.0595     9    .9619   .0000
 NATION(1)     1.7068  130.4966    .0002     1    .9896   .0000    5.5114
 NATION(2)     -.6976  163.5953    .0000     1    .9966   .0000     .4978
 NATION(3)     2.9082  130.4950    .0005     1    .9822   .0000   18.3239
 NATION(4)     4.4016  130.4992    .0011     1    .9731   .0000   81.5776
 NATION(5)      .2525  137.7086    .0000     1    .9985   .0000    1.2872
 NATION(6)    -2.5373  151.4391    .0003     1    .9866   .0000     .0791
 NATION(7)     1.5708  130.4911    .0001     1    .9904   .0000    4.8106
 NATION(8)     1.5925  130.4864    .0001     1    .9903   .0000    4.9161
 NATION(9)     1.4776  130.4818    .0001     1    .9910   .0000    4.3823
FYEAR                               3.7412     4    .4422   .0000
 FYEAR(1)      -.1572     .9997    .0247     1    .8750   .0000     .8545
 FYEAR(2)       .1587    1.0844    .0214     1    .8837   .0000    1.1719
 FYEAR(3)     -1.4790    1.1009   1.8047     1    .1791   .0000     .2279
 FYEAR(4)       .7296     .8906    .6711     1    .4127   .0000    2.0743
SEASON                               .5079     3    .9172   .0000
 SEASON(1)    -9.2962   28.3265    .1077     1    .7428   .0000     .0001
 SEASON(2)    -1.0735    1.6627    .4169     1    .5185   .0000     .3418
 SEASON(3)     -.0636  148.6784    .0000     1   1.0000   .0000     .9384
TARGET                              6.5049     8    .5909   .0000
 TARGET(1)   -11.0542   68.9532    .0257     1    .8726   .0000     .0000
 TARGET(2)    -5.6795  142.2226    .0016     1    .9681   .0000     .0034
 TARGET(3)    -3.8453    1.5104   6.4814     1    .0109  -.1653     .0214
 TARGET(4)      .7270   77.9652    .0001     1    .9926   .0000    2.0689
 TARGET(5)   -11.7091  315.7234    .0014     1    .9704   .0000     .0000
 TARGET(6)   -14.9458  169.6856    .0078     1    .9298   .0000     .0000
 TARGET(7)   -11.9333  446.5059    .0007     1    .9787   .0000     .0000
 TARGET(8)    -9.9625  446.5093    .0005     1    .9822   .0000     .0000
GEAR(1)          .8857    1.5091    .3444     1    .5573   .0000    2.4246
LOGDUR           .8994     .7030   1.6368     1    .2008   .0000    2.4581
LOGWT            .1405     .2074    .4589     1    .4981   .0000    1.1508
Constant      -2.5387  130.5083    .0004     1    .9845

-------- Model if Term Removed ----------

Term        Log                              Significance
Removed     Likelihood    -2 Log LR    df    of Log LR
```

```
NATION         -70.128        2.925     9           .9672
FYEAR          -70.906        4.480     4           .3449
SEASON         -72.409        7.486     3           .0579
TARGET         -72.288        7.245     8           .5105
GEAR           -68.854         .376     1           .5396
LOGDUR         -69.583        1.835     1           .1756
LOGWT          -68.908         .484     1           .4867

Variable(s) Removed on Step Number
3.       NATION

Estimation terminated at iteration number 11 because
Log Likelihood decreased by less than .01 percent.

 -2 Log Likelihood      140.257
 Goodness of Fit        637.485
 Cox & Snell - R^2         .024
 Nagelkerke - R^2          .156

                     Chi-Square    df Significance

 Model                   23.811    26        .5868
 Block                   23.811    18        .1613
 Step                    -2.925     9        .9672

Note:  A negative Chi-Square value indicates that the Chi-Square
       value has decreased from the previous step.

Classification Table for SLION
The Cut Value is .50
                 Predicted
                  0       1      Percent Correct
                  0  I    1
Observed     +-------+-------+
   0     0   I  984  I    0  I  100.00%
             +-------+-------+
   1     1   I   16  I    0  I     .00%
             +-------+-------+
                        Overall  98.40%


----------- Variables in the Equation ------------

Variable            B      S.E.     Wald    df      Sig       R    Exp(B)
FYEAR                             3.0484     4    .5498   .0000
 FYEAR(1)     -.0387     .9036    .0018     1    .9658   .0000     .9620
 FYEAR(2)      .4691     .9288    .2551     1    .6135   .0000    1.5985
 FYEAR(3)    -1.0835     .9681   1.2525     1    .2631   .0000     .3384
 FYEAR(4)      .4934     .7399    .4447     1    .5048   .0000    1.6379
SEASON                             .3305     3    .9542   .0000
 SEASON(1)   -8.9025   28.6449    .0966     1    .7560   .0000     .0001
 SEASON(2)    -.7998    1.6185    .2442     1    .6212   .0000     .4494
 SEASON(3)    -.0317  134.5831    .0000     1   1.0000   .0000     .9688
TARGET                            7.1639     8    .5191   .0000
 TARGET(1)  -13.1684   45.3041    .0845     1    .7713   .0000     .0000
 TARGET(2)   -9.9016  121.4454    .0066     1    .9350   .0000     .0001
 TARGET(3)   -3.9273    1.4792   7.0485     1    .0079  -.1754     .0197
 TARGET(4)   -4.3347    1.9875   4.7567     1    .0292  -.1296     .0131
 TARGET(5)  -11.6958  315.7219    .0014     1    .9704   .0000     .0000
 TARGET(6)  -13.8142  168.1613    .0067     1    .9345   .0000     .0000
 TARGET(7)  -11.9537  446.5059    .0007     1    .9786   .0000     .0000
 TARGET(8)   -9.8930  446.5091    .0005     1    .9823   .0000     .0001
GEAR(1)         .1488     .7522    .0392     1    .8431   .0000    1.1605
LOGDUR         1.0156     .6692   2.3028     1    .1291   .0430    2.7610
LOGWT           .1607     .2042    .6187     1    .4315   .0000    1.1743
Constant       -.5792    2.2150    .0684     1    .7937

-------- Model if Term Removed ---------

Term        Log                               Significance
Removed     Likelihood    -2 Log LR    df     of Log LR

FYEAR          -71.911        3.564     4           .4682
```

```
SEASON         -73.289        6.322     3           .0970
TARGET         -74.500        8.743     8           .3644
GEAR           -70.148         .040     1           .8422
LOGDUR         -71.441        2.626     1           .1051
LOGWT          -70.456         .655     1           .4184

Variable(s) Removed on Step Number
4.       GEAR

Estimation terminated at iteration number 11 because
Log Likelihood decreased by less than .01 percent.

 -2 Log Likelihood      140.296
 Goodness of Fit        630.102
 Cox & Snell - R^2         .023
 Nagelkerke - R^2          .155

                    Chi-Square    df Significance

 Model                  23.772    17        .1258
 Block                  23.772    17        .1258
 Step                    -.040     1        .8422

Note:  A negative Chi-Square value indicates that the Chi-Square
       value has decreased from the previous step.

Classification Table for SLION
The Cut Value is .50
                 Predicted
                 0       1      Percent Correct
                 0   I   1
Observed      +-------+-------+
   0     0    I  984  I    0  I  100.00%
              +-------+-------+
   1     1    I   16  I    0  I     .00%
              +-------+-------+
                        Overall  98.40%

----------- Variables in the Equation ------------

Variable            B      S.E.     Wald    df      Sig        R    Exp(B)

FYEAR                              3.0081    4    .5565    .0000
 FYEAR(1)      -.0822     .8770     .0088    1    .9253    .0000     .9211
 FYEAR(2)       .4156     .8922     .2169    1    .6414    .0000    1.5152
 FYEAR(3)     -1.1329     .9385    1.4573    1    .2274    .0000     .3221
 FYEAR(4)       .4394     .6921     .4032    1    .5255    .0000    1.5518
SEASON                              .3492    3    .9505    .0000
 SEASON(1)    -8.9156   28.6545     .0968    1    .7557    .0000     .0001
 SEASON(2)     -.8273    1.6129     .2631    1    .6080    .0000     .4372
 SEASON(3)     -.0586  134.5702     .0000    1   1.0000    .0000     .9431
TARGET                             7.2049    8    .5147    .0000
 TARGET(1)   -13.2688   45.3291     .0857    1    .7697    .0000     .0000
 TARGET(2)   -10.0246  121.1020     .0069    1    .9340    .0000     .0000
 TARGET(3)    -3.9253    1.4812    7.0228    1    .0080   -.1750     .0197
 TARGET(4)    -4.4387    1.9165    5.3639    1    .0206   -.1432     .0118
 TARGET(5)   -11.6848  315.7207     .0014    1    .9705    .0000     .0000
 TARGET(6)   -13.8068  168.0045     .0068    1    .9345    .0000     .0000
 TARGET(7)   -11.9653  446.5059     .0007    1    .9786    .0000     .0000
 TARGET(8)    -9.8454  446.5090     .0005    1    .9824    .0000     .0001
LOGDUR         1.0091     .6698    2.2696    1    .1319    .0405    2.7431
LOGWT           .1699     .1995     .7251    1    .3945    .0000    1.1852
Constant       -.4101    2.0466     .0401    1    .8412

-------- Model if Term Removed ---------

Term        Log                             Significance
Removed     Likelihood    -2 Log LR    df    of Log LR

FYEAR          -71.928        3.560     4           .4688
SEASON         -73.315        6.333     3           .0965
TARGET         -74.705        9.113     8           .3329
LOGDUR         -71.444        2.592     1           .1074
```

```
LOGWT          -70.533         .769     1           .3805

Variable(s) Removed on Step Number
5.       FYEAR

Estimation terminated at iteration number 11 because
Log Likelihood decreased by less than .01 percent.

 -2 Log Likelihood      143.857
 Goodness of Fit        614.567
 Cox & Snell - R^2         .020
 Nagelkerke - R^2          .132

                     Chi-Square    df Significance

 Model                   20.211    16        .2109
 Block                   20.211    13        .0901
 Step                    -3.560     4        .4688

Note:  A negative Chi-Square value indicates that the Chi-Square
       value has decreased from the previous step.

Classification Table for SLION
The Cut Value is .50
                 Predicted
                  0       1      Percent Correct
                  0  I    1
Observed     +-------+-------+
   0     0   I  984  I    0  I  100.00%
             +-------+-------+
   1     1   I   16  I    0  I      .00%
             +-------+-------+
                        Overall  98.40%

----------- Variables in the Equation ------------

Variable            B      S.E.     Wald    df      Sig       R    Exp(B)

SEASON                           .0858     3    .9935   .0000
 SEASON(1)    -8.2121   29.2084    .0790     1    .7786   .0000     .0003
 SEASON(2)     -.1300    1.4286    .0083     1    .9275   .0000     .8781
 SEASON(3)     -.3545  134.6901    .0000     1    .9979   .0000     .7015
TARGET                          7.7160     8    .4617   .0000
 TARGET(1)   -12.3447   45.7997    .0727     1    .7875   .0000     .0000
 TARGET(2)    -9.5700  124.0746    .0059     1    .9385   .0000     .0001
 TARGET(3)    -3.9043    1.4418   7.3326     1    .0068  -.1803     .0202
 TARGET(4)    -4.6617    1.8466   6.3727     1    .0116  -.1633     .0095
 TARGET(5)   -11.6680  315.7275    .0014     1    .9705   .0000     .0000
 TARGET(6)   -12.6488  172.7638    .0054     1    .9416   .0000     .0000
 TARGET(7)   -11.7706  446.5058    .0007     1    .9790   .0000     .0000
 TARGET(8)   -10.3763  446.5084    .0005     1    .9815   .0000     .0000
LOGDUR         1.0630     .6622   2.5766     1    .1085   .0593    2.8951
LOGWT           .0973     .1951    .2487     1    .6180   .0000    1.1022
Constant      -1.1563    1.8694    .3826     1    .5362

-------- Model if Term Removed ---------

Term        Log                            Significance
Removed     Likelihood    -2 Log LR    df    of Log LR

SEASON         -74.823        5.790     3           .1223
TARGET         -75.986        8.115     8           .4223
LOGDUR         -73.414        2.972     1           .0847
LOGWT          -72.057         .258     1           .6116

Variable(s) Removed on Step Number
6.       LOGWT

Estimation terminated at iteration number 11 because
Log Likelihood decreased by less than .01 percent.

 -2 Log Likelihood      144.114
 Goodness of Fit        620.919
 Cox & Snell - R^2         .020
 Nagelkerke - R^2          .131

                     Chi-Square    df Significance

 Model                   19.954    12        .0680
 Block                   19.954    12        .0680
 Step                     -.258     1        .6116

Note:  A negative Chi-Square value indicates that the Chi-Square
       value has decreased from the previous step.

Classification Table for SLION
The Cut Value is .50
                 Predicted
                  0       1      Percent Correct
                  0  I    1
Observed     +-------+-------+
   0     0   I  984  I    0  I  100.00%
             +-------+-------+
   1     1   I   16  I    0  I     .00%
             +-------+-------+
                        Overall  98.40%

----------- Variables in the Equation ------------

Variable            B      S.E.     Wald    df      Sig       R    Exp(B)

SEASON                             .0824     3    .9939   .0000
 SEASON(1)    -8.1911   29.2414    .0785     1    .7794   .0000     .0003
 SEASON(2)     -.1027    1.4273    .0052     1    .9427   .0000     .9024
 SEASON(3)     -.3705  134.8933    .0000     1    .9978   .0000     .6904
TARGET                            7.6832     8    .4650   .0000
 TARGET(1)   -12.2040   45.7859    .0710     1    .7898   .0000     .0000
 TARGET(2)    -9.5701  123.8916    .0060     1    .9384   .0000     .0001
 TARGET(3)    -3.8436    1.4315   7.2092     1    .0073  -.1782     .0214
 TARGET(4)    -4.7218    1.8386   6.5952     1    .0102  -.1674     .0089
 TARGET(5)   -11.7049  315.7263    .0014     1    .9704   .0000     .0000
 TARGET(6)   -12.4077  176.6772    .0049     1    .9440   .0000     .0000
 TARGET(7)   -11.5754  446.5057    .0007     1    .9793   .0000     .0000
 TARGET(8)   -10.5996  446.5081    .0006     1    .9811   .0000     .0000
LOGDUR         1.0785     .6549   2.7118     1    .0996   .0659    2.9403
Constant      -1.1503    1.8653    .3803     1    .5374

-------- Model if Term Removed ---------

Term        Log                             Significance
Removed     Likelihood    -2 Log LR    df    of Log LR

SEASON         -74.946        5.777     3          .1230
TARGET         -76.047        7.981     8          .4354
LOGDUR         -73.646        3.178     1          .0746

Variable(s) Removed on Step Number
7.       TARGET

Estimation terminated at iteration number 10 because
Log Likelihood decreased by less than .01 percent.

 -2 Log Likelihood      152.095
 Goodness of Fit        735.359
 Cox & Snell - R^2         .012
 Nagelkerke - R^2          .079

                     Chi-Square    df Significance

 Model                   11.973    11        .3657
 Block                   11.973     4        .0176
 Step                    -7.981     8        .4354

Note:  A negative Chi-Square value indicates that the Chi-Square
       value has decreased from the previous step.
```

```
Classification Table for SLION
The Cut Value is .50
                 Predicted
                  0       1      Percent Correct
                  0  I    1
Observed     +-------+-------+
   0     0   I  984  I    0  I  100.00%
             +-------+-------+
   1     1   I   16  I    0  I     .00%
             +-------+-------+
                        Overall  98.40%

----------- Variables in the Equation ------------

Variable            B      S.E.     Wald    df      Sig       R    Exp(B)

SEASON                            1.4589     3    .6918    .0000
 SEASON(1)    -6.4525   17.9576    .1291     1    .7194    .0000     .0016
 SEASON(2)     1.1753    1.0416   1.2732     1    .2592    .0000    3.2392
 SEASON(3)    -6.2300   77.7780    .0064     1    .9362    .0000     .0020
LOGDUR          .6661     .5560   1.4352     1    .2309    .0000    1.9467
Constant      -5.7567    1.3115  19.2657     1    .0000

-------- Model if Term Removed ---------

Term        Log                              Significance
Removed     Likelihood    -2 Log LR    df    of Log LR

SEASON        -81.699        11.303     3           .0102
LOGDUR        -76.888         1.682     1           .1947

Variable(s) Removed on Step Number
8.       LOGDUR

Estimation terminated at iteration number 10 because
Log Likelihood decreased by less than .01 percent.

