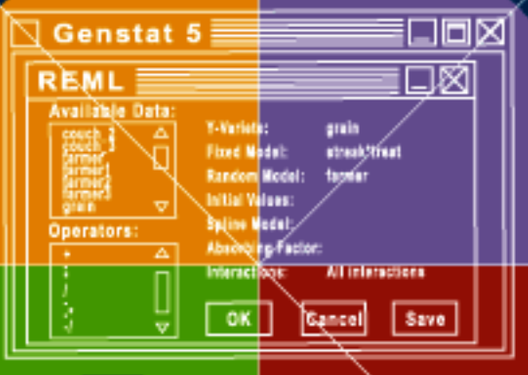
# Data analysis of agroforestry experiments




1

## Workshop overview & session summaries

World Agroforestry Centre
TRANSFORMING LIVES AND LANDSCAPES

# Data analysis of agroforestry experiments

1

## Workshop overview & session summaries


*by*

*Richard Coe, ICRAF, Kenya*

*Roger Stern, SSC, UK*

*Eleanor Allan, SSC, UK*

*edited by*

*Jan Beniest, ICRAF, Kenya*

*Janet Awimbo, ICRAF, Kenya*



The World Agroforestry Centre (ICRAF) is the international leader in Agroforestry - the science and practice of integrating 'working trees' on smallholder farms and in rural landscapes.

Agroforestry is an effective and innovative means to reduce poverty, create food security, and improve the environment. The Centre and its many partners provide improved, high quality tree seeds and seedlings, and the knowledge needed to use them effectively. We combine excellence in scientific research and development to address poverty, hunger and environmental needs through collaborative programs and partnerships that transform lives and landscapes, both locally and globally.

Founded in 1983, the Statistical Services Centre (SSC) is a not-for-profit body within the School of Applied Statistics at The University of Reading, UK. The SSC provides training and consultancy in both statistics and data management in the international arena. We aim to encourage good statistical practice, and the use of modern statistical methods in applied problems.

The SSC currently has nine statisticians, plus computing professionals and administrative staff.

ICRAF
The World Agroforestry Centre
United Nations Avenue
PO Box 30677
Nairobi, Kenya
Tel: + 254 2 524 000
Fax: + 254 2 524 001
Contact via the USA
Tel: + 1 650 833 6645
Fax: + 1 650 833 6646
E-mail: icraf@cgiar.org
Internet: www.worldagroforestrycentre.org

Statistical Services Centre
The University of Reading
Harry Pitt Building Whiteknights Road
P.O.Box 240
Reading RG6 6FN, UK
Tel: +44 (0) 118 378 8025
Fax: +44 (0) 118 975 3169
E-mail: statistics@rdg.ac.uk
Internet: www.rdg.ac.uk/ssc/


ISBN 92 9059 145 5

Design: Mariska Koornneef
Printed by: Kul Graphics Ltd, Nairobi, Kenya

# The Training Materials

These Training Materials were developed to help us present a series of courses on the analysis of data from agroforestry experiments. They are published here to assist others give similar training in the future.

The course is very practical and built around the analysis of real data sets. Concepts are explained largely without using mathematics. The computer software takes care of calculations and hence formulae are not used. Instead the course emphasises understanding of the analyses it is sensible to use, and the interpretation of results. We distinguish between learning to use the statistical software (buttons to press or commands to use) and understanding the statistical concepts, models and methods.

The course was designed initially to help with analysis of agroforestry experiments, and the examples given are from agroforestry trials. However both the statistical and teaching ideas can be applied to trials from agriculture, forestry and other application areas. Only one out of 17 sessions is dedicated to peculiarities of agroforestry research, and it should be easy to substitute other examples when using the materials. The materials refer to both on-station and on-farm trials. Emphasising the distinction between on-station and on-farm experiments is not necessary or helpful for this course. The approaches and methods for the analysis of a trial depend on its objectives, treatments, layout and measurements, not on where it was carried out.

The materials are presented in four printed parts together with a computer CD.

**Part 1** contains an overview of the course and teaching approaches, with suggestions on how the materials may be used and adapted. It also contains a summary of each of 17 teaching sessions.

**Part 2** contains the lecture notes, one for each of the sessions. They form a useful and readable resource in their own right and hence are presented as a separate document.

**Part 3** contains suggested exercises for each session. These are presented as a separate document as they are most likely to be adapted and modified to use local examples.

**Part 4** contains a protocol describing each of 16 experiments, the data from which are used in examples.

The CD contains

- a data file (in Microsoft Excel format) for each of the 16 example experiments
- files (in pdf format) for each of the 4 parts, so that further copies can be printed
- the original word processor files of all the text (in Microsoft Word format), so users may modify and adapt the text
- some additional documents (in pdf format) that are referred to in the materials

We encourage the copying and modification of these materials as long as the original source is acknowledged, and resulting products are not sold without our permission. We would appreciate being informed of any use and developments of these materials.

The materials were produced through a long term collaboration between the World Agroforestry Centre (ICRAF) in Nairobi, Kenya and the Statistical Services Centre of the University of Reading, UK.

# Table of contents

# The course structure and strategy

## Introduction

These course notes are on the analysis of data from experiments. They result from a series of statistics training courses organized by ICRAF/World Agroforestry Centre. These courses were originally on the design and analysis of agroforestry experiments, but they have been used more widely than this.

The first component was on the design of experiments. This analysis course assumes familiarity with the main concepts from the design course. A brief review is given in Session 1.

The second component was on data entry and management. This is a key area because poor data management often limits the processing of data. In this analysis course the examples provided have been 'managed' so that the concepts related to the analyses could be illustrated easily. We anticipate that an initial phase in the course preparation will be to organize datasets from participants similarly. Hence, the data management component though normally undertaken prior to this component, is not a necessary prerequisite.

This course is divided into two parts. The first part is entitled **The Everyday Toolkit** and covers the concepts that we believe scientists should be able to understand fully and the corresponding analyses that they should be able to undertake unaided following the training.

The second part is called **Handling Complexities**. This examines how experimental data can be processed where there are complications. These complications are divided into three broad types. The first is due to complexities in the design which may either be due to a complex treatment structure, to difficulties in the layout of the trial, or to the way data were measured. For example, a measurement of farmers' responses may be on a 5-point scale ranging from very good to very poor. The analysis of this type of 'categorical data' is described here.

We do not consider on-farm trials as a special category and hence examples of them will be used throughout the course. However, their analysis is often complicated because of their combination of a complex layout (many farmers, with few plots per farm) and the nature of measurement. The complexities arise from the lack of control of factors that would be within the treatment structure in an on-station trial, and the fact that this lack of control occurs both within and between farms. The handling of these complexities is discussed in the course.

The second type of complexity is that which is due to the particular field of application. Particular features of agroforestry trials include 'repeated measures', both in time and space, and difficulties that arise from the need to measure multiple components (e.g. concerning both trees and crops) within each plot. Courses for other audiences need to replace this section, as each subject area has a set of problems and methods specific to it. As an example, we provide a parallel session that considers some of the complications that are commonly encountered in livestock experiments.

Finally we consider complexities that arise because of the nature of the data. Coping with zeros in the data and missing values are among the topics considered here.

Our main aim in this second part of the course is for scientists to be aware of the methods that now exist to handle complex data. These are methods where scientists, at least initially, might want to work jointly with statisticians.

## Audience

This course is intended primarily for scientists undertaking agricultural research.



In targeting the course for scientists we are assuming some prerequisites. We assume scientists have some practical experience in the design and analysis of trials. They may have felt diffident in their write-up of an experiment for which they have been responsible, but they have an awareness of the process of conducting and processing an experiment.

We assume basic computing skills. Scientists who do not have regular access to a computer and who are not comfortable with the use of a word processor and a spreadsheet should not take this course. Most scientists will already have some experience in data processing using a statistical package. Those without this experience should make themselves aware of the capabilities of this type of software before the course.

We assume some basic statistical knowledge, usually from statistics courses taken while participants were students. We do not assume that they liked the course, or that they understood all the content. We hope that at least something was understood and a little is still remembered! The design course is also a good preparation for this training and we hope that participants who have followed it still remember (and use!) some of the concepts that were covered in that course.

We mentioned above that this course is divided into two parts, 'The Everyday Toolkit,' and 'Handling Complexities'. There is no reason why both parts have to run sequentially. If the material is too demanding, then the training could be split into two parts, with the first course concentrating on the Everyday Toolkit only. We consider this point further in the section on the duration of the course, where we suggest different ways that the training could be given.

We assume that participants are in posts where they are able to use the materials that are covered here regularly during the following year. This workshop should be seen as the start of the learning process, rather than as an end in itself. Even after an MSc, most statisticians find that they learn more in the subsequent years by practicing what they have learned. Similarly here, participants who follow the course by avoiding data analysis for the next year are unlikely to profit much from the training.

As with the design course, we anticipate that the materials and the approach to training might also be of interest to the supporting statisticians. They may be in the agriculture faculty of a university or in a research institute. They will typically find that the topics are familiar, but the practical way they are introduced and the way they are covered and justified may be new.

## Resource persons

At least one of the resource persons should be a statistician and preferably there would be at least two statisticians involved for the full duration of the training.

The approach adopted in this course is sufficiently different to conventional textbooks or other courses on statistics that we suggest that a statistician should first attend this, or a similar course, as a participant, before being involved as a resource person.

Ideally the resource persons should not all be statisticians; there should also be one or two agriculture professionals. Sometimes someone in an institute has become a de-facto 'semi-statistician', in that they are willing to help colleagues on their statistical problems occasionally. Having initially attended this type of course, or the course on design, they can support the resource team, which consolidates their statistical skills, while providing excellent support to participants.

Resource people who are skilled at scientific data management and in data processing can help participants with their data preparation and software use. They would not normally be resource persons for the statistics sessions.

## Datasets and data management

In the design course, one requirement was that participants provide a protocol for an experiment with which they are involved. The same is required here, but additionally participants should bring the data. Ideally they should bring the data as computer files and also a photocopy of the data on the original data collection sheets. The protocols at least should be sent ahead of the training so sessions can make maximum use of the data from participants.

The materials presented here include datasets from ICRAF, ICRISAT and ILRI experiments, and part of the preparation for the course involves substituting the examples from participants wherever possible.

The penultimate session in this course describes part of the course that is devoted to the analysis of the participants' own data. Usually, in addition to the sessions described here, there is an introductory session for participants to check that their data are organized, so they can be analysed during the course.

In ICRAF courses, a key component of this introductory session was for participants, working in pairs, to exchange the sets of data and protocols that they have brought. Here the recipient is a participant who knows nothing of the trial other than the data and protocols given. Each participant then seeks explanation of any elements that are not clear. Queries are specifically related to the data, rather than whether the objectives are clearly specified.

Examples of queries are as follows:

- What are the meanings of particular columns of data?
- What are the units of measurement?
- Are the data in Excel?
- Can they be easily transferred to GenStat, or to another statistics package?
- Is there sufficient information of the experimental design to complete an analysis?

Support staff are on hand to help in the organization of the data.

This session was often held on the first afternoon of the workshop and resulted in the participant's data being available for analysis, from day 2. This helped both the individual participants and the resource persons.

Further details of this session are given in the notes on the data management component. The examples of datasets that we have provided are intended to demonstrate 'good-practice'. For consistency they are all provided as Excel files, though some were originally in Microsoft Access.

## Duration

The design course was typically of 2 weeks duration and we assume here that this may be the case for the course covering both data management and analysis components. It is anticipated that the data management component would be for about 3 days. In this case the basic and further sections of statistics would take about 4 days each.

This would be highly intensive and other scenarios are possible. Where participants are relatively inexperienced the course might run for 3 weeks; the first week could be devoted to data management, the second to the basic analyses, and the third to the further methods.

If yet a further week were possible, making 4 weeks in total, we suggest that the additional week be devoted to the analyses and write-up of the scientists' own datasets. Here the first two weeks might be as above (in the three-week course), with the final 2 weeks shared between the further methods and the scientists' data.

If scientists are relatively inexperienced, but a course of only 2 weeks is possible, we suggest that it is preferable to cover one part of the material well rather than to attempt to describe all the topics. One scenario is to devote the first week to the data management plus an introduction to the statistical software. The second week would be devoted to the basic statistical analysis. Within this second week there might be a single session (half-day) reviewing the contents of the further statistical methods part.

Where training is within a country (so travel costs are low) then we suggest that dividing a four-week training course into two sections of 2 weeks each would be of even more use. In this case, we suggest that the first training might be on the data management component and the second could be on analysis.

Even within the 4 weeks scenario, it is important for participants to be realistic in their objectives from this training. We aim for participants to be comfortable with basic statistical methods. They should be able to conduct their own analyses when their problems are simple. They should know when a problem is simple and when it is complicated. For complex problems they should have some ideas about the methods of analysis and be comfortable in interactions with a statistician where one is involved. This is an ambitious list, but it is not the same as trying to turn participants into statisticians.

If training is for longer than 4 weeks then it will usually be broader than the course envisaged here. It is likely to include design, data management and analysis and also to consider

the analysis of survey and monitoring data, as well as data from experiments. In this case, these notes could be used within the corresponding sessions.

Notes from individual sessions have also been used within statistics training courses for MSc and PhD students, who are starting their career in agricultural research.

If 4 weeks is the longest time that might be planned for this training, then what is the minimum? There is no minimum. The analysis course includes a one-hour review of the design course. This was a two-week course, so perhaps a one-hour seminar on analysis is a practical minimum for a formal presentation. One session at the end of the 'Everyday Toolkit' section is a 'Where are we now?' review, that is about this length. We also have a booklet on analysis that serves as a brief review of the main concepts in the course as a whole.

A one-hour seminar could be for a participant's own institute, or for a particular person. Within one institute a scientist gave his Director General a one-hour review of the importance of the data management and statistical issues for the effectiveness of the research of the organization. It is useful to keep your director informed, but not realistic to expect a director to spend one week on a training course.



## Teaching style

Lecturing is kept to a minimum in this course.

It is planned that a typical half-day session would start with a lecture/demonstration of between 20 and 45 minutes. The main part of the session would then be devoted to practical and discussion work. Participants would usually divide into groups for this work, with each group having a defined set of tasks. Finally there would be a review period with a representative from each group presenting their findings.

Normally two participants are nominated each day to prepare a brief summary of the work undertaken and particularly the key points that were covered in the day. This is used to review the key points that have been introduced and also to build a report of the work actually covered in the workshop.

In the basic part of the course participants will normally conduct the full analyses of the data themselves. Most of the practical work is generally specified and participants have to translate the tasks into instructions for the statistics package. Reference documents on the statistical software are available to help with this translation.

In the further analyses, the style of the practical work is different, because the limited time does not allow for participants to become embroiled in computational difficulties. Where this is likely, an alternative is for participants to dictate the analyses, which either one of them, or a resource person then undertakes.

The notes in this manual are extensive, but this does not mean that all sessions must be included in the training. On the contrary, the existence of the notes permits resource people to omit coverage of certain topics, while still providing the necessary reference materials for the participants. We discussed different general scenarios in the section on the duration of the training and here we consider the potential problem that some sessions may take longer than foreseen and hence other sessions may have to be omitted.

The obvious sessions to omit are those in the 'Handling Complexities' part of the course. One possibility is to include all in the basic part and then move straight to Session 16 on the 'own data analysis' and Session 17 on the tasks following the workshop. If there is time for one further session then Session 15, which is on complications in the data, can also be included in the teaching.

In one course, where participants were not very experienced with computers and were finding many ideas to be new, they were split into groups, depending on how fast they wished to proceed. The slowest group only managed up to Session 4, plus Session 9, so they were at least able to present some results from their research.

Unlike the design course, this training needs computers and there must be sufficient time and computers for participants to have considerable hands-on practice. One computer between 2 participants is a minimum and often better than one each, because it encourages more discussion during practical work.

In versions of the training where participants have considerable time to process their own data it becomes useful to have one machine each, though this is not always practical.

There should also be a projection device that allows the whole group to view a computer screen, so analyses can be discussed jointly.

Participants can use an overhead projector for presentations. They can be prepared using presentation software, such as PowerPoint, or they can use the statistics package, or other software directly. Ideally the computers would be networked so the presentations that use a computer can use the same computer on which the materials were prepared.

Some of the work is carried out in small groups. Groups can work in different parts of the same room. Ideally there would also be additional small rooms for these discussions where needed. Again an ideal would be for one computer to be available per group, within each of these rooms.

## Software

A powerful statistics package is required. The requirement is also for a package that minimizes the time that must be devoted to mastering its use. The aim is to teach the concepts and the practice of data analysis, not the skills of using a particular package.

We currently use GenStat as the most effective package for the analysis of experimental data. The current version (5th release for Windows at the time the documentation was written) is also much easier than previous versions for data exploration and descriptive statistics, but it still remains more difficult than Excel for some of these tasks.

The course notes have been written to be as generic as we felt possible without making them more difficult for the reader. We expect resource persons to adapt the notes in general and adaptation for a different statistics package should not be difficult. Ideally the chosen package should include facilities to handle multilevel models (REML in GenStat and PROC Mixed in SAS). However this is a minor part of the course and therefore other packages could be considered.

In addition to a statistics package, Microsoft Excel is used here for data management, plus Microsoft Access and the ICRAF produced Logbook software for some courses. A word processor (possibly Word) and software for presentations (possibly PowerPoint) are also needed.

## Course Content

### Summary

Here the content of the course is summarized. This summary should be explained to participants at the start of the course as part of the introduction to motivate their involvement. This motivation is important for participants who have attended training courses on data analysis before and are still unsure of their capabilities.

If this training is put into perspective with the other components we find that:

a) With the design course, many participants felt that initially it missed their main priority, which was analysis. By the end of the design workshop they realied the importance of the design phase and had mastered the main concepts.

b) With the data management, some scientists felt initially that this was an underlying problem, but not as pressing as analysis. Others felt it was not a great issue and was something mainly for the technicians. The training period was again sufficient to impart the key concepts.

So now we start the training on analysis, which most scientists feel is their real need.

The purpose of this course is to emphasize concepts. It minimizes the formulae, they are needed more by the statistical software than the user. You should also minimize the time devoted to the statistical computing. This point may seem curious, because extensive use will be made of statistical software for the analysis. However, the software has become so much simpler to use, that the emphasis can remain on the ideas of the analysis and not on how to use the statistical package.

In the design workshop we saw that the objectives of the study led to the treatments to be applied, to the layout of the trial and to the measurements that needed to be taken. These 4 components (objectives, treatments, layout and measurements) are also the key to this workshop. Our aim here is to analyse the measurements, taking account of the layout and treatments, so we satisfy the objectives.

We will see that the requirements of an effective analysis are clear and often simple, if we concentrate on the analysis that is required to satisfy the objectives. Too many courses in the past have concentrated on helping participants to understand the analysis that is supplied by the statistical package. This has to be understood, but should not be the major concern of the course. We should assess what is needed for the analysis and demand it from the software, not look at what is provided and hope that it is needed.

By Session 3 we will see that the types of presentation of the data that are needed to satisfy the objectives are usually quite simple and also easy to provide. In Session 4, which is devoted to exploratory analysis, we show that while simple summaries are important, it is also necessary to look for features of the data that might alter the way we tackle analysis or even modify objectives.

We will find that the objectives rarely imply the need for an analysis of variance table. So what is the role of an analysis of variance? We answer this question in Sessions 5, 6 and 7. Perhaps it is not surprising that many scientists do not make much use of the ANOVA table, since its role (in relation to the objectives of the trial) is unclear. Considering the measurements as:

data = pattern + residual

is a useful idea. The ANOVA table is a useful way of summarizing all the components of the 'pattern'. Part of this 'pattern' concerns the treatment effects and it is these treatments that relate to the objectives of the trial.

In Session 6 we introduce the idea of a 'standard error' and its role in relation to the objectives of a trial. Session 7 shows how the idea of writing

data = pattern + residual as a 'statistical model'

is helpful. It enables users to see that many different problems can be analysed in a similar way, because they have a similar 'model'.



The final main concept of Multiple Levels is introduced in Session 8. If 'everything' is at just one level, then the analysis is usually simple. To discuss what we mean by 'everything', we need to consider our three components of treatments, layout and measurements again. So, the analysis of an on-farm experiment is usually not so simple, because we take measurements at two levels (at least). Thus, in a typical on-farm trial we record some data at the farmer level - we ask them questions - and we take other measurements at the plot level.

The session on the presentation of the results concludes the first part of the workshop. We expect participants to try the ideas on their own data throughout the workshop and will be looking for suitable examples to include in the formal training. Session 17 is a direct challenge to the participants, to discuss whether the workshop will change anything once they leave.

The second part of the workshop, that is titled Handling Complexities, assumes the concepts mentioned above. It considers common complexities about which participants should be aware in order to do an effective analysis. It is not expected that participants will necessarily be able to handle all the complexities unaided. But they should be able to detect their existence. They should also know whether the particular complexity is one that they can handle unaided, or one that requires support.

We have divided the complexities into those that were planned, those that are common because of the particular field of application and those that were 'surprises'. In describing the complexities we again consider them in relation to the three components of our trial. Thus there may be complications in the way the treatments are applied, in the layout, or in the measurements.

When participants assess their own objectives from the workshop, it should be made clear that the sessions, particularly in the second section, could only introduce the subjects concerned. The aim of this workshop is to provide both skills and confidence in analysing data.

## Sessions

### Part 1: The Everyday Toolkit

1. Review of experimental design ideas
   A quick review of the main concepts and terminology of experimental design.
2. Objectives of analysis
   Identifying objectives for analysis, so that it proceeds in a focused and constructive way.
3. Software familiarization
   The remainder of the course relies on use of a statistical package, which is introduced at this point.
4. Descriptive analysis and data exploration
   Looking critically at data to reveal the important patterns, before any formal inference starts.
5. Analysis of variance
   Analysis of variance is introduced as a tool in exploratory statistics of data with complex structures, typical of experiments.
6. Statistical inference
   Making formal statistical inferences and explaining the role of ANOVA.
7. Models
   Introducing the idea of a statistical model, and the link between models and analysis of variance.
8. Multi-level data
   Data are harder to analyse when they are at multiple levels. The concepts are introduced with split-plot designs and within-plot samples.
9. Presenting the results
   Putting the information from analysis in a format suitable for various audiences.

### Part 2: Handling Complexities

10. Review
    We review the first part of the course to set the scene for the second part.
11. Complexities introduced by the design
    Handling common analysis complexities that arise from the design of the experiment.
12. Dealing with categorical data
    We use an on-farm trial in which most of the measurements were categorical to illustrate modern approaches to the analysis of this type of data.
13. Getting more out of on-farm trials and other multilevel problems
    Handling complexities common in data from on-farm trials. We concentrate here primarily on the complexities that arise because of the availability of information at multiple levels.
14. Complexities due to the area of application
    There are currently two parallel sessions and we expect more to be provided in the future.
    Agroforestry experiments typically involve measurements of multiple components over long time periods. Methods for analysing these are covered.
    Mixed models and the analysis of crossover designs are described. They are common in livestock trials.
15. Complexities due the nature of the data
    These are problems that could not be anticipated - missing values, many zeros, non-constant variance and so on.
16. Analysis of participants' data
    This is a session in which participants work on their data. This may well be spread out through the course.
17. Action plan
    Planning the action participants will take on return to their institute.

## Resource materials

For each main session we have provided a summary, a lecture note and a practical exercise. These are provided as printable (pdf) files that can be read on-line and can be printed when needed. These materials are also provided as Microsoft Word files for resource staff to adapt as needed.

The protocols for 18 trials are provided, and this includes all those that are used for illustration on the course. This document links to the Excel files that provide the datasets used on the course. The next table summarizes the trials.

| Title | Type | Used | Features |
|---|---|---|---|
| Relay planting of *Sesbania sesban* and maize. | Station | Le.: 2 | Factorial treatment structure. Unequal replication. Strong trend along field not accounted for by blocks. |
| Effect of *Tithonia diversifolia* and *Lantana camara* mulches on crop yields in farmers fields. | Type 2 | Le.: 1, 6, 8, 13, 15 Ex.: 1, 2, 3, 4, 6, 7, 8 | 3 treatments, 1 complete replicate per farm. |
| Screening of suitable species for three-year fallow. | Station | Le:. 7, 11,5 Ex.: 1, 2, 4, 5, 7 | Simple design and layout. Repeated in 2 years. Several responses and covariates. |
| Upperstorey/ understorey tree management trial. | Station | Le.: 4, 5 Ex.: 1, 2, 4, 5, 11 | Complex treatment structure. Tree and crop components measured. |
| *Leucaena trichandra* seed production trial. | Station | Le.: 8, 11, 15, Ex.: 8 | Large trial. Multi-level design and data. Spatial analysis possible. |
| Fruit trees survival. | Type 2 | Ex.: 11,12 | Treatment factors of different types. Repeated measures. Binary responses. |
| On-farm cropping with *sesbania* and *gliricidia*. | Type 2 | Ex.: 13 | Simple treatments. Not all farms have the same treatments. Multi-level covariates. |
| Roots and Competition (RAC). | Station | Le.: 14 Ex.: 14 | Large trial with simple layout but multiple observations per plot. |
| Prototype hedgerow intercropping systems. | Station | Ex.: 14 | Simple trial with responses measured in 12 successive seasons and a strong rainfall x treatment interaction. |
| Fertilizer, *Tithonia* and *Lantana* mulch as sources of phosphorus for maize. | Station | Le.: 11 | Complex treatment structure. Unequal replication, treatments not orthogonal to blocks. |
| *Calliandra* feeding trial. | Station | | Simple crossover trial. |
| Effects of organic and inorganic sources of nutrients on striga, weeds and maize. | Station | | Mix of qualitative and qualitative treatments. Split plot design. |
| The influence of improved fallows on soil phosphorus fractions. | Type 1 | Le.: 8, 11, Ex.: 1, 2, 4, 8 | Factorial treatment structure. Split-plot design. Covariates measured. |
| Improved fallows and rock phosphate: farmers' experiences. | Survey, Type 3 | Le.: 4, 12, Ex.: 12 | Part of a large study with irregular design and qualitative responses. |
| On-farm trial with improved fallow and inorganic fertilizer. | Type 2/ 3 | | Factorial treatment structure. Irregular arrangement of treatments to farms. |

There is also a series of about 20 'good-practice' guidelines. These are found on-line and printed booklets that are on a range of topics, some of which provide supporting materials for the course, or can be used as pre-course reading. Those on design and data management, relate to topics that are assumed in the present course. The guidelines that are particularly relevant to the analysis course materials are as follows:

**Confidence and Significance: Key Concepts of Inferential Statistics**

These are topics we would like to assume, but find they are often poorly understood. This guideline is an excellent candidate for pre-course reading.

**Modern Approaches to the Analysis of Experimental Data**

This provides a résumé of the topics covered in Sessions 1 to 8 of the course.

**Informative Presentation of Tables, Graphs and Statistics**

This provides an extended version of the material in Session 9, but more general than for experimental data.

**Modern methods of Analysis**

This introduces some of the key concepts from the second part of the course. It is more general than the course with suggested methods applying equally to the analysis of survey data.

**Mixed Models and Multi-Level Data-Structures in Agriculture**

This extends the materials introduced in Sessions 13 and 14.

## Strategy

In this final section we describe aspects of the workshop strategy highlighted when drafts of this training material were used previously. We hope that the lessons learned will help future resource staff in planning their workshops.

In 1999 the workshop was run in three different countries in Africa. The first was in Cotonou, Benin in September 1999. This was for staff from the national programme in Benin, INRAB. The second workshop took place at ICRAF headquarters in Nairobi and was a regional course to researchers of agroforestry. The third workshop was an in-country training in Conakry, Guinea for researchers from the national programme, IRAG. In Guinea the workshop was given in two phases, firstly to a key group from all stations, and then with members of the key group acting as resource persons, to scientists in individual stations.

In Benin a draft version of the notes was used. Comments on the notes were sent to ICRAF, so that changes could be made for the second workshop. The revisions following this second workshop resulted in the first version of the course notes being produced in early 2000.

This process of revisions is described because we expect further changes in the future. Indeed we would welcome any suggestions for changes, revisions of sessions and additional sessions. One major addition to the notes following the ICRAF course was a plan to include extra materials on a 'Type 3 on-farm experiment'. This is almost a survey and is a useful indication that the same strategy of analysis will also be of some use in the processing of survey type data.

One feature of the Benin workshop was that 1 or 2 participants were nominated each day to produce a report of the day's proceedings. This was presented and discussed as the first item on the succeeding day and was then written up, by the participant, to be included in the workshop proceedings. It was quickly established that these reports should summarize briefly what was covered but then concentrate on the discussion points that were raised, rather than the topics that were already in the notes. The full workshop only covered Sessions 1 to 9 formally and the discussions often raised concerns that are considered in the later sessions. Topics included the role of multiple-comparison procedures, as well as the importance of the coefficient of variation, coping with zeros in the data, and ways of incorporating climatic data into an analysis.

By week 2 these reports became a key element of the workshop. They were of a high standard and typically generated a good debate of whether the essential points of the previous day's proceedings had been understood. They gave considerable confidence to the participants and showed that some of them had sufficient understanding of the key concepts that they could now contribute to training courses themselves. This helped the discussions on future plans (Session 17) , which were also in the second week.

In the earlier design course, participants often divided into small discussion groups. They prepared short presentations that generated a discussion at the end of most sessions. This approach was equally effective here for the sessions that did not involve computer use. However it was clearly a distraction to the participants in the sessions that involved computers and we abandoned the attempt. This was partly because the computers encourage participants to work singly or in pairs, rather than in small groups. It was also because the shape of the tables, in a large rectangle, plus the lack of easy discussion areas close to the machines, did not lend itself to this approach.

Comments on some of the specific sessions are as follows:

In the Benin course the first two days were spent on Sessions 1 and 2, namely on a review of design and of the specification of the objectives for the studies. An important result from these sessions was a realization by the participants that a review of their procedures for evaluating protocols was necessary.

GenStat was re-introduced on the third day, as most participants were already familiar with an earlier version. The changes in the new version were popular to the extent that, by the end of the course, the participants suggested the setting up of a GenStat support group.

Sessions 4 and 5 were two sessions that benefited from the use of local examples instead of those used in the notes. We used a single example in each session, which was copied to each computer.

The ideas of simple inference were (as usual) found to be difficult by the participants. 'Do participants understand the difference between a standard deviation and a standard error' is a key question? Can they explain what is meant by a confidence interval for the mean? These ideas were introduced in Session 6. It was strongly suggested that these topics be prepared as a booklet for pre-course or pre-session reading. (This is now available.)

In week 2 participants spent a long time (1.5 days) on Session 17. They made detailed plans on the in-station training that was needed over the next year. These plans covered design and data management as well as data analysis. It was agreed that this training within the stations was of more immediate priority than further training of this key group.



## Acknowledgements

The European Union, World Bank and ICT have funded the courses that have motivated production of this material. Some of the development work at Reading was funded by DFID grants. Some development at ICRAF was funded by grants from the Netherlands Ministry of Foreign Affairs through their Direct Support to Educational Institutions Programme (DSO).