 -2 Log Likelihood      153.777
 Goodness of Fit        763.003
 Cox & Snell - R^2         .010
 Nagelkerke - R^2          .068

                     Chi-Square    df Significance

 Model                   10.291     3        .0162
 Block                   10.291     3        .0162
 Step                    -1.682     1        .1947

Note:  A negative Chi-Square value indicates that the Chi-Square
       value has decreased from the previous step.

Classification Table for SLION
The Cut Value is .50
                 Predicted
                  0       1      Percent Correct
                  0  I    1
Observed     +-------+-------+
   0     0   I  984  I    0  I  100.00%
             +-------+-------+
   1     1   I   16  I    0  I     .00%
             +-------+-------+
                        Overall  98.40%

----------- Variables in the Equation ------------

Variable            B      S.E.     Wald    df      Sig       R    Exp(B)

SEASON                            1.3086     3    .7271    .0000
 SEASON(1)    -6.3746   18.0825    .1243     1    .7244    .0000     .0017
 SEASON(2)     1.1034    1.0374   1.1313     1    .2875    .0000    3.0145
 SEASON(3)    -6.3746   78.1851    .0066     1    .9350    .0000     .0017
Constant      -4.8283    1.0040  23.1276     1    .0000
```

```
--------- Model if Term Removed ----------

Term         Log                             Significance
Removed      Likelihood    -2 Log LR    df    of Log LR

SEASON          -82.034       10.291     3           .0162


------- Variables not in the Equation ---------
Residual Chi Square     21.539 with     27 df     Sig =  .7604

Variable           Score    df      Sig       R

NATION            3.3108     9    .9507   .0000
 NATION(1)         .0297     1    .8632   .0000
 NATION(2)         .1061     1    .7446   .0000
 NATION(3)         .0059     1    .9389   .0000
 NATION(4)        2.2056     1    .1375   .0354
 NATION(5)         .3592     1    .5489   .0000
 NATION(6)         .1506     1    .6979   .0000
 NATION(7)         .0005     1    .9830   .0000
 NATION(8)         .0032     1    .9552   .0000
 NATION(9)         .0443     1    .8332   .0000
FYEAR             2.4202     4    .6590   .0000
 FYEAR(1)          .0977     1    .7546   .0000
 FYEAR(2)          .0004     1    .9847   .0000
 FYEAR(3)         1.3569     1    .2441   .0000
 FYEAR(4)         1.3995     1    .2368   .0000
TARGET           10.8091     8    .2128   .0000
 TARGET(1)        1.3030     1    .2537   .0000
 TARGET(2)         .1808     1    .6707   .0000
 TARGET(3)         .0646     1    .7994   .0000
 TARGET(4)         .0000     1    .9955   .0000
 TARGET(5)         .0484     1    .8259   .0000
 TARGET(6)         .0417     1    .8383   .0000
 TARGET(7)         .0242     1    .8765   .0000
 TARGET(8)         .0081     1    .9284   .0000
GEAR(1)            .3937     1    .5304   .0000
TDAY               .1333     3    .9876   .0000
 TDAY(1)           .0079     1    .9290   .0000
 TDAY(2)           .0037     1    .9514   .0000
 TDAY(3)           .1042     1    .7468   .0000
LOGDUR            1.3865     1    .2390   .0000
LOGWT              .1567     1    .6922   .0000