Many individuals have contributed to the materials in assisting with courses and reviewing documents, these include Jane Poole, Marie Rarieya, Jan Beniest, Ian Dale, Wim Buysse, Gaston Kokode and Sekou Beavogui.

# Session 1 Review of experimental design

## Session objectives

- To review the main concepts of experimental design - objectives, treatments, experimental layout and measurements - in the context of experimental data analysis.
- To remind workshop participants of the key terminology used in experimental design.

## Summary

The design of an experiment starts with a statement of **objectives**. These must be clear, complete, relevant to the problem being addressed and capable of being met by an experiment. The objectives determine the three other components of the experimental design:

1. The **treatments**. The treatments are the conditions being deliberately compared in the experiment. The objectives may call for a simple collection of treatments, a factorial set of treatments, a range of quantitative levels of some factor, or combinations of these. The objectives will also determine suitable control treatments.
2. The **layout** of the experiment. This includes the choice of site or sites, of plots within sites and of measurement locations (samples) within plots. It also includes the arrangement of plots in blocks and the allocation of treatments to plots.
3. The **measurements** to be taken. The measurements include those taken directly to meet the objectives of the experiment and those taken to help understand the variation or pattern in other measurements.

Both on-farm and on-station trials can be described in terms of objectives, treatments, layout and measurements.

## Activities

### Introductory lecture

A short lecture to review the main features of design and remind participants of the key terminology used in experimental design. Point out that this is a one-hour review of material that has previously been covered in a two-week course.

### Practical Exercises

Participants meet in small working groups and look at some experimental protocols; one provided by the workshop resource persons, and one, or several of their own. The purpose of this exercise is to see that all participants understand the basic concepts and key terminology of experimental design and that they can apply this to any particular trial.

Start with a simple protocol provided by the course. The following four are suitable:

1. 'Relay planting of *Sesbania sesban* and maize'
2. 'Effect of *Tithonia diversifolia* and *Lantana camara* mulches on crop yields in farmers fields'
3. 'Screening of suitable species for three-year fallow'
5. 'Upperstorey/understorey tree management trial'

The group should follow the instructions given on the practical sheet. The group reports back in a plenary session with a *brief* summary of the 4 elements of the design, and their comments on the completeness and clarity of them. Emphasize that it is not necessary to repeat the whole of the background to the trial. The group discussion needs to be no more than 10 minutes, plus a 3-minute report for each group. Plenary discussion will then be used to clarify any design points. This part should only take 30 minutes because the protocols provided by the course must be clear and simple.

Next the groups look at a participant's protocol again, following the practical sheet. The group reports back in a plenary to describe the problems they had understanding the experiment and *not* to present experimental details. This part will take longer than the first, since experiments tend to be poorly described for outsiders and thus may yield quite some discussion when it comes to understanding the design and its objectives. This part easily takes 1 hour to complete.

## Supporting documents

- R. Coe, L. Nelson and R. Stern. 1996. *Design of agroforestry experiments*. Guidelines for training workshop resource persons. ICRAF/World Agroforestry Centre.
- R. Coe, R. D. Stern, E. Allan. 2002. *Session 1 - Review of experimental design*. Lecture notes. (Part 2). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.
- R. Coe, R. D. Stern, E. Allan. 2002. *Experiments portfolio.* (Part 4). Data analysis of agroforestry experiments. ICRAF/ World Agroforestry Centre.
- R. Coe, R. D. Stern, E. Allan. 2002. *Session 1 – Review of experimental design*. Exercise guidelines (Part 3). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.

Session 2

# Objectives and steps in data analysis

## Session objectives

- To list and describe the sequence of steps involved in the analysis of an experiment.
- To determine objectives for analysis of data from an experiment based on the objectives of the experiment.
- To describe the tentative numbers, tables and graphs that will be needed to meet the analysis objectives.
- To describe the logic behind the training workshop structure and content, in the context of analysis objectives.

## Summary

The general steps in the analysis of an experiment are:

1. Determining the objectives of the analysis.

2. Preparing the data by making the data files suitable for the planned analysis.

3. Descriptive analysis:
   i. Calculation of summary tables and graphs as earlier defined when setting analysis objectives.
   ii. Exploratory analysis to identify any unexpected patterns or results.

4. Confirmatory or formal analysis:
   i. Ensuring the planned tables, graphs, etc. are appropriate.
   ii. Adding measures of precision.
   iii. Improving the estimates of critical quantities.

5. Interpretation of analysis results.

6. Reporting of experimental results.

Defining the objectives of analysis will involve:

- Identifying the exact comparisons to be made or relationships to be estimated.
- Determining the exact data that are needed to make them (e.g. do we need comparisons of yield for each season or totaled over all seasons?).
- Designing the tables and graphs that will be used to present the results.

The objectives of analysis are determined by the objectives of the trial. However the analysis objectives are distinct from trial objectives in that:

- The objectives of the trial may have been stated in a rather vague way.
- The objectives set out in the original protocol may have other, unstated, objectives added, if these can usefully be met with the data available.
- It may not be possible to meet all the original objectives of the trial, either because the trial design does not allow it or because something unexpected has happened to prevent it.

The analysis objectives will evolve as the analysis proceeds.



# Activities

## Introductory lecture

A short lecture in two parts to:

a) introduce the steps in experimental analysis, and
b) link the setting of analysis objectives to experiment objectives.

## Practical exercises

Participants meet in small working groups and look at some experimental protocols, one provided by the workshop resource persons and one, or several of their own. The purpose of these exercises is to make sure that all participants can describe analysis objectives based on an experimental protocol, and apply this to any particular trial. Groups and protocols should remain the same as for Session 1 so that participants are already familiar with these trial designs.

The first exercise uses a protocol provided by the course and should take about 1 hour to complete.

The second exercise is for participants to look at their own experimental protocols that will be used in the data analysis sessions and determine the analysis objectives for their own data.

Exactly how this second exercise will be completed depends on what participants have brought with them to the training workshop. If everyone has a suitable protocol and data set, participants can work in pairs, with each pair looking at the two protocols. A selection of participants are requested to present their findings in a plenary session through a brief presentation. Participants are encouraged to start a file for the analysis of their own data, which by the end of the course should contain a complete record of what they did with the data, starting with the analysis objectives. This second exercise should take approximately 1.5 hours to complete.

### Concluding remarks

In a final plenary session, workshop organizers explain how the remainder of the workshop is structured in terms of the steps in data analysis. The 'Everyday Toolkit' part of the training workshop (Sessions 1 – 9) closely follows the steps outlined; proceeding through descriptive statistics and formal statistics to the final reporting stage. The 'Handling Complexities' part (Sessions 11 – 16) merely extends the range of situations that can be handled by both the descriptive and formal analysis methods.

## Supporting documents

- R. Coe, R. D. Stern, E. Allan. 2002. *Objectives and steps in data analysis*. Lecture notes (Part 2). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.
- R. Coe, R. D. Stern, E Allan. 2002. *Objectives and steps in data analysis*. Exercise guidelines (Part 3). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.
- R. Coe, R. D. Stern, E. Allan. 2002. *Experiments portfolio*. (Part 4) Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.

Session 3

# Software familiarization

## Session objectives

- To become familiar with the main statistics software package (GenStat) that will be used during the training workshop.
- To list the advantages of this software package in the analysis of agroforestry experiments.
- To review the role and use of supporting software needed to facilitate data analysis using this main statistics software package.
- To transfer data and results between these software packages.



## Summary

The first part of this session is devoted to practical work, introducing 'GenStat for Windows' to the participants. This is the only introductory session on the software. The other sessions will deal with the practical use of GenStat for the different types of analyses.

The lecture describes the key points that make GenStat the obvious package to be used for this training workshop. This includes GenStat's powerful and flexible facilities for the analysis of simple experimental designs, facilities that are used in Sessions 4 to 8. The substance of GenStat is consistent with the structures discussed in the workshop. We have already mentioned the fact that an experiment is characterized by the treatments structure, the plot layout and the measurements that are taken. This translates into the treatments, blocks and data for analysis, within GenStat's ANOVA system.

Many datasets contain complications and this is the theme of the sessions in the second part of the workshop. It is therefore essential to use a powerful statistics package and GenStat's regression and REML facilities to demonstrate this power.

The workshop emphasizes the need for software that permits analyses that are dictated by the objectives of the research, without great effort by the user. GenStat is appropriate,

although its facilities for data manipulation and plotting are not as intuitive as those of Excel.

The third exercise of the session is for participants to look at their own data. This is to be imported into GenStat if possible and then used to consolidate the simple use of GenStat described earlier.

# Activities

## Introductory Lecture

A brief five-minute lecture is conducted, introducing GenStat to participants, and giving an outline of the session. The Introduction Section of the lecture note for Session 3 could be used for this.

## Practical exercises

### Exercise 1—Introduction to GenStat

The purpose of this exercise is for participants to follow the introductory tutorial on GenStat as described in the GWIM software manual.

## Review/demonstration

Answer any questions/problems participants may have encountered during the previous GenStat tutorial.

Resource person to give projected computer demonstration on importing data into GenStat and using the **'restrict'** command. Within the demonstration of data importing it would be useful to point out to participants that the easiest way to import from Excel to GenStat is to mark the data range initially in Excel. You highlight the required data, go to **Name Box** in the top left corner of the sheet and write in a name for the range (e.g. 'data'). Then in GenStat you just select the worksheet that contains this range (i.e. it will appear as a sheet named 'data').

Use the 'Screening of suitable species for three-year fallow' trial for the demonstration, restricting data to look at only one season (e.g. 1991). Demonstrate how data can be restricted using different menu commands, and by rows or columns.

## Practical exercises

### Exercise 2

In this exercise participants are able to have a go at importing data into GenStat and using the 'restrict' command, as demonstrated previously. This exercise may be omitted if time is short, as participants will be using these procedures throughout the rest of the course.

### Lecture/Discussion

A lecture and/or discussion on the structure and facilities of GenStat (using the lecture notes provided) will be conducted. This could include a projected computer demonstration of some of GenStat's features. The use of GenStat for data analysis and its main features that justify its use for analysis of data in agroforestry experiments are discussed.

### Practical exercises

**Exercise 3—Datasets for analysis**

Participants look at their individual datasets and discuss how these need to be prepared for further analysis in the context of the use of GenStat for this purpose. Their data should be appropriate for being imported into GenStat by the end of this session.

### Concluding remarks

The final part of this session provides an opportunity for workshop organizers and participants to discuss the use of the various software packages in the context of data analysis.

One final additional remark to make is that in previous workshops participants have requested more time on the specific use of GenStat. This should not usually be necessary, but could be accommodated by a second GenStat practical, given later in the course, which would cover the remaining sections of Part 1 of GWIM. However, the preferable option is for participants to use some of their spare time during the workshop to try all, or part of the ANOVA section of GWIM. The GWIM manual contains 'fast track' options, which could be tried by participants who are more proficient at computations .

## Supporting documents


- R. D. Stern. E. Allan, R. Coe. 2002. *Software familiarization.* Lecture note (Part 2). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.
- Statistical Services Centre, University of Reading, U.K. 1999. *GenStat for Windows Introductory Manual*.
- R. D. Stern, E. Allan, R. Coe. 2002. *Software familiarization.* Exercise guidelines (Part 3). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.
- R. Mead, R. N. Curnow, A. M. Hasted. Chapman and Hall. 1993. *Statistical methods in agriculture and experimental biology (2nd Edition).*

Session 4

# Descriptive analysis and data exploration

## Session objectives

- Demonstrate that tables and graphical summaries can reveal patterns and trends related to the objectives of the analysis and study.
- Show how data exploration can identify unexpected patterns, and/or observations.
- Encourage close inspection of the data and interpretation of exploratory methods, prior to formal statistical analysis.



## Summary

Data analysis is more than just formal methods such as analysis of variance. An important part, which is all too often ignored, is initial investigation of the data. This preliminary investigation allows the researcher to look closely at the information in his data. The necessary tools include tables and graphs of descriptive statistics and the use of exploratory methods such as **boxplots** and **scatterplots.**

With descriptive statistics the researcher can quickly explore the patterns of interest, such as the treatment effects. If the tables and graphs needed to address the objectives have already been identified it is easy to 'fill in the numbers' in the presentation, and observe the pattern of response.

Careful examination of the data also gives an insight into the variability of the treatment pattern, and highlights any **outliers** that need further clarification. Suspected patterns such as block effects or fertility effects can be explored, and sometimes, unexpected patterns will emerge which the researcher may want to explore further.

# Activities

## Introductory Lecture

A short lecture, showing the role of descriptive summaries and exploratory methods as useful tools for close preliminary investigation of data shall be given. The idea of data = pattern + residual is introduced and the methods used to examine both 'pattern' and 'residual' are discussed. The data from the 'Upperstorey/understorey tree management' trial are used to illustrate the different points.

## Practical Exercises

**Exercise 1** - Computer practical where participants work in pairs on one of the four example datasets used in previous sessions with a view to familiarizing themselves with the data structure, identifying patterns which are associated with the objectives of the experiment and other possible patterns which might be expected. Consideration should also be given to variability, and detection of 'strange' values.

This is followed by short presentations by four pairs of participants, one pair for each dataset. These presentations should cover:

(a) The questions related to the objectives of the study - completed tables or graphs, methods used and findings.

(b) Other interesting questions that one wanted to explore and what they showed. Also any unexpected findings.

(c) Some assessment of variability in the data.

Round off with a discussion to elicit what methods are useful for different types of investigations and data. There should also be some discussion as to which other questions should be addressed.

**Exercise 2 –** Descriptive and exploratory analysis of participants' own data. This is a continuation of the second practical of Session 2 where participants identified the objectives of analysis in relation to the objectives of the study.

## Concluding Remarks

Some wrap-up by resource person of techniques concerning 'problems encountered, where do we go from here?' is needed at the end of this session.

## Supporting documents

- R. D. Stern. E. Allan, R. Coe. 2002. *Descriptive analysis and data exploration.* Lecture note (Part 2). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.
- Statistical Services Centre, University of Reading, U.K. 1999. *GenStat for Windows Introductory Manual*.
- R. D. Stern, E. Allan, R. Coe. 2002. *Descriptive analysis and data exploration.* Exercise guidelines (Part 3). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.
- R. Coe, R. D. Stern, E. Allan. 2002. *Experiments portfolio*. (Part4). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.

Session 5

# Analysis of variance as a descriptive tool

## Session objectives

- To understand the role of the Analysis of Variance (ANOVA) table.
- To understand and use residuals.
- To investigate how to reduce unexplained variation in the data.

# Summary

The tables and graphs required to meet the objectives of an analysis rarely include an analysis of variance table. This session explains the role of the ANOVA table in the analysis of experimental data. It is introduced here, as part of descriptive statistics, as a technique to look at the whole model (blocks and treatments). It is used to identify the components that are important. If we consider that

$$\text{data} = \text{pattern} + \text{residual},$$

then the ANOVA table extracts all the pattern, and hence also provides an opportunity to look at the residuals, i.e. the part of the data that can not be explained by the model, or pattern, in the experiment. Furthermore the pattern may be broken down into parts so that we can understand the relative importance of different components of it.

## Activities

### Introductory Lecture

Lecture on the use of the ANOVA table, explained using a simple example. This lecture should be conducted as an interactive session.

### Practical

The first part of the practical repeats material covered in the lecture. It should not be necessary for participants to report findings from this.

The second part asks participants to use descriptive ANOVA on their own datasets. Reporting conclusions from this is valuable. It is likely that 'inference' results will be reported, giving an entry point for discussing the distinctive roles of exploratory and formal analysis and the extent to which their analysis objectives are met by the former.

## Supporting documents

- R. D. Stern, E. Allan, R. Coe. 2002. *The analysis of variance as a descriptive tool*. Lecture note (Part 2). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.
- R. D. Stern, E. Allan, R. Coe. 2002. *The analysis of variance as a descriptive tool*. Exercise guidelines (Part 3). Data analysis of agorforestry experiments. ICRAF/World Agroforestry Centre.

Session 6

# Ideas of simple inference

## Session objectives

- To review the main concepts of statistical inference.
- To show what statistical inference adds to the descriptive summary.
- To show how the use of statistical inference can assist in the realistic interpretation of experimental results.
- To introduce the assumptions necessary for inferences to be valid.

## Summary

There are a few key ideas that are needed to understand the basics of statistical inference. First is the idea that there are parameters, which describe important characteristics of populations. These have to be estimated from data. Estimates have properties, the most important being the precision of the estimate. The standard error is a measure of precision of an estimate. A confidence interval is a convenient way of interpreting an estimate and its standard error. Statistical tests are used to test hypotheses about parameters, and result in a significance level that indicates the amount of evidence against the hypothesis.

Standard output from running analysis of variance on data from an experiment allows us to make inferences by (a) providing information on precision, and (b) making it possible to test certain hypotheses about the 'pattern' part of the data. The hypothesis testing can be useful in helping to decide which components of the 'pattern' we wish to concentrate on. It usually has limited value in directly interpreting treatment effects and meeting analysis objectives.

Contrasts (comparisons of carefully selected treatment combinations) are introduced to help meet specific analysis objectives.

# Activities

## Lecture

The main concepts of statistical inference are reviewed in a short lecture of approximately 45 minutes. Only 5 to 10 minutes are given to the general ideas. The rest of the time is devoted to a demonstration and discussion of the concepts for an on-farm example. The lecture note describes this case study, so it can be used for reference afterwards.

## Practical exercise

**Exercise**

The practical has the same objectives as the lecture, namely to consolidate the ideas of statistical inference and to apply these ideas to some real examples.

## Presentation of results and discussion

The reports review the conclusions from each case study, with particular emphasis on the extra component that is added by the statistical inference, compared to the use of descriptive statistics.

# Supporting documents

- R. D. Stern, E. Allan, R. Coe. 2002. *Ideas of simple inference*. Lecture note. (Part 2). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.
- Statistical Services Centre, The University of Reading, UK. 2001. *Confidence & Significance. Key Concepts of Inferential Statistics*. Booklet.
- Practical sheet

Session 7

# An introduction to statistical modelling

## Session objectives

- To introduce the concept of a statistical model for describing data.
- To practice fitting models and interpreting the resulting output.
- To demonstrate the relationship between a modelling approach to data analysis and simple ANOVA methods for balanced data structures.

## Summary

The key points for this session are:

- Data can be explored and variation explained by building a model with design effects (representing the layout and treatments) and other potential sources of variation.

- The modelling approach is flexible and powerful, overcoming some limitations of simpler methods (ANOVA) and providing the basis of many more advanced methods used to handle various complexities.

- The steps in using a model for data analysis are: exploring data to find a suitable model type, fitting the model, checking the assumptions of the modelling, and interpreting the fitted model. The process is iterative.

Models with few effects are not difficult to fit with current software, but more complicated models require some careful model building. This session introduces a very broad and deep subject and all future sessions refer back to it.

## Activities

There is a lot in this session and many new ideas are introduced. It will therefore be long (maybe a whole day) and should be broken down into a number of sections. A reasonable sequence would be:

**Lecture/demo 1:** covers the simple regression fitting and interpretation and 'Steps in modelling' sections.

**Part 1** of the practical: this reinforces the material from the lecture.

**Lecture/demo 2:** Review the practical, then proceed to the sections on 'Models with factors ' and 'Analysing other designs'.

**Part 2** of the practical: this repeats the lecture material. Participants may well have their own examples to use here.

**Lecture/demo 3:** covers the remaining material.

**Part 3** of the practical: note that the methods needed to answer the first question have been covered. Those needed for the second part can be introduced through discussion once participants have worked out what they are trying to achieve. The ideas of 'comparison of regression lines' can be introduced in a general discussion at the end of the practical if participants are up to it!

It is worth getting participants to present results from each part of the practical.

## Supporting documents

- E. Allan, R. Coe, R. D. Stern. 2002. *An introduction to statistical modelling*. Lecture note. (Part 2). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.
- Statistical Services Centre, University of Reading, U.K. 1999. *GenStat for Windows Introductory Manual*.
- Practical instructions.

Session 8

# An introduction to multiple levels

## Session objectives

- To introduce the concept of hierarchical structure in data.
- To introduce the idea of variability at the different levels and variance components.
- To demonstrate these ideas and show how they affect the analysis through the use of a split-plot experiment.
- To recommend ways of dealing with different complexities associated with multiple levels.

## Summary

The layout of many of the experiments involves multiple levels or layers, for example sites, blocks, plots and trees. Treatments may be applied to one or more of these levels, and measurements made on one or more of the levels. Some of the problems of analysis of data from experiments arise because of the multiple layers in the design. However, there are situations in which the multiple layers need not lead to complex analysis. Two are considered here, others will be covered in Session 11.

In a split-plot experiment, treatments are applied at two levels but measurements are made at only one (the sub-plot). Provided the design is orthogonal the analysis is straightforward and the only complication is that the precision of treatment comparisons differs, depending on the particular comparison.

When treatments are applied at one level, but measurements are taken at a level lower than the treatment, then a split-plot type of analysis can also be performed. An alternative is to summarize the lower level data to the level at which treatments were applied and carry out an analysis at that level.

When the data can effectively be summarized to a single level the 'general linear model' approach to analysis is appropriate. Models that have more random (error) terms are needed for more complex situations.

# Activities

## Lectures

### Lecture 1

The first lecture of approximately 30 minutes covers the analysis of the split-plot experiment (via interactive demonstration) and through it, the ideas of hierarchical structure and variance components. This uses MCH example pages 133-136. The lecture is followed by the first practical.

### Lecture 2

The second lecture of 45 minutes deals with other scenarios involving measurements at different levels. and layout at different levels will all be discussed via interactive computer demonstrations.

## Practical exercises

### Exercise 1

Hands-on computer practical work so that participants can understand how to perform correct split-plot analyses using GenStat software of about 45 minutes.

### Exercise 2

Further practical work of one hour, dealing with multiple observations per experimental unit and layout issues.

# Supporting documents

- Statistical Services Centre, University of Reading, U.K. 1999. *GenStat for Windows Introductory Manual.*
- E. Allan, R. D. Stern, R. Coe. 2002. *An introduction to the ideas of multiple levels.* Lecture note. (Part 2). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.
- R. Mead, R. N. Curnow, A. M. Hasted. Chapman and Hall. 1993. *Statistical methods in agriculture and experimental biology (2nd Edition).*
- Practical guidelines.

# Session 9 Writing up and presenting results

## Session objectives

- To show how to present results of analyses in tables and graphs.
- To show how to describe the statistical methods used.
- To show how to include results of analyses in a text.
- To remind participants that the analysis, and write-up do not end with the presentation of statistical results.

# Summary

It will usually be necessary to convey the results of your analysis to other people. The traditional way to do this is through writing of a scientific paper or report. Some of the challenges in this are:

- Keeping the statistics in perspective. The statistical analysis is a means to meet the objective, not the reason for the work. However some detail of statistics is needed to validate your conclusions.
- Designing and drawing effective tables and graphs. These are included to 'tell a story' and therefore must be constructed so that the story is clear.
- Reporting on methods in sufficient detail to make it clear what you have done without taking up too much space.
- Describing the results in the text so as to add to, not repeat, what is in tables and graphs.
- Designing the whole report to make it suitable for the intended audience.

## Activities

### Lectures

**Lecture 1**

The first lecture of one hour, is given by a science writer or editor, covers key elements and common mistakes made in scientific writing.

**Lecture 2**

The second lecture of 30 minutes deals with the important points of including statistical results in reports, papers and presentations.

## Practical exercise

### Challenge

Challenge to participants which asks them to:

i. take one objective from one of the studies they have been analyzing, and
ii. prepare a report and presentation which meets that objective, but is limited to a single graph or table and a maximum of 100 words of text.

The aim of this challenge is to force participants to think about the most effective way of describing the results while being concise. The reports have to be complete, presenting all the information necessary to meet the objective but can only include necessary information. Participants are told that it is not necessary to include any of the background or methods used in the study. How hard the challenge is depends on the study and objectives chosen.

At this stage in the course it would be good if everyone were to complete the challenge and present their results, but that may not be possible due to time constraints. It will also be necessary to decide whether the challenge is issued using the participants' own data or examples provided during the course. Using participants' own data is preferable.

The output can be presented in two ways: a single printed page containing the text and table/graph or a single transparency. Each participant makes a presentation using the single transparency [and limited to 100 words???] and hands in the single page. The work could be assessed and small prizes awarded. It would certainly be useful to have some resource people other than statisticians (a presentation specialist and one or two senior scientists) in the session to comment on the output. Participants should be allowed 3 – 4 hours to complete this challenge.



## Supporting documents

- R. Coe, R. D. Stern. E. Allan. 2002. *Writing up and presenting results*. Lecture note. (Part 2). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre
- Statistical Services Centre. University of Reading, UK. 2000. *Informative presentation of tables, graphs and statistics*.

Session 10

# Where are we now?- Review of basic statistics

## Session objectives

- To review the concepts presented in the previous sessions.
- To develop a framework for the methods of analysis and to introduce the remaining session topics.

# Summary

The first part of the course has shown scientists that analysis of much of the data arising from many experiments can be handled effectively by them if:

- They have a clear idea of the objectives of the analysis.
- They have a well-prepared data set ready.
- They have access to and expertise in using a suitable statistics software.
- They are familiar with some key concepts and methods.

The second part of the course focuses on complexities – in layout, treatments or measurements – that may make it difficult or inappropriate to use the simple methods discussed so far. The aim of the course now changes, with the emphasis of helping scientists to recognize when they have a complex problem, whereby they understand the approaches, but not necessarily all the details that are appropriate to the analysis.

## Activities

### Introduction

A brief introduction of 5 – 10 minutes to explain the objectives of this review session in the context of what has been covered so far and what will be covered under the remaining sessions of the training workshop.

## Practical exercise

### Discussion groups

Divide into 7 groups each of 2 to 3 participants that will review the material of the following sessions of the training workshop:

- Session 2: Setting objectives
- Session 4: Descriptive statistics
- Session 5: Analysis of variance
- Session 6: Statistical inference
- Session 7: Modelling
- Session 8: Multiple levels
- Session 9: Presentation

### Presentations

One person in each group makes a five-minute presentation of the key points covered in each session using not more than three up to a maximum of five slides. A presentations chairperson is instructed to keep exact time for this session. Presentations should be made as interesting as possible.

For each of these presentations, a chief discussant, selected from another group, comments on the presentation using a single slide (one to two minutes) and leads a five minute discussion on the presentation. Chief discussants must have seen the slides of the presentation before it is presented and focus their presentation and the discussion on whether they agree with the key points and indicate which important ones have been omitted.

## Lecture/demonstration

At the end of the presentations and discussions, the outcomes of what has been covered until now are presented. Using some examples indicating how the methods introduced so far can be applied, this lecture further shows what remains to be covered in the remaining sessions of the training workshop if data are to be exploited fully. The time frame for presenting these outcomes should take about 30 minutes.

# Supporting documents

- R. Coe, R. D. Stern and E. Allan. 2002. *Where are we now? Review of basic statistics*. Lecture note. (Part 2). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.
- Practical sheet

Session 11

# Design and analysis complexity

## Session objectives

- To understand how features of the design can lead to complications in the analysis.
- To understand some methods for spotting these complications.
- To learn how to handle some complications with a statistical model.

# Summary

Design complexities - that is, complexities of analysis, which are a consequence of the design use, and could be predicted without looking at the data - are common in real experiments. They may be due to the layout, treatments or measurements.

It is necessary for the analyst to (a) understand that the feature of the design may lead to a problem in analysis, or at least require an analysis other than the simple ones used so far, (b) know how to spot if the problem exists, and (c) understand strategies for handling the problem.

In many examples we can find a model, which elaborates on the simple models used so far and which handles the complexity. In this session we introduce the examples of more advanced treatment structure models to handle complex sets of treatments, the 'generalized linear model' to handle data which is not measured on a continuous scale, and show how complex layout/ treatment combinations can be tackled with 'general linear models'.



## Activities

### Lecture

A lecture is given covering different complexities of measurements and layout. More complex models than the general linear model are introduced; in particular generalized linear modelling and REML models, both of which will be demonstrated by computer analysis.

There is a lot of material in the lecture note. Given 'as is' this would be more than a single lecture. If the material is interspersed with demonstrations and discussions of many of the points that are not covered in detail, then there is enough here for several sessions. Resource people will have to determine how much time to spend on the material according to the interests and skills of the participants.

### Practicals

There are two practicals, each aimed at addressing particular objectives.

**Practical 1** introduces some of the skills, which will be needed for complex analysis. All participants need to have some exposure to these.

This is a hands-on computer practical with participants working in pairs, taking approximately 1 hour.

**Practical 2** gives participants the opportunity to attempt a more complex analysis with the assistance of a resource person.

There are two objectives here:

i. to understand the complexity of the problem and how it can be tackled,

ii to be able to interpret output from complex analyses.

This work should therefore be done in groups, with each group including a resource person, so that participants can learn from discussion. Participants should be allowed 1.25 hours to complete this task.

## Supporting documents

- R. Coe, R. D. Stern, and E. Allan. 2002. *Design issues which add complexity to the analysis.* Lecture note. (Part2). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.
- Statistical Services Centre, University of Reading, U.K. 1999. *GenStat for Windows Introductory Manual.*

Session 12

# Dealing with categorical data

## Session objectives

- Explain what categorical data are and what the problems for analysis are.
- Explain the special case of binary data.
- Explain possibilities and shortcomings of inference from contingency tables.
- Give an introduction to logistic regression.
- Give options for analysis of other categorical data.

## Summary

The session looks at datasets in which the responses of interest fall into a small number of distinct categories (such as yes/no, dead/alive, or red/blue/green) rather than being measured on a continuous scale.

Summary and descriptive tables for such data were described in Session 4. In this session, models for describing variation in these data are developed and interpreted. The logistic regression model is appropriate for binary data. Other types of categorical response variables can sometimes be analysed by transforming them to one or more binary responses.

## Activities

Lecture/demonstration/discussion. The discussion is carried out following the lecture note.

Practical exercise.

Follow up discussion as needed.

## Supporting Documents

- E. Allan, R. Stern, R. Coe. 2002. *Dealing with categorical data.* Lecture note. (Part 2). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre.

Session 13

# Getting more out of on-farm trials and multilevel problems

## Session objectives

- To understand that the multiple levels of variation common in data from on-farm trials can lead to analysis complexities.
- To be able to diagnose the existence of these problems.
- To find a suitable analysis strategy, including the use of models that represent the multiple levels of variation.

## Summary

Data from simple experiments are analysed by comparing the amount of variation in the 'pattern' with the random variation, this being the plot-to-plot variance not due to treatments or other known factors. In on-farm trials and other complex designs there is more than one type of variation (e.g. within and between farms). The analysis is therefore likely to need to use an analysis that recognizes these.

The complexity requires the following steps.

1. Understanding that the problem might actually exist.
2. Spotting if it really is a problem. This might be by suitably chosen diagnostic tables, or perhaps from unexpected error messages that the analysis software produces.
3. Finding an analysis that either:
   - avoids the problem by 'moving the analysis to a single level', or
   - allows for the problem by using a model that includes several variation terms, fitted using REML.