No more variables can be deleted or added.
```

Appendix 3: Department of Conservation Data Sets

BACI Experiment on 1080 in Rangataua Forest

Survival and Breeding of Known Age Takahe

Monitoring of Dactylanthus Tayloili Near the Summit of Mt Pirongia

Monitoring of Vegetation in Whareorino Forest, 1995-99

Blue Cod Size Data from Paterson Inlet, 1994-98

Blue Cod Numbers Data from Paterson Inlet, 1994-98

Tree Conditions at Ryan Creek, Heaphy Valley in 1995 and 1999

Diets of Fish Collected from Otago Streams

Survey of Deer Pellets in the Takahe Special Area, Murchison Mountains, Fiordland National Park, in 1998

Angler Survey Results at Lake Taupo, 1992-99

Monitoring of Mistletoe in Eglington Valley, Fiordland, 1995-97

Hurunui Mainland Island 5-Minute Bird Count Data

Possum Trapping Data from Hurunui Mainland Island

Waikoropupu Springs Vegetation Transects, 1991-99

Kaimanawa Recreational Hunting Area Permanent Plot Biodiversity

Kaimanawa Recreational Hunting Area Plot Density of Nothofagus Fuscus

Notes:

All of the smaller data sets (one or two printed pages) are shown here. Only the first page is shown for the larger data sets.

The data are held in three Excel files with names that mean exactly what they say. These are DOC-1TO8.XLW, DOC-9&10.XLW, and DOC-11ON.XLW.

Data Set 1: BACI Experiment on 1080 in Rangataua Forest Contact: Shirley McQueen, Dunedin
Data format verified with her: 14-Aug-99

Data are number of pellets initially, and number fed on during the night.
There was an aerial distribution of 1080 pellets on 30/8/97 in the impact area.
Control area is Rotokura.
BACI experiment with proportion data.
There are also two types of pollard bait (toxic and non-toxic at the impact site).
Question: are the data independent on successive nights?

	Site		
Code	Type	Time	Toxic
0	Control	Before	No
1	Impact	After	Yes

Case	Date	Place	Session	Type	Time	Toxic	Day	Bait left	Fed on
1	07/08/97	Rotokura		0	0	0	1	98	33
2	07/09/97	Rotokura		0	0	0	2	94	54
3	07/10/97	Rotokura		0	0	0	3	87	24
4	07/13/97	Rotokura		0	0	0	1	100	33
5	07/14/97	Rotokura		0	0	0	2	83	35
6	07/15/97	Rotokura		0	0	0	3	75	27
7	08/17/97	Rotokura		0	0	0	1	99	49
8	08/18/97	Rotokura		0	0	0	2	90	61
9	08/19/97	Rotokura		0	0	0	3	84	43
10	08/17/98	Rotokura		0	0	0	1	98	65
11	08/18/97	Rotokura		0	0	0	2	97	79
12	08/19/97	Rotokura		0	0	0	3	87	36
13	08/30/97	Rotokura		0	1	0	1	99	38
14	08/31/97	Rotokura		0	1	0	2	83	35
15	09/01/97	Rotokura		0	1	0	3	70	11
16	08/30/97	Rotokura		0	1	0	1	97	53
17	08/31/97	Rotokura		0	1	0	2	83	44
18	09/01/97	Rotokura		0	1	0	3	75	2
19	09/02/97	Rotokura		0	1	0	4	70	17
20	09/04/97	Rotokura		0	1	0	6	60	16
21	09/06/97	Rotokura		0	1	0	8	45	8
22	09/02/97	Rotokura		0	1	0	4	70	21
23	09/04/97	Rotokura		0	1	0	6	56	14
24	09/06/97	Rotokura		0	1	0	8	31	6
25	09/09/97	Rotokura		0	1	0	1	100	37
26	09/10/97	Rotokura		0	1	0	2	90	29
27	09/11/97	Rotokura		0	1	0	3	88	26
28	09/09/97	Rotokura		0	1	0	1	100	46
29	09/10/97	Rotokura		0	1	0	2	89	38
30	09/11/97	Rotokura		0	1	0	3	86	40
31	07/08/97	Beech850E	5	1	0	0	1	98	24
32	07/09/97	Beech850E	5	1	0	0	2	26	13
33	07/10/97	Beech850E	5	1	0	0	3	0	0
34	07/13/97	650W	6	1	0	0	1	95	18
35	07/14/97	650W	6	1	0	0	2	33	5
36	07/15/97	650W	6	1	0	0	3	13	1
37	08/17/97	750W	8	1	0	0	1	97	34
38	08/18/97	750W	8	1	0	0	2	41	24
39	08/19/97	750W	8	1	0	0	3	11	3
40	08/17/97	550W	8	1	0	0	1	84	38
41	08/18/97	550W	8	1	0	0	2	24	13
42	08/19/97	550W	8	1	0	0	3	0	0
43	09/09/97	900E	9	1	1	0	1	100	18
44	09/10/97	900E	9	1	1	0	2	100	26
45	09/11/97	900E	9	1	1	0	3	100	33
46	09/09/97	500E	9	1	1	0	1	100	16
47	09/10/97	500E	9	1	1	0	2	100	29
48	09/11/97	500E	9	1	1	0	3	100	22
49	08/30/97	700E		1	1	1	1	98	30
50	08/31/97	700E		1	1	1	2	98	41
51	09/01/97	700E		1	1	1	3	98	14
52	08/30/97			1	1	1	1	98	25
53	08/31/97			1	1	1	2	98	24
54	09/01/97			1	1	1	3	97	13
55	09/02/97	700E		1	1	1	4	98	12
56	09/04/97	700E		1	1	1	6	98	8
57	09/06/97	700E		1	1	1	8	98	2
58	09/02/97			1	1	1	4	97	6
59	09/04/97			1	1	1	6	97	9
60	09/06/97			1	1	1	8	97	1

Data Set 2: Survival and Breeding of Known Age Takahe

Contact: Jane Maxwell, Te Anau
Data format verified with her: 23-Aug-99

Annual survival and productivity of study birds, individuals grouped by cohort.

Codes		
	0	Dead
	1	Alive
	2	Bird alive with chick that survived one year
	3	Not sighted in November or since, presumed dead

Birds in November 1998 are not old enough to have produced a yearling.
Captive birds are released at 1 year of age.

		Sex
Codes	Sex	Confirmed
0	Unknown	No
1	Male	Yes
2	Female	

						November Samples								
Case	Cohort	Type	Bird	Sex	Confirmed	1991	1992	1993	1994	1995	1996	1997	1998	Notes
1	1990-91	Wild	The Guide	1	1	1	1	1	1	1	1	1	1	
2		Wild	R29438	2	0	1	0	0	0	0	0	0	0	
3		Wild	R44253	0	0	1	0	0	0	0	0	0	0	
4		Captive	Spartacus	1	0	1	1	1	1	1	1	1	1	
5		Captive	Jezebel	2	0	1	1	2	1	1	1	1	1	
6		Captive	Mayo	2	0	1	1	1	1	3	3	3	3	
7		Captive	R34335	2	0	1	0	0	0	0	0	0	0	
8		Captive	R29411	1	0	1	0	0	0	0	0	0	0	
9		Captive	R29421	1	0	1	0	0	0	0	0	0	0	
10	1991-92	Wild	Eva	2	1		1	1	1	1	2	1	1	Chick from Eva x The Guide
11		Wild	R34986	2	0		1	0	0	0	0	0	0	
12		Captive	Nick	1	0		1	1	1	1	1	1	1	
13		Captive	Juliette	2	1		1	1	1	1	2	1	1	Chick produced by Juliette x Itty Bitty, a W.R. study bird
14		Captive	Jenny	2	1		1	1	1	1	2	1	1	
15		Captive	Gretch	2	1		1	1	1	1	1	2	1	Chick produced by Gretch x John, also C.R. study bird
16		Captive	John	1	1		1	1	1	1	1	1	1	
17		Captive	Teapot	1	1		1	1	1	1	1	1	1	
18		Captive	Margaret	2	1		1	1	1	1	3	3	3	
19		Captive	Katrina	2	1		1	1	1	1	3	3	3	
20		Captive	Blackadder	1	1		1	1	3	3	3	3	3	
21		Captive	Roger	1	1		1	0	0	0	0	0	0	
22	1992-93	Wild	Quan	0	0			1	1	1	2	2	1	
23		Wild	Tarn	0	0			1	3	3	3	3	3	
24		Wild	Roger roger	2	0			1	1	1	1	1	1	
25		Wild	Star	1	0			1	1	1	1	3	3	
26		Wild	Sky	2	0			1	0	0	0	0	0	
27		Wild	R34903	1	0			1	0	0	0	0	0	
28		Captive	Georgia	2	0			1	1	1	1	1	1	
29		Captive	Smart	2	0			1	1	1	1	1	1	
30		Captive	Fern	2	0			1	1	1	1	1	3	
31		Captive	Tramp	1	0			1	1	1	3	3	3	
32		Captive	Adrian Mole	1	0			1	1	1	3	3	3	
33		Captive	Hannibal	1	0			1	0	0	0	0	0	
34		Captive	Hobo	1	0			1	0	0	0	0	0	
35		Captive	Erin	2	0			1	0	0	0	0	0	
36		Captive	Bert	1	0			1	0	0	0	0	0	
37		Captive	Ninety-nine	1	0			1	1	3	3	3	3	
38	1993-94	Wild	Sage	2	0				1	1	1	1	1	
39		Wild	Sophie	2	0				1	1	1	1	1	
40		Wild	Hillary	2	0				1	1	1	1	1	
41		Wild	Aurora	2	0				1	1	1	2	1	Chick produced by Aurora x Rouge, a C.R. study bird.
42		Wild	Itty Bitty	1	1				1	1	2	1	1	Chick produced by Itty Bitty X Juliette, a C.R. study bird
43		Wild	Clyde	1	1				1	1	1	2	1	Chick produed by Clyde X Bonny, a C.R. study bird
44		Wild	Rosco	1	1				1	1	0	0	0	
45		Wild	Moho	1	0				1	1	1	0	0	
46		Wild	Skid	1	0				1	1	0	0	0	
47		Wild	R34937	2	1				1	0	0	0	0	
48		Wild	R34938	2	1				1	0	0	0	0	
49		Wild	Rustle	2	1				1	0	0	0	0	
50		Wild	Huxley	0	0				1	0	0	0	0	
51		Captive	Bonny	2	0				1	1	1	2	1	Chick produced by Bonny X Clyde, a W.R. study bird
52		Captive	Herman	2	0				1	1	1	1	3	
53		Captive	Lola	2	0				1	1	3	3	3	
54		Captive	Tom	2	0				1	0	0	0	0	
55		Captive	Oscar	1	0				1	0	0	0	0	
56		Captive	Picasso	1	0				1	0	0	0	0	
57		Captive	Minty	1	0				1	0	0	0	0	
58		Captive	Buckshot	1	0				1	0	0	0	0	
59		Captive	Sheila	2	0				1	0	0	0	0	
60	1994-95	Wild	Zimmi	1	0					1	1	1	1	
61		Wild	Phantom	2	0					1	0	0	0	
62		Wild	Jacco	0	0					1	0	0	0	
63		Captive	Simon	1	0					1	1	1	1	
64		Captive	Rouge	1	0					1	1	2	1	Chick produced by Rouge x Aurora, a W.R. study bird
65		Captive	Rach	2	0					1	1	1	1	
66		Captive	Matt	2	1					1	1	1	1	
67		Captive	Padre	2	0					1	1	3	3	
68		Captive	Mawson	1	0					1	1	3	3	
69		Captive	Jim	1	1					1	0	0	0	
70		Captive	Sally	1	1					1	0	0	0	
71		Captive	Peggy	2	0					1	0	0	0	
72		Captive	Bob	1	1					1	0	0	0	
73		Captive	Dianna	2	0					1	3	3	3	
74		Captive	Pavlov	1	0					1	3	3	3	
75	1995-96	Wild	Mana	1	0						1	0	0	
76		Wild	Venus	2	0						1	0	0	
77		Wild	Michelle	0	0						1	1	3	
78		Captive	Bean	2	0						1	1	1	

79		Captive	Rangi	1	0	1	1	1	
80		Captive	Aldrid	1	0	1	1	1	
81		Captive	Appollo	2	0	1	1	1	
82		Captive	Tosh	1	0	1	1	1	
83		Captive	Sunshine	1	0	1	1	1	
84		Captive	Grover	2	0	1	1	1	
85		Captive	Moko	2	1	1	1	1	
86		Captive	Whoopi	2	0	1	1	3	
87		Captive	Tiddles	1	0	1	1	3	
88		Captive	Geronimo	2	0	1	0	0	
89		Captive	Snag	1	0	1	0	0	
90		Captive	Jake	1	0	1	0	0	
91		Captive	Klumpfuss	2	1	1	0	0	
92		Captive	Marley	1	0	1	3	3	
93	1996-97	Wild	Misty	0	0		1	1	
94		Wild	Trapper	2	1		1	1	
95		Wild	Jem	2	0		1	1	
96		Wild	Rainbow	2	0		1	1	
97		Wild	Fred	2	0		1	1	
98		Wild	Moho rising	0	0		1	1	
99		Wild	Skimmer	1	1		1	1	
100		Wild	Merlin	1	0		1	3	
101		Captive	Thunder	2	1		1	1	
102		Captive	Rimu	2	1		1	1	
103		Captive	Eeffoc	2	1		1	1	
104		Captive	Taipa	2	1		1	1	
105		Captive	Pita	2	1		1	1	
106		Captive	Johnson	2	1		1	1	
107		Captive	Sharma	2	1		1	1	
108		Captive	Ziggy	1	1		1	1	
109		Captive	Curfew	1	1		1	1	
110		Captive	Gonzo	2	1		1	0	
111		Captive	Teabag	1	1		1	3	
112		Captive	Lucifer	1	1		1	3	
113		Captive	Gypsy	1	1		1	3	
114		Captive	Fozzie	1	1		1	3	
115		Captive	Zed	2	1		1	3	
116		Captive	Teacosy	2	1		1	3	
117	1997-98	Wild	Ursu	2	1			1	
118		Wild	Atawhai	2	1			1	
119		Wild	Doc	2	0			1	
120		Wild	Clag	1	1			1	
121		Wild	Feldspar	1	1			1	
122		Wild	Granite	0	0			1	
123		Wild	B45 yearling	0	0			1	
124		Captive	Summer	1	1			1	
125		Captive	Kilford	1	1			1	
126		Captive	Ibanez	2	1			1	
127		Captive	Calib	2	1			1	
128		Captive	Itchy	1	1			1	
129		Captive	McKenzie	2	1			1	
130		Captive	Rastus	2	1			1	
131		Captive	Glingii	1	1			1	

Data Set 3: Monitoring of Dactylanthus Taylorii Near the Summit of Mt Pirongia

Contact: Pim de Monchy, Waikato
Data format verified by him: 19-Aug-99

Count of male flowers, female flowers, buds and chewed buds on monitored D. taylorii, 1997-99.
Location and tag numbers as described by S.J. Moore (1999).
Cage status code is 1 for caged, 0 for uncaged.
Blanks are unknown data values in the count columns.
Results are probably not independent between plants at one location (e.g., by hut), but should be independent between locations.

			Location	Cage Status			Male flowers			Female flowers			Buds			Chewed buds			Seeds set		
Case	Location	Tag	Code	97	98	99	97	98	99	97	98	99	97	98	99	97	98	99	97	98	99
1	By hut	296	1		1	1		8	18		1	2		6	0		0	0		0	
2	By hut	974	1	1	1	0		4	7		0	1		5	0		0	6		0	
3	By hut	1720	1			0			5			0			2			0		0	
4	By hut	1.1.03	1	0	0	0			8			0			0			0			
5	By hut	1708	1	1	1	1		1	1		0	1		1	2		0	0		0	
6	Goblin Wood	1210	2	0	1	1	0	1	7	0	2	4	0	0	0	1	1	0	0	0	
7	Goblin Wood	1211	2	0	1	1	0	11	0	0	5	0	0	3	0	0	0	0	0	0	
8	Goblin Wood	1212	2	0	0	0	4	11	0	5	5	1	4	5	0	0	0	6	3	0	
9	Goblin Wood	1213	2	0	1	1	1	13	0	2	1	0	1	0	0	0	0	0		1	
10	Goblin Wood	1214	2	0	1	1	1	5	4	0	0	1	5	1	0	7	0	0		0	
11	Goblin Wood	1215	2	0	1	0	0	2	2	0	0	0	6	6	1	0	0	3		0	
12	Goblin Wood	1216	2	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
13	Goblin Wood	1217	2	0	1	1	2	6	0	3	0	9	2	4	1	0	0	0	0	0	
14	Goblin Wood	1703	2	0	1	1		0	2		0	1		1	0		0	0			
15	Hihikiwi	1107	3	0	0	0	0	0	2	0	0	1	0	0	0	3	8	12		0	
16	Hihikiwi	1195	3	0	0	0	0	0	0	0	0	5	0	1	1	0	6	1		0	
17	Hihikiwi	1196	3	0	0	0	0	0	0	0	0	0	0	1	0	2	4	57		0	
18	Hihikiwi	1197	3	0	0	1	0	0	26	0	0	7	2	2	0	0	10	0	1	0	
19	Hihikiwi	1198	3	0	0	1	0	0	0	0	0	0	0	0	2	0	0	8		0	
20	Hihikiwi	1199	3	0	0	0	0	0		0	0		5	2		4	1			0	
21	Hihikiwi	1200	3	0	0	0	1	0	1	2	0	7	7	0	1	1	3	17		0	
22	Hihikiwi	1201	3	0	0	1	0	0		0	0		0	0		4	6			0	
23	Hihikiwi	1202	3	0	0	0	0	0	0	0	0	0	4	0	0	3	1	2		0	
24	Hihikiwi	1695	3		0	1			4			0			0			2		0	
25	Hihikiwi	1696	3		0	1														0	
26	Hihikiwi	1697	3		0	1														0	
27	Hihikiwi	1699	3		0	1			2			10			0			0		2	
28	Hihikiwi	1701	3		0	1			11			3			1			0		0	
29	Hihikiwi	1702	3		0	1		0	0		0	0		2	0		1	0		0	
30	Hihikiwi	1719	3		0	1			1			3			6			1		0	
31	Hihikiwi	New	3		0	0															
32	Hihikiwi	1698	3		0	1			18			16			0			1		0	
33	Hihikiwi	1700	3		0	0			0			0			0			6		0	
34	Hihikiwi Sth	1709	4		0	1			2			4			0			0		0	
36	Hihikiwi Sth	1711	4		0	1			1			0			0			0		0	
38	Hihikiwi Sth	1713	4		0	0			0			0			0			19		0	
39	Hihikiwi Sth	1714	4		0	0			0			0			0			0		0	
40	Hihikiwi Sth	1715	4		0	1			7			9			1			0		0	
41	Hihikiwi Sth	1716	4		0	1			0			0			0			0		0	
42	Hihikiwi Sth	1717	4		0	1			31			1			0			0		0	
43	Hihikiwi Sth	1718	4		0	0			0			0			0			1		0	
44	Middle Bell Track	1209	5		1	0	2		7	1		2	2		1	2		0		0	
45	Middle Bell Track	1706	5	1	1	0			1			0			1			0		0	
46	Nth of Hihikiwi	1203	6	0	0	0	1	0	0	0	1	8	0	0	0	3	8	2	0	0	
47	Nth of Hihikiwi	1204	6	0	0	1	2	0	2	3	0	27	1	0	1	5	15	2	0	0	
48	Nth of Hihikiwi	1205	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
49	Nth of Hihikiwi	1206	6	0	0	1	0	0	1	0	0	0	0	2	0	0	0	0	0	0	
50	Nth of Hihikiwi	1207	6	0	0	1	0	0	0	0	0	7	4	0	0	1	11	2	0	0	
51	Nth of Hihikiwi	1208	6	0	0	1	0	0	5	0	0	0	3	3	0	0	3	1	0	0	
52	On Bell Tk	1106	7	0	0	0														0	
53	On Bell Tk	1705	7	0	0	1														0	
54	Saddle before Cone	1704	8	0	0	1														1	

Data Set 4: Monitoring of Vegetation in Whareorino Forest, 1995-99

Contact: Pim de Monchy, Waikato
Data format verified by him: 19-Aug-99

Foliage cover (%), possum browse score (0-4), possum trunk use score (0-3), flowering level (0-4), and fruiting level (0-4).
Blanks are unknown values for various reasons.
All trees that died have been removed.
Possum control (aerial 1080) occurred over whole area in September 1995. Possum trap rates: Before, 60.6%; After, 7.4%.
Animal Health Board TB vector control (aerial 1080) in July 1998 in the area for line 2 only.

Plant Species Codes (1 plot with another species removed)
1 Kohekohe (Dysoxylum spectabile)
2 Mangeao (Litsea calicaris)
3 Northern rata (Metrosideros robusta)
4 Kaikomako (Pennantia corymbosa)
5 Hall's totara (Podocarpus hallii)
6 Kamahi (Weinmannia racemosa)

						Foliage Cover (%)					Browse (W)					Use					Flowers					Fruit				
Case	Line	Plot	Diameter	Tag	Species	95	96	97	98	99	95	96	97	98	99	95	96	97	98	99	95	96	97	98	99	95	96	97	98	99
1	1	1		349	4	65	65	65	55	65		0	0	1	0	2	1	1	2	2	1	3	3	3	2	0	0	0	0	0
2	1	1	15	350	4	55	65	65	65	55		0	0	0	0	2	1	0	1	1	3	4	0	4	1	0	0	0	0	0
3	1	1	10	351	4	55	55	65	45	55		0	0	0	1	1	0	0	1	0	2	4	3	4	3	0	0	0	0	0
4	1	1	20	352	4	75	65	75	65	35		0	0	0	1	1	0	0	1	2	4	4	4	4	3	0	0	0	0	0
5	1	2	12	354	4	45	55	65	55	45		0	0	1	1	1	0	0	1	1	2	3	1	2	0	0	0	0	0	0
6	1	2	23	353	4	55	55	55	45	55		0	0	0	1	2	0	1	1	1	3	3	2	2	2	0	0	0	0	0
9	1	3	40	367	6	15	35	55	65	55	3	1	1	1	2	1	0	1	1	1	0	0	0	0	2	0	0	0	0	2
10	1	3	30	364	6	15	25	65	65	65	3	0	1	1	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1
11	1	3	18	366	6	5	5	5	15	15	4	0	1	1	1	2	0	1	2	1	0	0	0	0	0	0	0	0	0	0
14	1	4	25	362	6	65	65	65	55	65				1	1	2	1	1	1	1	0	1		2	2	0	0		0	3
16	1	4	16	361	6	5	15	25	15	25	4	2	1	2	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
17	1	5	17	358	2	5	25	45	15	25	4	3	0	4	3	2	1	0	1	1	0	0	0	0	0	0	0	0	0	0
18	1	5	17	357	2	25	35	55	25	45	4	2	0	4	2	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0
19	1	5	20	356	2	15	15	25	15	35	4	3	3	3	1	2	2	0	0	1	0	0	0	0	0	0	0	0	0	0
20	1	5	19	355	6	35	55	45	55	65	2	1	1	1	1	2	1	1	1	1	0	0	0	0	0	0	0	0	0	0
21	1	5	12	359	6	25	35	45	35	35	2	1	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
22	1	6	250	360	3	65	75	65	65	65					0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0
24	1	7	30	318	3	35	35	55	35	35					1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
25	1	7	40	317	3	25	15	35	5	25					1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
27	1	7	60	315	3	35	45	45	25	35					1	3	3	1	1	2	0	0	0	0	0	0	0	0	0	0
28	1	7	130	314	3	45	45	45	35	25					0	3	3	0	1	1	0	0	0	0	0	0	0	0	0	0
29	1	8	130	313	3	25	25	35	15	15				1	0	2	0	0	1	1	0	0	0	0	0	0	0	0	0	0
30	1	8	150	312	3	65	65	65	55	55					1	3	0	1	1	1	0	0	0	0	0	0	0	0	0	0
31	1	8	140	306	3	75	75	65	65	55					1	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0
32	1	8	120	305	3	65	65	65	55	55					1		0			2	0	0	0	0	0	0	0	0	0	0
33	1	9	40	311	2	35	75	85	75	85	4	1	0	0	0	2	1	0	1	1	0	0	0	0	0	0	0	0	0	0
35	1	10	22	308	6	25	35	55	35	25	4	1	1	2	4	2	0	1	0	1	0	0	0	0	0	0	0	0	0	0
36	1	10	20	307	6	35	55	65	65	55	3	1	2	1	3	2	1	1	1	1	0	0	0	0	0	0	0	0	0	0
37	1	10	16	309	6	35	55	45	35	45	1	1	1	0	1	2	1	0	0	1	0	0	0	1	0	0	0	0	0	0
40	1	11	60	304	6	15	35	35	25	15	4	1		1	4	2	0	0	1	1	0	0	0	0	0	0	0	0	0	1
42	1	12	25	302	6	75	75	75	65	45					3	2	0	1	1	1	0	0	0		0	0	0	0		0
43	1	12	25	301	6	65	75	75	55	35				1	3	2	0	1	2	3	0	0	0	0	0	0	0	0	0	0
44	1	12	45	293	6	15	55	45	35	25	4	1			1	3	2	1	1	2	0	0	0	0	0	0	0	0	0	3
45	1	13	20 & 50	300	6	75	75	75	55	65				1	1	3	1	1	1	1	0		0	2	0	0	0	0	0	3
46	1	13	35	299	6	75	75	65	55	55		0		1	1	2	0	1	0	0	0		0	3	1	0	0	0	0	3
47	1	13	20	298	6	75	75	75	55	65				1	0	2	1	1	1	1	0		0	2	0	0	0	0	0	3
48	1	13	40	297	6	15	45	45	45	35	4	1		1	1	3	1	1	1	1	0	0	0	3	0	0	0	0		3
49	1	13	30	294	6	35	55	65	55	55	3	1		1	1	3	1	1	1	0	0	1	0	2	2	0	0	0	0	3
51	2	1	35	285	6	35	45	55	55	55	1	1	1	0	0	2	0		0	0	0	2	2	4	3	0	0	0	0	3
52	2	1	10	284	6	75	65	65	65	75	1	1	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	2
53	2	2	15	295	6	55	55	75	55	55	0	1	0	0	0	1	0	0	1	0	0	0	0	2	0	0	0	0	0	0
54	2	2	40	292	6	75	85	75	65	75	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
55	2	3	17	290	6	65	65	55	45	55	1	0	0	0	0	1	0	1	0	1	0	0	1	1	0	0	0	0	0	0
56	2	4	10	291	6	65	55	55	45	55	1	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
57	2	4	10	283	6	55	55	55	35	65	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
58	2	4	20	289	6	75	65	65	65	65	0	0	0	0	0	0	0	0	1	0	2	2	1	2	0	0	0	0	0	0
59	2	4	16	281	6	75	85	75	65	75	1	1	1	0	1	0	0	0	0	0	1	2	0	2	1	0	0	0	0	3
60	2	4	15	282	6	75	75	65	65	75	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
61	2	4	20	1105	6	65	75	75	55	75	2	1	2	1	1	1	1	1	1	0	1	3	0	2	0	0	0	0	0	3
62	2	4	16	286	6	75	75	65	55	75	1	1	1	1	1	1	0	0	0	0	0	0	2	2	1	0	0	0	0	2
63	2	4	11	280	6	65	65	65	55	75	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1
65	2	5	16	287	6	45	65	65	55	65	0	0	1	1	0	0	0	0	0	0	2	2	2	3	1	0	0	0	0	3
66	2	5	12	279	6	45	55	65	65	75	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
68	2	6	350	263	3	35	45	55	65	65								0	1	0	0	0	0		0	0	0	0	0	0
70	2	7	20	264	6	55	55	65	55	55	1	0	0	0	0	2	1	1	1	0	1	1	1	3	1	0	0	0	0	3
71	2	7	32	265	6	65	65	65	65	65	1	0	0	0	0	2	0	1	1	0	0	3	2	3	0	0	0	0	0	4
74	2	8	45	278	6	35	35	55	35	15	3	1	1	0	1	2	0	1	1	1	2	3	3	3	3	0	0	0	0	2
76	2	9	12	275	6	45	65	65	45	65	1	1	1	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	3
77	2	9	22	277	6	45	55	55	45	65	2	1	1	0	0	3		0	1	0	0	2	1	3	1	0	0	0	0	3
78	2	9	50	276	6	55	65	65	45	65	2	1	0	0	0	3	1	1	1	0	0	2	1	4	1	0	0	0	0	3
80	2	10	23	274	6	75	75	75	55	65	1	1	0	0	1	3	2	0	1	0	0	1	2	1	1	0	0	0	0	3
81	3	1	18	262	6	75	75	75	65	75	1	1	0	0	0	1	0	1	0	1	1	4	0	4	2	0	0	0	0	3
82	3	1	15 & 18	261	6	75	75	75	65	75	1	1	0	0	1	3	1	0	1	0	0	4	0	3	3	0	0	0	0	3
83	3	1	30	260	6	15	15	55	65	75	4	2	0	0	1			0	0	0	0	1	0	4	2	0	0	0	0	4

Data Set 5: Blue Cod (Parapercis colias) Size Data from Paterson Inlet, 1994-98

Contact: Rob Davidson, Nelson
Data format verified by him: 23-Aug-99

Reserve Code: 1 in proposed reserve, 0 outside,
There are up to 126 measurements of individual cod for each site/year combination.
Site numbers are the same as for Data Set 6.

	Site					Summary		Length of Cod (mm) ----------------->															
Site	Number	Year	Month	Reserve	n	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
PR1	1	94	5	1	33	271.85	49.61	395	396	250	260	323	360	308	300	281	299	275	212	303	220	276	212
PR4	4	94	5	1	25	314.76	51.54	425	317	304	329	414	367	246	254	421	300	355	294	276	264	283	289
PR6	6	94	5	1	40	291.55	45.51	245	371	300	269	311	230	209	278	304	303	291	294	282	300	316	258
PR7	7	94	5	1	15	287.67	54.29	242	333	394	261	236	341	297	301	319	205	360	247	275	283	221	
PR8	8	94	5	1	77	266.23	41.70	312	281	287	290	267	275	253	240	351	210	382	267	256	320	230	247
PR9	9	94	5	1	53	269.85	36.75	355	223	257	219	250	304	273	263	410	363	324	318	259	279	300	268
PC11	11	94	5	0	40	268.80	35.71	301	352	231	302	246	268	268	293	313	258	256	226	258	257	284	260
PC12	12	94	5	0	65	294.95	46.37	235	255	288	257	252	255	379	292	392	312	308	290	277	271	291	253
PC13	13	94	5	0	6	280.83	50.51	226	265	272	325	240	357										
PC16	16	94	5	0	52	280.29	54.94	323	310	349	245	322	229	306	291	282	341	388	302	407	244	269	243
PC19	19	94	5	0	79	297.13	48.04	444	444	333	404	418	394	366	240	335	289	260	251	259	357	282	272
PC20	20	94	5	0	55	300.18	52.35	336	411	273	378	357	267	362	357	316	229	439	344	326	282	200	272
PR1	1	94	10	1	26	256.27	51.74	238	290	355	337	252	219	268	359	203	189	282	282	249	333	294	267
PR4	4	94	10	1	3	292.67	24.58	316	267	295													
PR6	6	94	10	1	37	320.14	46.52	390	420	425	323	326	275	317	394	318	290	331	306	282	332	304	352
PR7	7	94	10	1	6	294.33	75.71	263	304	336	416	201	246										
PR8	8	94	10	1	100	255.27	41.67	353	383	388	229	273	246	286	237	304	249	238	224	230	309	295	238
PR9	9	94	10	1	85	287.96	48.38	357	271	378	350	305	318	359	225	331	323	313	251	313	308	265	281
PC11	11	94	10	0	89	281.25	36.79	280	284	255	313	263	252	266	239	284	252	284	327	269	229	231	277
PC12	12	94	10	0	33	285.27	37.60	290	282	375	300	199	252	337	266	299	290	274	300	374	233	305	298
PC13	13	94	10	0	16	299.88	60.69	292	365	240	258	355	277	380	367	267	210	375	216	359	323	285	229
PC16	16	94	10	0	87	277.99	41.60	360	303	317	242	279	280	290	283	246	362	287	220	223	327	353	283
PC19	19	94	10	0	90	291.94	37.09	374	268	292	335	309	267	305	376	311	262	290	269	335	249	396	253
PC20	20	94	10	0	79	297.48	40.25	342	331	290	263	328	367	255	371	228	299	386	338	382	356	315	292
PR1	1	95	5	1	14	286.86	69.30	445	335	235	320	287	265	250	222	405	232	210	245	300	265		
PR4	4	95	5	1	10	351.30	71.84	305	458	379	336	301	276	460	306	416	276						
PR6	6	95	5	1	39	306.13	43.73	340	315	365	415	260	290	315	325	355	300	320	315	290	250	330	290
PR7	7	95	5	1	26	321.15	45.57	364	404	334	314	281	314	350	357	319	403	336	361	430	270	281	248
PR8	8	95	5	1	90	260.29	40.20	258	273	291	288	243	322	275	344	272	269	269	313	319	212	254	269
PR9	9	95	5	1	53	303.34	38.65	288	314	341	321	309	266	280	339	325	356	340	423	317	272	414	289
PC11	11	95	5	0	93	290.47	46.38	292	337	265	310	295	348	268	255	322	273	212	199	258	271	164	292
PC12	12	95	5	0	41	282.39	50.22	300	265	275	190	360	265	385	390	285	350	290	295	305	295	298	277
PC13	13	95	5	0	23	364.13	37.26	320	354	425	344	342	366	314	415	332	398	422	363	392	368	417	287
PC16	16	95	5	0	78	309.63	41.60	315	280	280	255	260	335	320	290	410	305	280	315	305	290	360	365
PC19	19	95	5	0	47	299.28	52.03	284	436	293	263	257	274	270	279	203	276	200	238	219	236	258	325
PC20	20	95	5	0	37	302.41	52.91	254	283	275	255	365	312	355	298	225	282	276	303	324	293	255	253
PR1	1	96	5	1	12	312.58	52.84	338	230	290	303	307	359	307	372	422	267	293	263				
PR4	4	96	4	1	5	364.80	64.50	452	370	389	335	278											
PR6	6	96	5	1	31	294.26	51.89	380	221	279	358	322	307	266	285	325	315	193	296	332	239	257	297
PR7	7	96	4	1	10	270.90	53.21	285	332	341	330	263	181	217	226	272	262						
PR8	8	96	5	1	53	263.68	42.22	261	185	315	225	259	296	242	312	193	186	236	292	281	286	198	315
PR9	9	96	5	1	68	307.94	50.92	449	292	375	265	358	370	312	340	343	305	276	390	232	225	322	418
PC11	11	96	5	0	120	291.97	50.12	292	318	271	245	230	396	242	318	337	286	342	282	387	279	203	252
PC12	12	96	5	0	57	305.91	51.26	320	390	345	283	315	342	332	290	314	344	256	328	312	360	283	233
PC13	13	96	4	0	14	317.86	46.45	442	275	371	273	314	326	305	299	299	285	358	327	303	273		
PC16	16	96	5	0	80	311.06	54.51	377	308	211	187	283	312	318	309	307	306	300	326	312	208	329	404
PC19	19	96	5	0	47	262.55	52.98	266	282	240	282	283	276	255	336	375	296	185	293	353	286	311	222
PC20	20	96	5	0	16	278.00	58.05	306	338	300	337	206	270	381	313	341	280	195	216	210	249	296	210
PR1	1	97	6	1	4	325.50	60.63	262	299	404	337												
PR4	4	97	4	1	21	320.86	63.73	330	330	413	306	435	388	378	219	302	336	424	314	343	321	278	360
PR6	6	97	6	1	62	276.42	54.38	285	238	296	242	316	258	280	271	319	231	228	210	231	360	256	232
PR7	7	97	5	1	97	309.22	38.01	289	306	284	257	302	310	360	250	328	304	336	338	319	283	285	260
PR8	8	97	6	1	12	260.25	63.76	215	333	396	245	204	236	281	324	189	198	277	225				
PR9	9	97	6	1	7	310.29	39.27	320	316	324	303	282	377	250									
PC11	11	97	5	0	126	292.79	48.68	348	309	271	290	241	256	218	260	308	291	278	303	275	285	274	225
PC12	12	97	5	0	59	284.93	45.96	341	246	287	237	278	272	245	340	268	298	268	285	272	410	338	310
PC13	13	97	5	0	14	342.14	56.26	395	353	362	248	381	385	317	334	378	311	407	254	260	405		
PC16	16	97	5	0	26	273.23	77.72	176	265	188	341	281	165	265	236	342	412	338	318	190	197	229	380
PC19	19	97	5	0	62	285.26	45.77	345	354	333	325	295	317	326	251	266	219	257	364	317	280	242	318
PC20	20	97	6	0	15	320.33	70.42	344	380	375	270	441	270	316	354	326	313	250	245	201	278	442	
PR1	1	98	4	1	67	300.87	75.27	214	431	238	208	218	225	221	299	212	410	322	197	355	386	325	335
PR4	4	98	4	1	24	307.58	50.40	419	294	328	311	310	234	345	366	264	342	324	246	293	359	342	297
PR6	6	98	4	1	97	320.29	48.34	317	397	343	379	298	381	352	345	284	410	343	380	331	267	322	379
PR7	7	98	4	1	80	299.69	33.56	284	210	296	303	327	298	301	284	391	298	364	291	324	337	295	334
PR8	8	98	4	1	53	271.00	46.93	227	233	244	218	242	287	266	274	278	274	239	193	255	303	268	278
PR9	9	98	4	1	63	276.98	47.99	403	214	259	202	271	214	258	288	243	228	277	343	235	307	420	283
PC11	11	98	4	0	91	285.91	44.67	365	315	255	255	296	281	398	247	237	298	349	276	366	265	260	269
PC12	12	98	4	0	76	290.49	47.98	355	237	299	442	382	271	398	330	258	296	228	280	293	332	295	245
PC13	13	98	4	0	9	301.11	80.40	248	393	215	246	262	250	454	292	350							
PC16	16	98	4	0	48	274.23	77.39	244	280	303	290	280	425	325	311	200	225	297	295	307	289	419	234
PC19	19	98	4	0	55	281.35	39.33	261	291	292	272	294	327	236	311	257	222	236	254	294	223	268	270