# Activities

## Lecture

A lecture is given that shows that multiple levels of variation may give rise to complexities in analysis and the strategies that can be used to overcome it. The simple strategy of 'move everything to one level' can be explained as one that is effective in many cases, but not universally applicable. The alternative that builds a model is similar in scope to the strategies used for design complexities in the previous session.

## Practical exercise

**Exercise**

The exercise sets out some precise questions to be tackled. Participants should be encouraged to analyse the root of difficulties they have and try to understand what the analysis options are doing to overcome them.

Participants will need help with GenStat, but should clearly articulate what they are trying to do before being shown the commands to accomplish it.

# Supporting documents

- *R. Coe, E. Allan, R. D. Stern. 2002.* Getting more out of on-farm trials and multilevel problems *Lecture note. (Part 2). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre*

Session 14

# Complications in agroforestry trials

## Session objectives

- To recognize some common characteristics of agroforestry trials that may complicate analysis and find strategies to cope with them.
- To understand the nature of 'repeated measures' in an experiment and simple methods to analyse them.
- To discuss approaches for dealing with the multiple components from some agroforestry trial data.

## Summary

Agroforestry trials have some common characteristics that may complicate the analysis compared with more 'traditional' agronomy experiments. Three are examined in this session.

1. They tend to be long term, giving rise to long series of observations from the same set of plots.
2. They often involve measurement and comparison of different parts of a single plot (e.g. under and away from a line of trees).
3. They generate multiple products.

The first two are aspects of the same statistical problem - measurements repeated in space or time. If this data structure is not recognized then invalid conclusions may result. There are several approaches to handling this problem. The simplest and most generally applicable requires calculating meaningful summaries of the repeated observations for each plot, then analysing these using standard methods. This is an example of the strategy of 'moving everything to a single level'.

# Activities

## Lecture

There will be an introductory lecture covering common complexities in agroforestry trials then focusing on analysis of repeated measures. Decide whether to dwell on the mechanics of doing the calculations (not covered in the lecture notes). Some of the data manipulations may be confusing and distract participants from the important statistical ideas.

## Practical exercise

**Exercise**

A practical discussing an example of repeated measures in space or in time.

**Lecture**

A brief lecture on handling multiple components.

# Supporting documents

- R. Coe . 2002. *Complexity in agroforestry trials*. Lecture Note. (Part 2). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre

Session 15

# Complications in data

## Session objectives

- To understand some of the common problems that become apparent when starting to analyse data from experiments.
- To take remedial action as to continue with an effective analysis.

# Summary

In the lecture we list the types of problem, such as, missing values that commonly occur. We then look at the possible strategies that scientists can employ for dealing with the problems. They range from ignoring the problem to making a separate research study of the problem itself.

The practical examines ways of solving problems in particular cases, and these solutions are then discussed.

## Activities

### Lecture

An introductory lecture of 45 minutes illustrated by a demonstration, using the contents of the lecture note and also using appropriate examples selected by workshop resource persons.

### Practical exercise

**Exercise**

There could be 5 or 6 groups, investigating different complications that occur in their own datasets or in those used in the workshop. Each group prepares a presentation on the complications that they have considered, the methods used for the solution, the ease with which they can be applied, and their success in resolving the problem. This exercise should take about 1.5 hours.

It is important that presenters are clear on the objectives of their presentation. It will be tempting for them to discuss their data analysis instead. While some discussion is needed, the objective is to describe methods of resolving complications.

Presentations, followed by discussion for 1 hour, (about 10 minutes for each group) should provide instruction to the whole group on practical ways of dealing with complications.

### Review

This would probably be by one of the resource persons for 15 minutes and would summarize the key points of the session and particularly of the discussion.

## Supporting documents

- R. D. Stern, R. Coe, E. Allen. 2002. *Complications in the data*. Lecture note. (Part 2). Data analysis of agroforestry experiments. ICRAF/World Agroforestry Centre
- Practical sheet

Session 16

# Data analysis

## Session objectives

- To analyse and present the results for participants own data.
- To consolidate the concepts introduced in the workshop by analysing these data with support and guidance from other workshop participants and resource persons.

# Summary

The session is concerned solely with participants analysing their own data and presenting the results.

## Activities

This session is one that should not necessarily wait until the end of the workshop. Scientists will have described the data they have brought, in the introduction to the workshop. In some course programmes there will be time within the workshop devoted specifically to the analysis of these problems. Alternatively, or additionally, a day towards the end of the workshop may be devoted to this area.

Some participants may also have made arrangements to continue their own work in the resource centre, following the formal close of the workshop. Others may wish to continue the analysis in their own institute, but seek a formal review of their work with a well-defined time scale to encourage the completion of the analysis.

Despite the many possibilities, it is important that each workshop defines the programme for these activities carefully. Realistic objectives must be set for what can be accomplished by each participant during the time permitted. There will normally be presentations by participants at the end of this session. These must be kept brief, for example 5 minutes per presentation, with 5 minutes discussion. Otherwise the presentations can become as ineffective as many conference sessions! Written information that can be circulated prior to the presentations, should be encouraged.

## Supporting documents

- Lecture note.

Session 17

# On your return

## Session objectives

- To develop realistic personal action plans for future data analysis work, based on the outcomes of the training workshop.
- To develop a strategy for research support in data management and statistics at either the individual, group, institutional or national levels.

# Summary

This session consists of presentations and discussions on what participants can do differently following this workshop. Changes could be on a personal level, or be for the institute or the national programme. They can affect data that will now be analysed and presented, a changed computing or software or training strategy, etc.

Scientists need to have impact. They need to promote technologies from their research that help farmers. The reason for this workshop is that the lack of skills to analyse data fully, is often quoted as a barrier to the completion of research projects. So the ideal is for scientists to feel that they may now be able to have more impact than before. However, increasing the skills of a few individuals who attend a course is unlikely to lead to much change, hence the need for this session.

Problems can be raised particularly if this workshop has suggested routes to a solution. Difficulties, where discussion is needed to suggest solutions should be raised as early as possible, so there is time for the necessary discussion. It is however, all too easy to raise problems rather than solutions. Participants should therefore be encouraged to concentrate on what they **can** do, rather than on what they can **not** do.

## Activities

We would like to split the discussion into two main themes. The first follows the main theme in this workshop, the improvement of scientists' capabilities for data management and statistical analyses, so that they can complete their research activities more effectively on their return.

Most scientists have data that could usefully be analysed and written up. We would like to know of scientists' plans over the next 6 months and one year. We encourage scientists to be reasonable in their ambitions and hope that we can contact them after these periods to see how much progress they have made and whether there is any support we can give.

The analysis of data and the preparation of reports and other publications are necessary, but not sufficient. The research only becomes useful when someone reads the report or paper and takes some action to put the findings into practice. We would like the subsequent use of the results to be considered at the same time. We accept that some research is not intended to be applied directly by farmers; it is perhaps to assist our understanding of part of the process. However we do think that concern by scientists on how their results might be used is helpful in the writing of them, as often these are written in different ways for the different types of reader.

The second theme is that of the development of a strategy for research support, in particular for the areas of data management and analysis that have been discussed in this workshop. But it could also relate to support on design, and possibly to other related topics that participants feel could improve their effectiveness to conduct and complete their research.



However, to be able to make concrete proposals we would like to encourage scientists to limit the range of topics that are included in this discussion of strategy. For example, we accept there are problems of computer access, of training for higher degrees, of hierarchical structures in research institutes and so on. We cannot solve all these problems!

We do feel, however, that there are possibilities for improvements in the 'system' for statistical support, both in aspects that you can do yourselves, on your return and in ways that you can have access to support from others. It is these aspects we would like to explore briefly here. They relate to the ways computers are used, the software that is employed, who does the different tasks and so on. The starting point is your current strategy.

The following table might be used:

| | **Now** | | **Future** | |
|---|---|---|---|---|
| | Software | Strategy | Software | Strategy |
| Data entry | | | | |
| Data management | | | | |
| Data analysis | | | | |
| Presentation of graphs | | | | |
| Report writing | | | | |

How do you get support from others? This could be from colleagues, from statisticians, etc. They could be in your institute locally or centrally. They could be employed by your programme or by another organization.

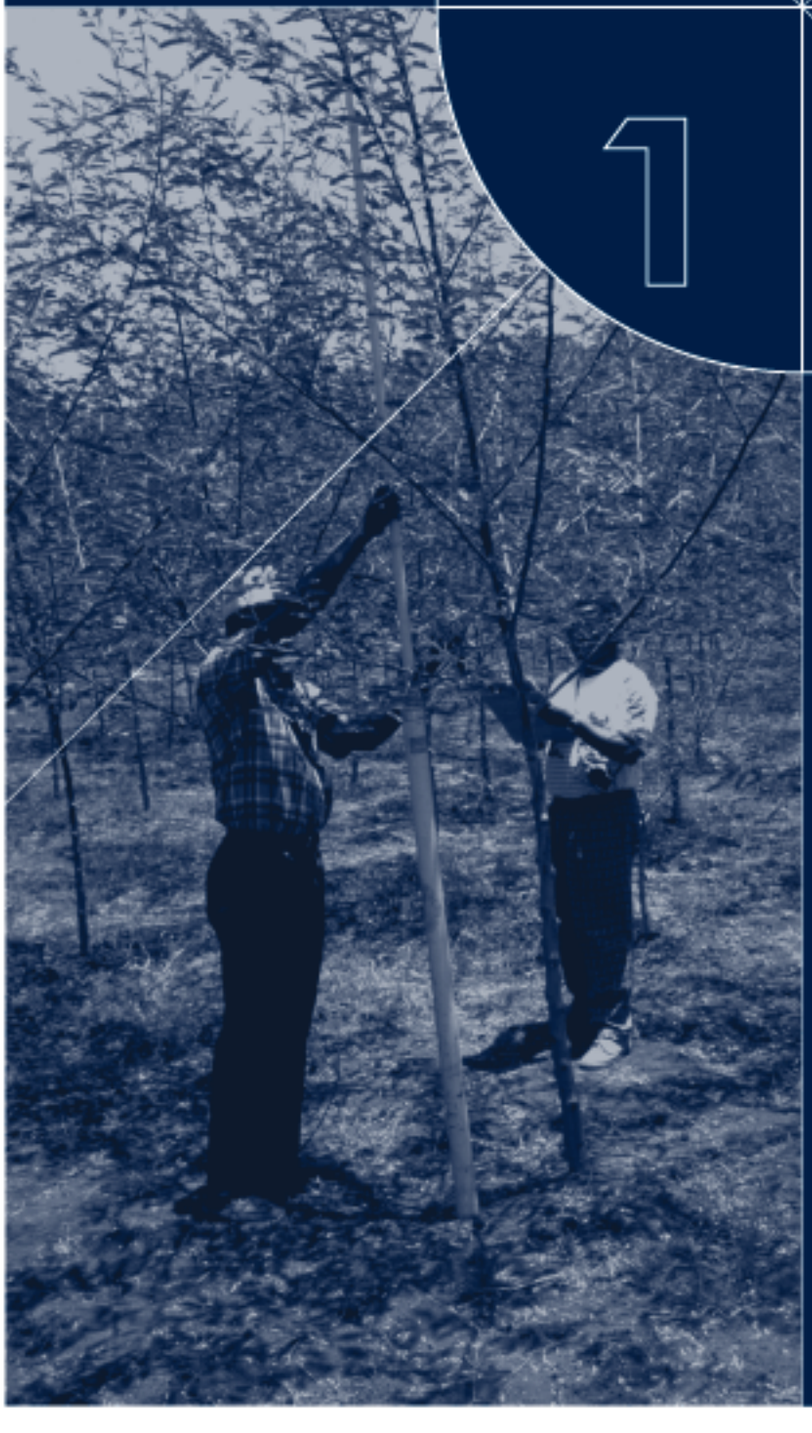
What is your current situation and what would you like that you think is feasible?

On this strategy area, we anticipate that there might be an initial 15 minute introduction and then a 45 minute discussion at a point that is perhaps halfway through the workshop. It should not be too early, because it would be helpful if participants were able to relate the question of support to the topics in the workshop. If there is a day for Session 16, 'Own analyses', it could be within this day.

This would set the scene for a full discussion of the topics in this session. This would be towards the end of the workshop, though not too late, giving enough time for action if clear proposals are identified. We anticipate that the project report would summarize the key points from this discussion.

ISBN 92 9059 145 5

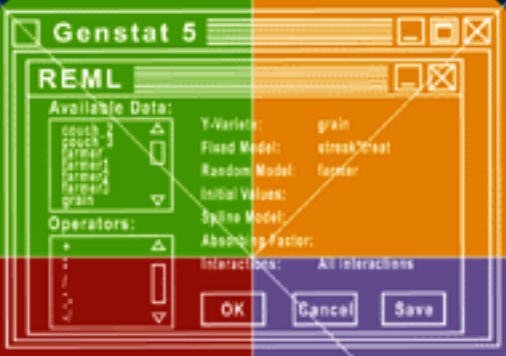
1



World Agroforestry Centre

TRANSFORMING LIVES AND LANDSCAPES

# Data analysis of agroforestry experiments




2

## Lecture notes

World Agroforestry Centre
TRANSFORMING LIVES AND LANDSCAPES

# Data analysis of agroforestry experiments

2

*Lecture notes*

*by*

*Richard Coe, ICRAF, Kenya*

*Roger Stern, SSC, UK*

*Eleanor Allan, SSC, UK*

*edited by*

*Jan Beniest, ICRAF, Kenya*

*Janet Awimbo, ICRAF, Kenya*



The World Agroforestry Centre (ICRAF) is the international leader in Agroforestry - the science and practice of integrating 'working trees' on smallholder farms and in rural landscapes.

Agroforestry is an effective and innovative means to reduce poverty, create food security, and improve the environment. The Centre and its many partners provide improved, high quality tree seeds and seedlings, and the knowledge needed to use them effectively. We combine excellence in scientific research and development to address poverty, hunger and environmental needs through collaborative programs and partnerships that transform lives and landscapes, both locally and globally.

Founded in 1983, the Statistical Services Centre (SSC) is a not-for-profit body within the School of Applied Statistics at The University of Reading, UK. The SSC provides training and consultancy in both statistics and data management in the international arena. We aim to encourage good statistical practice, and the use of modern statistical methods in applied problems.

The SSC currently has nine statisticians, plus computing professionals and administrative staff.

ICRAF
The World Agroforestry Centre
United Nations Avenue
PO Box 30677
Nairobi, Kenya
Tel: + 254 2 524 000
Fax: + 254 2 524 001
Contact via the USA
Tel: + 1 650 833 6645
Fax: + 1 650 833 6646
E-mail: icraf@cgiar.org
Internet: www.worldagroforestrycentre.org

Statistical Services Centre
The University of Reading
Harry Pitt Building Whiteknights Road
P.O.Box 240
Reading RG6 6FN, UK
Tel: +44 (0) 118 378 8025
Fax: +44 (0) 118 975 3169
E-mail: statistics@rdg.ac.uk
Internet: www.rdg.ac.uk/ssc/


ISBN 92 9059 145 5

Design: Mariska Koornneef
Printed by: Kul Graphics Ltd, Nairobi, Kenya

# The Training Materials

These Training Materials were developed to help us present a series of courses on the analysis of data from agroforestry experiments. They are published here to assist others give similar training in the future.

The course is very practical and built around the analysis of real data sets. Concepts are explained largely without using mathematics. The computer software takes care of calculations and hence formulae are not used. Instead the course emphasises understanding of the analyses it is sensible to use, and the interpretation of results. We distinguish between learning to use the statistical software (buttons to press or commands to use) and understanding the statistical concepts, models and methods.

The course was designed initially to help with analysis of agroforestry experiments, and the examples given are from agroforestry trials. However both the statistical and teaching ideas can be applied to trials from agriculture, forestry and other application areas. Only one out of 17 sessions is dedicated to peculiarities of agroforestry research, and it should be easy to substitute other examples when using the materials. The materials refer to both on-station and on-farm trials. Emphasising the distinction between on-station and on-farm experiments is not necessary or helpful for this course. The approaches and methods for the analysis of a trial depend on its objectives, treatments, layout and measurements, not on where it was carried out.

The materials are presented in four printed parts together with a computer CD.

**Part 1** contains an overview of the course and teaching approaches, with suggestions on how the materials may be used and adapted. It also contains a summary of each of 17 teaching sessions.

**Part 2** contains the lecture notes, one for each of the sessions. They form a useful and readable resource in their own right and hence are presented as a separate document.

**Part 3** contains suggested exercises for each session. These are presented as a separate document as they are most likely to be adapted and modified to use local examples.

**Part 4** contains a protocol describing each of 16 experiments, the data from which are used in examples.

The CD contains

- a data file (in Microsoft Excel format) for each of the 16 example experiments
- files (in pdf format) for each of the 4 parts, so that further copies can be printed
- the original word processor files of all the text (in Microsoft Word format), so users may modify and adapt the text
- some additional documents (in pdf format) that are referred to in the materials

We encourage the copying and modification of these materials as long as the original source is acknowledged, and resulting products are not sold without our permission. We would appreciate being informed of any use and developments of these materials.

The materials were produced through a long term collaboration between the World Agroforestry Centre (ICRAF) in Nairobi, Kenya and the Statistical Services Centre of the University of Reading, UK.

# Table of contents

Session 1

# Review of experimental design

*R. Coe, R. D. Stern, E. Allan*

## Introduction

It is necessary to understand both the design of a trial, and some of the ideas behind the design, before attempting to analyse the data. This session provides a brief review of these design ideas.

The key element in the design of a trial is the 'Statement of Trial Objectives'. We will use this session to show that the objectives of the trial determine the treatments to be applied in the experiment, as well as the measurements that are to be taken. Satisfying these objectives will also require an appropriate layout of the plots (or other units) in the experiment. In the remaining sessions we will often breakdown a trial into its component parts of treatments, layout and measurements.

Later, in Session 2, we will use these overall objectives of the trial to construct the objectives of the analysis. The idea that the analysis of the data must be able to satisfy the objectives of the trial is the dominant theme of this whole course.

## Objectives

Objectives drive the whole of the design of any experiment. For this reason the design process has to start by determining objectives.

The objectives for a trial must be:

- **Clear**. If the objectives are vague it will not be possible to decide on the rest of the design.
- **Complete**. Often the statement of objectives is incomplete so that when designing the experiment many questions can not be answered.

**Example**

An example of an incomplete objective statement would be something like:

"To evaluate the effectiveness of improved fallows for restoring soil fertility".

A trial cannot be designed until the objective is completed with such details as:

- Where is the evaluation needed? Over what range of environmental conditions?
- What range of 'improvements' need to be evaluated?
- What are the criteria for evaluation? Whose criteria are these? Over what time scale?

A more complete statement of objectives might be something like:

" Planted fallows, using *Sesbania sesban* seedlings, planted at 1m x 1m and grown for two years before clearing (removing wood and incorporating other biomass) have improved maize yields, compared to those obtained following a two-year natural fallow in on-station trials in E. Zambia. Now we wish to carry out a trial to determine:

- Grain yield increase (relative to natural fallows) in the first, second and third cropping season achieved using this technology on degraded (i.e. about to be put into fallow) clay and sandy soil maize fields in Eastern province of Zambia.
- Change in soil organic matter and N pools (compared to natural fallow) at the end of the improved fallow and after 1, 2 and 3 following maize crops."

- **Relevant**. In the area of development research, experiments are carried out to help solve development problems. The objectives of the experiment must be relevant to solving the problem. It must be clear how we will be a step nearer solving the problem once we have the results from the experiment.
- **Capable of being met by an experiment**. Not every research question needs an experiment.

## Treatments

The objectives will determine the treatments or conditions to be compared. The key concepts that will determine the choice of treatments are:

### Contrast or comparison between treatments

Many objectives simply require the comparison of the mean results of two treatments. In certain cases, more complex comparisons are needed. In general, each hypothesis corresponds to a contrast. The hypothesis or question determines the contrast and the contrast determines the treatments that will be incorporated in the experiment and not the reverse, as often happens.

For example, a trial is conducted to determine whether farmers find the addition of mulch from hedgerow species a useful addition to the management of fertility in vegetable plots, and if so, which of two common hedgerow species (*Lantana* or *Tithonia*) is superior. The first objective requires a contrast 'mulch vs no mulch'. The second requires a contrast '*Lantana* vs *Tithonia*'. Together these define three treatments which are;

1. *Lantana* mulch,
2. *Tithonia* mulch,
3. no mulch.

### Controls

Very often, treatments require some sort of standard against which they can be compared. Choice of these control treatments is determined by the experimental objectives. Control does not necessarily mean zero input or do nothing or use farmer practices. Any of these might be useful for certain objectives, but others may be equally or more appropriate.

The *Tithonia* and *Lantana* mulch example refers to 'mulch as addition to management of fertility', meaning the appropriate control, should be whatever the farmer usually does.

### Factorial treatment structure

Factorial treatments refer to sets of treatments defined by the combination of two or more treatment factors, each with two or more levels.

For example, in a fodderbank trial, two cutting heights combined with three plant densities gives a total of six treatments.

Such sets of treatments often arise naturally from the proposed hypotheses but they can also be used to test several unrelated hypotheses in the same experiment more efficiently than in separate ones. Understanding factorial structure and the idea of interaction can further suggest other hypotheses that should be tested.

Complications may arise if:

- not all factorial combinations are distinct or can be included (they may not be possible),
- control treatments have to be included in addition to a factorial set,
- the total number of treatments will be too large if all possible combinations are included in the experiment.

### Quantitative factors

When treatment factors are quantitative (e.g. amount of manure, tree density), the aspects to be decided are the range of levels, the number of different levels, the spacing of the levels and the replication at each level. There are many options in the choice of levels when two or more quantitative factors are involved. The choice of levels is not usually defined by objectives that may refer more to the identification of rates, slopes or optima.



## Layout

The layout of the trial describes both the 'objects' (plots) of the experiment and the way the treatments are allocated to them. In agroforestry experiments the basic 'objects' or units are usually plots of land. However the same ideas apply to other types of experiments, such as those on animals.

When describing experiments on plots of land there is a hierarchy in the layout. At the top level are the sites where the trial is conducted. There might be one or several. A single site may be chosen to match the conditions in the objectives. Alternatively a number of sites may be required, for example covering a range of environments. If the trial is done on-farm, then each farm is a site, and they may be grouped according to criteria relevant to the objectives.

Within sites there will often be known (or suspected) patterns of variation that define blocks. The size and shape of the blocks should be determined by the variation expected.

Within each block will be a number of plots. The layout of a single plot (its size, shape, position of trees and crops within it, the size of guard areas and so on) will depend on objectives and practical considerations of conducting the trial.

In some trials it is necessary to split plots into sub-plots, with different treatments applied to different sub-plots.

Within plots or sub-plots there are then measurement positions, perhaps individual trees, rows of a crop or layers of soil that are sampled.

Treatments are applied to plots or sub-plots. The design includes specifying;

i. The treatments used at each site (often all treatments occur at all sites, but this need not be the case, particularly in on-farm trials in which treatments reflect farmers interests).
ii. The treatments applied in each block. It is common, but not necessary, for each treatment to occur in each block.
iii. The allocation of treatments to plots (or sub-plots) should be random, and it is important to know whether this is the case.

## Measurements

The objectives determine the measurements that are taken. The measurements include those which define 'responses', typically variables such as crop yield, tree growth, disease levels and farmers rating of a practice, and those which are used to explain or understand variation in the responses, perhaps the rainfall at each site, the soil depth in each plot or the labour available on a farm.

Measurements can be made at any level in the layout hierarchy. There will be characteristics of sites (perhaps rainfall, soil type and average farm size) and blocks that can be measured. Some measurements can be made on whole plots (e.g. grain yield). Others require samples (e.g. grain moisture).

When measurements are made on part of the plot we can distinguish cases in which i) the actual location of the sample is irrelevant, and ii) the location is critical to the interpretation of the data, for example when we want to know how something changes with distance from a tree or depth in the soil.

## Practical considerations

Many details of an experimental design reflect practical constraints that occur in any field situation, and might mean the design cannot follow the guidelines for an 'ideal experiment'.

Examples in agroforestry experiments include:

- **Treatments**: farmers may not have space to evaluate all the desired treatments, so the set of treatments is not the same for each farm.
- **Layout**: randomization may not be complete, so that the cost of preventing interference between tree and non-tree plots is kept reasonable.
- **Measurement:** varying harvest times might mean that it is not feasible to measure grain yield in farmers' fields. Instead farmers are asked to score crop performance on a 5-point scale.

A common problem in agroforestry experimentation is that 'distance from trees' is often considered a treatment even though this is usually more a problem of choice of measurement, rather than treatment.

## Types of design

There are many ways in which designs and the ideas behind them can be described and classified. The way chosen here (in terms of treatments, layout and measurements) is unusual, but is used because:

- It helps keep the review brief!
- It turns out that much of the analysis can be described in the same terms.

The latter is particularly useful when we move from the simplest cases to the sort of sets of data that experimenters in the real world often generate, and it is necessary to understand why some of the simpler analysis methods are inadequate.

Trying to name designs (e.g. randomized block design, split-plot design with repeated measures) is not very helpful because there are many more important characteristics of a design than can be incorporated into a brief name. Therefore rather than trying to give a name to an experimental design, it is important to understand how it is put together.

Often on-station and on-farm trials are handled separately, because it is thought that design and analysis ideas for the two are different. However the underlying principles are actually the same for both and it is the different objectives, constraints and opportunities that can give rise to differing problems in design and analysis. For this reason we do not distinguish between on-farm and on-station experiments, but rather describe how to handle differing treatment, layout and measurement structures.

The training workshop is concerned with analysis of experiments. However some ideas from survey design and analysis are relevant for some experiments.

For example, in an on-farm experiment farmers might be selected following a well-defined sampling scheme, and inferences made about population responses on the basis of the trial results.

## The analysis training workshop

It will not be possible to complete an effective and valid analysis of the data if you do not understand these 4 aspects of the trial; objectives, treatment, layout and measurements. The details of the analysis, and in particular the extent to which the statistics is 'easy' or 'hard', depend on them and the way they interact. The same ideas are used to structure the analysis-training workshop.

The first part of the training workshop concerns experiments in which:

- The treatments form simple sets, such as levels of a single factor or, complete factorial arrangements.
- The layout does not have too many layers in it (blocks, plots within blocks and possibly split plots) and the treatment allocation is simple (a factor is allocated to either all or none of the units in a layer).
- The measurements result in data, which are complete and (approximately) normally distributed.

The second part of the training workshop then covers experiments in which complications arise, such as:

- Treatments that have complex structures because, for example, some combinations are not present, or extra control treatments are added to a factorial set.
- The layout has more layers, such as several sites or many sample locations within each plot, and in which treatments are not allocated 'evenly' to the units in a layer.
- The measurements are made on categorical scales, are made on several different 'layers' in the hierarchy, or contain many missing values or zeros.

Session 2

# Objectives and steps in data analysis

*R. Coe, R.D. Stern, E. Allan*

## Introduction

In the previous session we reviewed experimental design and the importance of stating clear trial objectives. In this session we consider the common steps taken in any data analysis. We want to carry out an analysis of the data that will enable us to meet those trial objectives. The first step in data analysis is therefore determining the objectives of the analysis, and this is covered in detail within this lecture note. Another prerequisite for the analysis is preparation of a data file. This step should have already been covered in data management, a preliminary session to the analysis course.

The remaining steps in analysis are outlined in this session. They are covered in detail in further sessions of the course.

## Steps in the analysis

Every experiment is unique. Trials vary in all aspects of the design because their objectives differ. They will therefore need different analyses to meet these objectives. Trials differ in their position within a research programme, the interests of the researcher, and the time (or money) available for analysis, as well as in the more obvious features of design and data collected. Even a trial, which is an exact replica of an earlier one, will have as one objective, confirmation of the results of that earlier one, and therefore is different.

So why try to write down steps in analysis? The reason is that the analysis of many experiments does not proceed as smoothly or efficiently as it might because some important points are forgotten. It is thus helpful to define some generic steps, even though the details change from trial to trial.

### 1. Defining the objectives of analysis

**It is impossible to do a sound analysis without knowing what you wish to achieve.** Too often an analysis is started without a clear idea of where it is going. The result is usually a lot of wasted time and an inadequate analysis. Avoid this by deciding on the objectives of the analysis before starting it.

## 2. Preparing the data

The computer files of raw data will already have been prepared and checked. However there will still be some work to do in preparing the files for statistical analysis. For example, we may need to:

- Construct the variables needed for analysis (e.g. find plant N content from concentrations and biomass data).
- Summarize them to the right 'level' (e.g. average multiple samples per plot-to-plot values). This is often easier to achieve in the database environment or spreadsheet (Access or Excel) than in the statistical analysis environment (GenStat or SAS).
- Make sure the column names are suitable for the statistics package used.
- Save in a format that can be read by the statistics software.

## 3. Descriptive analysis

- Calculation of summary tables and graphs, as defined when setting objectives.

- Exploratory analysis to identify any unexpected patterns or results.

## 4. Confirmatory analysis

This is the formal analysis, mainly aimed at:

- adding measures of precision (e.g. standard errors and results of significance tests) to the results found in the descriptive analysis,
- improving the estimates of various critical quantities.

## 5. Interpretation

This is not just about understanding the meaning of a statistical test, but the whole job of integrating the new knowledge with the existing body of knowledge on the problem. This will involve comparing results with those from other studies, building predictive models and formulating new hypotheses.

## 6. Reporting

Reporting the analysis and presenting the final tables and graphs.

Along with the report, consider the archive of the data and the analysis methods used. The analysis methods need careful recording in case anything is questioned or needs repeating. The data need archiving both so that any analysis can be repeated, and further use of the data can be made.

## Defining the objectives of analysis — more details

**It is impossible to do a sound analysis without knowing what you wish to achieve**. Too often an analysis is started without a clear idea of where it is going. The result is usually a lot of wasted time and an inadequate analysis. Avoid this by deciding on the objectives of the analysis before starting it.

The objectives of the analysis are primarily determined by the objectives of the trial. These should have been stated clearly when the trial was designed. However, it might be necessary to clarify some of them as you think about analysis. Often the objectives of experiments reduce to:

- Simple comparisons between two treatments. For example, what is the difference in crop yield between an agroforestry system and a monocrop system? What proportion of farmers finds the agroforestry system preferable?
- Comparisons involving several treatments. For example, is there a production advantage in an agroforestry system compared with monocrop plots of all the components?
- Identifying the rate at which a response changes with a change in stimulus. For example how does crop performance increase with increasing soil organic matter? How does survival depend on the level of farmer management?
- Determining how the response to a stimulus changes with different conditions. For example, how does the addition of rock phosphate change response to added organic matter?
- Identifying optimal management or conditions. For example, what pruning frequency and timing will maximize fodder production?

Almost certainly there will be several related objectives, that is why most experiments have more than two treatments. The objectives should also indicate which measurements are needed in the analysis. These may be 'responses' (e.g. tree height, farmer's rating of a practice, economic return from a cropping system) or measurements that distinguish different conditions (e.g. number of times a farmer weeded, soil depth).

The definition of the objectives can be taken further in two important ways.

## Deciding on which variables to base conclusions

Most data sets from experiments contain several measurements. Which measurements will be used in analysis to meet the objectives?