PC20	20	98	4	0	44	338.05	70.29	242	403	335	184	362	360	402	270	415	239	294	404	329	298	340	354

Data Set 6: Blue Cod (*Parapercis colias*) Numbers Data from Paterson Inlet, 1994-98

Contact: Rob Davidson, Nelson
Data format verified by him: 23-Aug-99

Site numbers are the same as for the cod size data.
Reserve Code: 1 in proposed reserve, 0 outside,
Replicates (Rep) were obtained from 30m by 5 m strips along a line, separated by buffer zones.
For some sampling times there are missing sites or less than 6 replicates.

						Number of Fish with Size			
	Site					cm			Water
Case	Number	Year	Month	Reserve	Rep	< 10	10-33	> 33	Depth (m)
1	1	94	5	0	1		6	2	4
2	1	94	5	0	2	1	4		5
3	1	94	5	0	3		7	1	6
4	1	94	5	0	4		2	2	4
5	1	94	5	0	5		2		5
6	1	94	5	0	6	1	3		5
7	2	94	5	0	1	1		1	3
8	2	94	5	0	2		2		2
9	2	94	5	0	3		5	2	5
10	2	94	5	0	4		11	7	5
11	2	94	5	0	5		7	2	5
12	2	94	5	0	6		8	2	3
13	3	94	5	0	1			1	5
14	3	94	5	0	2		1	1	4
15	3	94	5	0	3		1	3	4
16	3	94	5	0	4				4
17	3	94	5	0	5				3
18	3	94	5	0	6				3
19	4	94	5	0	1				7
20	4	94	5	0	2				6
21	4	94	5	0	3		2	1	6
22	4	94	5	0	4		1		6
23	4	94	5	0	5			1	7
24	4	94	5	0	6			1	7
25	5	94	5	0	1		2	1	5
26	5	94	5	0	2		1	1	5
27	5	94	5	0	3		5		5
28	5	94	5	0	4		1	2	5
29	5	94	5	0	5		3	2	5
30	5	94	5	0	6		6	1	5
31	6	94	5	0	1		3	2	8
32	6	94	5	0	2		3	1	9
33	6	94	5	0	3		3		10
34	6	94	5	0	4		2		9
35	6	94	5	0	5		5		10
36	6	94	5	0	6		3		7
37	7	94	5	0	1		5		7
38	7	94	5	0	2	2			9
39	7	94	5	0	3		1		7
40	7	94	5	0	4		3	1	5
41	7	94	5	0	5		5	1	4
42	7	94	5	0	6		5	3	5
43	8	94	5	0	1		8		8
44	8	94	5	0	2	3	11		9
45	8	94	5	0	3	1	11		8
46	8	94	5	0	4	3	14		8
47	8	94	5	0	5	6	16		7
48	8	94	5	0	6		10		9
49	9	94	5	0	1		5	5	5
50	9	94	5	0	2		7	1	4
51	9	94	5	0	3	1	5		5
52	9	94	5	0	4		13	2	7
53	9	94	5	0	5		3	1	7
54	9	94	5	0	6		13	1	7
55	10	94	5	0	1		1	2	7
56	10	94	5	0	2			2	6
57	10	94	5	0	3			1	5
58	10	94	5	0	4		3	1	7
59	10	94	5	0	5		2	1	7
60	10	94	5	0	6		4		5
61	11	94	5	1	1		7		8
62	11	94	5	1	2		7	2	7
63	11	94	5	1	3		5	2	6
64	11	94	5	1	4		23	1	9
65	11	94	5	1	5		15		8
66	11	94	5	1	6		7	1	8
67	12	94	5	1	1		3	1	7
68	12	94	5	1	2		4	1	6
69	12	94	5	1	3		14		9
70	12	94	5	1	4		12	1	10
71	12	94	5	1	5	1	2		11
72	12	94	5	1	6		10		12
73	13	94	5	1	1		1		5

Data Set 7: Tree Conditions at Ryan Creek, Heaphy Valley in 1995 and 1999

Contact: Phil Knightsbridge, Hokitika
Data format verified by him: 25-Aug-99

Possum control occurred in 1997.
There are three lines of plots in 1995 and these plus another line in 1999.
The tag numbers for trees are the same in 1995 and 1999, but all trees were not sampled in both years.
The plot is considered the sample unit rather than the tree.
Codes for Dbtop-Brwho changed in 1999 to agree with those used in 1995.

Variables:
Folcov = foliage Cover in 10 categories from 5-95%.
Dbtop = dieback in top third of tree.
Dbwho = dieback in whole canopy.
Brtop = possum browse in top third of canopy.
Brwho = possum browse on the whole canopy.
Stuse = possum marking on the lower 2m of trunk.
Flower = flowering.
Fruit = fruiting.

Codes	Species	Dbtop	Dbwho	Brtop	Brwho	Stuse	Flower	Fruit
0		None	None	None	None	None	None	None
1	Melicytus ramiflorus (mahoe)	1-25%	1-25%	1-25%	1-25%	Light	Rare	Rare
2	Metrosideros umbellata (southern rata)	26-50%	26-50%	26-50%	26-50%	Moderate	Occasional	Occasional
3	Myrsine salicina (toro)	51-75%	51-75%	51-75%	51-75%	Heavy	Common	Common
4	Podocarpus totara (totara)	>75%	>75%	>75%	>75%		Abundant	Abundant
5	Pseudopanax crassifolius (lancewood)							
6	Raukaua simplex (haumakaroa)							
7	Weinmannia racemosa (kamahi)							

Case	Year	Line	Plot	Tag	Species	Folcov (%)	Dbtop	Dbwho	Brtop	Brwho	Stuse	Flower	Fruit
1	1995	1	1	1073	2	85	2	1	1	2	0	0	0
2	1995	1	1	1072	3	55	2	2	0	0	0	0	0
3	1995	1	1	1077	3	45	1	3	1	2	0	0	0
4	1995	1	1	1080	7	75	1	1	1	1	0	0	0
5	1995	1	1	1078	2	85	2	2	1	1	1	0	0
6	1995	1	1	1076	3	65	1	2	1	1	0	0	0
7	1995	1	1	1075	7	55	3	2	1	1	2	0	0
8	1995	1	2	1074	7	45	1	1	1	1	0	0	0
9	1995	1	2	1071	7	55	1	1	1	1	0	0	0
10	1995	1	2	1070	3	45	3	2	1	1	0	0	0
11	1995	1	2	1069	2	55	3	2	2	2	0	0	0
12	1995	1	2	1068	3	65	3	1	1	1	0	0	0
13	1995	1	2	1067	3	55	1	2	1	1	0	0	0
14	1995	1	3	1066	7	55	1	2	2	2	0	0	0
15	1995	1	3	1065	7	55	1	3	1	2	0	0	0
16	1995	1	3	1064	3	55	1	2	1	1	0	0	0
17	1995	1	3	1063	7	75	1	1	1	1	0	0	0
18	1995	1	3	1062	7	45	1	2	1	1	0	1	0
19	1995	1	3	1061	3	55	1	3	1	1	0	0	0
20	1995	1	3	1060	7	85	3	2	1	1	0	0	0
21	1995	1	3	1059	3	65	1	2	1	1	0	0	0
22	1995	1	3	1058	7	75	1	1	1	1	0	0	0
23	1995	1	3	1057	3	45	2	3	1	1	0	0	0
24	1995	1	4	1056	3	55	2	1	2	2	0	0	0
25	1995	1	4	1055	7	75	1	1	1	1	0	1	0
26	1995	1	4	1054	3	55	1	2	1	1	0	0	0
27	1995	1	4	1053	5	75	1	1	1	1	0	0	0
28	1995	1	4	1052	5	55	0	0	0	0	0	0	0
29	1995	1	4	1051	2	55	3	2	1	1	0	0	0
30	1995	1	5	1050	3	55	1	3	1	1	0	0	0
31	1995	1	5	1049	7	45	1	2	1	1	0	0	0
32	1995	1	5	1048	5	45	1	1	1	1	0	0	0
33	1995	1	5	1047	7	65	3	2	1	1	0	0	0
34	1995	1	5	1046	5	65	2	2	1	1	0	0	0
35	1995	1	5	1045	3	65	3	3			0	0	0
36	1995	1	5	1040	7	55	2	2	1	1	0	0	0
37	1995	1	5	1039	3	75	1	1			0	0	0
38	1995	1	5	1038	3	65	3	3	2	1	0	0	0
39	1995	1	5	1037	7	75	1	1	1	1	0	0	0
40	1995	1	5	1036	3	65	1	3	1	1	0	0	0
41	1995	1	6	1044	7	65	2	2	2	2	0	1	0
42	1995	1	6	1043	7	75	3	2	1	1	0	1	0
43	1995	1	6	1042	7	75	1	1	1	1	0	0	0
44	1995	1	6	1041	7	75	1	1	1	1	0	0	0
45	1995	1	6	1014	7	75	1	2	1	2	0	0	0
46	1995	1	7	1015	7	45	3	3	2	2	0	0	0
47	1995	1	7	1016	7	45	3	2	1	1	0	0	0
48	1995	1	7	1017	3	75	3	2	1	1	0	0	0
49	1995	1	7	1018	7	35	3	2	1	1	0	0	0
50	1995	1	7	1019	3	35	1	2	1	1	0	0	0
51	1995	1	7	1020	3	55	1	2	1	1	0	0	0
52	1995	1	7	1021	7	55	1	2	1	1	0	1	0
53	1995	1	8	1022	3	65	1	1	1	1	0	0	0
54	1995	1	8	1023	7	65	1	1	1	1	0	0	0
55	1995	1	8	1024	3	75	0	0	0	0	0	0	0

Data Set 8: Diets of Fish Collected from Otago Streams

Contact: Eric Edwards, Invercargill
Data format verified by him: 23-Aug-99

Where possible at least 10 fish were sampled for each site. Numbers are not sequential because some had empty guts.
Head = the maximum width measurement across an invertebrate head capsule. Scale is microscope units (0.0425mm).
Length = length of an invertebrate (mm).
Head and Length are measurements used to estimate the size of the prey item or its food value.
Length can be estimated from Head measure and biomass (food value) can be estimated from the Head or Length.

		Predator		Prey
Codes	Site	Species	Source	Type
1	Canton	Galaxiid	Stream	Larva
2	Lee	Trout	Land	Adult
3	Shepherd		Other	Pupa
4	Silver Trib A			Other
5	Stony			
6	Sut			
7	Upper KyeBurn			

Case	Site	Species	Fish	Source	Type	Head	Length	Prey Family
1	1	1	2	1	2	9		Chironomidae
2	1	1	2	1	1	5		Chironomidae
3	1	1	2	1	1	4		Chironomidae
4	1	1	2	1	1	5		Chironomidae
5	1	1	2	1	1	30		Plecoptera
6	1	1	2	1	1	16		Plecoptera
7	1	1	2	1	1	22		Plecoptera
8	1	1	2	1	1	26		Plecoptera
9	1	1	2	1	1	24		Plecoptera
10	1	1	2	1	1	20		Plecoptera
11	1	1	2	1	1	26		Plecoptera
12	1	1	2	1	1	25		Plecoptera
13	1	1	2	1	1	16		Plecoptera
14	1	1	2	1	1	25		Plecoptera
15	1	1	2	1	1	27		Plecoptera
16	1	1	2	1	1	25		Plecoptera
17	1	1	2	1	1	21		Plecoptera
18	1	1	2	1	1	23		Plecoptera
19	1	1	2	1	1	22		Plecoptera
20	1	1	2	1	1	24		Plecoptera
21	1	1	2	1	1	21		Plecoptera
22	1	1	2	1	1	20		Plecoptera
23	1	1	2	1	1	25		Plecoptera
24	1	1	2	1	1	23		Plecoptera
25	1	1	2	1	1	16		Scirtidae
26	1	1	2	1	1	16		Scirtidae
27	1	1	2	1	1	40		Zygoptera
28	1	1	2	1	1	20		Zygoptera
29	1	1	2	3	4		10.0	Grass blade
30	1	1	2	3	4		10.0	Grass blade
31	1	1	3	1	1	5		Chironomidae
32	1	1	3	1	1	5		Chironomidae
33	1	1	3	1	1	30		Ephemeroptera
34	1	1	3	1	1	27		Plecoptera
35	1	1	3	1	1	27		Plecoptera
36	1	1	3	1	1	21		Plecoptera
37	1	1	3	1	1	25		Plecoptera
38	1	1	3	1	1	22		Plecoptera
39	1	1	3	1	1	20		Plecoptera
40	1	1	3	1	1	21		Plecoptera
41	1	1	3	1	1	24		Plecoptera
42	1	1	3	1	1	23		Plecoptera
43	1	1	3	1	1	15		Plecoptera
44	1	1	3	1	1	23		Plecoptera
45	1	1	3	1	1	22		Plecoptera
46	1	1	3	1	1		3.0	Unid aq
47	1	1	4	1	2	10		Coleoptera
48	1	1	4	1	1	25		Plecoptera
49	1	1	4	1	1	24		Plecoptera
50	1	1	4	1	1	30		Plecoptera
51	1	1	4	1	2		4.0	Trichoptera
52	1	1	4	1	2	17		Trichoptera
53	1	1	4	1	2	14		Trichoptera
54	1	1	4	1	4	27		UNID aq
55	1	1	4	2	4		4.5	Araneae
56	1	1	4	3	4		2.0	Sand
57	1	1	4	3	4		0.5	Sand
58	1	1	5	1	1	5		Chironomidae
59	1	1	5	1	1	42		Ephemeroptera
60	1	1	5	1	1	21		Ephemeroptera
61	1	1	5	1	1	25		Plecoptera

Data Set 9: Survey of Deer Pellets in the Takahe Special Area, Murchison Mountains, Fiordland National Park, in 1998

Contact: Jane Maxwell, Te Anau
Data format verified with her: 25-Aug-99

Some lines are short (10 plots) and others long (more than 10). The short line plots are 15m apart on transects starting from the river, the long lines continue up to the treeline or an obstacle.
Plots are classified as Bottom if they are on short lines, or one of the first 10 on long lines. Otherwise they are classified as Side.
Pellets are recorded as present or absent in a 1.14m radius circle in each plot.
The number of groups is in a 2.5m radius circle in each plot.
This is a repeat of surveys carried out in 1978 (for pellet frequency of presence) and 1986 (for both frequency of presence and pellet group density).
In 1978, 2.5% of 1019 plots contained pellets. In 1986 1.3% of 1111 plots contained pellets.
Reference for earlier surveys: Nugent, G. & Sweetapple, P., New Zeal. J. Ecol. 12: 33-8.

Variable Codes	Plot Type	Aspect	Slope	Landform	Ground Cover	Type of Forest	Pellets
0							Absen
1	Bottom	East	0-9	Faces	Bare	Seral (a) dominated by Hoheria glabrata and shield fern	Presen
2	Side	South	10-19	Gullies	Crownfern	Seral (b) dominated by Schefflera digitata, Melicytus ramiflorus and/or Fuchsia excorticata.	
3		West	20-29	Ridges, etc.	Grass	Silver beech (a) - mid slope to treeline with varied understorey	
4		North	30-39	Slips	Forest Litter	Silver beech (b) - mixed silver beech/hardwood, often with Weinmannia racemosa or Metrosideros umbellata as a codominant	
5			40-49	Terrace	Moss	Mountain beech forests	
6			50+	Toeslope	Shield Fern	"S" types - short subalpine shrubland grading to low forest	
7						"O" type - forest clearings below timberline, including slips	
8						"A" type - subalpine grassland dominated by Chionocloa spp	

Site: Chester Burn

Case	Line	Plot	Type	Aspect	Slope	Landfon	Cover	Forest	Pellets	Groups
1	1	1	1	1	1	5	5	5	0	1
2	1	2	1	1	1	5	5	5	0	0
3	1	3	1	1	1	5	5	5	0	0
4	1	4	1	1	1	5	5	5	0	0
5	1	5	1	1	1	5	5	5	0	0
6	1	6	1	1	1	5	5	5	0	0
7	1	7	1	1	1	5	5	5	0	0
8	1	8	1	1	2	6	4	5	0	0
9	1	9	1	3	2	6	4	3	0	1
10	1	10	1	1	1	5	4	5	0	0
11	2	1	1	3	5	1	4	4	0	1
12	2	2	1	3	3	1	5	4	0	0
13	2	3	1	3	6	1	4	4	0	0
14	2	4	1	3	6	1	4	4	0	0
15	2	5	1	3	5	1	5	5	0	0
16	2	6	1	3	6	1	4	5	0	0
17	2	7	1	3	6	1	4	4	0	0
18	2	8	1	3	6	1	4	5	0	0
19	2	9	1	3	6	4	4	7	0	0
20	3	1	1	1	1	5	5	4	0	0
21	3	2	1	1	1	5	5	4	0	0
22	3	3	1	1	2	1	5	4	0	0
23	3	4	1	1	2	5	5	4	0	0
24	3	5	1	1	1	5	5	4	0	0
25	3	6	1	1	1	5	5	4	0	0
26	3	7	1	1	1	5	5	5	0	0
27	3	8	1	1	1	5	5	4	0	0
28	3	9	1	1	1	5	4	5	0	0
29	3	10	1	1	1	5	5	5	0	0
30	4	1	1	3	3	1	4	4	0	0
31	4	2	1	3	4	1	4	4	0	0
32	4	3	1	3	6	1	4	4	0	0
33	4	4	1	3	6	1	4	5	0	0
34	4	5	1	3	6	1	4	5	0	0
35	4	6	1	3	6	1	4	5	0	0
36	4	7	1	3	5	1	4	5	1	1
37	4	8	1	3	4	1	2	5	0	0
38	4	9	1	3	4	1	4	5	0	0
39	4	10	1	3	6	1	5	5	0	0
40	5	1	1	1	1	5	4	4	0	0
41	5	2	1	1	1	5	5	4	1	0
42	5	3	1	1	1	5	5	4	0	0
43	5	4	1	1	1	5	5	4	0	0
44	5	5	1	1	1	5	5	4	0	0
45	5	6	1	1	1	5	5	4	0	0
46	5	7	1	3	2	1	2	4	0	0
47	5	8	1	1	1	5	5	4	0	0
48	5	9	1	3	2	2	5	5	0	0
49	5	10	1	1	2	2	5	1	1	0
50	6	1	1	3	1	5	5	3	0	1
51	6	2	1	3	1	5	4	3	0	0
52	6	3	1	3	4	1	4	3	0	0
53	6	4	1	3	4	1	4	3	0	0

Site: Ettrick Burn

Case	Line	Plot	Type	Aspect	Slope	Landfon	Cover	Forest	Pellets	Groups
1	1	1	1	2	6	1	5	3	0	0
2	1	2	1	1	3	1	5	3	0	0
3	1	3	1	1	3	1	5	3	0	0
4	1	4	1	1	5	1	5	3	0	0
5	1	5	1	1	6	1	5	3	0	0
6	1	6	1	1	3	1	5	3	0	0
7	1	7	1	1	6	1	5	3	0	0
8	1	8	1	1	6	1	5	3	0	0
9	1	9	1	1	6	1	5	3	0	0
10	1	10	1	1	6	1	5	3	0	0
11	2	1	1	3	1	5	4	3	0	0
12	2	2	1	3	1	5	4	3	0	0
13	2	3	1	3	1	5	4	3	0	0
14	2	4	1	3	1	5	4	3	0	0
15	2	5	1	3	1	5	4	3	0	0
16	2	6	1	3	1	5	4	3	0	0
17	2	7	1	3	1	5	4	3	0	0
18	2	8	1	3	1	5	4	3	0	0
19	2	9	1	3	1	5	4	3	0	0
20	2	10	1	3	1	5	4	3	0	0
21	3	1	1	2	1	1	5	3	0	0
22	3	2	1	1	2	1	5	3	0	0
23	3	3	1	1	1	1	5	3	0	0
24	3	4	1	2	1	1	1	3	0	0
25	3	5	1	2	1	1	5	3	0	0
26	3	6	1	2	1	1	5	3	0	0
27	3	7	1	2	1	1	5	3	0	0
28	3	8	1	2	2	1	5	3	0	0
29	3	9	1	2	3	1	5	3	0	0
30	3	10	1	2	3	1	5	3	0	0
31	4	1	1	4	1	5	5	3	0	0
32	4	2	1	4	1	5	5	3	0	0
33	4	3	1	4	1	5	5	3	0	0
34	4	4	1	4	1	5	5	3	0	0
35	4	5	1	4	1	5	5	3	0	0
36	4	6	1	4	2	6	5	3	0	0
37	4	7	1	4	2	6	5	3	0	0
38	4	8	1	4	2	6	5	3	0	0
39	4	9	1	4	3	6	5	3	0	0
40	4	10	1	4	3	2	5	3	0	0
41	4	11	2	4	3	2	5	3	0	0
42	4	12	2	4	3	2	5	3	0	0
43	4	13	2	4	3	1	5	3	0	0
44	4	14	2	4	3	1	5	3	0	0
45	4	15	2	4	3	1	5	3	0	0
46	4	16	2	4	5	1	5	3	0	0
47	4	17	2	4	6	1	5	3	0	0
48	4	18	2	4	6	1	5	3	0	0
49	4	19	2	4	6	1	5	3	0	0
50	4	20	2	4	6	1	5	3	0	0
51	4	21	2	4	6	1	5	3	0	0
52	4	22	2	4	6	1	4	3	0	0
53	4	23	2	4	5	1	4	3	0	0

Site: Snag Burn

Case	Line	Plot	Type	Aspect	Slope	Landfon	Cover	Forest	Pellets	Groups
1	1	1	1	4	2	1	5	3	0	0
2	1	2	1	4	2	1	5	3	0	0
3	1	3	1	4	2	1	5	3	0	0