For example, a trial was designed to compare maize production in an agroforestry system, with that from a sole maize system. The trial ran for 5 years, producing a maize crop in each of the seasons. Any of the following may be appropriate variables on which to base the analysis:

- The yield in each season, which is analysed separately (the comparison may well depend on the time the system has been running, and on the weather in each season).
- The total yield over 5 years, to look at cumulative benefits.
- The cumulative yield, ignoring yield in the first season when no benefits are expected.
- Cumulative discounted financial benefits.



## Table and graph design

Designing the tables and graphs that will be used to display the results before any numerical analysis is started, can help focus the statistical work. For example, you may decide you need a table that shows the mean maize yield for each treatment for each season. However, it may not be quite that simple; perhaps only some treatments need including; perhaps others need combining by averaging; maybe there were subplot treatments, to be averaged, displayed in a separate table; and so on. Furthermore, we need to be clear about the actual variables to be displayed. In the case of a crop yield, that may be obvious, but may not be obvious for a more complex measurement. For example, root length density was measured at 10 different depths at 6 times during the growing season. The combination of depths and times to compare between treatments, and included in the tables or graphs, may be fixed by the objectives. The design of the final tables or graphs presented may well be modified depending on the results of the analysis. For example, if the response to treatments turns out to be much the same each season, there is no point in giving the results for each season separately. The useful information is all in the average, across seasons. It is still worth thinking of the tentative table and graph design at the start.

As an example, consider the example experiment 'Relay planting of *Sesbania sesban* and maize'. Two of the objectives are:

- To determine a good time for planting *S. sesban* in relay with maize.
- To determine the interaction between *Sesbania* and fertilizer applied to the maize.

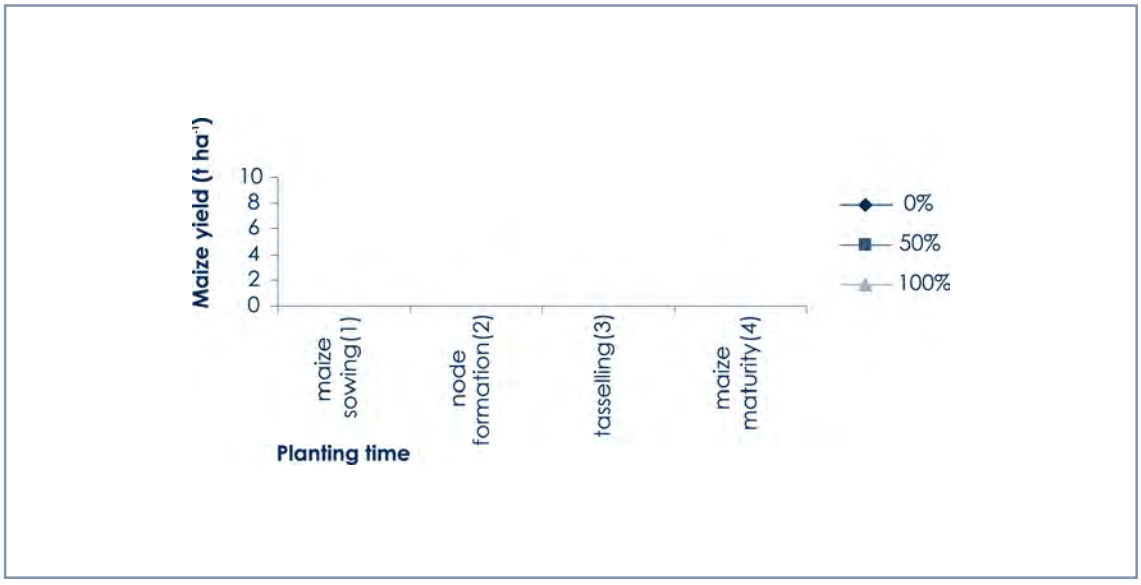
Data available:

- Dry maize grain yield (t $ha^{-1}$) for the 1994 season.

It seems likely that the differences we find in maize yield will depend on both the planting time and the amount of fertilizer applied. We can combine our two objectives above to form a third objective; investigating the interaction between planting time and fertilizer application. We might use the following Table 1 and Graph 1 to answer all three objectives (Note that in the graph we have not included the 'no *Sesbania sesban*' treatment, though the yield for that treatment could perhaps be indicated with horizontal lines).

| Dry maize grain yield (t $ha^{-1}$)<br>Planting time | Fertilizer (% of recommended amount)<br>0 50 100 |
|---|---|
| maize sowing(1)<br>node formation(2)<br>tasselling(3)<br>maize maturity(4)<br>*no* Sesbania sesban | |
| SED | |

Table 1. 1994 season, average maize yield (t $ha^{-1}$) for every combination of planting time and fertilizer application.




Graph 1. Average dry maize yield for 1994 season (t $ha^{-1}$). Four *Sesbania sesban* planting times and three different fertilizer application levels.

The primary objectives set out at the start of the trial are likely to have others added before analysis starts. For example:

- Understanding **why** the results of the primary objective are the way they are.
- Investigating unexpected observations made during the trial.

Furthermore, the objectives may well be modified during the analysis as unexpected patterns in the data emerge. This does not detract from the value of starting, by defining analysis objectives.

Session 3

# Software familiarization

*R. D. Stern, E. Allan, R. Coe*

## Introduction

In this lecture note we describe some characteristics of the statistics package, GenStat, which we will be using for the remainder of the course. We will also discuss the reasons for its choice as our main analysis package. This session will give you an opportunity to begin using GenStat and to learn about its uses in statistical data analysis.

This is the only session that is devoted entirely to learning how to use the statistics package GenStat. In the past, similar workshops have devoted far longer than we have planned, to the training in the use of the statistics package. We found that although participants could become reasonably adept in the mechanics of using the package, they lacked the statistical skills in knowing which analyses were appropriate considering the objectives and the structure of their experimental data.

More recently however, the ease of use of Windows software, plus the greater computing experience of most participants, permits us to concentrate more on statistical concepts, while still practicing our use of GenStat. Most of the remaining course sessions will make use of GenStat and extend participants' experience in the software, whilst at the same time covering the ideas of data analysis.

For the practical part of this session you will go through the tutorial within the GenStat for Windows Introductory Manual (GWIM). Exercises in later sessions of the course will occasionally refer back to other sections of GWIM. Participants who need more time to practice the mechanics of GenStat are welcome to follow other sections of the manual, either within the course or afterwards.

## Statistical software – GenStat for Windows

There are many statistics packages. We use GenStat for Windows in this training course because it provides a flexible system for analysing experimental data. The main analyses use three methods, as described by the following parts, each of which corresponds to a Windows dialogue.

These are called ANOVA, regression and REML. We also use a wide range of facilities for data manipulation, exploration and descriptive statistics.

The ANOVA system in GenStat analyses all orthogonal experimental designs. It is the main dialogue in GenStat that will be used in the basic part of the course. The structure of the ANOVA dialogue is shown in Figure 1 below.

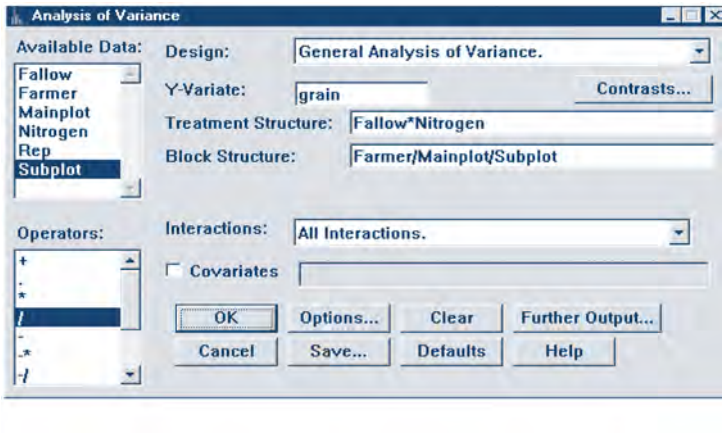



Figure 1. The ANOVA dialogue box.



It fits well with the philosophy of distinguishing between the;

- treatments,
- layout (called blocks in GenStat),
- measurements.

GenStat's ANOVA system is very powerful and presents the results in a format that is both clear and easy to use as shown in Table 1.

```
***** Analysis of variance *****
Variate: grain

Source of variation      d.f.          s.s.        m.s.      v.r.     F pr.

Farmer(Rep) stratum         8       55.1600      6.8950      4.73

Farmer.Mainplot stratum
Fallow                      3       14.0076      4.6692      3.20     0.041
Residual                   24       35.0173      1.4591      3.06

Farmer.Mainplot.Subplot stratum
Nitrogen                    1       16.6317     16.6317     34.90     <.001
Fallow.Nitrogen             3        6.7683      2.2561      4.73     0.008
Residual                   32       15.2498      0.4766

Total                      71      142.8346

Variate: grain

Grand mean              1.696

   Fallow    Continuous maize    Crotolaria   Tephrosia  Tithonia
                        1.050         1.825       2.279     1.630

 Nitrogen                  No           Yes
                         1.215         2.176

                      Fallow      Nitrogen          No       Yes
               Continous maize         0.700       1.400
                    Crotolaria         1.785       1.864
                     Tephrosia         1.492       3.065
                      Tithonia         0.883       2.376


*** Standard errors of differences of means ***

Table                   Fallow      Nitrogen      Fallow
                                    Nitrogen
rep.                    18           36             9
s.e.d.                   0.4026       0.1627        0.4638
d.f.                    24           32            39.11
Except when comparing means with the same level(s) of Fallow
                                                   0.3254
 d.f.                   32
```

Table 1. Output from GenStat's ANOVA system

GenStat also encourages a critical approach to data analysis, as is illustrated by the message concerning large residuals above.

Not all experimental designs have the neat arrangements of treatments and layout that allow GenStat's ANOVA commands to be used. GenStat's facilities for regression are introduced in Session 7. The regression facilities are also useful for data analysis when observations, such as counts or disease ratings are not normally distributed. This extra complexity is discussed in the second part of this course, in Session 11.

We show, in Session 8, that experiments can be more complicated to analyse when information is at more than one level. A simple example is a split-plot design, where treatments are applied to different-sized plots. Similarly lattices are a common design for variety trials, they have two levels of blocking. These simple cases of multi-level can be handled by GenStat's ANOVA system. More complicated designs require special facilities for multi-level modelling and this involves the third main system within GenStat for the analysis of experimental data, which is called REML.

Despite the possible complexity of the analyses, their use within GenStat is straightforward and simply involves the completion of a Windows dialogue in each case. The workshop can therefore concentrate on the appropriate use of these facilities and the interpretation of the results in relation to the objectives of the trial.

An additional strength of GenStat for the analysis of experimental data, results from the types of data structures that can be defined (though we see below, that this is a 'two-edged sword'). A statistics package is essentially a 'column calculator' and ordinary columns of data are called **variates** in GenStat. Observations from an experiment (e.g. measurements, farmers' scores) are normally stored as variates.



In the analysing experimental data we emphasize that the method of analysis is dictated by the **structure** of the experiment. The structure is defined by the **block** and **treatment** effects, and these are defined to be **factor columns** in GenStat. This is the same as a **class** variable in SAS, but is defined in the GenStat dataset rather than for each procedure used. Factor columns may have any number of **levels** and the levels may take any numeric value, this helps in the modelling of experiments with **quantitative factors** (such as spacing levels, fertilizer levels). Text strings may be attached as **labels** to **qualitative factors**, such as species names, to enable results to be presented clearly. Although GenStat makes a clear distinction between factors and variates, it is simple to change a factor into a variate and vice-versa.

Other types of structure may also be defined, and one that is used in the workshop is a table. In analysing experimental data, tables are usually used to display means and sometimes frequencies and percentages. Tables in reports are often one-way, (this then looks like a column of summary numbers), or two-way (perhaps a table of treatment means for each level of spacing and variety). Examples of a one-way and a two-way table are shown in Table 2 below. Note that tables in GenStat can be up to nine-way!

| Mean | |
|---|---|
| Trt | |
| c | 1.727 |
| g | 2.470 |
| s | 2.503 |

| Mean | | | |
|---|---|---|---|
| slope | 0.00 | 1.00 | 2.00 |
| trt | | | |
| c | 2.258 | 1.434 | 1.447 |
| g | 3.274 | 2.078 | 2.310 |
| s | 2.920 | 2.255 | 1.570 |

Table 2. Example of a one-way and two-way table

We describe the use of GenStat's language in the GWIM guide (Part 2, 'Moving from menus to commands') and a feature of the language is that it can be extended by (experienced) users. Many scientists have contributed extra facilities for the analysis of experimental data. These additions are incorporated into libraries of procedures and some are being used automatically as you complete the Windows dialogues for the analyses. The fact that GenStat's language is a mixture of built-in commands and extra procedures is almost invisible to the user, because they are used in the same way. But the existence of these extra facilities explains the rich set of facilities provided for the analysis of experimental data in GenStat.

These positive points about GenStat explain why it is the main statistics package for the workshop, as written. However, there are some negative points in relation to the analyses that we have mentioned. The main one is that GenStat is not as intuitive for data manipulation and simple descriptive statistics as Excel, for example. Hence experienced Excel users may find they wish to conduct some of the simple analyses in Excel. For the tasks considered here, we define an experienced Excel user as someone who is confident in the use of Excel's pivot tables.

The difficulties in these basic areas are primarily for two reasons. The first is that the different types of structure in GenStat (variates, factors, tables, etc), mentioned above, that help users in conducting a logical analysis, also enforce a discipline that is sometimes counter-intuitive. For example, you might want to plot data that are in a GenStat table, rather than in columns of variates. You can, but first you must put the data back into columns of variates, because data stored in GenStat's tables, cannot be plotted directly.

The second reason for complications is that the Windows version of GenStat is in two parts. The first part of GenStat is the Windows 'front end'. This includes all the parts you see, such as the dialogues and the GenStat spreadsheets. Data are simplest to manipulate in a GenStat spreadsheet. The resulting values are then copied to the second part of GenStat, the **server**, which is the part of GenStat that does the actual analyses. The values, once copied into the server, are then ready for the analysis, when you have left the spreadsheet. Similarly, when you complete a dialogue box the 'front end' translates the contents into GenStat commands that are then sent to the server. The server obeys the commands and sends the results back to the Windows 'front end' where you can see the results of your analyses.

GenStat is simple to use once you follow the logic imposed by the two parts of the program, but perhaps not as simple as software that is totally confined within Windows.

Despite these weaknesses we consider GenStat to be the obvious package for the workshop if both the basic (Session 1 to 9) and the further sections (Sessions 11 to 16) are taught. The software is reasonably easy to learn and we would encourage all participants to use GenStat even if they are already familiar with another package.

The case is not so clear if the training emphasizes the material from just some of the sessions. For example a course devoted primarily to the more advanced topics from Sessions 11 to 16, which was to be given to scientists who already use SAS, might continue with that package. SAS includes PROC GLM instead of GenStat's regression facilities and PROC MIXED instead of REML. In contrast, a course that just emphasized the descriptive and exploratory methods of analysis might consider Minitab, Systat or S-PLUS.



## Other software

The other main package used in this workshop is Excel. All the standard datasets are in Excel. Excel's facilities for data manipulation and simple plotting may be used in preference to GenStat for those who find it easy. An alternative strategy is to use GenStat for these tasks as a way of consolidating your knowledge of the package.

Workshops that are designed to include additional sessions on data management may also use MsAccess.

We hope that most participants have prior experience in using a word processor, possibly Word, for report writing and also in PowerPoint to construct presentations. If necessary, both packages are easy to learn to the standard needed for the workshop. A short guide may be provided on each to be studied during participants' spare time. These packages are peripheral to the objectives of the workshop and we cannot therefore allocate sessions to the learning of these packages.

# Session 4 Descriptive analysis and data exploration

*E. Allan, R. Coe, R. D. Stern, J. De Wolf*

## Introduction

In Session 2 we saw that the formation of the objectives of the analysis will enable us to construct outline tables and graphs needed to meet the objectives. One of the aims of this session is to provide the data summaries in the form we specified in Session 2. All the analyses described may be done in GenStat. However this initial descriptive analysis may, in some cases, also be done in Excel, which has powerful tools for calculation in tables. However Excel will not be useful for any further analyses.

Simple exploratory and descriptive analysis will often provide much of what is needed in a complete data analysis. Hence, for some studies, you will then be ready to move straight on to Session 9 after this session, and to consider how you will write up and present your results.

The descriptive analysis and data exploration provides the following:

1. The first estimates and summaries, arranged in tables and graphs, to meet the objectives.
2. Information about the variability or uncertainty in the data.
3. Indications of unexpected patterns and observations that need to be considered when doing formal analyses.

Scientists sometimes ignore this descriptive stage of data analysis and proceed immediately to conduct an ANOVA, which is described in Sessions 5 and 6. They may be concerned that by carrying out exploratory and descriptive analysis they are adding substantially to the time needed to do their analyses. We set their minds at ease by explaining that for simple experiments this preliminary stage usually does not take long; it takes minutes, rather than days. As it focuses directly on the analysis in relation to the objectives of the trial, it will often save time overall, when considering the total project time including the writing of the report. For larger experimental studies or surveys this exploratory analysis does take time. This reflects the complexity and richness of the data sets. Scientists who begrudge the time spent on the analysis should reflect on the proportion of time that is used for the analysis in relation to the total 'person years' that are devoted to the study as a whole. It is a waste of all this time if the data are not then effectively analysed.

In this session we look at two contrasting examples.

## Example 1: a field experiment

The first example uses data from the trial 'Upperstorey/understorey tree management'. The objectives are:

a) To determine the production potential of the upperstorey trees grown at different intra-row spacings and in association with Napier grass and *Calliandra* as understorey hedgerows.
b) To determine the effect of the upperstorey trees and understorey species on adjacent crops.
c) To determine the effect of the upperstorey and understorey species on each other.

Notice that the other three components of the design, the treatments, layout and measurements, will also be crucial in this exploratory analysis. The experiment is typical in that the layout is very 'regular' (3 replicates of 10 treatments) and the main responses (tree dimensions and crop yields) are measured on continuous scales.



### What do the data tell you?

Preliminary investigations allow you to look closely at the data collected in your experiment. They can be described as:

**data = pattern + residual**

**Pattern** is the result of factors such as the experimental treatments, and other characteristics such as soil fertility, which cause the response to change. Identifying the pattern, or at least that due to treatments, is therefore an important part of the analysis. Knowledge of the treatments will be crucial in identifying this pattern. Other parts of the pattern may well be induced by the layout, such as differences between blocks, for example. The layout information will therefore also be of importance in identifying patterns. Finally, the way we summarize a pattern will depend on the nature of the measurement.

**Residual** is the remaining unexplained variation. This might just represent 'noise' or variation about which we cannot do anything. But it may contain further patterns due to explanations we have not looked at yet.

Both 'pattern' and 'residual' can, and should, be studied in initial investigations of the data. Useful tools include descriptive statistics and graphs such as boxplots and scatterplots.

## Pattern

Pattern can be easily summarized by descriptive statistics such as mean values.

The most obvious pattern to look for is the effect of the treatments since this effect relates to the experimental objectives. We will already have identified an appropriate table or graph, whose objective is to provide an answer to the questions raised by the experiment's objectives.

Considering objective (b) in our example, the effect of the upperstorey trees and the understorey hedgerows on the adjacent crop, can be deduced from the mean crop yield for the different spacings and different hedgerows. A table which gives mean values for the combination for upper and understorey treatments, and includes the overall means for both the different tree spacings and the different hedgerow conditions, would be a desirable table to produce. Notice in Table 1 that to create this table we need to know what the treatments are, not just that there are 10 treatments. The common way of coding data that simply lists treatments 1 to 10 would miss crucial information in this analysis.

| Mean under spacing | CAL | NAP | NO | *Mean* |
|---|---|---|---|---|
| 1 | * | * | 23.60 | *(23.60)* |
| 3 | 19.10 | 17.30 | 24.83 | *20.41* |
| 5 | 22.53 | 20.87 | 25.03 | *22.81* |
| no trees | 22.17 | 19.77 | 27.20 | *23.04* |
| *Mean* | *21.27* | *19.31* | *(25.17)* | *22.24* |

Table 1.

The table is laid out in increasing width of spacing (metres). ['No trees' is the widest possible spacing (i.e. infinity), but is coded 0 in the dataset.]

The figures in italics are mean yields for the different spacings, averaged across the understorey treatments, and mean yields for the different understoreys averaged across the different spacings. These would be a useful summary of the effects of spacings and of understorey if:

(1) These two affected production independently.
(2) The take was complete, with all combinations present.

In this case condition (2) is not met, so the marginal means are of limited use.

Crop yield is highest (27.2 kg/plot) when there are no trees and no hedgerow. So how much does the introduction of either of these affect the crop yield? Adding a line of trees to the plot decreases the yield by 2 or 3 kg/plot, and the reduction is influenced slightly by the closeness of the spacing. Adding a hedgerow, but no trees, has a slightly more serious effect since the yield now decreases by about 5 or 7 kg/plot, depending on whether *Calliandra* or Napier grass is grown.

The picture is less clear when different tree spacings and understorey treatments are used together, since the effect of the trees and their different spacings is now slightly different when introduced along with a hedgerow. If anything, they seem to have a slight antagonistic effect on the yield reductions. It now appears that, in the presence of a hedgerow, whilst 3 m spacing may give a slightly decreased yield compared with no trees, the yield with 5 m spacing is just as good as when there are no trees.

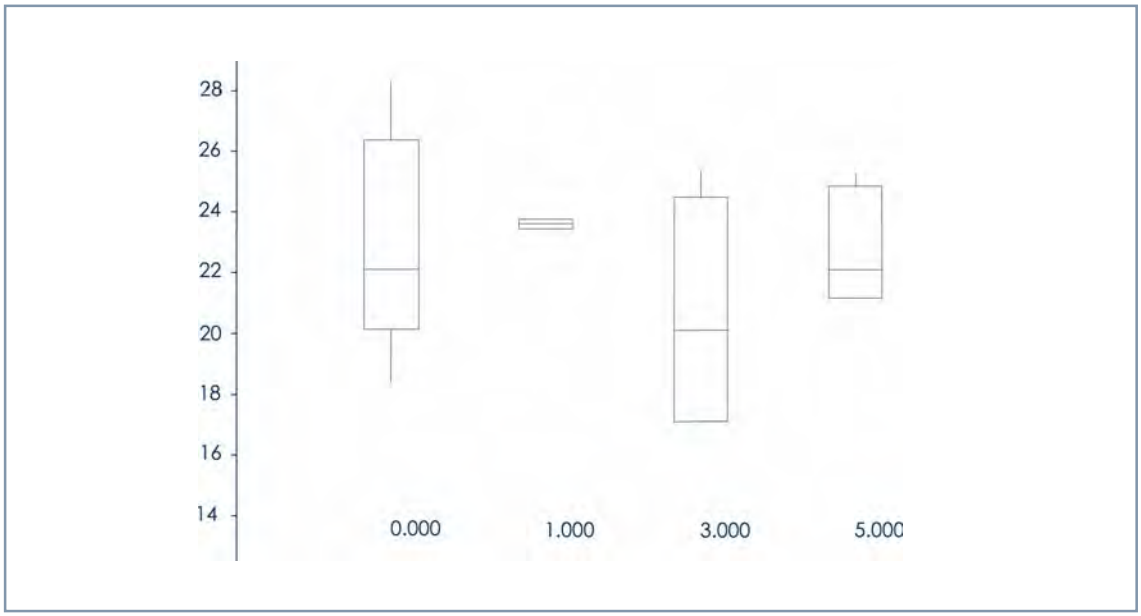


## Residual

The residual is the variation in the data around the pattern. Looking at this variation during the exploratory analysis is important because:

- You need to be aware of the variation in the data as it impacts on much of the analysis.
- If the variation is large it might contain further pattern-forgotten effects due to treatments, layout or measurements.
- There may be a pattern in the variation! For example, variation that is very different for different treatments.
- Odd values which may have undue influence on conclusions are revealed.
- The variation may be an important part of conclusions if you want to discuss risk and uncertainty.

To explore the variability surrounding the pattern associated with our treatments, we could use boxplots. Below  in Graph 1 are boxplots for crop yield for the different tree spacings in our example.



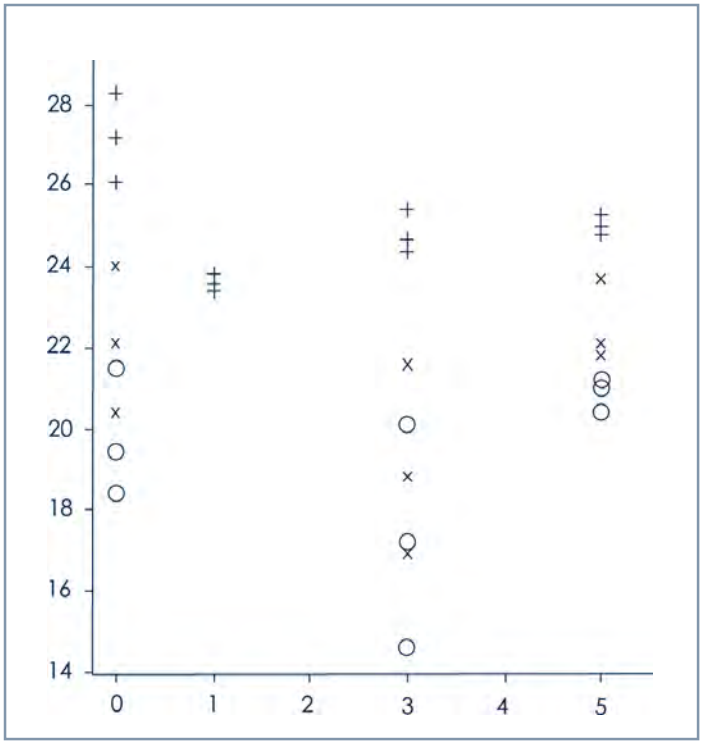

Graph 1. Boxplot of crop yield for different spacings

Boxplots (or box-and-whisker plots) are useful for visualizing both pattern and residual. The horizontal lines in the middle of the box indicate the median yield. The boxes range from the lower quartile to the upper quartile, and therefore give an indication of spread in the yields. The whiskers (vertical lines extending out from the boxes) identify minimum and maximum crop yields.

The first thing to notice in the above plot is that the variability for 1 m spacing is so much less than all the other spacings. Why should that be so?

Remember that in this trial the 1 m spacing was only studied without a hedgerow treatment, whilst the others were all studied with three different understorey conditions. The increased variability with the other spacings is due to the differences between the hedgerow conditions. **This plot is therefore not a good summary from which to draw conclusions about variability,** since some of the pattern associated with the understorey treatments is hidden within the overall variability.

A better exploration here would be a scatterplot as seen in Graph 2, where the individual plot yields are plotted against spacing (on the x-axis) with different symbols, or colours, indicating the different understorey treatments. Here we use + = none, o = napier, x = *calliandra*.

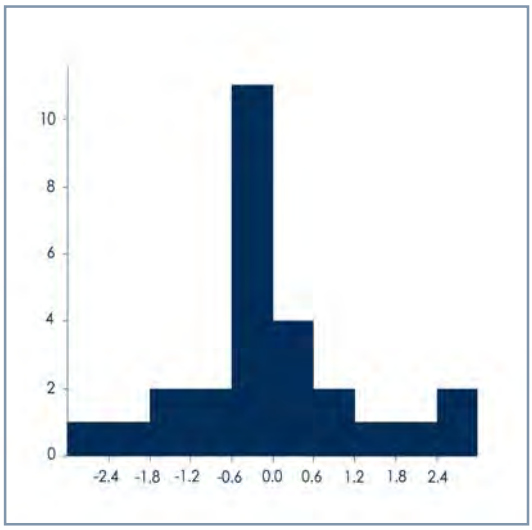



Graph 2. Scatterplot showing yield v. spacing (symbols show different understorey treatments)

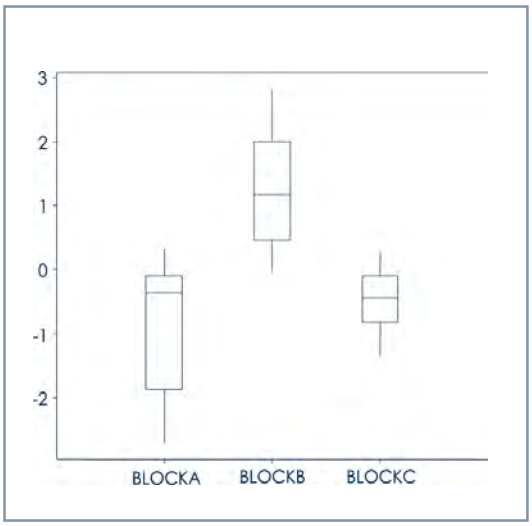
This plot shows quite clearly that there is a reduction in crop yield when Calliandra or Napier grass hedgerows are used. It also shows the effects of tree spacing, which have already been discussed. With only 3 observations per experimental treatment, it is difficult to say much about the residual variation surrounding the treatment pattern. If anything there is a slight suggestion of increased variability in the crop yields when the hedgerow treatments are present.

The above plot (or even the boxplot on the previous page) is also a useful tool for detecting outliers i.e. strange values in the data which need to be checked further. There are none in this example.

An alternative approach is to actually calculate residuals. In this example we could summarize the pattern simply by the mean for each treatment. If these means are subtracted from the original observations, any remaining variation is not 'caused' by the treatments and might reveal other patterns that need allowing for, in the formal analysis. If the bean yield residuals are calculated in this way then we have 27 numbers, some negative (observations less than the treatment mean) and some positive (observations greater than the treatment mean). The histogram (Graph 3) below shows them to be spread from about –2.5 to +2.5 with a standard deviation of 1.3, but with most of the values close to zero. The boxplot (Graph 4) on the right shows the variation in these same residuals separately for each block. It looks as if observations in block B tend to be greater than their mean while those in blocks A and C tend to be less than their mean, something to remember in further analyses.




Graph 3.




Graph 4.

## Other exploratory analyses of pattern and residual

Having identified effects due to treatment, what other patterns might we be interested in? Essentially, anything which explains some more of the variation in the data. Pattern explained by block differences is an obvious one to examine, since experiments are usually blocked because of heterogeneity in the experiment. Other patterns of potential interest would be those associated with concomitant information, e.g. fertility recorded as previous yield, or some other factor recorded as the experiment progressed.

The choice is limited in our example. The experiment was laid out in blocks, and so we should look at the pattern associated with the block differences. Again we use mean values, or boxplots as in Table 2. The mean values are:

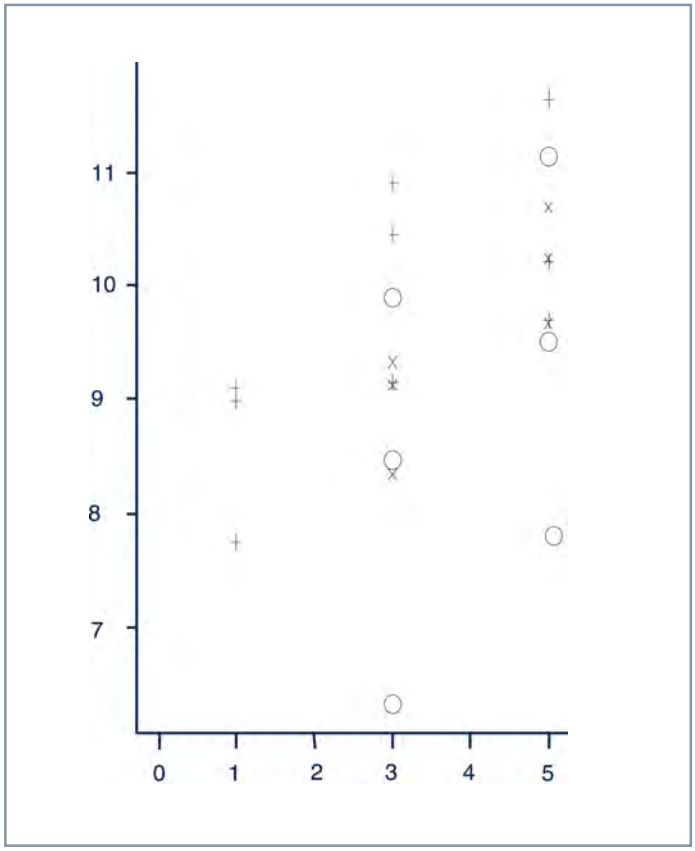
| | Mean |
|---|---|
| block | |
| BLOCK A | 21.47 |
| BLOCK B | 23.44 |
| BLOCK C | 21.81 |

Table 2.

Crop yields are slightly higher in block B as opposed to A or C, as was also shown by looking at the residuals earlier. These differences will explain some of the residual variation about the treatment mean yields, which we observed earlier.

Let us continue the theme of investigating the data for patterns. We will consider the third objective of the trial, which concerns the effect that the upperstorey trees and understorey species have upon each other.

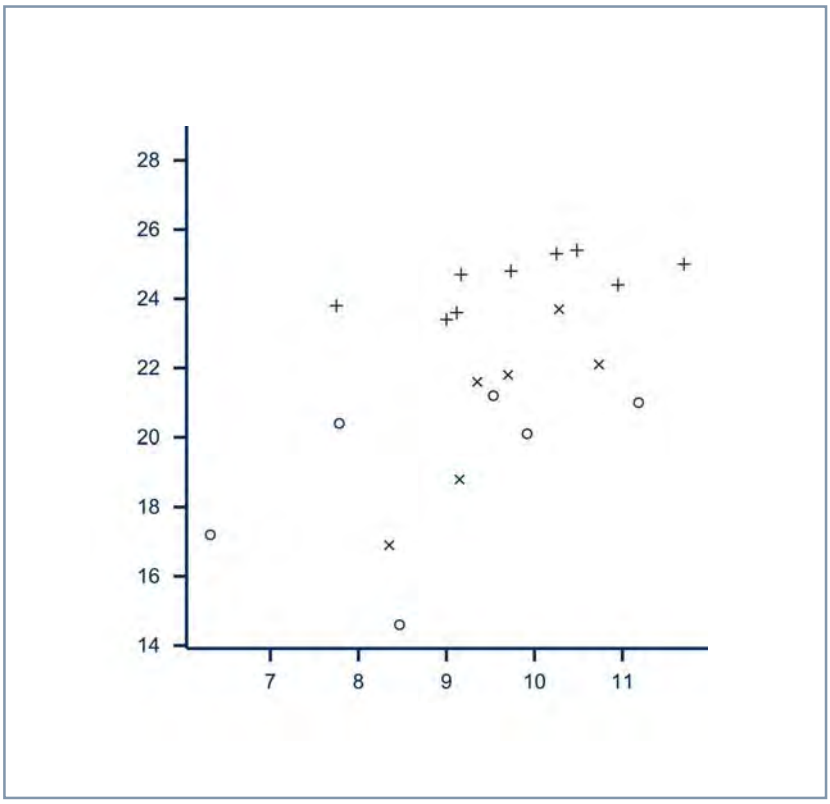
First of all, how are the trees affected by their spacing, and by the understorey conditions? We will look at a scatterplot of tree diameter breast height (dbh) versus tree spacing, using different symbols for each type of understorey treatment (+ = none, o = Napier grass, x = *Calliandra*) in Graph 5 below.




Graph 5. Scatterplot of dbh vs spacing

The plot of dbh (a measure of tree size) shows that (i) trees tend to be larger when the spacing is farther apart, and (ii) their growth is reduced when hedgerows are grown in the intervening space between the trees.

What other aspects of the data could we look at in our investigation of how a combination of trees and hedgerows might affect adjacent crop yield? We have seen above, that trees tend to be larger when the spacing is farther apart. But how does this affect the crop yield? The plot below in Graph 6 depicts the relationship between yield and tree size (dbh).




Graph 6. Relationship between yield and tree size (dbh)

Yield is positively related to the height of the trees when there are hedgerows; but there is barely any relationship when there are no hedgerows (+ = none, o = Napier grass, x = *Calliandra*).

## Example 2: monitoring of farmers experiences

Here we consider the types of presentation that are often a useful start when analysing data from monitoring a large on-farm trial. The example we use is described in the protocols as 'Improved fallows and rock phosphate: farmers' experiences'. The study is about experiences of farmers with improved fallows and the subsequent crop. These farmers had been taught what these fallows could do to improve their soils and were able to visit demonstration fields in their neighbourhood. The researchers made seed for planting the improved fallow, and fertilizer (rock phosphate), available. More than a year later a sample of the 121 farmers that had at least planted the fallows were visited and asked to answer a questionnaire. The questionnaire was carried out in two phases. First farmers were asked some general questions. Later the interviewer and the farmer visited each of the fields planted with improved fallow. A total of 219 such fields were visited.

One characteristic of this type of study is that the information is largely in the form of qualitative variables. Farmers are asked to give their opinion about a treatment in terms of better or worse than a control, rather than for actual yields. The information about whether a farmer is practicing a certain method is a simple 'yes' or 'no'. The different nature of these variables compared to continuous variables (e.g. yield in kg/ha and height in cm) discussed until now, requires a different analysis. We consider the descriptive analysis in this session. A separate session is devoted to further analysis of this type of data (Session 12).

The general objectives of this study are given in the protocol. Typically, as here, there are many objectives, and this means that the descriptive presentations are a more time-consuming and key part of the analysis, compared to an on-station trial. We will focus on a few objectives, of which the first concerns whether farmers would use rock phosphate in the future if they would have to pay for it themselves.

Sometimes objectives in this kind of study are absolute, rather than a comparison. This is typical in a survey, but not common in an experiment where the objectives usually relate to comparisons between treatments. Some objectives in type 3 trials are also comparative (e.g. 'Do female farmers adopt improved fallows as frequently as male farmers?'), but absolute results (e.g. 'How many farmers adopted improved fallows?', 'What area is currently under improved fallow?') are also important.

We consider first, the way in which tables of counts can be given. They may be useful in their own right and are also needed as the initial step if results are to be given as tables of percentages. We devote a section of this lecture note to the ways in which percentages can be provided. In the session we show that the main complication arises when it is not clear what constitutes 100%. Depending on the objective it may be appropriate to consider the percentage of farmers (121 were interviewed), or the percentage of the 219 plots, or the percentage of the area that is represented by the plots.

Despite the differences in the way results are presented from this type of study, compared to an on-station trial, the key point is one of similarity. In each case, the analysis can start by users deciding on the structure of the tables and graphs that correspond to the objectives of the trial. When they have the data, they can complete these presentations and begin writing the report.

## Tables of counts

Our first objective concerned whether farmers would be prepared to use rock phosphate if it had to be paid for. This was answered 'yes' by 82 farmers and 'no' by 39 farmers.

To look at this objective in more detail it may be useful to investigate how male and female farmers responded to the same question. This leads to a two-way table (Table 3) and the following counts were observed. The numbers within the rectangle are the actual count in each sub-category. The last row and last column give the totals over a column or a row and are called 'margins'.

| | Use rock P? | | |
|---|---|---|---|
| | Yes | No | |
| Female | 46 | 20 | 66 |
| Male | 36 | 19 | 55 |
| | 82 | 39 | |

Table 3.

Female and male farmers do not seem to differ very much in their opinion about the future use of rock phosphate.

Obviously, two-way tables can have more than two levels for one or both factors. For instance, the farmers' prior experience with the use of fertilizers has three categories of 'experience', based on the frequency they used fertilizer in the past. Cross-tabulation with the future use of rock phosphate, leads to the two-way table (Table 4) below.

| Experience with | Use rock P? | | |
|---|---|---|---|
| fertilizer | Yes | No | |
| frequent | 14 | 5 | 19 |
| little | 22 | 13 | 35 |
| none | 43 | 21 | 64 |
| | 79 | 39 | |

Table 4.

From this table of counts we deduce that experienced farmers tend to want to continue with rock phosphate more often than less experienced farmers. Notice that the totals in the 'Use rock P?' margin have changed. A total of 82 said 'Yes' but now it is only 79. This is because 3 farmers did not respond to the question on fertilizer experience. They could be included in a forth category of 'Unknown' or 'Missing'. Such missing information can be confusing as the observations used in analysis change according to the variables being looked at. However in large data sets it will nearly always occur and methods and interpretations have to be robust to it.

Tables of counts can be expanded to represent more than two factors. A three-way table (Table 5) can accommodate the three factors involved in the two previous questions.

| | | Yes | No | |
|---|---|---|---|---|
| Female | Frequent | 13 | 1 | 14 |
| | Little | 8 | 5 | 13 |
| | None | 23 | 14 | 37 |
| Male | Frequent | 1 | 4 | 5 |
| | Little | 14 | 8 | 22 |
| | None | 20 | 7 | 27 |
| | | 79 | 39 | |

Table 5.

This presentation of the counts is more complex and not entirely satisfying (e.g. we miss the margins that summarize the counts for gender irrespective of experience and vice versa). It is, however, also much more informative. Only now we notice that the conclusion of the previous table, namely that experience with fertilizers influences future plans with rock phosphate farmers, depends on the sex of the farmer. Female experienced farmers tend to want to continue much more than their male counterparts. Considering only inexperienced farmers, however, the non-continuers form a larger proportion among women than among men. More complex tables will often be necessary to reflect the information in the data adequately. It makes presentation difficult and requires more advanced methods for analysis.

Earlier in this session we introduced the idea of looking at data as 'pattern + residual'. Is this useful here? At an individual level it is not. If we know that about 80 out of 120 farmers intend to use rock phosphate (the pattern) then we can not say whether an individual who does not (the residual) is a surprise or not. However it is a useful idea when looking for pattern in the tables. The overall pattern is about 80 out of 120 intending to use rock phosphate. Is this pattern about the same in each row of the tables (i.e. for farmers of different experience or gender)? This analysis can be seen as calculating a residual from the overall pattern for each row and making a judgment as to whether it is large.

## Percentages

Each of the counts in the tables given above can be expressed in percentage terms and this is often useful for readers to be able to draw conclusions. For the first objective of whether farmers would be prepared to continue to use fertilizer, we could state that:

'68% of the 121 farmers answered yes'.

Similarly the information in the two-way table (Table 6) entering the factor 'gender' can be given as shown below:

| | Yes | No | Total | Sample size (Farmers) |
|---|---|---|---|---|
| Female | 70% | 30% | 100% | 66 |
| Male | 65% | 35% | 100% | 55 |
| Total | 68% | 32% | 100% | 121 |

Table 6.

The percentages shown in this table are easier to interpret than the numbers in the corresponding table of counts. Percentages show the results on the same scale independently from the sample size, which was different for female and male farmers.

Presenting counts as percentages is thus a useful thing to do. However, we should point to the fact that when transforming to percentages, information gets lost. It is not entirely clear anymore from where the percentage calculation started. What constituted 100%? To compensate for this loss, the percentages have to be accompanied by additional information. Consider again the initial result at the start of this section, namely:

'68% of the 121 farmers answered yes'.

In this sentence we have given the percentage, but we have also added 'of the 121 farmers'. This phrase indicates the sample size, which gives the reader an indication of accuracy (68% can be obtained also with 41 out of 60 farmers or 820 out of 1210). It makes it explicit that 100% equals all farmers and not a subset. Extending this idea to the 2-way table (Table 6) given above, the results could be phrased as:

'The percentages were similar for females and males, with 70% of the 66 females giving a positive reply, compared to 65% of the 55 males.'

Indicating what constitutes 100% is also important when missing values or cases of non-response are part of the data. They reduce the effective sample size. This becomes particularly important when an observation is not relevant for a subset of the entire sample. For instance, questions about effects on crops after the fallow are relevant only for farmers that completed the entire fallow-crop cycle. This could be considerably less than the original 121 farmers.

Now change the question slightly (Table 7) from 'What percentage of farmers use rock P' to 'What percentage of land uses rock P?' (or, more precisely 'What percentage of land is farmed by farmers who use rock P?'). Now 100% is not all the 121 farmers but all the hectares they farm.

| Use rock P? | Count | Average farm size (ha) | Total farm size (ha) | % of farmers | % of land |
|---|---|---|---|---|---|
| Yes | 80 | 1.9 | 154.7 | 69% | 67% |
| No | 36 | 2.1 | 75.4 | 31% | 33% |
| | 116 | 2.0 | 230.1 | 100% | 100% |

Table 7.

69% of the farmers use rock P and they farm 67% of the land. In this case the two percentages are similar, as the average farm size for the two groups (1.9 ha and 2.1 ha) are very similar. We can think of other meaningful percentages, which can sound similar (e.g. % of fields with RP). Such percentages can differ considerably and it is important to distinguish carefully between them, usually by stating explicitly what constitutes 100%.

The difference in method of calculating % of land and % of farmers is worth looking at carefully. The % of farmers is calculated as:

% farmers = 100 (number responding yes)/(number responding yes + number responding no)

$$\%\,farmers = \frac{100\sum_{yes} 1}{\sum_{yes} 1 + \sum_{no} 1}$$

The % of land is similarly calculated as:

$$\%land = \frac{100\sum_{yes} a_i}{\sum_{yes} a_i + \sum_{no} a_i}$$, in which $a_i$ is the area of farm i.

The difference between them can now be seen that in the first case we count '1' for each farm, in the second we count '*a*'. Each farm is given a different importance or 'weight' according to its area. If there happened to be one huge farm of 1000 ha that would dominate the calculations. The idea of weighting observations in analysis arises in many different places and is important to understand.

## Counts and percentages at multiple levels

Analyzing the results of large on-farm trials very often involves combining observations done at different levels in a hierarchy of experimental units. The observations of the examples we used until now were all 'farm level' observations, for which farm or farmer is the experimental unit. The gender, past experience with fertilizer, or plans of a farmer, are measured in this study once per farm. On the other hand, information about the experience with a particular fallow species originates from the field within the farm where the species was planted. Several such fields could be situated on one farm. These field level observations constitute a lower level in the hierarchy. In the on-farm trial example, two separate questionnaires were conducted at these two levels. Combining information gathered at the farm level and at the field level causes difficulties, which we will refer to as multiple level complexities. Multiple levels complicate confirmatory analysis, as we will discuss in Session 8. For the descriptive analysis, we will have to be careful and explicit about units, what we are counting and 'What is 100%?' when calculating percentages.

Suppose we want to investigate the hypothesis that views on the effect of improved fallow on the subsequent crop yield, depend on the gender of the farmer. The gender of the farmer was observed at the farm level. The effect on yields was noted for each field, sometimes several times for each farm. If we look at the gender of the farmers of the 171 cropped fields and tabulate this with the effect the fallow had, we obtain the table (Table 8) on the next page.

| gender | negative | none | positive | unknown | |
|---|---|---|---|---|---|
| female | 16 | 1 | 59 | 23 | 99 |
| male | 4 | 6 | 44 | 18 | 72 |

Table 8.

Or, expressed as percentages

| gender | negative | none | positive | unknown | |
|---|---|---|---|---|---|
| female | 16% | 1% | 60% | 23% | 100% |
| male | 6% | 8% | 61% | 25% | 100% |

Table 9.



The results for men and women are very similar, except that when the effect was not positive, women tended to be more extreme in their opinion.

Lets rephrase the question to 'Do both sexes have the same opinion about the effect on crop yields'. This question is similar to the previous one and could be answered with the data in the above table. However, if we think carefully about the new question, the problem is situated on the farm level. We cannot claim that 16 out of 99 female farmers were negative about the technology, since we only asked 66 women this question. We could see for each farmer what opinion he or she had over all the fields in the farm. This implies that we come up with a category of 'mixed' opinion, to indicate a combination positive, none or negative within a farm. The results are given in Tables 10 and 11 below.

| gender | negative | none | positive | mixed | unknown | |
|---|---|---|---|---|---|---|
| female | 7 | 1 | 34 | 3 | 14 | 59 |
| male | 2 | 2 | 34 | 1 | 9 | 48 |

Table 10.

or

| gender | negative | none | positive | mixed | unknown | |
|---|---|---|---|---|---|---|
| female | 12 % | 2 % | 58 % | 5 % | 24 % | 100 % |
| male | 4 % | 4 % | 71 % | 2 % | 19 % | 100 % |

Table 11.

Men seem now to be more positive than women about effects on yields.

## Doing the exploratory analysis

We started out by saying that this exploratory analysis is often omitted from the analysis of a trial. By now you should be convinced that it is worth doing. How long it takes depends on the complexity of the data and of the analysis objectives. For many experiments it will be quick, once you have a clear idea of the objectives of analysis, your data are well organized and you are familiar with the software, the analyses need actually only take a few minutes! What will take longer is looking at the results and working out what they are telling you. The time taken need not add much to the total analysis time, yet it will be time very well spent.

Most researchers start off with the data in Excel (or a similar package) and move it to a statistics package such as GenStat for the analysis. Which environment is best for the exploratory analysis? As an unskilled GenStat user you will probably find it easier to do much of the exploratory analyses in Excel. Some of the facilities there are powerful and flexible, the cross-tabulations (pivot tables) and the way graphs of these can quickly be drawn, for example. As you become more familiar with GenStat or another statistics package you will probably find it easier to do the analyses in that environment.

Later in the course we will see that the whole analysis is iterative, it may be necessary to come back to the exploratory analysis after doing some formal analysis. It is then much easier to do everything in one computing environment, and it has to be the statistics package.

## Concluding comments

In our spacing trial example, we were able to see the 'pattern' in crop yield associated with the experimental treatments, by studying mean crop yield values averaged over the replicates in the study. We were also able to examine other patterns. For instance we saw that the tree size was affected by the width of the spacing, wider spacing being associated with larger trees. We also observed that there was a positive relationship between crop yield and tree size, particularly in the presence of hedgerows.

We were able to examine the 'residual', i.e. the variability, which was not explained by the treatment-related pattern, using scatterplots. We also suspected, and saw, that some of this residual variability could be explained by differences between the blocks in the experiment. We have therefore been able to deduce quite a lot about our treatments and their effects from this preliminary analysis.

What we have not been able to do though is look at all of the above together. We have only looked at 'slices' of the data, i.e. we separately looked at pattern due to treatment and the residual variability ignoring any effects of blocks, and then looked at the pattern due to blocks. We could not look at pattern due to blocks and to treatments together, and then inspect the resulting residual variability. To do this we need to take the analysis further and move on to more complex methods such as analysis of variance and modelling, which are discussed later in the following three sessions.

The examples were very different in terms of treatments, layout and measurement. The first experiment had a clear treatment structure and a very regular layout, with each treatment occurring once in each block. The measurements were on a continuous scale so familiar summaries such as means and variances are useful. In the second example the layout was very irregular, and many of the measurements were qualitative. The tools needed to complete the descriptive statistics are therefore different but its importance is not changed. The same is true when we move on to formal analysis. The purpose and steps in the analysis will be similar to the example above, but the complex layout and the qualitative measurements mean that some of the standard techniques (e.g. analysis of variance and regression) have to be modified and extended before they are useful.

Session 5

# Analysis of variance as a descriptive tool

*R. D. Stern, E. Allan, R. Coe*

## Introduction

The standard analysis of experimental data is called the 'Analysis of Variance' (ANOVA) and includes an ANOVA table. However, in Session 2 and 4 on the presentation of the results, to satisfy the objectives, you probably did not include an ANOVA table. What then is the role of the ANOVA table? We shall explore the answer to this question in this session.

Our aim in the analysis is to summarize the data in relation to the objectives. To do this we have to understand the sources of variation in the data. In the last session we considered various methods of data exploration. The main limitation in the analyses of Session 4 is that there is often a lot of structure in the data (blocks, and various treatment factors) and it is difficult to look at these different sources of variation together. Analysis of variance aims to do just this. The general ANOVA is introduced and in this session may be made more specific to meet particular objectives in Session 6.

There is no strict order to the tasks of either data exploration, general ANOVA or specific summaries. We will find that the ANOVA table can also help in our initial exploration of the data. We also find that the importance of some of the specific summaries, which correspond to the objectives, can be assessed using the ANOVA table.

In this session we consider the ANOVA table simply as a summary of the data. We are still in the domain of descriptive statistics. We will add the ideas of statistical inference (e.g. F-tests) in Session 6.

## An example

We will use an example dataset to describe some of the elements of an analysis of variance, and its use in descriptive analysis. The data we will use comes from the 'Upperstorey/ understorey' tree management trial that we have discussed in previous sessions. The actual treatment structure of this trial is fairly complicated, because not every spacing treatment occurs with every understorey hedgerow. For the purposes of this session we will take a subset

of the data, omitting the data for 1 m spacing. By removing this spacing, we are left with a factorial treatment structure with two factors. One treatment factor is spacing with three levels, and the other is understorey treatment, also with three levels. We will see later that it is convenient to think of the spacing as three levels of tree density of 0, 0.2 and 0.33 trees per metre.

We are left with a trial with 9 treatments described by factors 'density' and 'under', each of these has three levels. How can we analyse this data using ANOVA? Shown below in Figure 1 is the GenStat dialogue box for the General Analysis of Variance. Note that we have told GenStat the measurement being analysed, (crop yield labelled '**y**'), the treatments and something about layout (the blocks).

The treatments are specified as **density*under** . This is GenStat shorthand for **Density + Under + Density.Under** . This simply defines the pattern as having three parts, the density effect, the understorey effect and their interaction. The '+' symbol does not indicate things are added together in the usual sense.

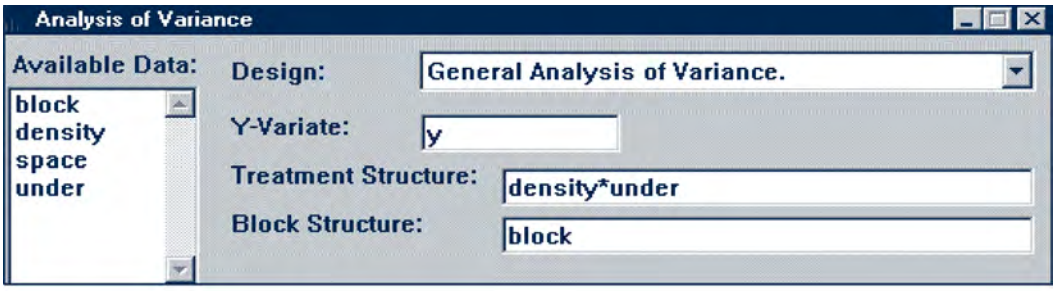



Figure 1. GenStat dialogue box for the general analysis of variance

In Table 1 below, we show some of the printout from GenStat for the analysis of variance for this experiment.

```
***** Analysis of variance *****

Variate: y

Source of variation          d.f.        s.s.        m.s.   v.r.

block stratum                   2      23.776      11.888   9.73

block.*Units* stratum
density                         2      38.247      19.123  15.66
under                           2     192.169      96.084  78.66
density.under                   4      13.431       3.358   2.75
Residual                       16      19.544       1.222

Total                          26     287.167
```

```
* MESSAGE: the following units have large residuals.

block BLOCKA *units* 5      -1.87  s.e. 0.85
```

```
***** Tables of means *****

Variate: y

Grand mean   22.09

 density      0.00         0.20      0.33
             23.04        22.81     20.41

  under        CAL          NAP        NO
             21.27        19.31      5.69

 density     under          CAL       NAP          NO
              0.00        22.17     19.77        27.20
              0.20        22.53     20.87        25.03
              0.33        19.10     17.30        24.83
```

Table 1. Printout from GenStat for analysis of variance (ANOVA)

There are three parts to the display above. The first is the ANOVA table. Then there is a message about one observation that has a large residual. Finally there are three tables of treatment means. You are already familiar, from the previous sessions, with the display of tables of means. So it is the first two elements of the display, namely the ANOVA table and the message about the residuals which are of most interest here.

The display above is not the default from GenStat. We have chosen not to give F probabilities in the ANOVA table, nor standard errors for the tables of means, because that is part of statistical inference and we concentrate here on descriptive statistics. Statistical inference will be discussed in Session 6.

Our general premise in constructing the ANOVA table is the idea that

data = pattern + residual

which you have met before. The **pattern** is the part of the structure of the data that we know about. Here we know about the blocks and the treatment factors. The residual is the remaining variation in the data. It is the 'noise' in the system.

Before assessing whether the ANOVA table is useful, check that you understand all elements of the table as follows:

- Each line in the ANOVA table corresponds to a possible **source of variation** of the data. In this experiment these lines correspond to the **layout** and **treatment factors** in the experiment and to the residual.
- Thus the **structure of the experiment** determines the structure of the ANOVA table.
- Now check the column giving the **degrees of freedom (d.f.)**. Do you understand this? (Hints: degrees of freedom are pieces of information. In the example there were 27 plots, so (n-1)=26 degrees of freedom in total.)
- Then there is the column giving the **sum of squares** of each component. Here, with the total sum of squares of 287.2, as much as 192.2 is explained by one of the treatment factors (under) and all except 19.5 is explained by the layout and treatment factors together.
- The final two columns in the ANOVA table above give the **mean squares (m.s.)** and the ratio of each component, relative to the **residual mean square (v.r.)**.

The total sum of squares reflects the fact that the smallest value is 14.6 g $m^{-2}$ and the largest is 28.3 g $m^{-2}$. The spread, i.e. the variability of the residuals is far less, and the largest is –1.87 and is noted in the summary above.

Now we try to assess whether the ANOVA table is useful. There are two aspects to pursue; the first is to continue in the critical examination of the data that was part of the data exploration considered in the last section. The second is in an assessment of the importance of the different tables of treatment means.

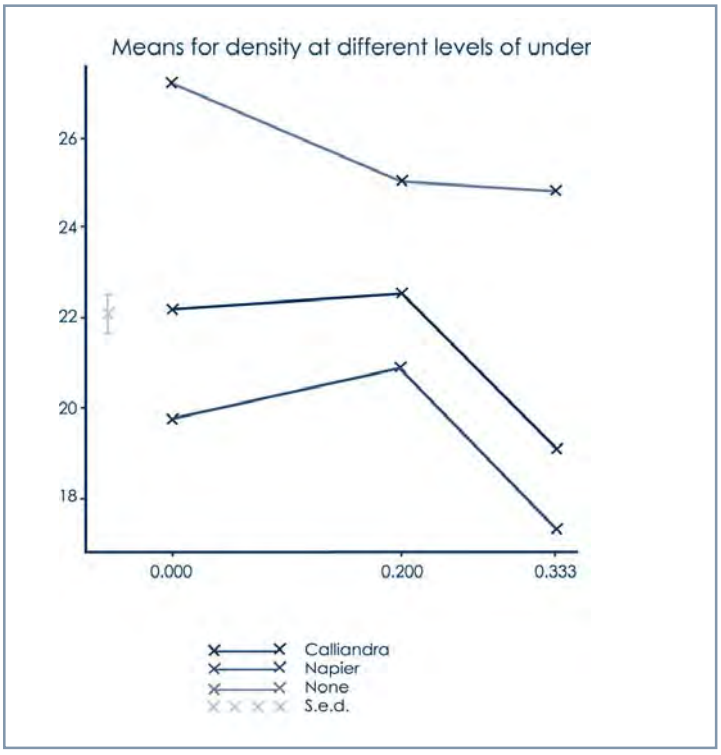
We start with the treatment means, because that aspect is simpler. There are three lines in the ANOVA table that correspond to the three tables of means. So we can use the ANOVA summary to assess the importance we should attach to each of the tables of means. Here we see that the most important is the table of means for the different levels of the understorey treatment,

because that factor explains 192 of the total of 287 in the sum of squares. For each degree of freedom (i.e. looking at the mean square), it is almost 80 times as large as the residual (v.r.=78).

The table of means confirms this. The means for the three types of hedgerow vary from 19.3 to 25.7 g $m^{-2}$.

The contribution to the sum of squares from the second factor, the tree density, is less and we see that the three means only vary from 20.4 to 23.0 g $m^{-2}$.

What, however of the more complicated table of the interactions? These means enable us to see the extent to which the effect of tree density depends on the hedgerow type. The means of this table vary greatly, from 17.3 to 27.2, but the ANOVA summary (v.r.) informs us that most of this variation is due to the main effects particularly the effect of hedgerow. So here, this table is probably not of great importance. Graph 1 below probably shows this table more clearly:




Graph 1.

Now we return to the role of ANOVA in the critical assessment of the data. One term of importance is the residual mean square, which has a value of 1.22 here. This is the residual variance, so the residual standard deviation, s = $\sqrt{1.22}$ = 1.1 g $m^{-2}$. This value is the sort of difference that we might record on any plot if, hypothetically, we were able to repeat the experiment. It is the deviation that an experimental value can have that is not part of the 'pattern', i.e. we cannot explain it by the blocks or treatments in our experiment.

If this value is small, then we can study the 'pattern' easily. Here we are particularly interested in studying the treatment effects, because the objectives all relate to some aspects of the treatments – that is why we did the experiment! If the 'noise', (the residual) is large it becomes harder to study the treatment effects.

We will now go back a step. Table 2 below shows the ANOVA table when we have omitted 'under' from the model. Now the residual variation is much larger. The means for the different levels of density remain as before, but the 'noise' is so large that the data do not indicate that these means are interestingly different. Remember that we saw above that interesting differences are those where the variance ratio (v.r.) is large.

```
***** Analysis of variance *****

Variate: y

Source of variation      d.f.      s.s.      m.s.     v.r.

block stratum               2     23.78     11.89     1.16

block.*Units* stratum
density                     2     38.25     19.12     1.87
Residual                   22    225.14     10.23

Total                      26    287.17
```

```
***** Tables of means *****

Variate: y

Grand mean 22.09

 density                0.00      0.20      0.33
                       23.04     22.81     20.41
```

Table 2. ANOVA table where 'under' had been omitted

From this example we can see that one important detail to check is whether we have included all terms in the 'model' or in the 'pattern' that we need.