4	1	4	1	4	2	1	5	3	0	0
5	1	5	1	4	2	1	5	3	0	0
6	1	6	1	4	2	1	5	3	0	0
7	1	7	1	4	2	1	5	3	0	0
8	1	8	1	4	2	1	5	3	0	0
9	1	9	1	4	2	1	5	3	0	0
10	1	10	1	4	2	1	5	3	0	0
11	2	1	1	4	1	1	5	3	0	0
12	2	2	1	4	2	1	5	3	0	0
13	2	3	1	4	1	1	5	3	0	0
14	2	4	1	4	1	1	5	3	0	0
15	2	5	1	1	2	1	5	3	0	0
16	2	6	1	1	2	1	5	3	0	0
17	2	7	1	1	1	1	5	3	0	0
18	2	8	1	1	2	1	5	3	0	0
19	2	9	1	1	2	1	5	2	0	0
20	2	10	1	1	3	1	6	2	0	0
21	3	1	1	4	1	1	5	3	0	0
22	3	2	1	4	1	1	5	3	0	0
23	3	3	1	4	2	1	5	3	0	0
24	3	4	1	4	2	1	5	3	0	0
25	3	5	1	4	2	1	5	3	0	0
26	3	6	1	3	2	1	5	3	0	0
27	3	7	1	3	1	1	5	3	0	0
28	3	8	1	4	2	1	5	3	0	0
29	3	9	1	4	2	1	5	3	0	0
30	3	10	1	3	3	1	5	3	0	0
31	4	1	1	4	1	6	5	3	0	0
32	4	2	1	4	1	6	5	3	0	0
33	4	3	1	4	1	6	5	3	0	0
34	4	4	1	4	1	6	6	2	0	0
35	4	5	1	4	2	6	6	2	0	0
36	4	6	1	1	2	6	6	2	0	0
37	4	7	1	1	2	6	6	2	0	0
38	4	8	1	1	2	6	6	2	0	0
39	4	9	1	1	3	6	6	2	0	0
40	4	10	1	1	3	6	6	2	0	0
41	5	1	1	3	2	1	6	2	0	0
42	5	2	1	3	2	1	6	2	0	0
43	5	3	1	3	2	1	5	3	0	0
44	5	4	1	4	1	1	6	3	0	0
45	5	5	1	4	1	3	6	3	0	0
46	5	6	1	4	2	3	5	3	0	0
47	5	7	1	3	1	5	6	3	0	0
48	5	8	1	3	1	1	5	3	0	0
49	5	9	1	3	2	1	5	3	0	0
50	5	10	1	3	2	1	6	3	0	0
51	6	1	1	1	1	6	6	2	0	0
52	6	2	1	1	1	6	6	2	0	0
53	6	3	1	1	1	6	5	2	0	0

Site: Point Burn

Case	Line	Plot	Type	Aspect	Slope	Landfon	Cover	Forest	Pellets	Groups
1	1	1	1	1	1	5	3	7	0	1
2	1	2	1	1	1	5	3	7	0	0
3	1	3	1	1	1	5	6	3	0	0
4	1	4	1	1	1	5	6	3	0	0
5	1	5	1	1	1	6	4	3	0	0
6	1	6	1	1	1	6	6	3	0	0
7	1	7	1	1	2	6	4	3	0	0
8	1	8	1	1	2	6	5	3	0	0
9	1	9	1	1	2	6	6	3	0	0
10	1	10	1	1	2	6	4	3	0	0
11	2	1	1	2	1	6	5	3	0	0
12	2	2	1	2	2	6	6	3	0	0
13	2	3	1	2	2	6	6	3	0	0
14	2	4	1	2	2	6	5	3	0	0
15	2	5	1	2	3	6	6	3	0	0
16	2	6	1	2	3	6	5	3	0	0
17	2	7	1	2	3	6	6	3	0	0
18	2	8	1	2	3	6	6	3	0	0
19	2	9	1	2	3	6	5	3	0	0
20	2	10	1	2	2	6	5	3	0	0
21	3	1	1	4	1	5	6	3	0	0
22	3	2	1	4	1	5	6	3	0	0
23	3	3	1	4	1	5	4	3	0	0
24	3	4	1	4	2	6	5	3	0	0
25	3	5	1	4	2	6	5	3	0	0
26	3	6	1	4	2	6	5	3	0	0
27	3	7	1	4	2	6	5	3	0	0
28	3	8	1	4	2	6	4	3	0	0
29	3	9	1	4	2	6	4	3	0	0
30	3	10	1	4	2	6	4	3	0	0
31	4	1	1	2	2	4	3	7	0	0
32	4	2	1	2	3	4	3	7	0	0
33	4	3	1	2	3	4	6	7	0	0
34	4	4	1	2	2	4	3	7	0	0
35	4	5	1	2	3	4	3	7	0	0
36	4	6	1	2	3	4	3	7	0	0
37	4	7	1	2	3	4	3	7	0	0
38	4	8	1	2	3	4	3	7	0	0
39	4	9	1	2	3	4	6	7	0	0
40	4	10	1	2	3	4	3	7	0	0
41	5	1	1	4	1	5	6	3	0	0
42	5	2	1	4	1	5	6	3	0	0
43	5	3	1	4	2	6	4	3	0	0
44	5	4	1	4	3	6	5	3	0	0
45	5	5	1	4	3	6	5	3	0	0
46	5	6	1	4	4	1	5	3	0	0
47	5	7	1	4	5	1	5	3	0	0
48	5	8	1	4	4	1	4	3	0	0
49	5	9	1	4	4	1	4	3	0	0
50	5	10	1	4	4	1	4	3	0	0
51	6	1	1	2	1	5	6	3	0	0
52	6	2	1	2	1	6	6	3	0	0
53	6	3	1	2	1	6	6	3	0	0

Data Set 10: Angler Survey Results at Lake Taupo, 1992-99

Contact: Michel Dedual, Turangi
Data format verified with him: 27/8/99

Echo-sounder counts of legal-size fish (generally in November)

Year	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Count	89.9			108.0	115.0	145.0	205.2	144.7	117.8	186.8	112.5

CPUE = (Fish Kept + Over-Size Fish Returned)/Hours.
Size, Quality, Success and Enjoyment are assessed by the anglers on a 5-point scale, with 5 being best.
Rating scales changed to Size&Quality, Satisfaction, and Enjoyment during 1996.

Code	Summer	Method
0		Unknown
1	1991/92	Deep Trolling Downrigger
2	1992/93	Deep Trolling Lead Line
3	1993/94	Deep Trolling Wire Line
4	1994/95	Fly Fishing Floating Line
5	1995/96	Fly Fishing Sinking Line
6	1996/97	Shallow Trolling (Harling)
7	1997/98	
8	1998/99	

					Fish Returned			Angler Ratings of Fishing					
		Fishing	Fishing	Fish	Over	Under						Size &	
Case	Summer	Method	Hours	Kept	Size	Size	CPUE	Size	Quality	Success	Enjoyment	Quality	Satisfactior
1	1	2	2.00	1	0	0	0.50	5	4	5	5		
2	1	6	0.50	0	0	0	0.00						
3	1	2	1.00	0	0	0	0.00						
4	1	2	1.50	0	0	0	0.00	3	4	2	4		
5	1	2	1.50	0	0	0	0.00	3	4	2	5		
6	1	2	1.00	0	0	0	0.00						
7	1	2	0.50	0	0	0	0.00						
8	1	2	1.50	0	0	0	0.00						
9	1	2	1.50	0	0	0	0.00						
10	1	2	2.45	0	0	0	0.00						
11	1	2	2.50	1	0	0	0.40	3	5	3	5		
12	1	2	1.00	0	0	0	0.00						
13	1	3	0.50	0	0	0	0.00						
14	1	6	2.00	0	0	0	0.00	3	3	2	4		
15	1	6	1.50	0	0	0	0.00						
16	1	6	1.50	0	0	0	0.00						
17	1	6	1.50	0	0	0	0.00			1	5		
18	1	6	2.50	0	0	0	0.00	3	2	2	4		
19	1	6	2.50	3	0	0	1.20	3	2	2	4		
20	1	6	2.50	0	0	0	0.00	3	2	1	5		
21	1	6	1.25	0	0	0	0.00						
22	1	6	1.25	0	0	0	0.00						
23	1	6	1.50	0	0	0	0.00			3	5		
24	1	6	2.50	1	0	0	0.40						
25	1	6	2.50	0	0	0	0.00						
26	1	6	2.45	1	0	0	0.41						
27	1	6	0.25	0	0	0	0.00	4	4	4	4		
28	1	6	0.25	0	0	0	0.00						
29	1	6	0.25	0	0	0	0.00	2		2	2		
30	1	2	1.50	2	0	0	1.33	4	4	3	4		
31	1	2	1.50	0	0	0	0.00	4	4	3	4		
32	1	2	2.00	0	0	0	0.00	3	4	2	4		
33	1	2	2.00	0	0	0	0.00	3	4	3	4		
34	1	2	3.00	0	0	1	0.00	4	3	4	5		
35	1	2	3.00	1	0	0	0.33	3	4	2	4		
36	1	2	3.00	0	0	0	0.00	3	4	2	4		
37	1	2	5.00	0	1	0	0.20	3	3	3	3		
38	1	2	1.00	0	0	0	0.00	3	2	5	5		
39	1	2	1.00	0	0	0	0.00	3	2	5	5		
40	1	2	1.50	0	0	0	0.00	3	4	2	4		
41	1	2	1.50	0	0	0	0.00	3	4	2	4		
42	1	2	2.00	1	0	0	0.50						
43	1	2	2.00	0	0	0	0.00						
44	1	2	2.00	0	0	0	0.00	5	4	5	5		
45	1	2	1.00	1	0	0	1.00	3	4	4	4		
46	1	2	1.00	0	0	0	0.00	4	3	3	3		
47	1	2	1.00	0	0	0	0.00	4	3	3	3		
48	1	2	2.00	1	1	0	1.00	3	4	3	3		
49	1	2	2.00	1	1	0	1.00	3	4	3	3		
50	1	3	2.50	0	0	0	0.00	4	5	5	5		
51	1	3	2.00	1	0	0	0.50	3	4	4	4		
52	1	6	8.00	1	0	0	0.13	4	4	4	4		
53	1	6	3.00	2	0	0	0.67						
54	1	6	3.00	0	0	0	0.00						
55	1	6	3.00	0	0	0	0.00						
56	1	6	1.00	1	1	0	2.00						
57	1	6	1.00	0	0	0	0.00						
58	1	6	3.00	0	1	2	0.33	4	5	2	4		

Data Set 11: Monitoring of Mistletoe in the Eglington Valley, Fiordland, 1995-97

Contact: Brian Rance, Invercargil
Date format verified with him: 30/8/99

Baseline surveys done in December 1994 and March 1995

Code	Area	Site	Treatment	Host	Species	Damage Condition
1	Deer Flat	Deer Flat	None	Silv Beech	A flavida	None
2	Dore Pass	Dore Pass	Treated	Mt Beech	P colensoi	1-10% Light
3	Knobs Flat	Knobs Flat			P tetrapetala	10-25% Med
4	NZDA	Totara Flat				25-50% Heavy
5	Totara Flat					50-75% V. Heavy
6	Totara Flat Is					75-99% Severe
7						100% Appears Dead
8						Disappears

					Possum				Condition in Later Samples					
	Mistletoe				Control		Mistletoe	Baseline	1995	1996	1996	1997	1997	1998
Case	Tag	Tree	Area	Site	Treatment	Host	Species	Condition	Sep	Feb	Sep	Jan	Sep	Nov
1	DFPT09a	1143	1	1	2	1	3	2	2	2	2	2	2	
2	DFPT18	T732 (1144)	1	1	2	1	3	7	7	7	7	7	7	
3	DFPT16	1147	1	1	2	1	3	1	1	1	1	2	2	2
4	DFPT15	1146	1	1	2	1	3	3	3	4	4	3	2	3
5	DFPT14	1145	1	1	2	1	3	2	3	7	4	6	6	6
6	DFPT13	1145	1	1	2	1	3	6	7	7	7	7	7	7
7	DFPT12	T732 (1144)	1	1	2	1	3	7	7	7	7	7	7	
8	DFPT10	T732 (1144)	1	1	2	1	3	6	6	6	6	7	7	
9	DFPT21	1148	1	1	2	1	3	2	2	2	2	2	2	2
10	DFPT09	1143	1	1	2	1	3	2	2	2	2	2	2	
11	DFPT08	1141	1	1	2	1	3	2	2	2	3	3	3	3
12	DFPT07	1135	1	1	2	1	3	6	7	7	7	7	7	7
13	DFPT05	B1149 (1133)	1	1	2	1	3	6	6	7	7	7	7	7
14	DFPT03	1132	1	1	2	1	3	6	7	7	7	7	7	8
15	DFPT02	1131	1	1	2	1	3	5	6	6	6	6	4	4
16	DFPT11	T732 (1144)	1	1	2	1	3	7	7	7	7	7	7	
17	DFPC06	1138	1	1	2	1	2	2	2	6	6	6	6	6
18	DFPC12	1145	1	1	2	1	2	1	1	1	1	2	2	2
19	DFPC13	1144	1	1	2	1	2	1	1	1	1	2	2	
20	DFPC14	1145	1	1	2	1	2		1	1	1	2	2	2
21	DFPC11	1142	1	1	2	1	2	5	4	4	4	4	4	2
22	DFPC10	1140	1	1	2	1	2	5	6	6	7	7	7	7
23	DFPC09	1139	1	1	2	1	2	1	2	2	2	2	2	4
24	DFPT19	1148	1	1	2	1	3	2	2	2	2	2	2	2
25	DFPC07	1138	1	1	2	1	2	2	2	2	2	2	2	2
26	DFPT20	1148	1	1	2	1	3	5	6	7	7	7	7	7
27	DFPC05	1138	1	1	2	1	2	1	2	2	3	3	2	2
28	DFPC04	1138	1	1	2	1	2	1	1	1	2	2	2	2
29	DFPC03	1137	1	1	2	1	2	2	1	1	1	2	2	2
30	DFPC02	1136	1	1	2	1	2	3	3	3	4	4	4	4
31	DFPC01	1134	1	1	2	1	2	1	1	2	2	2	2	2
32	DFPT22	T751	1	1	2	1	3		2	2	3	3	3	
33	DFPT04	B1149 (1133)	1	1	2	1	3	2	2	2	2	2	2	2
34	DFPC08	1138	1	1	2	1	2	2	2	2	2	2	2	2
35	DFPT01	1131	1	1	2	1	3	6	7	7	7	7	7	7
36	DFPT06	B1149 (1133)	1	1	2	1	3	5	6	6	6	7	7	7
37	DPPC10	G29	2	2	2	1	2	2	2	1	1	2	2	2
38	DPPC05	G28	2	2	2	1	2	5	6	5	6	5	5	5
39	DPPCO6	G28	2	2	2	1	2	6	7	7				
40	DPPC07	G29	2	2	2	1	2	2	2	2	2	2	2	2
41	DPPC02	G26	2	2	2	1	2	4	6	7	7	7	7	7
42	DPPC09	G29	2	2	2	1	2	5	6	6	7	7	8	8
43	DPPC08	G29	2	2	2	1	2	2	2	2	2	2	2	3
44	DPPC04	G28	2	2	2	1	2	2	2	2	2	2	2	3
45	DPPC01	G25	2	2	2	1	2	6	6	6	7	7	7	7
46	DPPT05	G26	2	2	2	1	3		5	6	6	6	5	4
47	DPPT04	G26	2	2	2	1	3		6	7	7	7	7	7
48	DPPT03	G26	2	2	2	1	3	5	6	7	7	7	7	7
49	DPPT02	G24	2	2	2	1	3		2	2	2	2	8	8
50	DPPT01	G24	2	2	2	1	3	5	6	5	5	5	8	8
51	DPPC03	G27	2	2	2	1	2	6	8	8	8	8	8	8
52	DPPC06	G28	2	2	2	1	2	6			7	7	7	8
53	DPPC11	G29	2	2	2	1	2					2		
54	KFPC02	G44	3	3	2	1	2	1	1	1	1	1	1	1
55	KFPT08	G46	3	3	2	1	3	2	2	2	2	2	2	2
56	KFPT01	G40	3	3	2	1	3	2	2	2	2	2	3	2
57	KFPT07	G45	3	3	2	1	3	1	1	1	1	1	1	1
58	KFPT06	G45	3	3	2	1	3	1	1	7	8	8	8	8
59	KFPT05	G43	3	3	2	1	3	2	2	2	2	2	2	2
60	KFPT04	G42	3	3	2	1	3	1	1	2	2	2	2	2
61	KFPT03	G41	3	3	2	1	3	2	2	2	2	2	2	2
62	KFPT02	G40	3	3	2	1	3	2	2	2	2	2	2	2
63	KFPT10	G47	3	3	2	1	3	2	2	2	2	3	2	8

64	KFPT11	G48	3	3	2	1	3	1	1	2	2	2	2	2
65	KFPT12	G48	3	3	2	1	3	1	1	1	1	1	1	1
66	KFPT13	P832	3	3	2	1	3		3	3	2	2	1	1
67	KFPC01	G41	3	3	2	1	2	2	2	2	7	7	7	7
68	KFPT09	G46	3	3	2	1	3	2	1	1	1		8	8
69	NZDAPT07	G35	4	3	2	1	3	2	2	2	2	3	3	3
70	NZDAPC01	G34	4	3	2	1	3	1	1	2	1	2	2	2
71	NZDAPT11	P831	4	3	2	1	3		1	2	2	2	2	2
72	NZDAPT10	G39	4	3	2	1	3	2	1	1	1	1	1	2
73	NZDAPT09	G38	4	3	2	1	3	4	3	3	3	4	3	3
74	NZDAPT08	G38	4	3	2	1	3	4	4	4	4	4	4	4
75	NZDAPT06	G33	4	3	2	1	3	3	3	3	3	3	3	2
76	NZDAPT05	G32	4	3	2	1	3	2	2	2	2	2	2	3
77	NZDAPT04	G32	4	3	2	1	3	3	2	3	2	3	3	3
78	NZDAPT03	G31	4	3	2	1	3	4	5	5	5	5	5	3
79	NZDAPT01	G30	4	3	2	1	3	2	2	2	2	2	2	2
80	NZDAPT02	G30	4	3	2	1	3	2	2	2	2	2	2	2
81	TFAF05	B1159	5	4	1	2	1	1	3	4	4	4	6	
82	TFAF02	B1158	5	4	1	2	1	1	1	2	2	2	6	
83	TFAF04	B1159	5	4	1	2	1	5	6	7	6	7	7	
84	TFAF06	B1159	5	4	1	2	1	1	1	6	7	7	7	
85	TFAF07	B1160	5	4	1	2	1	4	3	3	6	6	7	
86	TFAF08	B1161	5	4	1	2	1	4	3	3	6	6	7	
87	TFAF09	B1161	5	4	1	2	1	5	5	5	6	6	7	
88	TFAF10	B1162	5	4	1	2	1	5	6	7	7	7	7	
89	TFAF11	B1163	5	4	1	2	1	4	4	5	6	6	7	
90	TFAF12	B1164	5	4	1	2	1	5	6	6	6	6	8	
91	TFAF01	B1157	5	4	1	2	1	5	6	6	7	7	7	
92	TFAF03	B1158	5	4	1	2	1	1	1	1	2	2	6	
93	TFIAF06	B1150	6	4	1	2	1	5	6	6	6	7	8	
94	TFIAF15	B1155	6	4	1	2	1	4	4	4	4	5	8	
95	TFIAF14	B1154	6	4	1	2	1	5	5	5	5	5	6	
96	TFIAF13	B1154	6	4	1	2	1	5	6	6	6	6	6	
97	TFIAF12	B1153	6	4	1	2	1	5	6	6	6	7	7	
98	TFIAF11	B1153	6	4	1	2	1	5	7	8	8	8	8	
99	TFIAF01	B1150	6	4	1	2	1	3	2	2	7	7	7	
100	TFIAF10	B1152	6	4	1	2	1	5	6	6	6	7	7	
101	TFIAF09	B1151	6	4	1	2	1	2	6	6	7	7	8	
102	TFIAF07	B1150	6	4	1	2	1	6	6	6	6	6	7	
103	TFIAF05	B1150	6	4	1	2	1	4	4	4	4	4	5	
104	TFIAF04	B1150	6	4	1	2	1	4	3	5	6	7	7	
105	TFIAF03	B1150	6	4	1	2	1	3	3	2	2	3	3	
106	TFIAF02	B1150	6	4	1	2	1	3	3	3	4	4	4	
107	TFIAF17	B1156	6	4	1	2	1	6	7	7	7	7	8	
108	TFIAF18	B1156	6	4	1	2	1	3	5	5	8	8	8	
109	TFIAF08	B1150	6	4	1	2	1	3	4	5	5	6	7	
110	TFIAF19	B1156	6	4	1	2	1	3	5	3	3	4	5	
111	TFIAF20	B1156	6	4	1	2	1	4	4	5	6	6	7	
112	TFIAF16	B1155	6	4	1	2	1	3	6	6	6	7	7	

Data Set 12: Hurunui Mainland Island 5-minute bird count data

Contact: Andrew Grant, Christchurch
Date format verified with him: 8-Sep-99

There are two areas, treatment and non-treatment, and a two operator/data collection sequence. Operator 1 in non-treatment stations 1 - 10 in the morning, then stations 10 - 1 in the afternoon. Operator 2 is the same in treatment areas. Next day the operators switch areas and repeat. There are 8 data sets per collection phase over two days. There are two data phases per season, one pre treatment and one post treatment.

						Station	
Codes	Date	Season	Area	Poison	Time	Sequence	Operato
1	3/11/95	95/6	Treated	Pre	AM	1-10	JK
2	5/11/95	96/7	Untreated	Post	PM	10-1	JY
3	16/2/96	97/8					AVD
4	17/2/96	98/9					AS
5	2/11/96						
6	3/11/96						
7	20/2/97						
8	21/2/97						
9	31/10/97						
10	1/11/97						
11	22/2/98						
12	24/2/98						
13	3/11/99						
14	4/11/99						
15	21/11/99						
16	22/11/99						

Field		Summer					Station																														
Sheet	Date	Season	Area	Poison	Station	Time	Sequence	Operator	Bellbird	Blkbird	rwncreep	angoose	Chaffinch	Dunnock	Fantail	Grywarb	Kaka	Kea	.Tcuckoo	Pariduck	Redpoll	Rifleman	Robin	hcuckoo	Silvereye	Starling	Thrush	Tomtit	Tui	Mohua	akeetspp	Magpie	WPlover	Falcon	SBBGull	SIPO	BF Tern
1	1	1	1	1	1	2		2	1	1	2	0	6	0	0	1	0	0	0	0	0	4	1	0	0	0	3	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	2	2		2	3	1	0	0	3	0	0	0	0	0	0	0	0	3	0	0	0	2	1	0	0	2	3	0	0	0	0	0	0
1	1	1	1	1	3	2		2	3	2	2	0	5	0	0	0	0	0	0	0	2	4	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
1	1	1	1	1	4	2		2	1	1	1	0	3	0	0	1	0	0	0	0	0	3	1	0	0	0	0	0	0	1	4	0	0	0	0	0	0
1	1	1	1	1	5	2		2	2	0	0	0	3	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	2	3	0	0	0	0	0	0
1	1	1	1	1	6	2		2	0	0	0	0	3	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	7	2		2	2	0	0	0	4	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	8	2		2	2	1	0	0	3	0	0	1	0	0	0	0	0	4	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0
1	1	1	1	1	9	2		2	2	0	0	0	1	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
1	1	1	1	1	10	2		2	2	1	0	0	3	0	0	1	0	0	0	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
2	1	1	1	1	1	2		2	3	0	0	0	6	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