If we find that our residual mean square is large, then a second possibility is that there is a problem in the data. We show this by 'mistyping' the first observation in Block B (0 spacing, NO hedgerow) unit 1, which was 28.3, as 14.5. Notice how the conclusions have changed. The residual mean square is now much larger (8.64) and there does not now seem to be a reason to look at the table of means for the different levels of tree density.

```
***** Analysis of variance *****

Variate: y

Source of variation  d.f.     s.s.     m.s.    v.r.

block stratum           2      1.696    0.848    0.10

block.*Units* stratum
density                 2     25.980   12.990    1.50
under                   2    106.916   53.458    6.19
density.under           4     26.311    6.578    0.76
Residual               16    138.224    8.639

Total                  26    299.127


* MESSAGE: the following units have large residuals.

block  BLOCKA  *units*1    4.92          s.e.  2.26
block  BLOCKB  *units*1   -8.39          s.e.  2.26
```

Table 3. Mistype in the first observation in Block B

Notice also that the message about large residuals has alerted us to the need to look critically at the observation where we made a 'deliberate error'. There the residual is stated to be –8.39 with a standard error of 2.26. From experience we normally find that when a residual is more than 3 to 4 times the standard error then there is something wrong. Here it is 3.7 times greater. The 'mistake' that we have made here is so extreme that it would probably have been detected earlier in the exploration of the data. But less extreme errors or problem observations, for example had 25.3 been typed instead of 28.3, it would not have been detected. Such errors only become clear once an attempt is made to include all terms that contribute to the pattern in the model and then examine the residuals.

The standard printout from GenStat merely notes the values of any residuals it considers extreme. It is often useful to look at all the residuals and you can instruct GenStat to save and print these residuals.

This is a 'text-book' example. When analysing real data sets it is often even more useful to display the residuals in their (randomized) field positions. Had this been the case above, we would have wondered whether the fact that the 2 largest residuals were adjacent (both are first plot of the block), indicated some pattern to be investigated further.

Another section of output from the ANOVA command, which GenStat can be requested to give (see Table 4), and many packages print automatically, is the coefficient of variation (c.v.). What is it useful for?

```
***** Stratum standard errors and coefficients of variation *****
Variate: y

Stratum         d.f.      s.e.        cv%

block              2     0.307        1.4
block.*Units*     16     2.939       13.6
```

Table 4. Coefficient of variation (c.v.)

The c.v. is simply the standard deviation (square root of the RMS) from the 'residual' line for each stratum in the ANOVA table divided by the overall mean for the data, written as a percentage.

$$\%C.V. = \frac{\sqrt{residual.m.s.}}{overall.mean} \times 100$$

Most packages only print this for the lowest stratum, GenStat however provides a %c.v. for each stratum.

So the c.v. is simply the residual variation scaled by the mean. The scaling means that it is 'dimensionless', the value will be the same whatever units the data are measured in.

There is no value of c.v. for an experiment that is 'acceptable' or 'correct', when inspected in isolation it is of no use to us. At best the c.v. can be compared with other trials of the same type. If you have enough experience in your field of research you may well know that, for example, the c.v. of annual fodder biomass production in simple management trials is about 20%. If you carried out a similar trial and you found that the c.v. was substantially larger than this, you would then know that your trial was much more variable than typical trials of the type, and you need to investigate possible causes.

Note that the c.v. will be high if the residual variation is high or the overall mean is low. If the mean gets close to zero the c.v. becomes arbitrarily large. It is therefore not a useful measure when the variate being analysed is something that can take zero or negative values.

In summary, the c.v. is sometimes useful when used as a diagnostic but is rarely a key value needed to meet the objectives.

## Extending the Analysis of Variance, given the objectives

We have seen that the analysis of variance as shown above can be utilized as part of the exploration of the data. This is well recognized, and the results, such as those shown above, are given by all statistical packages. Once the data have been examined critically the tables of means, such as are given earlier are sometimes all that is required to satisfy the objectives of the trial.

Usually though, it is useful to proceed further and how this is done depends on the objectives. In our 'Upperstorey/Understorey' example, the tree density factor is quantitative, hence the objectives are likely to relate more to the shape of the response curve than to the mean response at a particular density level. Look back at our initial graph of the data (Graph 1). From this plot, the response to the changing tree density treatment is not clear, but could be curved.

We can investigate this 'hunch' within the ANOVA, the results are shown below in Table 5.

***** Analysis of variance *****
Variate: y

| Source of variation | d.f. | s.s. | m.s. | v.r. |
|---|---|---|---|---|
| block stratum | 2 | 23.776 | 11.888 | 9.73 |
| block.*Units* stratum | | | | |
| density | 2 | 38.247 | 19.123 | 15.66 |
| Lin | 1 | 27.509 | 27.509 | 22.52 |
| Deviations | 1 | 10.738 | 10.738 | 8.79 |
| under | 2 | 192.169 | 96.084 | 78.66 |
| density.under | 4 | 13.431 | 3.358 | 2.75 |
| Lin.under | 2 | 0.320 | 0.160 | 0.13 |
| Deviations | 2 | 13.111 | 6.555 | 5.37 |
| Residual | 16 | 19.544 | 1.222 | |
| Total | 26 | 287.167 | | |

Table 5. Investigating a 'hunch' within ANOVA

The general message here is that the objectives of the trial will dictate how the treatment term in the ANOVA is to be broken down. One way is to consider what are called **contrasts.** The example in Table 5 above may be considered as an example of this type of breakdown, more details are given in GWIM Part 2, Regression. Sometimes the breakdown of the treatment effects is not into single degrees of freedom, but into further factors. For example, we may know that some treatments can be grouped together using additional criteria. In Session 6, Exercises, we

will look at another example, when the treatment structure suggests intuitive treatment comparisons, and we can use 'user-defined' contrasts in these situations. The analysis of variance table above suggests that the effect of changing tree density of crop yield cannot be described by a straight line.

To sum up, the ANOVA table provides a concise summary of the experimental data. It provides a breakdown of the different parts of the model and also indicates the magnitude of the residual. It is primarily a tool for the scientist to help in the analysis of the data rather than as part of the presentation of the results.

# Session 6 Ideas of simple inference

*R. D. Stern, E. Allan, R. Coe*

## Introduction

In this session we review the role of statistical inference in the analysis of experimental data. In our final presentation of the results, described in Session 9, we will find that statistical inference appears to play only a small role. We are already able to prepare the necessary tables and graphs, using the methods described in Sessions 2 to 5. The role of statistical inference often appears to just add an additional line to the tables, giving the standard errors. Many scientists and readers find it difficult to understand this additional line!

The ideas of statistical inference are not simple. We assume that the phrases 'statistical significance', 'standard error', 'confidence limits' and so on are familiar, even if their meaning is not entirely understood. We review the essential ideas in the first section of this lecture note and include a booklet entitled 'Review of basic ideas of inference' for those that need a more detailed refresher. If you are not clear of the difference between 'standard error' and 'standard deviation' then you need to read this booklet.

Statisticians are sometimes approached by scientists with the claim that their data are already analysed but a proper 'statistical' analysis is now needed, because the results are to be included in a conference proceedings, an article or as part of a thesis. The implication is that the 'statistical bits', the standard errors etc. that are described here, are irrelevant for the analysis. As you might expect, we do not agree with this interpretation. Indeed we do not recognize that there is a difference between an 'analysis' and a 'statistical analysis'. We will show in this session that the ideas of statistical inference can often play a central role in the analysis of (experimental) data.

In Sessions 2 to 4 you reviewed the presentation of the results that would be appropriate for the objectives of the trial. At the end of this session you will be able to decide whether the presentations that you would like to give are now appropriate. Perhaps a table that you proposed to give is actually too simple. For example, in the '*Lantana/Tithonia* mulch' trial, if the fertility level alters the effects of a type of mulch, then a simple table giving mean values for each mulch, might be misleading.

What is required instead is a table giving the effects of each mulch, at the different fertility levels. We actually designed this table in Session 2. The contrary might also be the case, namely that a complicated presentation is not required, and that a simple table will suffice.

In Session 5 the ANOVA indicated which tables were appropriate, by considering how much of the 'pattern' in the data could be attributed to each component of the experiment. The ideas of statistical inference build on this analysis of the important factors.

Once the choice of appropriate tables and graphs has been made it is then important to assess with what precision your conclusions can be made. This is also part of statistical inference. These ideas are introduced simply in this session. They also form a key part of the idea of 'statistical modelling', which will be introduced in Session 7.

## Concepts of statistical inference

Reference is made to the booklet on 'Review of basic ideas of statistics', which includes a description of the key concepts of statistical inference. They are as follows:

- We consider a (hypothetical) **population** from which we have a **sample**.
- We wish to use our sample to **estimate** or to **test** hypothesis concerning properties of the population.

Let us use a simple example to look at some of the terms used in statistical inference. Suppose that we have a sample of 50 maize farmers from a district in Zimbabwe, of whom 10 use a new maize variety. This is 20% of the farmers. We wish to estimate the percentage of all maize farmers in this district using the new variety. If we have a **random sample** then our value of 20% is also our estimate of the percentage of farmers in the district who use the new variety. We need a **measure of precision** of this estimate. Our value of 20% is more precise than an estimate based on 5 farmers (of whom 1 uses the new variety) but less precise than an estimate based on 5000 farmers. The measure of precision used here is the **standard error**. In this case, with 50 farmers, the standard error is about 6%. To interpret the standard error we can form a **confidence interval** for the population percentage. In this case the 95% confidence interval is about 20 ± 2 * (standard error) or 8% to 32%. If an experiment gives sufficiently precise estimates we are ready to take decisions. If not, then we may have to repeat an (adapted) experiment.

We may wish to **test a hypothesis**. For example, suppose FAO claims that half (i.e. 50%) of the farmers currently use the new variety. This is inconsistent with our data, because we think that the percentage is between 8% and 32%. So we reject the claim; the difference between the hypothesized and observed proportion can be said to be **statistically significant**. On the other hand, a claim by a research institute that a quarter of the farmers, (25%) use the new variety is consistent with our data. This difference is not significant.

We must distinguish between **statistical significance** and **practical importance**. If a result is not statistically significant, there might still be something of practical importance in the data. In the example above we might really want to be able to distinguish between 20% and 25% of the

farmers using the new variety. Our sample of 50 farmers does not allow us to make that distinction.

We will now examine what these ideas of statistical inference add to the analysis of experimental data. In the descriptive statistics that we have been using so far, we have normally calculated summary results like the mean yield for each variety, that relate to the objectives of our study. In statistical inference we can calculate these same results and use them as estimates of the means from some population of values.

## An example

In the example in Table 1 on the next page, we use the data from the **Central district** of the on-farm study 'Effect of *Tithonia diversifolia* and *Lantana camara* mulches on crop yields in farmers' fields', considered in earlier sessions. The output shown is roughly as provided from GenStat. Note that parts of the output relating to inference (e.g. F pr.) omitted in Session 5, are now included.

```
***** Analysis of variance *****
Variate: grain

Source of variation      d.f.         s.s.       m.s.      v.r. F pr.

farmer stratum             15     154.4270    10.2951          14.31

farmer.*Units* stratum
treat                       2      37.2980    18.6490     25.92 <.001
Residual                   30      21.5869     0.7196

Total                      47     213.3118

***** Tables of means *****
Variate: grain

Grand mean 3.80

 treat            Control        Lantana        Tithonia
                     2.76           3.74           4.91

*** Standard errors of means ***
                      Table    treat
                       rep.       16
                       d.f.       30
                     e.s.e.      0.212

*** Standard errors of differences of means ***
                      Table    treat
                       rep.       16
                       d.f.       30
                     s.e.d.      0.300

*** Least significant differences of means (5% level) ***
                      Table    treat
                       rep.       16
                       d.f.       30
                     l.s.d.      0.612
```

Table 1. ANOVA of the on-farm study 'Effect of *Tithonia diversifolia* and *Lantana camara* mulches on crop yields in farmers' fields'

What are your conclusions from the ANOVA table?

We have included three ways of giving the standard error. It can be presented for each mean (e.s.e.), or the difference between two means (s.e.d.), or the least significant difference (l.s.d.) can be given.

How is the standard error of the mean calculated? How are these three values related to each other? The equations for each of these measures, for this example, are given below.

e.s.e. = $\sqrt{\frac{residual.m.s.}{n}}$ where n is the number of replicates for the treatment

s.e.d. = $\sqrt{\frac{2 \times residual.m.s.}{n}}$

l.s.d. = t-value (at 5%, using residual d.f.) x s.e.d.
(N.B. When the residual d.f. is large the t-value will be approximately 2.)

These are the calculations we use for this experiment, which is balanced with equal number of replicates for each treatment. However, when the design does not have equal replication, the calculation of s.e.d. will need to be adjusted to take this into account. The s.e.d. for comparing treatment 1 versus 2 then becomes;

s.e.d. = $\sqrt{\frac{res.m.s}{n1} + \frac{res.m.s.}{n2}}$,

where n1=number of replicates for treatment 1, n2 = number of replicates for treatment 2. Chapter 3 of MCH contains more details on this topic.

You would not normally present more than one of these measures. In this session you should evaluate which is the most useful for your work. In Session 9 on the presentation of results, we will consider which measure is the most appropriate to use when writing reports.

The analysis in Table 2 below gives a slightly more complicated ANOVA table. It is of the same set of data as above, but using the data from both districts. We can also see the dialogue box we would have used in GenStat to produce this output. Note that we have included a location.treatment interaction term. This is because in addition to looking at the effect of the treatment we are also interested to see how this effect might differ between the Central and West districts. Check you understand all the output shown.

```
***** Analysis of variance *****

Variate: grain

Source of variation      d.f.      s.s.       m.s.       v.r.      F pr.

location stratum           1      7.4934     7.4934      0.99

location.farmer stratum   26    196.9269     7.5741      9.39

location.farmer.*Units* stratum
treat                      2     41.9281    20.9641     25.99      <.001
location.treat             2      3.5249     1.7624      2.19      0.123
Residual                  52     41.9403     0.8065

Total                     83    291.8135

* MESSAGE: the following units have large residuals.
location Central farmer 5 3.299  s.e. 1.531

location Central farmer 11  *units* 1     1.663   s.e. 0.707
location Central farmer 11  *units* 2    -1.861   s.e. 0.707
location West    farmer 12  *units* 1     2.524   s.e. 0.707
location West    farmer 12  *units* 3    -2.083   s.e. 0.707

***** Tables of means *****
Variate: grain

Grand mean 3.542

treat        Control          Lantana          Tithonia
               2.685            3.527            4.415

location treat      Control        Lantana        Tithonia
Central               2.497          3.477          4.653
    rep.                 16             16             16
West                  2.936          3.593          4.098
    rep.                 12             12             12

*** Standard errors of differences of means ***
Table          treat          location
                                 treat
rep.              28          unequal
d.f.              52               52
s.e.d.                        0.3666          min.rep
            0.2400            0.3430          max-min
                              0.3175          max.rep
```

Table 2. ANOVA table using the same set of data from Table 1, but using the data from both the Central and West districts

In the session on descriptive statistics you may have studied the two-way table of mean yields and concluded that the effect of the mulch was different in the two districts. However, in the ANOVA table the F-probability (F pr.) for the location.treat interaction is 0.123. If we take around 5% (i.e. 0.05) as giving reasonable evidence that an effect is present, we can conclude that there is not much evidence for an interaction. We have found little evidence to suggest that the effect of the mulch is different in the two districts. Let us take a more detailed look at the two-way table.

The s.e.d.'s reported for this table look more complex, so an explanation is needed. In this experiment the sample sizes (reps) are unequal in the two-way table of means for the treatments by location. This is because there were a different number of farmers in the two districts. As we mentioned previously, when the sample sizes are unequal then we need to adjust our calculation of the s.e.d. GenStat has done this for us and has produced three different s.e.d.'s for the table. It tells us which s.e.d. to use for each comparison. For example, we use 0.37 (min.rep) when comparing two means which both have the minimum number of reps (12), or to put it in another way, if we are comparing two treatments at the West location. When comparing two treatments at our Central location, we note that we have 16 replicates for each mean so we use the s.e.d. for max.rep (0.32). We will use the max-min s.e.d. of 0.34 when comparing two means which come from different locations.

The F-probability has 2 degrees of freedom and is calculated by looking at two treatment contrasts, such as the difference between *Tithonia* (T) and Control (C) and between *Lantana* and Control. Using our s.e.d. we can see that the difference between *Lantana* and Control is roughly the same at both locations. However, the effect of *Tithonia* compared to Control in the Central district (2.1 t ha$^{-1}$) is much greater than the difference in the West district (1.1 t ha$^{-1}$). Indeed, if we were to calculate an s.e.d. for this comparison we would see that the difference is significant. The new s.e.d. is calculated using the s.e.d. for max-rep and min-rep in the equation;

$$\text{s.e.d. [T-C (Central) – T-C (West)]} = \sqrt{s.e.d(\max.rep)^2 + s.e.d.(\min.rep)^2}$$
$$= 0.48 \text{ t ha}^{-1}$$

The F-probability 'masks' this significant effect because the *Lantana* – Control comparison at different locations is non-significant. We can see that there are times when our F-probability is not going to explain all of the information in the data.

When presenting the results of our analysis we need to decide which of our tables of means form the most useful summary. The non-significant interaction effect implies that we need only present the table of means for each treatment, along with its corresponding s.e.d. (0.24). However, if we want to further investigate the relationship between *Tithonia* and location, and highlight this information, then our two-way table of means is best.

## Looking critically at the data

The analysis above also indicates that some observations need to be examined critically. Two of the largest residuals come from Farmer 12 in the West district. The analysis in Table 3 below is a repeat of the one above, but this time we have omitted this one farmer.

```
***** Analysis of variance *****

Variate: grain

Source of variation       d.f.       s.s.       m.s.      v.r.      F pr.

location stratum             1      12.4629   12.4629     1.71

location.farmer stratum     25     181.9040    7.2762    12.11

location.farmer.*Units* stratum
treat                        2      33.7331   16.8666    28.07    <.001
location.treat               2       6.7427    3.3714     5.61    0.006
Residual                    50      30.0428    0.6009

Total                       80     264.8856

* MESSAGE: the following units have large residuals.

location Central farmer      5        3.299           s.e. 1.499

location Central farmer     11    *units* 1    1.662  s.e. 0.609
location Central farmer     11    *units* 2   -1.861  s.e. 0.609

***** Tables of means *****

Variate: grain

Grand mean 3.476

            treat          Control        Lantana      Tithonia
                             2.686          3.474         4.267

  location     treat       Control        Lantana      Tithonia
  Central                    2.430          3.410         4.587
         rep.                16             16            16
    West                     3.059          3.567         3.802
         rep.                11             11            11

*** Standard errors of differences of means ***

Table                        treat       location
                                            treat
rep.                        27                unequal
d.f.                        50                50
s.e.d.                                         0.3305     min.rep
                             0.2110            0.3036     max-min
                                               0.2741     max.rep
```

Table 3. A repeat of the analysis in Table 2, but omitting Farmer 12 in the West district (who had 2 of the largest residuals)

We can see that by removing Farmer 12 in the West district our F-probability for the interaction term now becomes significant (0.006). Note that the s.e.d. for the two-way table are now smaller because we have removed some of the outlying values, we now have more chance of identifying differences between treatments and locations.

We return, in a later session, to the subject of whether it is valid to omit observations as has been done above. Here our aim is primarily for you to use this example to be able to understand the basic topics of statistical inference.

As we are now doing statistical inference, it is also important to consider what assumptions are necessary for our inferences to be valid. This topic is considered in more detail in the next session, which is on modelling. One assumption that is necessary for the F-probabilities to be valid is that the data are normally distributed. A second assumption, the most important, is that the data are equally variable, whatever the treatment.

In the practical we briefly consider how to look at the validity of these assumptions. In Session 15 we will consider the types of remedial action that can be taken when the assumptions are not valid.



## Satisfying the objectives

Continue with the mulch trial and look at the data from the Central district only. The first objective of our study was stated as follows:

'To determine the effect of *Tithonia* and *Lantana* mulch on crop yield.'

Our table of means is as follows:

| Control | *Lantana* | *Tithonia* |
|---|---|---|
| 2.76 | 3.74 | 4.91 |

From our GenStat output we see that the standard error of the difference between any two of the means is 0.3. So, the mean effect of *Lantana*, compared to the Control is estimated to be 0.98 t $ha^{-1}$ of maize, with a standard error of 0.3 t $ha^{-1}$. Perhaps a confidence interval would be more clear.

A 95% confidence interval for the difference between two treatments is given as 95% C.I. = Difference in means ± l.s.d.

The l.s.d was given in the output as 0.61 and shows that the 95% confidence interval for the mean increase is 0.98 ± 0.61 t $ha^{-1}$, simply 0.37 to 1.59 t $ha^{-1}$.

The confidence interval is an interval for the **mean** increase. It does not mean that all farmers will have seen an increase within this range. On the contrary you probably saw, when doing the descriptive statistics that 3 of the 16 farmers had a lower yield with the *Lantana* mulch than without it.

Similarly the effect of *Tithonia* is estimated as (4.91 – 2.76) = 2.15 t $ha^{-1}$, with the same standard error and width of confidence interval as for *Lantana*. That is the 95% confidence interval for *Tithonia* – Control is 1.54 to 2.76 t $ha^{-1}$.

So, we have perhaps satisfied this first objective just by looking at the corresponding mean treatment differences, together with a measure of their precision. This is useful, but it is not the end of our analysis of this trial. Look at the overall objective for the trial, which is as follows:

'The overall objective of this trial is to find out if the good results obtained in station with *T. diversifolia* and *L. camara* will be confirmed in farmers' fields and conditions.'

The standard errors in our trial arise from the randomization of the treatments on the farmers' plots. The conclusions for the Central district apply to our 16 farmers in the year during which the measurements were made (1995). If we look above at the overall objective it is much more general in terms of 'farmers' fields and conditions'. In Session 14 we will look at the extent to which our conclusions can be given more generally than just the 16 farmers in the trial.

Notice that in satisfying the first objective above we have not used a 'significance test' nor have we stated the results as 'xxx is significant'. Many researchers would automatically test hypotheses with this type of data, particularly focusing on the 'null hypothesis' of no difference in mean grain yield between the three treatments. This would not be incorrect (although the null hypothesis that mulch has no effect on grain yield is scientifically foolish), it would just not help us to meet the objectives. There are two outcomes of such a test:

1. 'A significant difference' between treatments, in which case we have to go on to describe what that difference is.
2. 'No significant difference'. This case does NOT mean 'there are no differences between treatments' or 'mulch has no effect on mean yield'. It simply means that our experiment has not been able to detect a difference. In order to understand if this is because the actual effect of the mulch is very small or because the experiment was imprecise we still have to look at confidence intervals of mean treatment effect.

The overall trial objective of 'confirming the good results found on station' might be interpreted as testing the hypothesis that the mean yield increase, due to mulch used on-farm is the same as the mean yield increase observed on-station. However such a test would still have to be followed up with a look at confidence intervals to ensure that a 'non significant' result was not just due to doing a poor on-farm experiment.

## Breaking up the treatment effects

In this experiment there were just three treatments, so the analysis following the standard ANOVA was brief. In other trials the objectives imply a breakdown of the treatment effects into smaller components. We have previously discussed a breakdown of quantitative factors into linear and quadratic terms (Lecture note in Session 5). In the practical following this session we will look at a breakdown into 'user-defined' contrasts where we can tell GenStat which treatment comparisons we want to make. This topic is also covered in GWIM, which describes different ways of examining the treatment effects in Part 2, Analysis of Variance, Further Topics.

Some readers will be surprised that this session does not devote time to the subject of multiple comparison tests, for example Neuman Keuls, or Duncan's multiple range tests. They are mentioned briefly in the booklets on statistical inference and on presentation of the results. Our reason for omitting them is that they do not relate to any of the objectives of the case studies that we have included. This may be a bitter blow to some who are used to these tests as a standard routine. But there is so much to cover in this workshop that we hope you will eventually be pleased that some topics can be omitted in this way!

Session 7

# An introduction to statistical modelling

*E. Allan, R. Coe, R. D. Stern*

## Introduction

So far in the course we have seen exploratory data analysis and analysis of variance for simple situations. With the exploratory investigations we could look at the effect of one or two factors such as blocks and treatments on our data, and try to identify other possible patterns, but it was not possible to look at the effects of all sources of variation simultaneously. The sessions on analysis of variance explained how to look at the effects of blocks and treatment factors together, but only for simple data structures. What should we do with the designs that the ANOVA commands cannot handle?

Statistical modelling is a more general approach than the ANOVA for separating data into pattern+residual, quantifying and describing both. The main ideas of statistical modelling can be described with very simple examples, as we will do here. However these same ideas are the key to building models for many other cases, and will be used through the remainder of the course.

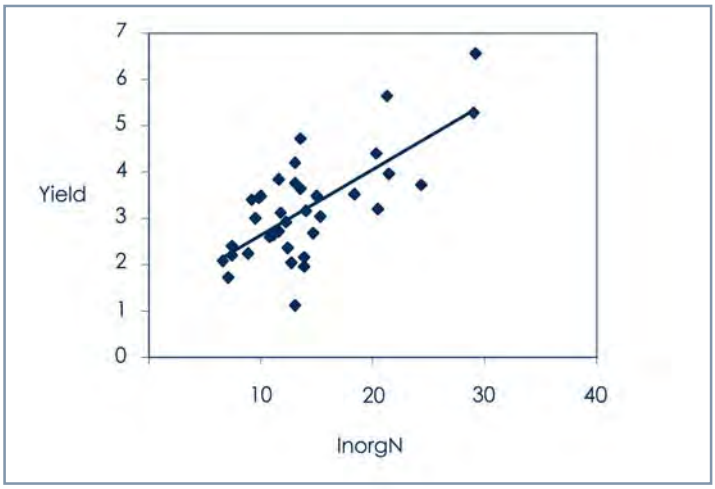
## A simple model

Our example here uses the trial 'Screening of suitable species for three-year fallow'. One of the objectives of this trial concerns the relationship between crop yield and the soil nitrogen and this is investigated here. As a start we consider the inorganic soil nitrogen (labelled INORGN) and data from the first experiment (1991) only.

Exploring the data is straightforward: we are in interested the way yield is affected by changes in inorganic N so it is natural to graph the data. The scatter diagram (Graph 1) below shows results much as expected, a clear relationship between yield and inorganic N, but with a fair amount of scatter. Yield clearly increases with increasing N, but for any given level of N there is a wide range of yields, maybe ±1 t $ha^{-1}$.

Thus we can see the data as
data = pattern + residual

This description is now formalized by defining a specific model. This means describing precisely, both components, the pattern and the residual.




Graph 1. Relationship between inorganic N and yield

The **pattern** could be described by many different curves, which are consistent with the data. For this data there is no reason not to use the simplest, a straight line. Thus we can write

yield = $a$ + $b$.INORGN + residual

The 'a' and 'b' are 'parameters', as yet unknown. The residual is as usual, the difference between the pattern and the data, and can be found once 'a' and 'b' are known.

The next step is to find suitable values for 'a' and 'b'. There are many ways of choosing these, but the most common (for various mathematical reasons) is to choose them as the values which give the 'best fit', defined as minimizing the sum of squared residuals or, equivalently, variance of the residuals.

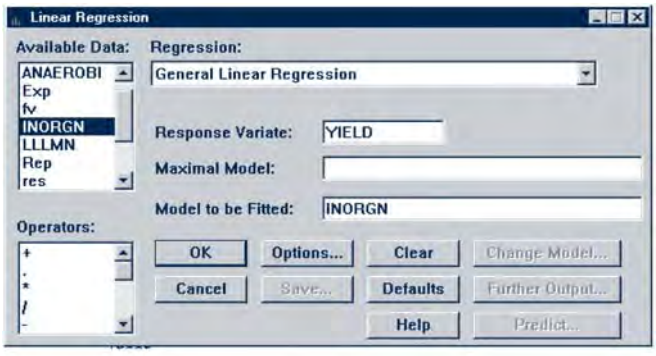



Figure 1. GenStat Linear Regression dialogue box

GenStat needs to know the 'response' or the left-hand side of the model and the pattern, in this case just INORGN. In Figure 1 above notice that the constant is included automatically.

GenStat's regression commands produce the following estimates:

| | estimate | s.e. |
|---|---|---|
| Constant | 1.228 | 0.366 |
| INORGN | 0.1419 | 0.0241 |

Table 1.

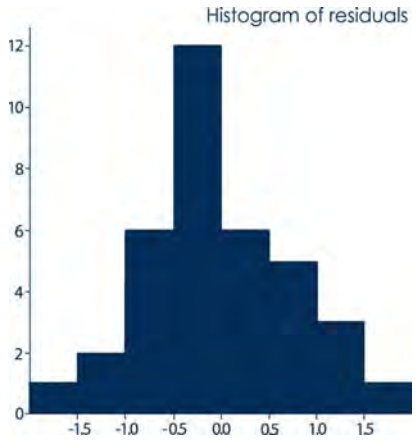
The 'a' value is estimated as 1.23 t $ha^{-1}$ and the 'b' as 0.14 t $ha^{-1}$ (mg $kg^{-1}$ )$^{-1}$. Thus the pattern part of the data is estimated to be

yield = 1.23 + (0.14 x INORGN)

The residuals can now be calculated (Table 2), and GenStat does this. We need to check that there is no more 'pattern' in the residuals. Plotting them in various ways does not reveal any more pattern. The residuals can be summarized by giving their variance or standard deviation (0.81 t $ha^{-1}$), and perhaps their distribution, revealed in a histogram (Graph 2).

| INORGN | YIELD | Pattern = 1.23+0.14INORGN | residual |
|---|---|---|---|
| 21.32 | 5.64 | 4.25 | 1.384 |
| 20.36 | 4.39 | 4.12 | 0.268 |
| 9.98 | 3.48 | 2.64 | 0.838 |
| 13.03 | 3.75 | 3.08 | 0.674 |
| 6.66 | 2.09 | 2.17 | -0.082 |
| 12.79 | 2.06 | 3.04 | -0.986 |
| 8.93 | 2.22 | 2.50 | -0.273 |
| 13.09 | 1.14 | 3.09 | -1.947 |
| 11.55 | 2.74 | 2.87 | -0.130 |
| 13.51 | 4.72 | 3.14 | 1.572 |
| 11.58 | 3.85 | 2.87 | 0.983 |
| 13.61 | 3.62 | 3.16 | 0.464 |
| 9.78 | 3.45 | 2.62 | 0.830 |
| 7.43 | 2.39 | 2.28 | 0.103 |
| 24.41 | 3.72 | 4.69 | -0.970 |
| 9.58 | 3.00 | 2.59 | 0.416 |
| 14.67 | 2.66 | 3.31 | -0.648 |
| 10.78 | 2.62 | 2.76 | -0.142 |
| 29.07 | 5.29 | 5.35 | -0.062 |
| 20.54 | 3.19 | 4.14 | -0.953 |
| 9.13 | 3.39 | 2.52 | 0.871 |
| 14.11 | 3.16 | 3.23 | -0.074 |
| 7.4 | 2.21 | 2.28 | -0.066 |
| 15.39 | 3.03 | 3.41 | -0.384 |
| 12.29 | 2.91 | 2.97 | -0.066 |
| 13.9 | 2.15 | 3.20 | -1.049 |
| 12.48 | 2.34 | 3.00 | -0.658 |
| 29.16 | 6.54 | 5.37 | 1.177 |
| 21.46 | 3.95 | 4.27 | -0.324 |
| 13.13 | 4.21 | 3.09 | 1.123 |
| 15.05 | 3.48 | 3.36 | 0.121 |
| 7.15 | 1.74 | 2.24 | -0.505 |
| 18.39 | 3.52 | 3.84 | -0.320 |
| 11.74 | 3.12 | 2.89 | 0.222 |
| 13.94 | 1.95 | 3.21 | -1.255 |
| 11.09 | 2.65 | 2.80 | -0.155 |

Table 2. Calculation of residuals




Graph 2. Histogram of residuals

This look at residuals suggests our description of the pattern is acceptable, so we can go ahead and interpret the model. For example, it suggests that a technology that can increase the inorganic N content of the soil by 10 mg/kg should give an increase in yield of about 10*0.14 = 1.4 t $ha^{-1}$. The 'a' parameter mathematically represents the yield at 0 inorganic N. However we have to be cautious when interpreting it as this, since this is an extrapolation beyond the range of the data. The minimum value observed was 6.6 mg/kg and we can not be sure that the straight line is a good representation for the relationship from here back to zero, indeed biological considerations would suggest this would be most unlikely.



## Further information about the model

The full output produced by the default settings of GenStat on fitting this model is shown in Table 3 below.

```
Response variate: YIELD
   Fitted terms: Constant, INORGN


*** Summary of analysis ***

             d.f.        s.s.          m.s.    v.r. F pr.
Regression      1       22.46       22.4579   34.63 <.001
Residual       34       22.05        0.6485
Total          35       44.51        1.2716

Percentage variance accounted for 49.0
Standard error of observations is estimated to be 0.805
* MESSAGE: The following units have large standardized residuals:
             Unit    Response      Residual
                8       1.138         -2.45
* MESSAGE: The following units have high leverage:
             Unit    Response      Leverage
               15       3.720         0.123
               19       5.290         0.228
               28       6.542         0.230

*** Estimates of parameters ***

                    estimate          s.e.      (34) t pr.
Constant               1.228         0.366     3.36  0.002
INORGN                 0.1419        0.0241    5.88  <.001
```

Table 3. Full output produced by the default settings of GenStat

The analysis of variance table can be interpreted much as in earlier examples. The total variation in yield (SS = 44.51) is broken down into two components; that explained by the regression line (SS = 22.64), and the variation in the residuals (22.05). Thus the 'pattern' accounts for about half the overall variation in the data and this is stated below the table as 49%. This is sometimes referred to as '$r^2$'.

The m.s. for residual is the variance of the residuals, so $\sqrt{0.6485} = 0.805$ is their standard deviation. The residuals appear to have roughly a normal distribution, so we expect most of them to lie in the range ±2 standard deviations = ±1.6. This is confirmed by looking at the graph or list of residuals. GenStat picks out one residual as being larger than expected (note GenStat residuals are scaled as (data-pattern)/(standard deviation) so differ from those calculated simply by subtraction). Looking at the graphs suggests that this just happens to be the largest (-ve) residual but it is not so much larger than the rest to cause suspicion.

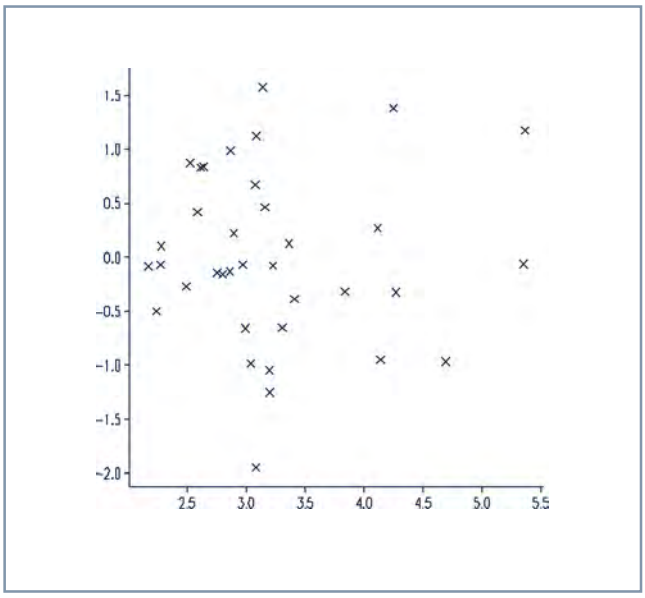
GenStat also picks out three points with 'High leverage'. These are points that could have a large effect on the estimates obtained and will not be discussed further here.

The output also includes information to help in formal inference. The most useful are standard errors of parameter estimates. The slope parameter estimate has standard error of 0.024. Thus the estimated increase in yield with an increase of 10 mg/kg inorganic N has a standard error of 0.24. An approximate 95% confidence interval for this increase is thus 1.4 ± 2*0.24 = 0.9 t $ha^{-1}$ to 1.9 t $ha^{-1}$. In the ANOVA table results of an F-test are presented and in the list of parameter estimates the results of t-tests are shown. The F-test and t-test for INORGN both test the null hypothesis of no linear relationship between yield and inorganic N (i.e. $b = 0$). It is clear from the graph that this null hypothesis is not of interest, so the test results do not add anything. The t-test for Constant tests the hypothesis that $a = 0$. Unless there is a specific scientific reason for looking at this value, it is also not of interest.

These inference results can be justified mathematically if certain conditions are true:

1. The model structure (a straight line plus independent residuals) is appropriate for the data.
2. The variance of the residuals is constant.
3. The residuals are roughly normally distributed.

Assumptions 1 and 3 have already been checked and seen to be reasonable. Assumption 2 is checked by looking at the variation of the residuals relative to anything that might change them. A common problem that occurs is when the variation of residuals increases with the mean, and this can be checked by plotting residuals vs 'fitted values' (i.e. the values on the fitted line for any value of inorganic N). Graph 3, shown below, does not reveal any strong pattern in variance across the range of means.




Graph 3. Residuals vs fitted values

Note that GenStat can produce these various 'diagnostic plots' using the 'Further Output' button of the Regression dialogue.

## Steps in statistical modelling

The steps that have been followed in the previous section are general for any statistical modelling and are therefore worth emphasizing. They are:

1. Explore the data to detect patterns and relationships.
2. Choose a possible model, based on the patterns seen, knowledge of the design that generated the data, and any information about the underlying scientific processes involved.
3. 'Fit' the model (estimate unknown parameters).
4. Check that the model is fitting well. Go back to Step 2 if not.
5. Interpret the fitted model to satisfy the analysis objectives.

## A model with factors

Yield data from the experiment 'Screening of suitable species for three-year fallow' was analysed in the previous section by looking at its relationship with inorganic soil N. However there is another way of looking at this data, corresponding to another of the objectives; determining which fallows give good yields.

The design of this trial is 'orthogonal'; each block has every treatment once. Differences between treatment means are useful for comparing treatments, since block differences cancel

out. An ANOVA with the pattern described by replicate and treatment effects is the sensible way to analyse the data.

Here there were 9 treatments and 4 replicates. The standard way to analyse these data is as a two-way ANOVA with blocks and treatments, and to summarize the results using treatment mean values, treatment differences and standard errors. These are presented in Table 4 below, only so that they can be compared with the results from the general linear model which we will look at next.

```
***** Analysis of variance *****

Variate: YIELD

Source of variation     d.f.          s.s.          m.s.         v.r.    F pr.

Rep stratum                3        1.0783        0.3594         1.53

Rep.*Units* stratum
Treat                      8       37.7811        4.7226        20.07    <.001
Residual                  24        5.6463        0.2353

Total                     35       44.5058


* MESSAGE: the following units have large residuals.

Rep 1  *units* 6     -0.847  s.e. 0.396
Rep 2  *units* 1     -0.934  s.e. 0.396


***** Tables of means *****

Variate: YIELD

Grand mean 3.232

  Treat     1 S.sesban     10 C.calo    2 G.sepium    3 L.leuco     4 F.congesta
               5.546          2.585         3.844         3.678          3.459

  Treat     5 C.siamea    6 groundnut  7 nat.fallow     8 M.only
               2.106          3.080         2.812         1.975


*** Standard errors of differences of means ***

Table       Treat
rep.            4
d.f.           24
s.e.d.     0.3430
```

Table 4. Two-way ANOVA with blocks and treatments

Now we see how the same analysis can be carried out, by fitting a model to the data.

First, what is the form of the model to be fitted? Exploring the data, for example by boxplots, suggests there are certainly treatment differences. The design of the trial allows for block (replicate) differences so we should include them as well. We will assume additive block and treatment effects, something which has not been explicitly justified, but is the same assumption as is used in the ANOVA approach.
The model can be written down as

Yield = constant + replicate effect + treatment effect + residual

(i.e. data = pattern + residual)

Note that the formulation of this model is different from the previous regression model. The term 'treatment effect' means that the yield is increased or decreased by a fixed amount depending on which treatment was applied. The term thus represents a series of quantities describing the effect of each different treatment. It does not represent a slope of a regression line. The same is true of the 'replicate effect'. This can be represented as

$$Yield_{ij} = constant + rep_i + treat_j + residual$$

where $Yield_{ij}$ is the crop yield for the i'th replicate of treatment j, $rep_i$ are 4 replicate effects and $treat_j$ are 9 treatment effects. The dialogue is shown in Figure 2 below.

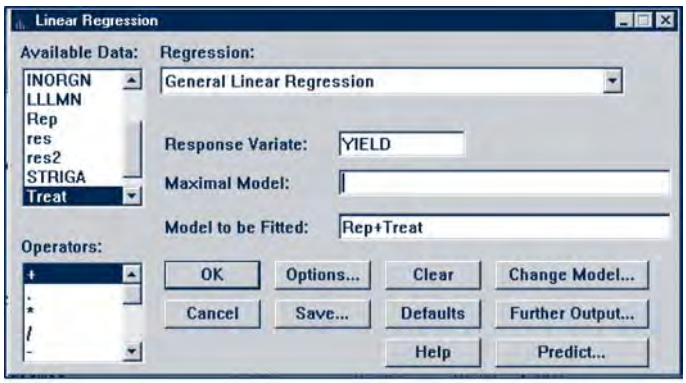



Figure 2.

The output from fitting the model in GenStat is shown in Table 5 below.

```
***** Regression Analysis *****

Response variate: YIELD
  Fitted terms: Constant + Rep + Treat


*** Summary of analysis ***

            d.f.     s.s.     m.s.     v.r.    F pr.
Regression  11     38.859    3.5327   15.02   <.001
Residual    24      5.646    0.2353
Total       35     44.506    1.2716

Percentage variance accounted for 81.5
Standard error of observations is estimated to be 0.485
* MESSAGE: The following units have large standardized residuals:
            Unit      Response  Residual
            6         2.057     -2.14
            10        4.716     -2.36

***** Regression Analysis *****

*** Accumulated analysis of variance ***

Change      d.f.        s.s.      m.s.      v.r.
+ Rep          3        1.0783    0.3594    1.53
+ Treat        8       37.7811    4.7226   20.07
Residual      24        5.6463    0.2353

Total         35       44.5058    1.2716
```

Table 5. The output from fitting the model in GenStat

The summary of analysis table just tells us the total amount of variation in the yield, which can be explained by both the replicates and treatment effects.

The accumulated analysis of variance table is the same as the one produced earlier where we calculated an analysis of variance using the ANOVA commands. Note, though, the way it is written here. The sources of variation are written as +Rep and +Treat. What does this mean? To understand this we introduce briefly the concept of model building, which is what makes modelling such a flexible analysis tool.

In constructing the above ANOVA table we have taken the total variation in the yield, and looked to see how much of that variation was due to the fact that plots were in different replicates. Here we see that 1.08 of the total sum of squares (44.51) was due to the differences between replicates. This left 43.43. Of this remaining unexplained variation we then looked to see how much could be explained by the different effects of treatment. It turned out that 37.78 could be attributed to the treatments, leaving 5.65 unexplained.

Would it have made a difference if we had fitted treatment effects first, and then looked to see how much of the remaining variation could then be explained by replicates? The answer to this question is no, not in this case. When data structures are orthogonal, the order of fitting is unimportant. The sums of squares are the same no matter what order the terms are fitted. (The results are not reproduced here, the reader is left to do this on his own if so interested.) This is not so when data structures are non-orthogonal, then the order of fitting does matter and consideration needs to be given as to how that is done.

So far the output is similar to the previous ANOVA analysis. From now on it will be different. We now get parameter estimates (Table 6), rather than treatment means. These correspond to the terms fitted in the model.

*** Estimates of parameters ***

| | estimate | s.e. | t(24) | t pr. |
|---|---|---|---|---|
| Constant | 5.370 | 0.280 | 19.18 | <.001 |
| Rep 2 | 0.280 | 0.229 | 1.23 | 0.232 |
| Rep 3 | 0.018 | 0.229 | 0.08 | 0.937 |
| Rep 4 | 0.406 | 0.229 | 1.78 | 0.088 |
| Treat 10 C.calo | -2.961 | 0.343 | -8.63 | <.001 |
| Treat 2 G.sepium | -1.702 | 0.343 | -4.96 | <.001 |
| Treat 3 L.leuco | -1.868 | 0.343 | -5.45 | <.001 |
| Treat 4 F.congesta | -2.087 | 0.343 | -6.08 | <.001 |
| Treat 5 C.siamea | -3.440 | 0.343 | -10.03 | <.001 |
| Treat 6 groundnut | -2.466 | 0.343 | -7.19 | <.001 |
| Treat 7 nat.fallow | -2.734 | 0.343 | -7.97 | <.001 |
| Treat 8 M.only | -3.571 | 0.343 | -10.41 | <.001 |

Table 6. Parameter estimates

Here we see a constant parameter, three Rep parameters and eight Treat parameters. There is no estimate for Rep 1 or Treat 1. These values are not printed as they are fixed to be zero.

These parameter estimates can be used to estimate, or predict, what yield to expect for each treatment in each replicate, i.e. using the model pattern

yield = constant + replicate + treatment

The estimated yields for the treatments in each replicate are shown in Table 7 below:

| | | | | Predicted yield |
|---|---|---|---|---|
| Treatment 1 | Replicate 1 | 5.370 | | = 5.370 |
| Treatment 1 | Replicate 2 | 5.370 + 0.280 | | = 5.650 |
| Treatment 1 | Replicate 3 | 5.370 + 0.018 | | = 5.450 |
| Treatment 1 | Replicate 4 | 5.370 + 0.406 | | = 5.776 |
| Treatment 2 | Replicate 1 | 5.370 | - 1.702 | = 3.668 |
| Treatment 2 | Replicate 2 | 5.370 + 0.280 | - 1.702 | = 3.948 |

Table 7. Estimated yields calculated from parameter estimates

But what has happened to our mean values? When presenting results we normally use treatment mean values. With a general linear model it is possible to construct mean values using the parameter estimates, most software packages will calculate them for you automatically. In GenStat they can be produced using the PREDICT command.

For the nine treatments in our example they are (Table 8):

```
Response variate: YIELD
              Prediction  S.e.
    Treat
 1 S.sesban     5.546      0.243
10 C.calo       2.585      0.243
 2 G.sepium     3.844      0.243
 3 L.leuco      3.678      0.243
 4 F.congesta   3.459      0.243
 5 C.siamea     2.106      0.243
 6 groundnut    3.080      0.243
 7 nat.fallow   2.812      0.243
 8 M.only       1.975      0.243
```

Table 8.

These mean values are sometimes called 'adjusted mean values'. This is because they are averaged over the levels of the other factor or factors in the model; in this example this is only the replicates. Referring back to the parameter estimates and the estimated treatment yields for each replicate we can see how these mean values have been calculated.

Note too that in this example, the mean yields (and their standard errors) are the same as the ordinary treatment means (and standard errors). This only happens when we have orthogonal data. If the data were non-orthogonal the adjusted means would be different from the ordinary means.

Consider now a comparison of two treatments. Say we were interested in the difference in yield between treatments 10 and 1. How do we estimate this difference? One way is to look at the difference between the adjusted means i.e. 2.585-5.546 giving us an estimate of -2.961 t $ha^{-1.}$

Alternatively we can look at the way the adjusted means are constructed from the parameter estimates. The mean for treatments 1 and 10 are:

constant + (Rep1 + Rep 2 + Rep 3 + Rep 4)/4 + Treat 1
constant + (Rep1 + Rep 2 + Rep 3 + Rep 4)/4 + Treat 10

Hence the difference in means is just:

Treat 10 - Treat 1

The parameter estimate for Treatment 1 is zero, so the difference is just the parameter labelled Treat 10 = -2.961 again. Hence the estimate Treat 10 is also the estimate of the difference between treatments 1 and 10. Similarly Treat 2 estimates the difference between treatments 1 and 2, etc. The difference between treatments 10 and 2 would be estimated by (Treat 10 - Treat 2).



*** Estimates of parameters ***

| | estimate | s.e. | t(24) t pr. |
|---|---|---|---|
| Treat 10 C.calo | -2.961 | 0.343 | -8.63 <.001 |

Table 9. Estimate of a parameter

The parameter estimate line for Treat 10 (Table 9) gives us quite a lot of information about the comparison between treatments 1 and 10. Not only does it give us the estimate of the difference, it also gives us the standard error of this difference i.e. 0.343, and a t-test for testing the hypothesis that there is no difference between these two treatments. [ t = -8.63, compared against the t-distribution with 24 d.f., is highly significant $p<0.001$.] The equivalent information for (Treat 10 - Treat 2) is a little harder to obtain. However GenStat can be asked to calculate it.

## Analysing other designs

The previous section introduced a harder way of doing something we could do simply before! What is the point? A little experience shows us that the ANOVA commands often fail. Whenever the design is 'incomplete' in some way, for example, we do not have every treatment occurring equally often in each block, GenStat (and other packages) give a message such as:

```
Design unbalanced - cannot be analysed by ANOVA
Model term Treat (non-orthogonal to term Rep) is unbalanced, in the Rep.*Units*
stratum.
```

However the modelling approach works for these examples in exactly the same way as before. In the following example (Table 10) the data from the first year of the 'Screening of suitable species for three-year fallow' trial are analysed again but this time the second row of data (Treatment 2 from replicate 1) has been removed. The output from fitting the model has of course changed, but can be interpreted in exactly the same way as before.

```
 Response variate: YIELD
   Fitted terms: Constant + Rep + Treat


*** Summary of analysis ***

               d.f.        s.s.        m.s.      v.r. F pr.
Regression       11      38.263      3.4784     16.41 <.001
Residual         23       4.875      0.2119
Total            34      43.137      1.2687

Percentage variance accounted for 83.3
Standard error of observations is estimated to be 0.460
* MESSAGE: The following units have large standardized residuals:
     Unit   Response  Residual
     5        2.057      -2.03
     9        4.716      -2.57
```

```
*** Accumulated analysis of variance ***

Change    d.f.       s.s.       m.s.     v.r.
+ Rep        3     1.6990     0.5663     2.67
+ Treat      8    36.5637     4.5705    21.56
Residual    23     4.8746     0.2119
Total       34    43.1374     1.2687

*** Estimates of parameters ***

                  estimate      s.e.      t(23)      t pr.
Constant             5.280     0.270      19.56      <.001
Rep 2                0.400     0.226       1.77      0.090
Rep 3                0.138     0.226       0.61      0.548
Rep 4                0.526     0.226       2.33      0.029
Treat 10 C.calo     -2.961     0.326      -9.10      <.001
Treat 2 G.sepium    -1.971     0.355      -5.56      <.001
Treat 3 L.leuco     -1.868     0.326      -5.74      <.001
Treat 4 F.congesta  -2.087     0.326      -6.41      <.001
Treat 5 C.siamea    -3.440     0.326     -10.57      <.001
Treat 6 groundnut   -2.466     0.326      -7.57      <.001
Treat 7 nat.fallow  -2.734     0.326      -8.40      <.001
Treat 8 M.only      -3.571     0.326     -10.97      <.001
```

Table 10. Data from 'Screening of suitable species for three-year fallow trial with the second row of data (Treatment 2 from replicate 1) removed

Notice that the s.e. of the Treat 2 parameter is larger than the others. This should be expected as we have less information now about treatment 2 than about the other treatments.

## One further model

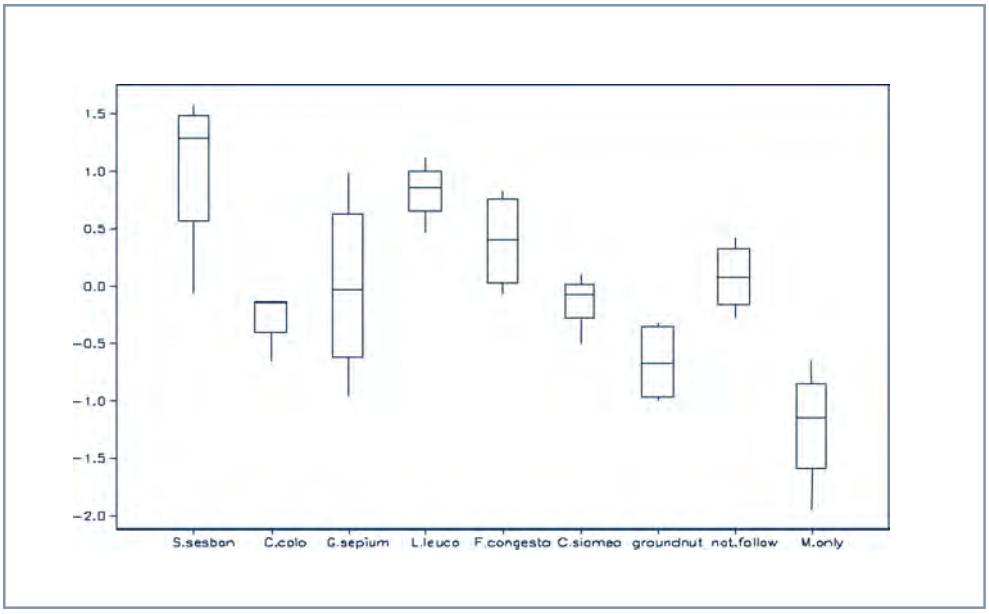
In the first section we fitted a model with a variate on the right-hand side of the equation to explain the relationship between a response and a continually varying 'explanatory' variable. In the second example we fitted a model with factors on the right-hand side in order to estimate difference between factor levels, in this case treatments. The modelling framework used is so general, that we can combine both these. An example is shown in this section.

Still using the 'Screening of suitable species for three-year fallow' trial, consider the following question. The yield is strongly determined by the inorganic N in the soil. The yield also varies between different treatments. Treatments can change the level of inorganic N. But does the change in inorganic N account for all the effect of treatment on yield, or are the treatments having some effect in addition to that due to increasing inorganic N?

To investigate this question we start with the simple regression analysis, fitting

$$yield = a + b.INORG$$

The residuals from this model are variations in yield not explained by a simple straight-line relationship with inorganic N. A boxplot of these by treatment suggests that there are large differences. The *Sesbania* plots (trt 1), for example, have positive residuals (higher yield than expected from the model) whereas the maize-only plots (trt 8) have negative residuals as shown in Graph 4.




Graph 4. Residuals after fitting INORG by treatment

Hence we need a model that accounts for this additional 'pattern' that was previously not noticed. The dialogue (Figure 3) below shows how a model with a linear effect of inorganic N and additive treatment effects can be fitted.


Figure 3.

*** Accumulated analysis of variance ***

| Change | d.f. | s.s. | m.s. | v.r. |
|---|---|---|---|---|
| + INORGN | 1 | 22.4579 | 22.4579 | 121.20 |
| + Treat | 8 | 17.2300 | 2.1537 | 11.62 |
| Residual | 26 | 4.8179 | 0.1853 | |
| Total | 35 | 44.5058 | 1.2716 | |

Table 11. Accumulated analysis of variance

This ANOVA (Table 11) shows that a substantial part of the residual variation from the simple regression is actually explained as treatment variation. There is very strong evidence that the treatments have some effect on yield, which is not accounted for by change in the inorganic N level. The parameter estimates in Table 12 show what these effects are:

*** Estimates of parameters ***

| | estimate | s.e. |
|---|---|---|
| Constant | 3.826 | 0.578 |
| INORGN | 0.0739 | 0.0231 |
| Treat 10 C.calo | -2.089 | 0.408 |
| Treat 2 G.sepium | -1.349 | 0.324 |
| Treat 3 L.leuco | -0.995 | 0.408 |
| Treat 4 F.congesta | -1.327 | 0.386 |
| Treat 5 C.siamea | -2.249 | 0.480 |
| Treat 6 groundnut | -2.057 | 0.330 |
| Treat 7 nat.fallow | -1.801 | 0.421 |
| Treat 8 M.only | -2.878 | 0.373 |

Table 12. Parameter estimates

We see that for a fixed level of inorganic N the yield following a *Sesbania* fallow exceeds that of a natural fallow by (Treat 1 - Treat 7) = (0 - -1.801) = 1.801 t $ha^{-1}$ with a standard error 0.421.

The model

yield = constant + b.INORG + Treat

is identical to the model

yield = constant + Treat + b.INORG

However the accumulated analysis of variance produced when fitting the two (Table 13) are different.

```
*** Accumulated analysis of variance ***

Change     d.f.       s.s.      m.s.      v.r.
+ Treat       8    37.7811    4.7226     25.49
+ INORGN      1     1.9067    1.9067     10.29
Residual     26     4.8179    0.1853

Total        35    44.5058    1.2716
```

Table 13. Accumulated ANOVA

Remember how the earlier ANOVA was found: identify how much variation is explained by the first term (this time Treatment) and then how much additional variation is explained by the second term (now inorganic N). Thus the line in the ANOVA table for INORGN now shows how much variation can be attributed to changes in inorganic N once we have accounted for differences between the treatments. It is not surprising this is small, since there is not much variation between levels of inorganic N within treatments. The previous ANOVA showed how much variation is due to INORGN when treatments are ignored. These two quantities are clearly different. The two ANOVA tables are both correct but address different questions and are suitable for different analysis objectives.

There are further steps we could not take with this modelling. For example, we have ignored blocks. Should they also be included in the model? Can we explain the treatment effects using, in addition to the inorganic N, any of the other variables that were measured? No new ideas are needed to pursue these objectives.

## Assumptions underlying the model

As with earlier analyses, there are some assumptions underlying this modelling approach. There are two parts to the model, the pattern and the residual components. The assumptions made about both parts can be checked.

The pattern part of the model contains effects of continuous variates and/or factors specifying a series of levels. Graphs can be used to check that a simple relationship is a straight line. Non-linear forms will require a different model. If there are several variates in the pattern, other methods still allow us to check whether the relationship is a straight line. Likewise there are some graphical methods for checking if the block+treatment form of the model is appropriate. Notice that much of the important information about pattern will have been detected during the exploratory analysis of the data.

The assumptions made about the residual part of the model should be checked. They play an important role in the statistical inference. The key ones are:

- Residuals are independent (knowing one residual will not tell you anything about the next). Check by looking at the residuals in field (or perhaps time) order.

- Residuals have a constant variance. Check by looking at the variance of residuals for different groupings of the data (e.g. treatment groups). Since one of the commonest ways this assumption is violated is for the variance to increase with the mean value, try plotting the residuals against the predicted value. GenStat produces these residual plots automatically.

- Residuals have roughly a normal distribution. Look at a histogram or (better) a probability plot.

- There should be no 'pattern' left in the residuals. Any systematic pattern is a sign that something is missing from the first part of the model.

## ANOVA or model fitting?

We have seen that the ANOVA and model fitting approaches to analysis of data from an experiment are the same. Yet most statistical software has two different facilities. Why? The ANOVA approach works for designs, which have certain properties, this means that the calculations can be done in a particularly simple way, which also has an impact on how easily results can be displayed. The software writers exploit this wherever possible. The table (Table 14) below summarizes differences between the two approaches. Looking down the table will probably convince you that it is worth using the ANOVA approach whenever your design allows it. Note that some of these differences in the approaches will be explained in later sessions.

| | ANOVA approach | Modelling/Regression approach |
|---|---|---|
| GenStat commands | BLOCKS TREATMENTS ANOVA | MODEL… FIT… |
| Scope | Any 'orthogonal' design, including split-plot type designs. | Works for any design with a single error term. There may be problems with experiments with many factor/levels. |
| Presentation of results | Clear, straightforward ANOVA table, means and s.e.d.'s. | Involves extra work to get tables of treatment effects and s.e.d.'s. |
| Use of contrasts | Straight-forward for standard or custom contrasts. | Involves extra work. |
| Adding covariates | Limited covariate models handled straightforwardly. | Any form of covariate model handled, but results may be complex to interpret. |
| Other data types (counts, proportions,…) | Does not generalize to other data types. Approximate methods by transformation. | Leads naturally to models for other data types. |
| Missing values | Approximate methods good for a few missing values. | Exact method is always valid. |

Table 14.

## Specifying models

So far the models fitted have been of the form

**response = constant + effects + residual**.

When using GenStat's Linear Regression facility we specify both the response variable and the effects. The constant is automatically included. The residual is automatically included. The inference results given are also based on the assumptions listed earlier.

The effects are specified as a list of terms separated by '+' symbols.

If the term is a **variate**, GenStat fits a regression, i.e. includes in the model an effect 'b.variate' where b is a parameter to estimate.

If the term is a **factor,** GenStat fits a series of constants, with a separate parameter to estimate for each level of the factor.

# Session 8 An introduction to multiple levels

*E. Allan, R. D. Stern, R. Coe*

## Introduction

Scientists who remember experiencing difficulties in their analyses prior to this training workshop and who have followed the previous sessions may wonder whether their difficulties are over. We hope so. However, this session introduces another topic that often causes complications. Data are often available at more than one level. The topic of multiple level data is often poorly understood. In this session we introduce the subject and show how, despite the complication, the analysis can often be handled in a relatively straightforward manner. We will also return to this subject in the second part of this course, in Sessions 11, 13 and 14.

To give hope to scientists we would claim that an understanding of the topic of multiple levels together with the subjects introduced so far, provides the necessary basic tools for a standard data analysis. We are therefore reaching the end of Part 1 of this workshop material.

A hierarchical, or nested, structure is common in data from experiments. The Randomized Complete Block Design (RCBD) is a simple example. We can even think of this as three levels. There is the experiment level. Within the experiment are blocks and within the blocks are plots. The analysis is usually simple, because the treatments are applied to the plots and the measurements are also taken on the plots. The analysis is therefore at a single 'level', namely the plot level. In such simple cases we may also have some measurements at the 'experiment' level, such as climatic data and soil type, date of planting and so on. These values are reported but there is no formal analysis at this level, because we have just a single experiment.

This is not always the case. In some studies there may be a whole series of experiments and we might then be interested in a multi-site analysis, which would involve 2 levels; the plots within each site, and the sites themselves. In Session 13 we will look at an example of this type of analysis using data from an on-farm trial where treatments were applied to plots but characteristics of the farms were measured at the farm level.

It is common for **measurements** to be taken at a level different from that to which treatments were applied. Here we will look at some of the ideas underlying measurements made at different levels, using an example of treatments applied to 4-tree plots and measurements made on the individual trees.