2	1	1	1	1	2	2		2	1	0	9	0	3	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0
2	1	1	1	1	3	2		2	1	0	2	0	4	0	0	1	0	0	0	0	1	4	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0
2	1	1	1	1	4	2		2	2	2	0	0	4	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	4	2	0	0	0	0	0	0
2	1	1	1	1	5	2		2	2	0	0	0	3	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	3	0	0	0	0	0	0	0
2	1	1	1	1	6	2		2	3	0	0	0	3	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	1	1	7	2		2	2	1	0	0	2	0	0	1	0	0	0	0	0	5	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
2	1	1	1	1	8	2		2	1	1	0	0	1	0	0	1	0	0	0	0	0	4	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
2	1	1	1	1	9	2		2	1	2	0	0	1	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	2	0	0	0	0	0	0
2	1	1	1	1	10	2		2	1	0	0	0	3	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	1	1	1	1	1		2	2	3	0	0	6	0	0	0	0	3	0	0	1	2	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0
3	1	1	1	1	2	1		2	1	0	0	0	5	0	0	1	0	0	0	0	2	4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
3	1	1	1	1	3	1		2	3	0	0	1	3	0	0	1	0	1	0	0	0	3	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0
3	1	1	1	1	4	1		2	3	0	4	0	4	0	0	1	0	0	0	0	0	2	1	0	0	0	0	0	0	3	0	0	0	0	0	0	0
3	1	1	1	1	5	1		2	1	0	0	0	4	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	1	1	1	6	1		2	1	0	0	0	2	0	0	0	0	0	0	0	0	6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	1	1	1	7	1		2	1	0	0	0	3	0	0	1	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	1	1	1	8	1		2	2	1	0	0	5	0	0	1	0	0	0	0	0	4	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0
3	1	1	1	1	9	1		2	0	0	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
3	1	1	1	1	10	1		2	3	0	0	0	4	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
4	1	1	1	1	1	1		2	2	1	0	1	7	0	0	1	0	0	0	0	0	3	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0
4	1	1	1	1	2	1		2	2	1	0	0	5	0	0	1	0	0	0	0	0	4	0	0	0	1	0	0	0	0	5	0	0	0	0	0	0
4	1	1	1	1	3	1		2	1	1	5	0	3	0	0	1	0	0	0	0	0	3	1	0	0	0	0	0	0	2	4	0	0	0	0	0	0
4	1	1	1	1	4	1		2	1	0	0	0	3	0	0	0	0	0	0	0	0	4	2	0	0	0	0	0	0	0	3	0	0	0	0	0	0
4	1	1	1	1	5	1		2	2	0	0	0	5	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	1	1	6	1		2	1	0	0	0	3	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	1	1	7	1		2	2	0	0	0	4	0	0	1	0	0	0	0	1	5	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
4	1	1	1	1	8	1		2	2	0	0	0	2	0	0	2	0	0	0	0	0	5	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0
4	1	1	1	1	9	1		2	1	2	0	0	2	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
4	1	1	1	1	10	1		2	3	0	0	0	5	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	1	2	1	1	2		1	0	2	0	0	3	0	0	0	0	0	0	0	1	1	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0
5	1	1	2	1	2	2		1	1	1	0	0	2	0	0	0	0	0	0	0	1	3	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
5	1	1	2	1	3	2		1	1	0	0	0	3	0	0	0	0	0	0	0	0	1	1	0	0	2	1	0	0	0	0	0	0	0	0	0	0
5	1	1	2	1	4	2		1	0	1	0	1	3	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	1	2	1	5	2		1	1	1	0	0	1	0	0	0	0	0	0	1	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
5	1	1	2	1	6	2		1	1	1	0	0	1	0	0	0	0	0	0	0	1	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5	1	1	2	1	7	2		1	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
5	1	1	2	1	8	2		1	1	1	0	0	2	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
5	1	1	2	1	9	2		1	2	1	0	0	3	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	1	2	1	10	2		1	2	1	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	1	2	1	1	2		1	1	1	0	1	1	0	0	0	0	0	0	2	0	1	0	0	0	2	1	0	0	0	1	0	0	0	0	0	0
6	1	1	2	1	2	2		1	0	1	0	0	4	0	0	0	0	0	0	0	1	4	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0
6	1	1	2	1	3	2		1	0	0	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
6	1	1	2	1	4	2		1	1	0	0	0	2	0	0	1	0	0	0	0	1	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
6	1	1	2	1	5	2		1	1	0	2	0	2	0	0	2	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6	1	1	2	1	6	2		1	2	0	0	1	3	0	0	0	0	0	0	1	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Data Set 13: Possum Trapping Data from Hurunui Mainland Island

Contact: Andrew Grant, Christchurch
Date format verified with him: 8-Sep-99

Pre-poisoning 5 sites were run (4 treatment and 1 non-treatment). Post-poisoning 5 different sites were run.
Trap sprung mean possum caught.

Code	Season	Date	Line	Treatment	Poison	Sex	Age	Sprung
0						Unknown		
1	95/6	27/2/96	A	Pre	Treated	Male	Adult	Yes
2	96/7	28/2/96	B	Post	Untreated	Female	Juvenile	No
3	97/8	29/2/96	C					
4	98/9	1/3/96	D					
5		25/4/96	E					
6		26/4/96	G					
7		27/4/96	H					
8		28/4/96	I					
9		7/3/97	J					
10		8/3/97	K					
11		9/3/97						
12		10/3/97						
13		29/4/97						
14		30/4/97						
15		1/5/97						
16		2/5/97						
17		9/3/98						
18		10/3/98						
19		11/3/98						
20		12/3/98						
21		28/4/98						
22		29/4/98						
23		30/4/98						
24		31/4/98						
25		5/3/99						
26		6/3/99						
27		7/3/99						
28		8/3/99						
29		26/5/99						
30		27/5/99						
31		28/5/99						
32		29/5/99						

									Kg	Trap
Case	Trap	Season	Date	Line	Treatme	Poison	Sex	Age	Weight	Sprung
1	1	1	1	1	1	1				2
2	2	1	1	1	1	1				2
3	3	1	1	1	1	1				2
4	4	1	1	1	1	1				2
5	5	1	1	1	1	1				2
6	6	1	1	1	1	1				2
7	7	1	1	1	1	1				2
8	8	1	1	1	1	1				2
9	9	1	1	1	1	1				2
10	10	1	1	1	1	1				2
11	11	1	1	1	1	1				2
12	12	1	1	1	1	1				2
13	13	1	1	1	1	1				2
14	14	1	1	1	1	1	2	1	2.85	1
15	15	1	1	1	1	1				2
16	16	1	1	1	1	1				2
17	17	1	1	1	1	1				2
18	18	1	1	1	1	1				2
19	19	1	1	1	1	1				2
20	20	1	1	1	1	1				2
21	21	1	1	1	1	1				2
22	22	1	1	1	1	1				2
23	23	1	1	1	1	1				2
24	24	1	1	1	1	1				2
25	25	1	1	1	1	1				2
26	1	1	2	1	1	1				2
27	2	1	2	1	1	1				2
28	3	1	2	1	1	1				2
29	4	1	2	1	1	1				2
30	5	1	2	1	1	1				2
31	6	1	2	1	1	1				2
32	7	1	2	1	1	1				2
33	8	1	2	1	1	1				2
34	9	1	2	1	1	1				2
35	10	1	2	1	1	1				2
36	11	1	2	1	1	1				2
37	12	1	2	1	1	1				2
38	13	1	2	1	1	1	2	1	3.1	1
39	14	1	2	1	1	1				2
40	15	1	2	1	1	1				2

Data Set 14: Waikoropupu Springs Vegetation Transects 1991-99

Contact: Rhys Barrier, Nelson
Date format verified with him: 10-Sep-99

There are three randomly placed transect, with the percentage of non-vegetated area (bedrock, cobble, gravel, sand, silt/debris recorded from 1991 to 1999. Monitoring is to see that divers in the water are not causing the amount of vegetation to change. There was weeding in 1993 to remove watercress, which may have caused an increase in the non-vegetated area.

Code	Transect
1	A-A'
2	B-B'
3	B'-A

		Start	Percentage of Quadrat Not Covered in Vegetation								
Case	Transect	Dist (m)	1991	1992	1993	1994	1995	1996	1997	1998	1999
1	1	0	87	55	68	75	45	80	70	32	20
2	1	5	15	10	5	0	10	10	20	0	5
3	1	10	90	60	90	95	90	100	100	75	40
4	1	15	70	60	65	97	80	80	100	40	45
5	1	20	30	60	35	60	65	40	65	60	29
6	1	25	20	25	10	50	5	10	35	0	0
7	1	30	8	10	30	10	20	0	0	0	0
8	1	35	0	0	0	0	35	0	20	0	10
9	1	40	2	10	0	10	15	0	0	0	2
10	1	45	0	0	0	0	0	0	0	0	0
11	2	0	45	70	40	75	40	75	55	30	10
12	2	5	30	35	10	40	30	40	20	40	10
13	2	10	60	65	70	90	80	55	55	80	20
14	2	15	46	20	60	30	0	35	60	50	30
15	2	20	0	10	0	5	0	65	0	0	0
16	2	25	52	75	50	95	90	90	80	50	55
17	2	30	99	83	84	80	80	80	90	100	
18	2	35	100	100	100	100	100	100	100	100	100
19	2	40	100	100	100	100	100	100	100	100	100
20	3	0	50	50	32	65	80	100	60	55	0
21	3	5	45	40	25	11	10	10	15	10	5
22	3	10	60	19	5	10	15	15	10	5	5
23	3	15	40	25	10	10	20	20	20	0	10
24	3	20	15	10	5	10	10	10	10	0	10
25	3	25	0	5	5	0	5	0	0	0	1
26	3	30	0	10	5	20	0	10	0	0	15

Data Set 15: Kaimanawa Recreational Hunting Area Permanent Plot Biodiversity

Contact: Sean Husher, Turangi
Date format verified with him: 12 Oct 99

A representative sample of 20x20 m forest monitoring plots was established in January 1979 and remeasured in 1987 and 1998.
It might be better to separate these into densities of palatable and non-papatable species.
One row of suspicious data removed for plot 33N42.

The data here are the number of species in five size classes as follows:
1 = 16-45 cm seedlings
2 = 46-75 cm seedlings
3 = 76-105 cm seedlings
4 =105-135 cm seedlings
5 = >135cm saplings
Blanks indicate no data collection

	Numbers of species in each size class														
	1979					1987					1998				
Plot	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
13N2	7	3	2	2							5	3	3	2	7
13N12	7	3	2	3							5	3	3	2	4
13N22											5	2	1	0	7
13N32	2	3	3	1							1	1	1	0	4
13N42	5	3	2	1							8	1	1	2	6
13N52	4	1	0	0							1	0	1	0	9
17N2	2	0	0	0							10	4	3	0	16
17N12	2	1	1	0							5	2	1	1	7
17N22	0	1	1	0							6	0	1	0	3
17N32	5	2	2	0							6	4	4	2	13
17N42	7	3	2	3							4	2	2	0	16
17N52	6	3	4	1							7	4	4	3	13
17N60	4	2	1	0							8	5	1	3	14
1N2	3	0	0	0		10	4	2	0	7	10	5	5	5	18
1N12	5	0	1	0		5	3	3	1	5	4	4	2	2	6
1N22	13	2	4	0		13	1	2	0	8	8	4	1	2	18
21N2	3	3	1	0		7	4	3	2	9	9	4	5	3	14
21N12	5	3	1	0		8	5	3	1	9	6	7	3	4	15
21N22	4	1	1	0		10	7	3	1	2	8	3	3	1	10
21N32	3	1	1	1		4	2	3	0	3	5	2	1	2	8
21N42	4	1	0	0		8	4	3	3	8	5	2	1	1	12
21N52	1	1	2	0		5	2	2	2	2	3	1	1	1	8
25N2	3	2	1	0		3	3	0	0	1	5	2	1	0	3
25N22	5	0	0	0		7	0	0	0	8	4	3	1	0	11
25N32	3	1	0	1		12	3	2	2	5	6	2	0	0	8
25N42	5	2	0	0		14	6	1	0	5	4	2	1	0	8
25N52	6	2	1	0		9	2	1	1	5	8	1	0	1	11
27N2	9	7	3	3		4	6	6	3	12	4	3	4	3	15
27N12	10	5	7	4		9	6	4	2	11	3	2	1	1	10
27N22	5	3	2	0		6	3	2	1	6	2	2	0	0	14
27N32	4	3	1	0		4	2	1	1	6	2	2	0	1	5
27N42	8	7	0	0		9	6	4	4	9	5	3	0	2	12
27N52	4	2	1	0		6	3	2	1	2	4	1	1	1	8
29N2	7	8	1	2		16	7	1	2	6	2	1	0	0	14
29N12	5	1	0	0		8	2	3	1	3	0	0	0	0	5
29N32	3	1	1	0		3	4	2	2	4	3	1	1	1	10
29N42	4	2	2	1		5	3	1	2	3	3	1	1	1	6
29N52	2	1	1	0		5	1	1	1	2	1	1	1	0	3
34N2	4	3	2	0		7	3	3	1	4	6	3	0	0	7
34N12	4	2	2	1		8	2	1	0	5	5	4	3	1	7
34N22	3	4	0	2		4	4	0	0	4	4	4	2	1	6
34N32	9	3	2	1		8	3	2	1	4	8	4	1	1	7
34N42	4	2	0	0		11	2	0	0	7	8	4	2	1	11
37N2	8	6	2	3		14	5	3	1	10	8	5	5	3	16
37N12	4	1	1	0		10	5	3	4	3	9	4	4	3	11
37N22	7	3	1	0		14	2	2	1	4	9	4	4	4	7
37N32	3	2	1	0		5	2	3	1	3	6	2	1	2	4
37N42						12	2	2	1	4	8	2	0	0	8
37N52	4	1	0	0		12	2	3	2	4	8	1	1	2	7
5N2	5	2	0	1		7	2	2	0	5	8	3	3	2	12
5N12	3	1	1	1		8	3	2	1	6	5	3	3	4	12
5N22	4	1	1	1		3	3	1	0	6	3	3	2	2	2
9N2	6	6	0	0							4	6	3	3	16
9N12	14	4	2	4							11	6	4	4	15
9N22	8	4	2	1							6	2	1	1	4
9N42	5	1	1	1							3	1	0	1	14

Data Set 16: Kaimanawa Recreational Hunting Area Permanent Plot Density of Nothofagus fuscus

Contact: Sean Husher, Turangi
Date format verified with him: 12 Oct 99

A representative sample of 20x20 m forest monitoring plots was established in January 1979 and remeasured in 1987 and 1998.
This data set gives the density of Nothofagus fuscus (an indicator species chosen by the conservancy) in the understorey in five size classes, and basal area of the overstorey.

Size classes are:
1 = Notfus seedlings 16 - 45 cm
2 = Notfus seedlings 46-75 cm
3 = Notfus seedlings 76 - 105 cm
4 = Notfus seedlings 105 -135 cm
5 = Notfus seedlings >135cm

Density in in seedlings per hectare.
Basal areas are the cover (in square metres) per hectare at breast height, not corrected for slope.

	1979						1987						1998					
	Density in Size Class					Basal	Density in Size Class					Basal	Density in Size Class					Basal
Plot	1	2	3	4	5	Area	1	2	3	4	5	Area	1	2	3	4	5	Area
13N 2					0.00	7.3											0.00	0.4
13N12						16.6												17.3
13N22													0.28	0.06	0.06		0.05	
13N32						1.8												2.3
13N42						39.4							0.17					40.2
13N52						<0.1												
17N1						39.4												
17N2						72.1							1.16	0.33	0.17		0.06	56
17N3																		25.8
17N4																		41.7
17N5																		1.4
17N6																		35.1
17N7																		62.9
17N12						54.6							0.11				0.00	
17N22						32							0.11					
17N32	0.11					30.9							0.06	0.06			0.03	
17N42	0.5					0.6											0.03	
17N52	1.11					16.1							0.44	0.33	0.22	0.11	0.46	
17N60	0.17					73.3							0.88	0.28	0.22	0.17	0.34	
1N 2						7.7	0.17	0.06	0.11		0.00	7.7	0.44	0.28	0.11	0.22	0.15	6.4
1N12						41.8	0.17					17.3	0.44	0.06		0.33	0.06	14.3
1N22	0.17					23.5	0.11				0.01	30.8	0.72	0.11			0.02	31.6
21N 2	0.28	0.17				54.3	0.33	0.06				83.5	0.22					84.4
21N12	0.06					99.1	0.17					104.1						94.8
21N22						45.8	0.06					51.3	0.22					51.3
21N32						37.3	0.06					39.8	0.33	0.06			0.00	43.8
21N42						89.1	0.11				0.00	79.1	0.11				0.04	75
21N52						104.3						103.8						104.3
25N2						0.7						0.9						1.2
25N12						4.1	0.17				0.01	5.3	0.17					4.6
25N22						0.3						0.2					0.00	0.3
25N32						2.2						2.8						3
25N42						10.3						10.8						11.6
27N 2	2.49	1.49	0.33	0.39		29.3	1.38	0.44	0.28	0.39	0.53	34	0.28	0.28	0.06	0.11	0.38	38.7
27N12	1.82	0.22	0.17			50.3	0.66	0.06	0.06		0.08	52.5					0.04	55.6
27N22	0.06		0.06			145.7					0.00	158.7					0.01	185.6
27N32						49.8					0.00	53.7						56
27N42	0.77	0.17				133.4	0.88	0.28	0.06	0.11	0.02	140.9	0.06	0.06			0.02	145.1
29N 2						<0.1					0.01	0.2					0.00	0.5
29N12						33.8						38.1						40.9
29N32						54.4					0.01	57.9	0.17				0.00	9.1
29N52						1						1.1						
34N 2	0.06					2						2.8						
34N12	0.11					64.6	0.06					65.3	0.22	0.06				61.9
34N22		0.44				51.8	0.06					60.7	0.06					59
34N32	0.39		0.06			12.1	0.06					12.1	0.17					23
34N42						4.9	0.22				0.01	6.4	1.11	0.06	0.06		0.01	7.9
37N 2						34.9	0.06					5.1	0.06				0.01	6.8
37N12						23.9	0.28					24.6	0.33					26.4
37N22	0.11						0.94						0.33					
37N32	0.06					74.3	0.22					77	0.88					74
37N42						22.5	0.11				0.00	24					0.00	26.3
37N52							0.22						0.17					
5N 2	0.06					19.8						19.6	0.22				0.01	21.6
5N12						86.7	0.06					87.9	0.06				0.01	88.3
5N22	0.11					0.6						0.7	0.11					
9N1																		25.1
9N 2						21.8											0.05	11.6
9N12	0.17					56.1							0.33	0.11			0.03	
9N22	0.06												0.11					

Appendix 4: Where the Answers to the Questions in the Modules can be Found

Module 1

Question 1:	Read section 1.3.
Question 2:	Read section 1.4.
Question 3:	Read section 1.4.
Question 4:	Read section 1.5.
Question 5:	Read section 1.6.
Question 6:	Read section 1.7 and look at the example on 1080 poison pellets.
Question 7:	Think about changing the algorithm to produce confidence intervals instead of t-statistics.
Question 8:	Read section 1.9.
Question 9:	Read section 1.10.
Question 10:	Read section 1.11.
Question 11:	A good question, difficult to answer.
Question 12:	Read section 1.13. The relevance is a matter of opinion.

Module 2

Question 1:	Read section 2.1.
Question 2:	Read section 2.5 and rework equation 8.
Question 3:	Read section 2.14.
Question 4:	Read section 2.8 and use equation 15.
Question 5:	Read section 2.8 and use equation 16 and Neyman allocation.
Question 6:	Read section 2.7 and 2.10.

Module 3

Question 1:	Read section 3.1 to 3.4.
Question 2:	Read section 3.8 and 3.9.
Question 3:	Read section 3.10 to 3.12.

Module 4

Question 1:	Read sections 4.2 and 4.3.
Question 2:	Read sections 4.2 and 4.3.
Question 3:	Read section 4.4.
Question 4:	Read section 4.5.
Question 5:	Read section 4.5.
Question 6:	Read section 4.6.
Question 7:	Read section 4.7.
Question 8:	Read section 4.8.

Module 5

Question 1:	Read section 5.2.
Question 2:	Read section 5.2.
Question 3:	Look at the regression model example in section 5.3.
Question 4:	Read the part about this test in section 5.3.
Question 5:	Read sections 5.4 and 5.5.
Question 6:	Read section 5.5.
Question 7:	Read section 5.6.

Module 6

Question 1:	Read section 6.2.	
Question 2:	Read section 6.1.	
Question 3:	Read section 6.6.	This is an example of BACI with one control and one impact site and with repeat before and repeat after surveys.
Question 4:	Read section 6.6.	

Glossary

Accuracy	A measure of how close an estimated statistic (e.g. the sample mean) is to the equivalent population parameter (e.g. the population mean).
Anova	Analysis of variance: a method to analysis data to assess the amount of variation associated with different known factors or variables.
Autocorrelation	Data show autocorrelation when observed value of a variable is correlated to adjacent values, usually in time or in space. E.g. a tree's height measured over time is autocorrelated data.
Average	Also called the mean. This is a sample statistic that describes .the centre of a group of measurements.
Bonferroni	Bonferroni adjustments are made to correct for the inflated type 1 errors that result from conducting several simultaneous tests of significance. C. E. Bonferroni was an Italian working in the area of probability theory.
Chi-squared distribution	A distribution often used in analysing count data, and for comparing sample variances with population variances. The distribution has only positive values.
Coefficient of variation	A measure to describe the relative variation in data by dividing the standard deviation by the mean. Also sometimes the standard error of an estimate divided by the estimate.
Composite sampling	A useful sample design where the cost of collecting large samples in the field is relatively low, but the cost of analysing all of the samples is high.
Confidence level	The probability that the statistical method produces a confidence interval that includes the parameter of interest.
Continuous	Continuous variables are usually measurements, e.g. weights, temperature.
Control	Control groups are necessary in experiments to find out what might have happened in the absence of the treatment.
Correlation	The correlation coefficient measures the strength of the linear relationship between two quantitative variables.
Degrees of freedom	Often this is just the sample size minus one. However, the expression degrees of freedom is also used more generally for the number of independent comparisons that can be made, e.g. in an analysis of variance table.
Dependent	A dependent variable is one for which the values are related to the values for certain other variables, which are sometimes called the independent variables.
Discrete	A discrete variable has gaps between each of the values that it can take, e.g. numbers of birds, age (in years).
Effect	The deviation in the value of the variable of interest from what would be expected in the absence of the treatment, e.g. the reduction in possum index after a control operation is the effect of possum control.
Error	The level of uncertainty in the sample statistic or the inaccuracy of a sample statistic due to a less than ideal sample design.

Experiment	A planned, rigorous approach to a study of e.g. a biological system. It often involves the control or allowance for the effects of some variables whilst manipulating others.
Factor	Usually a qualitative variable, e.g. gender, blood type.
F-distribution	Used in comparing variances, e.g. for comparison of the variability of the sample means with the variability within the samples.
Fixed effect	An effect where the interest is in the specific levels of the factor that are observed, as compared with a random effect.
Hypothesis	The conjecture that the research is designed to investigate.
Independent	Independent variables are those where the value of one does not depend on the value of the other.
Interaction	In experimental design an interaction of factor A and B means that the observed effect of A depends on the level of B, and vice-versa.
Mean	The average of the α observations, or of the population.
Median	The middle number when the observations are placed in rank order.
Meta-analysis	The statistical synthesis of results of separate, independent studies.
Mode	The most frequently occurring observation(s).
Model	A descriptive device for simplifying and characterising processes, e.g. biological systems, often expressed by an equation showing how the variable observed depends on other variables and/or factors.
Monitoring	A process to review the changes in a process, generally over time.
Multiple comparison	A method to investigate specific differences when there are more than two factors, or treatments, being investigated.
Multivariate	A type of statistical analysis where the interest is in more than one dependent variable.
Nested	A nested design is used when replicate units are completely contained within one treatment.
Non-parametric	Methods that do not require making assumptions about the specific parametric form of the distribution that the data are sampled from.
Normal distribution	A commonly used distribution that is symmetrical, and has a smooth, single peaked, bell-shaped density curve.
Observation	A sample is a collection of observations or unit from a population. The actual measurement from a unit is also referred to as an observation.
Parameter	A numerical characteristic of the population, e.g. the average, the variance.
Parametric	Methods that require making an assumption about the specific mathematical form of the distribution the data are sampled from.
Poisson distribution	A discrete distribution often assumed for count data.
Population	The complete set of units that are of interest, e.g. the population of trout in a lake at a point in time.

Power	The probability of detecting the existence of an effect from a test of significance, given that the effect does exist.
Precision	The accuracy with which a variable is measured or a parameter is estimated.
Probability	The likelihood that an event will occur. The simplest interpretation is in terms of a long run frequency; e.g. the probability of obtaining a head from tossing a fair coin is 0.5 because a head is expected to occur 50% of the time from many tosses.
Pseudo-replication	Assuming that observations are a random sample from a population when in fact they are only randomly selected from part of the population, or they are not selected independently.
P-value	The probability of observing a test statistic value as extreme as that observed when a null hypothesis is true.
Quartile	The quartiles of a sample or population divide the sample or population into four parts. The first quartile exceeds 25% of all values; the second quartile (also called the median) exceeds 50% of all values, etc.
Random	A process is described to be random if what is observed from the process is affected by chance to some extent.
Random effects	In analysis of variance, a factor is considered to have random effects if the levels of the factor observed are randomly selected from a population of possible levels. The alternative is that effects are fixed.
Random sampling	Sampling in such a way that each unit in a population has the same chance of being selected whenever a draw is made.
Range	The range of a sample or population is the largest value minus the smallest value.
Repeated measures	When observations are taken on a sample unit at several different times then the study is said to have repeated measurements.
Replication	Repeating experimental conditions to obtain several observations differing only by unexplained random errors is called replication. Values randomly sampled from the same population are also sometimes called replicates.
Residual	What is left over after all known effects are accounted for. This is typically used to describe the part of an observation that is estimated to be due to unknown random causes.
Sample	A selection of some of the units from a population.
Sample size	The number of units (observations) in a sample.
Serial correlation	The ordinary correlation coefficient calculated by pairing up the values in a time series X_t with values k time steps away. For example the first serial correlation is the correlation between X_t and X_{t+1}.
Significance level	The probability that a test of significance will give a significant result when the null hypothesis is true.
Standard deviation	A measure of the amount of variation in the values in a sample or population.

Standard error	The standard deviation of an estimator of a parameter. Statistical inference The process of drawing conclusions on the basis of samples from populations.
Stratification	The division of a population into non-overlapping sets of sample units, usually with the idea that the units within strata will be relatively similar.
Stratified sampling	Sampling separately from each of the strata in a population.
Sums of squares	Quantities used in regression analysis, analysis of variance, etc., which measure the variation related to specific factors. E.g. the regression sum of squares measures the amount of the variation in the data that is accounted for by the regression equation. They are sums of squared differences between observations and mean values.
Survey	The process of sampling a population.
Systematic	A systematic sample is one where every kth observation in time or space is selected.
t-distribution	A distribution that occurs frequently in statistical analyses, e.g. in the comparison of the means of samples from two normally distributed populations with the same variance.
Temporal versus spatial	Temporal is to do with time and spatial is to do with space.
Transformation	The process of using a mathematical equation to change the scale of measurement. E.g. the log transformation Y=log(X) is often used and an analysis done on the Y values rather than the corresponding X values.
Treatment	A type of experimental manipulation on the units used in a study.
Trends	Changes (usually with time or space) that are generally in the same direction.
Type 1 error	The probability of getting a significant result by mistake for a test of significance.
Type 2 error	The probability of getting a non-significant result by mistake for a test of significance.
Univariate	An analysis on one variable is called univariate.
Variables	The different quantities measured in a study, e.g. the height and weight of the subjects in a nutrition survey.
Variance	A measure of the amount of variation in data. Also the square of the standard deviation.