The other two main components of an experiment (**treatments** and **layout**) can also involve multiple levels. A common situation of multiple treatment levels is where the treatments are applied at more than one level as in split-plot experiments. We will discuss multiple levels of both the treatment and layout factors in this session. Session 11 contains more detail on layout multiple levels.

## Treatments at different levels

The split-plot experiment is an example of a design where there are treatments at more than one level. We illustrate the structure and the appropriate analysis using one of our example data sets. It is the 'Influence of improved fallows on soil P fractions on-farm' trial.

Example: The experiment looks at the effect of 4 different fallow treatments, in combination with N applications, on maize yield. There were a total of 9 farms (reps) with 4 fallow treatments (main-plots) within which there were 2 different nitrogen treatments, 'N' and 'no N' (sub-plots).

Nitrogen treatments were randomly allocated to the sub-plots and fallow treatments were allocated to the main-plots within each farm. Yield of maize grain, from the season following the harvesting of the fallow treatment, was recorded for each sub-plot. The treatment structure is a factorial one, but the two factors (fallow and N) are applied at different levels. This is an example of an experiment where the measurements are made at only one level, namely the sub-plot level, but the treatments are applied at two.

When treatments are applied to one type of plot, the variation between these (after allowing for block and treatment effects) is the residual mean square (residual m.s.) found in the ANOVA, and is critical in interpreting the results. We have two types of plot, and it seems reasonable that the variation of each type need not be the same. It is likely that the variation between main-plots will be different, probably larger, than the variation between sub-plots within main-plots. So we might expect two different residual variances to be relevant to interpreting these results, and indeed this is the case.

The ideas of different levels of variation in a split-plot experiment and how to interpret the ANOVA table are all given in MCH (Pages 130–136). Here we are specifically interested in one or two particular features of the analysis, which are discussed below. The dialogue box (Figure 1) for analysing this trial is shown below, along with the GenStat output, which we will use for illustration.

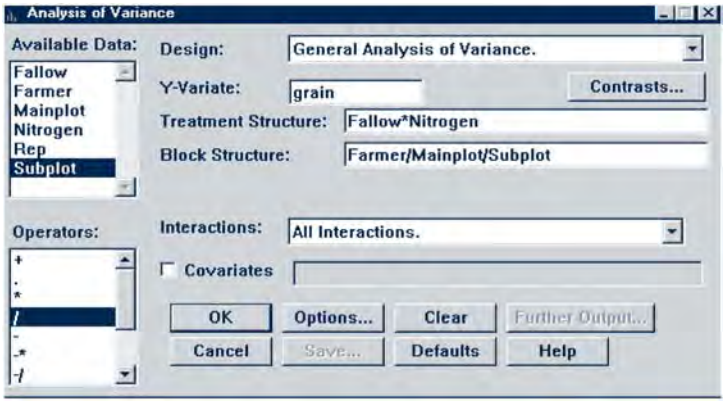



Figure 1. Dialogue box for analyzing the trial

The layout can be expressed as 'sub-plots nested within main-plots nested within Farmer', written as Farmer/Mainplot/Subplot.The analysis has recognized that the Nitrogen treatment is a subplot of the Fallow treatment and that these treatments have been applied within each farm. The treatment structure can be expressed as 'Fallow effect + Nitrogen effect + Fallow.Nitrogen interaction'. This is the usual way of analysing a factorial treatment structure, which we learnt in Session 5.

Discussion centres first on the ANOVA table (Table 1), shown below, and then on the comparison of mean values (Table 2).

```
***** Analysis of variance *****
Variate: grain

Source of variation    d.f.        s.s.      m.s.    v.r.    F pr.

Farmer   stratum          8     55.1600    6.8950    4.73

Farmer.Mainplot stratum
Fallow                    3     14.0076    4.6692    3.20    0.041
Residual                 24     35.0173    1.4591    3.06

Farmer.Mainplot.Subplot stratum
Nitrogen                  1     16.6317   16.6317   34.90    <.001
Fallow.Nitrogen           3      6.7683    2.2561    4.73    0.008
Residual                 32     15.2498    0.4766

Total                    71    142.8346
```

Table 1. ANOVA table produced by previous dialogue

First notice that there are two residual lines. The fallow treatments are main-plot treatments and so are assessed at the main-plot level, relative to variation between different main-plots. N treatments are applied to the sub-plots and are assessed relative to the sub-plot variation. Note too that the Fallow.Nitrogen interaction is also assessed at the sub-plot level because the interaction depends on differences between N treatment effects, each of which is measured at the sub-plot level.

```
Variate: grain

Grand mean 1.696

Fallow  Continous maize      Crotolaria        Tephrosia     Tithonia
             1.050             1.825             2.279        1.630

 Nitrogen              No          Yes
                    1.215        2.176

  Fallow   Nitrogen      No          Yes
  Continous maize     0.700        1.400
  Crotolaria          1.785        1.864
  Tephrosia           1.492        3.065
  Tithonia            0.883        2.376


*** Standard errors of differences of means ***

Table     Fallow     Nitrogen          Fallow
                                     Nitrogen
rep.       18          36               9
s.e.d.      0.4026      0.1627          0.4638
d.f.       24          32              39.11
Except  when  comparing  means  with  the  same  level(s)  of  Fallow
                                      0.3254
  d.f.                               32
```

Table 2. Table of means from previous dialogue

The tables of means are presented in just the same way as in earlier examples. The standard errors for the main effects tables are also similar to earlier examples. Notice that the s.e.d. for nitrogen treatments is smaller than that for fallows, because the sub-plot variation is only about 1/3 that of the main-plots. GenStat gives two s.e.d.'s for assessing the two-way table of means. This reflects the fact that some comparisons in that table are between main-plots and others are between sub-plots.

We observe in this example, that although the treatments are at two levels, the data have an orthogonal structure and therefore the analysis is straightforward. Reasonable comparisons of the main-plot treatments can be made, by simply comparing treatment means. This would not be the case if, for example, each main-plot did not have all the sub-plot treatments represented. Because of this simple structure it is possible, in most statistical software, to construct an ANOVA table and test effects relative to the correct level of variation. Care is needed in the comparison of treatment means, particularly the interaction effects. For these we have seen that the standard errors are different depending on the comparison of interest. The precision of comparison of main-plot treatments at some fixed level of the sub-plot treatment is dependent on both main-plot variation and sub-plot variation.

It is also necessary, of course, to fully understand that the layout was actually a split-plot design! If we had ignored this, then the analysis of variance and the interpretation following it, would have been wrong.

## Layout at different levels

A second type of multilevel experiment involves complexities related to the layout. At times, trials may be designed with multilevel layout structures. Examples of this type of layout complication include trials that have two or more levels of 'blocking'. We will look at an example of this situation.

Example: We have the protocol and data for a '*Leucaena trichandra* seed production' trial (see Part 4, Experiments portfolio, of the course documents). The trial looks at 20 leucaena families. There were 20 replicates of 20 plots and each family was represented once in every replicate. Four trees were grown in each plot and variables such as tree height and cross-sectional area were measured for each tree in the plot.

This design sounds quite straightforward, however, the trial has an additional layout complexity. The designer of the trial thought that blocks of 20 plots would be too large and wanted to use much smaller blocks for what appeared to be a very variable site. Each of the 20 replicates was arranged as 5 blocks of 4 plots with four different families in a block. There are now several different strata to the data; replicates, blocks within replicates, plots within blocks and trees within plots, and two of the strata are blocking factors. What effect does this have on the analysis? Should we treat the experiment as an RCBD with 20 replicates or as an incomplete block design with 100 blocks of 4 plots? Both analyses are possible, but which should we choose?

If we treat the experiment as 100 blocks of 4 plots, the design is non-orthogonal (i.e. not every family appears in each block). It is then not possible to carry out an analysis on the individual tree data using the ANOVA commands. The ANOVA approach will not work because of the non-orthogonality of the blocks and families; and a general linear model approach also fails because it assumes there is only one level of variation in the data. Our only approach would be to analyse the data averaged up to the plot level. How then does this approach compare with treating the data as a RCBD with 20 replicates and 20 families per replicate? This issue is explored in the practical.

## Measurements at different levels

Measurements are sometimes taken at a different level from that at which the treatments were applied. In some cases these measurements may be at a higher level. For example, in the '*Lantana/Tithonia* mulch on-farm' trial that we analysed in Sessions 5 and 6, there were 28 farmers, each of which had 3 plots. Our analysis so far has been carried out on the plot yields. However, the full dataset also includes some measurements (e.g. striga counts), which were recorded at

the field rather than the plot level. Analysis of the striga data alone is straightforward, but considering the striga data together with the plot yields is more complicated, because of the two levels of measurement.

The opposite situation occurs when measurements are taken at a level lower than the treatments, but no other factor is applied at this level. For example, striga is sometimes recorded in quadrats with several quadrats per plot. How does this then affect the analysis?

We can look again at the '*Leucaena* seed production' trial, which we have previously considered in terms of its multilevel layout structure. Treatments, i.e. the different families, are applied to the plots, but several measurements are made within the plot, i.e. on the trees. This effectively gives us four times as many observations as we would expect: 20 x 20 x 4 = 1600 values from 400 plots. However this does not mean that we have 4 times as much information. Again you need to recognize the structure of the data. For instance it could be viewed at a split-plot with no sub-plot treatment.



## Sensible approaches to the analysis

The individual trees are giving us information about the variability amongst trees in the plots; but in terms of our 20 families we want to assess family differences relative to the plot-to-plot variation. There are two ways that this can be done.

The first is to follow the suggestion above; treat the data like a split-plot. Plots are our 'main-plots' and families are the 'main-plot treatment' whilst the trees are the 'sub-plots' to which no treatment is applied.

The GenStat output of this analysis is presented in Table 3 below.

```
***** Analysis of variance *****

Variate: height

Source of variation  d.f.(m.v.)   s.s.         m.s.        v.r.    F pr.

rep stratum               19        289337.      15228.      4.48

rep.plot stratum
family                    19        848762.      44672.      3.13   <.001
residual                 361       1228383.       3403.      1.38

rep.plot.tree stratum   1181(19)    2904719.       2460.

Total                   1580(19)    5232265.
```

Table 3.

In the ANOVA in Table 3 above, the rep.plot residual of 3403 is estimating variation at the plot level, whilst the rep.plot.tree mean square of 2460 is estimating variation between trees. The plot-to-plot variance is larger than that of tree-to-tree.

Families are tested at the plot level, which is correct since this is the level at which the factor is applied.

```
***** Tables of means *****

Variate: height

Grand mean 201.7

family      1         2         3         4         5         6         7
        212.4     206.6     205.5     243.4     209.6      22.6     240.7

family      8         9        10        11        12        13        14
        177.3     196.6     202.6     176.1     201.0     207.6     212.9

family     15        16        17        18        19        20
        225.6     174.0     164.2     159.8     179.1     217.5

*** Standard errors of differences of means ***

Table     family
rep.       80
d.f.      361
s.e.d.      9.22
```

Table 4. Table of means of previous analysis

The second approach to the analysis that we could use would be to summarize the tree-level data to the level at which the treatments are applied. We therefore need a value at the plot level that combines the information observed at tree-level. In this case we are interested in tree height, thus average height over all 4 trees in a single plot summarizes this information well and creates an observation at the plot level. In other examples a total may be more appropriate.

The analysis of the average tree height per plot is presented in Table 5 below.

```
***** Analysis of variance *****

Variate: m_height

Source of variation   d.f.         s.s.       m.s.     v.r.  F pr.

rep stratum             19      72334.3     3807.1     4.48

rep.*Units* stratum
family                  19     212190.4    11167.9    13.13  <.001
Residual               361     307095.9      850.7

Total                  399     591620.6

(The table of treatment means is identical to the one above and is not
presented here).

*** Standard errors of differences of means ***

Table               family
rep.                 20
d.f.                361
s.e.d.               9.22
```

Table 5. Analysis of the average tree height per plot



The two analyses give identical results in terms of significance tests, treatment means and standard errors of differences. And indeed so they should. In any standard split-plot analysis the results at the main-plot level are based on averaging over the information at the sub-plot level. This is shown by the sums of squares and mean squares in the two tables. Everything in the split-plot analysis which uses all the data is scaled up by a factor of 4, the number of trees in the plot.

When confronted with this type of data, either approach can be taken and the choice is left to the analyst. However, summarizing data up to the next level seems a natural and fairly easy way to deal with the complexity of measurements at a level lower than the treatment. We will see later that this can include other summary measures such as a difference or a contrast.

The problem of missing values is discussed in Session 15. In this example there were just 19 trees missing out of 1600 and so we would not expect them to have much impact on the results. However, it would not be unusual for many more of the trees in a trial to die and hence for their heights and weights to be missing. We then have to think carefully about possible differences between the two approaches of summarizing to one level and analysing individuals.

## The consequences of using the wrong analysis

A common mistake which people make with this type of data, is to ignore the structure completely. What effect does this have on the analysis? What would have happened here had we analysed all 1600 values (less missing ones), ignoring the hierarchical structure? The results are presented in Table 6 below; again treatment means are not given, as they are the same as those above.

```
***** Analysis of variance *****

Variate: height

Source of variation    d.f.(m.v.)    s.s.     m.s.     v.r.    F pr.

rep stratum              19       288768.   15198.     5.71

rep.*Units* stratum
family                   19       851257.   44803.    16.83   <.001
Residual               1542(19)  4106052.    2663.

Total                  1580(19)  5232265.


*** Standard errors of differences of means ***

Table     family
rep.        80
d.f.      1542
s.e.d.     8.16

(Not adjusted for missing values)
```

Table 6. Incorrect analysis of tree heights ignoring plots

The residual mean square on 1542 d.f. is estimating some weighted average of variation between plots, and variation between trees within the plots, and so for this reason it is not a sensible 'residual variance' against which to compare the treatments. Compared with the split-plot analysis of these data (see Table 3), we can see that this residual is lower than the plot-mean-square and higher than the tree-mean-square. Hence if we use it as our residual for significance testing then the testing is too liberal. The standard errors of family differences are now wrong, being too low, with corresponding degrees of freedom, which are too large.

How serious a mistake is this? Here it does not seem to make much difference whether the right or wrong analysis is used, the differences between families are clear regardless of method and the s.e.d. does not greatly alter (from 9.22 down to 8.16). However this is only happening because the variation due to plot differences is not much larger than that due to tree differences. This is not always the case, and in many other instances, failure to account for the hierarchical structure can have serious consequences leading to erroneous conclusions.

How could the 'General linear model' method of analysis (Session 7) be used with this data? If the 1600 tree-level values were modelled with a simple model including replicate and treatment effects, the analysis would be the same as the incorrect 'tree level' analysis described above.

It is therefore important that the structure of the data is correctly identified so that the correct analysis can be performed. For instance, if the experiment above had been carried out as 1600 plots in which individual trees were grown, then the analysis illustrated above using all 1600 measurements with no structure, would have been correct. Make sure you understand the difference between a trial with 1600 trees as individual plots and one with 1600 trees but 400 plots of 4 trees. They can look identical in the field!

We have now come to the end of simple approaches to multiple levels. Later in the course we will look at additional simple approaches we can take to this type of complication. However, there will also be situations in which we can't take a simple approach. For example, we have investigated how to analyse split-plot experiments when every sub-plot treatment occurs in every main-plot. But what can we do if we have an 'incomplete' split-plot design? We have also looked at how to analyse experiments where measurements have been taken at different levels, in the '*Lantana/Tithonia* mulch on-farm' trial we have measurements at both the plot and farm level. But what happens if our measurements are sometimes taken at one level and at other times at a different level, e.g. in our on-farm trial, if our striga was a farm level variable for most farms, but on a few farms it was measured at the plot level.

For these, more complicated multilevel experiments, we need to move on to random effects models and the ideas of REML. These are discussed later in the course.

Session 9

# Writing up and presenting results

*R. Coe, R. D. Stern, E. Allan*

## Introduction

Statistical analysis of an experiment has no value unless the results can be conveyed to the relevant audience. This can be difficult. The researcher knows just how complex, and perhaps confusing, the analysis has been, has discovered many nuances to the data, can think of many caveats and conditions that perhaps apply, yet must simply convey the results that meet the objectives of the study.

The style in which the results are presented will depend on:

- the intended audience,
- the medium used.

Here we assume the audience to be other scientists. The style of presentation would be rather different if the information is being provided for policy makers, farmers or fund providers.

Most of the comments in this session refer to written presentations, scientific reports and papers, but the final section covers some ways in which the presentation should be adapted for other media.

There are professionals in the area of writing and presenting results who should be consulted whenever possible.

## Statistics in perspective

Remember the role of the statistics in the report is to provide the numerical evidence to support the arguments being made. The detailed results of the analysis, the numerous tables of means, standard errors and c.v.'s, regression results with parameter estimates, analyses of variance, $r^2$ etc., may or may not be pertinent to the argument. They should be included only when they are necessary to support the conclusions. The results of statistical analyses are presented because they show why you reached your conclusions, but are rarely the conclusions themselves and certainly should not be presented just because you calculated them.

Occasionally results are also included in a report to act as an archive of the data obtained. In that case it may be appropriate for a more complete reporting, even if some of the numbers are not discussed in the text.

The role of measures of precision (such as standard errors) is to quantify the uncertainty in other numbers. The uncertainty is important because it shows how much confidence can be placed on the other numbers reported, and hence how firm or clear the conclusions are. Thus, they too are part of the essential information that supports the argument, and should be included.

## Tables and graphs

The tables and graphs are the first things that a reader of any report notices, and often they are the first thing studied. For this reason it is essential that tables and graphs given are:

- necessary (don't include tables if they do not contribute to the conclusions),
- complete with all the relevant data,
- concise (do not include details that are not needed),
- unambiguous,
- easy to read (tables and graphs should be giving a message, so make the message obvious).

The booklet 'Informative Presentation of Tables, Graphs and Statistics' provides much sound advice on how to draw tables and graphs.

Computers make it quick and simple to layout tables and graphs. Remember that the aim of a graph is not to show how many gimmicks are included in your software, but to 'tell a story' with the data. Often the most effective graph is a simple one from which excessive 'clutter' of shading, writing, grid lines and pseudo-3D effects have been removed. Similarly, a table can be made more readable by paying attention to spacing, lines under headings, the relative sizes of fonts etc. Fortunately, computers also make it easy to experiment with different layouts. Try a few and ask colleagues to evaluate them for ease of recognizing the important results.

## Describing statistical methods used

When writing a report it is often necessary to include a description of the statistical methods that have been used. In scientific papers this is done for two reasons, so that anyone can see that the methods are appropriate, and so that the analysis can be repeated later. Journal editors often insist on a brief description of the statistical method used. In simple cases it is possible to give a clear description very briefly (e.g. 'The average soil nitrate was calculated for each plot from the 10 samples. Logs of the plot averages were taken to stabilize the variance before calculating an analysis of variance based on the randomized block design'). However, in

other cases it is difficult to be brief, because the analysis involved many steps or was otherwise non-standard. It is then up to you to find a compromise between a long detailed description and something incomplete. It is necessary to be clear about any data that were omitted from the analysis, and the reasons for that. Do not reproduce details of methods that have been published, but simply give a reference. It is usual to name the statistical software used and give a reference for it. If the analysis was particularly complex or novel, it may be appropriate to describe the software commands used to calculate it.

However, much detail of statistical method goes into the report and it is important that you keep a careful note of what you did with the data. That way you can give more details to anyone who needs the data and can repeat the analysis if questions arise later.

## Statistical results in the text

It is necessary to discuss statistical results in the text. However, avoid repeating lists of results that can be seen more clearly in the table. Whole paragraphs that just reproduce in longhand the data in a table, are just a waste of effort and space. Of course, some results in each table or graph must be referred to in the text. A table, or part of a table, that is ignored in the text can usually be omitted.

Avoid long lists of 'effects that are significant'. It is common to find strings of sentences of the type 'Soil nitrate was significantly different ($p<0.05$) between treatments in all seasons except 1997 when it was nearly significant ($p=0.07$). Soil ammonia-N was significantly different in the first two seasons but not in 1997 or 1998. However the effect…'. These points can be seen from a well-drawn table. Instead the text should draw attention to the important patterns which will be interpreted, e.g. 'Soil nitrate and ammonia N were clearly higher in agroforestry plots than the control in all years except the driest, when the difference was only half that seen in normal years'.

The term 'significant' is often over-used in statistical reports of experiments. See how far you can go without using it. The benefits of being able to do this well are:

- 'Significant' has a common usage quite different from the statistical one. Some readers may be confused.
- There is no clear boundary between 'significant' and 'not significant', using phrases such as 'clear evidence for..', 'some evidence for…' and 'little reason to believe that…' enables you to reflect the real state of your knowledge more fairly.

Remember also that many study objectives need estimates not of whether a treatment difference exists, but whether the difference is large enough to be useful in some way. It is often more appropriate to give estimates of the size of an effect (e.g. with a confidence interval) than statements about significance.

## The rest of the analysis

In most experiments the analysis does not finish with the statistical conclusions based on the raw experimental data. The real objectives of the trial will only be met when the results are used in further analyses. This can take many forms, a farm level financial analysis, estimated impact of adoption on regional productivity, mapping of potential adoption domains, comparison of results with those from a simulation model, and many others. These further analyses typically involve combining information from the current experiment with information from other sources, other experiments, surveys, remotely sensed data, models etc. A new range of statistical techniques may well be needed to complete these analyses.

## Other presentations

It is increasingly important for researchers to be able to present their work verbally, backed up by slides containing tables, graphs, and a few key points of text. These presentations require a different layout. For example, it may be best not to present any s.e.d.'s on the tables to prevent them getting too cluttered (but have them to hand in case anyone asks). Large tables are difficult to interpret when they are briefly shown on a screen (I suspect about 16 numbers is the maximum it is sensible to include). So perhaps large tables can be redrawn as several small ones, each of which highlights a point you wish to make. Colours can be used effectively in both tables and graphs, but used carelessly can ruin an otherwise good presentation.

The message is always: remember the audience and medium you are using.

Session 10

# Where are we now?- Review of basic statistics

*R. Coe, R. D. Stern, E. Allan*

The problem of analysing data from an experiment, was introduced by first identifying the objectives. Meeting the objectives started with simple summaries presented in graphs and tables. Formal analysis was then introduced because the simple summaries did not handle:

- Complex structures from which we had to disentangle several different effects.
- Uncertainty.

The formal analysis depends on the features of the design, its layout, treatments and measurements, and the way these interact. In Sessions 5 and 6 we discussed analysis of variance for designs with a 'simple' layout; only one layer in the design, with different effects of interest 'orthogonal'. In Session 7 we showed that the same analysis could be obtained by fitting a model using regression. Session 8 extended the analyses to designs, which had data or treatments in several levels or layers of the design, but were still orthogonal. Thus so far, the various techniques of analysis have depended on the layout and treatment structure of the design. The methods introduced have been appropriate for data, which are (approximately) normally distributed (a feature of the measurement). The range of methods introduced, and their suitability for designs with different layout, treatment and measurement characteristics, are summarized in Table 1 below.

Here the methods introduced already are shaded grey. Methods for other situations are also named.

| | Simple layer structure | | Multiple layer structure | |
|---|---|---|---|---|
| | Orthogonal | Non-orthogonal | Orthogonal | Non-orthogonal |
| Normal distribution | ANOVA | Regression | ANOVA | REML |
| Other distributions | Generalized linear models | | Generalized mixed models | |

Table 1.

The remainder of the course now aims to:

- Show some simple applications of the methods in the black boxes.
- Introduce some more advanced applications of methods already described.
- Show how some problems not directly falling into any of these categories can be detected and tackled using tools already available.

The first part of the course contained material, which we believe any experimenter carrying out field experiments should become familiar with, and able to implement unaided, with confidence. Now the course changes a little. The techniques being introduced are undoubtedly harder to understand. In fact we find that they are only slightly harder to understand and use, given the power of modern software. But the added complexity means they may not be in every scientist's tool bag. Instead we expect scientists to:

- Be aware that these methods exist.
- Understand when their problem needs these methods.
- Understand the results of an analysis.
- Implement the analysis with a bit of help.

Of course there will be some individuals for whom these methods will be needed regularly in their work and they will become competent in their use.

Session 11

# Design and analysis complexity

*R. Coe, R. D. Stern and E. Allan*

## Introduction

A theme of the course so far has been that the analysis is determined by the objectives and the design of the trial; the treatments, layout and measurements. We saw that in many cases an analysis that meets the objectives and is valid for the design can be accomplished with relatively simple statistical techniques. However the practical analyst knows that complexities often arise. In this session we look at complexities which result directly from the design. In other words, these are complexities that can be predicted from knowledge of the design, without even looking at the data.

Here we cover three types of complexity that are associated with the main components of the trial, namely the treatments, layout and the measurements. A common feature of this and later sessions, which cover other types of complexity, is that there is a range of approaches to the problems. There may well be a solution that, though clearly not optimal, allows us to make progress, using methods that are already familiar. An example is the idea of omitting observations, so that the remaining data have a simple structure that we know how to handle.

In many cases satisfactory solutions require the building and interpretation of statistical models. The process of choosing, fitting and interpreting the model to handle these complexities is identical to that used for earlier examples, but the complexities mean that the models have to be more refined. The models used so far can be summarized as

response = effects + random residual

So far the 'response' has been assumed to be measured, on a continuous scale, the 'effects' have been simple combinations of factors and variates, and the 'random residual' has been assumed to have a normal distribution with a single variance. Many of the complexities that arise can be handled by modifying each of these model components. For example, we can use different effects to describe complex treatment structures. We can allow other sorts of response variables and other distributions than the normal, or we can allow more than a single random term in the model. These last two developments are relatively recent and have not yet been well integrated into standard agricultural research practice, but are powerful and important developments.

In this session we provide some motivation and simple examples of each of these changes to the models. Further details are provided in later sessions.

## Measurement complexities

The measurements we have looked at so far have been on a continuous scale (height, weight, concentration, etc.) and a natural way to analyse these has been by looking at means. In the simplest situations we met objectives by comparing two means. More complex problems were reduced to describing how the mean changed with, for example, different treatments or continuously varying quantities. The linear models put this into a consistent framework. Making additional assumptions (e.g. the variance of the residual part is constant, residuals have a normal distribution) allow us to make statistical inferences.

There is no reason why data on height or yield has to follow the patterns required for the analysis of variance or regression to be valid and useful. In Session 15 we look at options for analysis when some of these assumptions turn out to be unreasonable. However we can often predict that they will be unreasonable for certain types of measurements. Examples include:

- Counts (e.g. number of insects on a plant).
- 'Success/failure' data (e.g. survival of a plant at 6 months, presence/absence of a disease). Typically several individuals per plot are assessed and the data summarized as '6 out of 10 surviving' or '7 out of 100 diseased'.
- Proportions (e.g. estimates of the proportion of the plants in a plot that are flowering, or the proportion of a leaf surface which is diseased). These data are similar to the previous example, but without knowing the total number of individuals assessed.
- Scores (e.g. disease level scored as 0, 1, 2, 3, 4 for none to very severe, value of a technology scored as 0 to 5 for useless to essential).
- Categories (e.g. seed classed as yellow, white or purple).
- Time until an event (e.g. time to flowering), which may be 'censored', that is we only know that the time is greater than some value (the last time we looked), not the actual time.

## What problems might there be from different types of data?

The previous methods may be inappropriate for these new data types for a number of reasons.

1. The mean may not be a useful summary (What is the average colour of two white seeds and a yellow one?).
2. Simple additive effects of treatments and other variables may be unrealistic, giving results which lie outside the permissible range (% survival has to lie in the range 0 to 100%, for example).
3. Critical assumptions needed for inference may be unrealistic (If the mean number of aphids per plant for treatment 1 is 4 and for treatment 2 is 400 you would not expect the variance of both to be the same).
4. The normal distribution assumption may be very unrealistic.

## What are the options?

As with all data analyses, start with a clear statement of the objectives and understand the nature of the data you are trying to analyse. Suppose seed colour was recorded in a variety trial with the aim of comparing the seed colour of two varieties. What summaries of the data are going to be useful? To some extent that will depend on the actual values recorded. If it turns out that all variety A seeds were yellow and all B seeds were white, no further analysis is necessary. The table of results (Table 1) would be as simple as:

| Variety | Seed Colour |
|---|---|
| A | yellow |
| B | white |

Table 1.

If we find some white and yellow in each variety then probably the proportion of seeds that are yellow is a useful summary and we might construct a table (Table 2) such as:

| Variety | % yellow seed |
|---|---|
| A | 23 |
| B | 46 |

Table 2.

If there are more than two colours then again the summary may have to be changed to something like (Table 3):

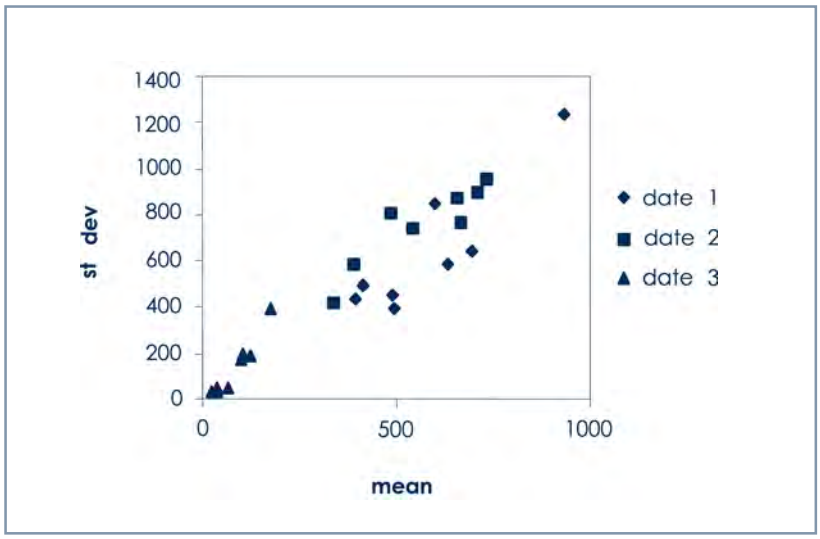
| Variety | Seed colour | | |
|---|---|---|---|
| | % yellow | % white | % purple |
| A | 23 | 73 | 2 |
| B | 46 | 15 | 35 |

Note: percentages do not add to 100 due to damaged and rotten seed.
Table 3.

What would clearly be incorrect would be to code yellow as '1', white as '2' and purple as '3', then find the average of these codes for each variety. Starting the formal statistical analysis without thinking about the objectives and summaries that make sense, can lead to such foolishness.

A useful strategy to simplify analyses is to construct new variables that are of a type you can handle, and which are tuned to your analysis objectives. For example, if working with three or more colours is confusing, maybe a new variable 'yellow/not yellow' is easier, and is acceptable if assessing the yellowness is the main objective. This is discussed further in Session 12. If you do not know how to analyse 'time to flowering' when some of the values are recorded as '>92 days', then maybe we can work with a new variable which just records whether flowering occurred by day 85 or not. Such an analysis may not formally be 'optimal' as some information in the data is not used. However if the strategy allows you to make progress with an informative and valid analysis, it is worth doing, and may well give you all the information needed if the new variables are constructed with careful attention to the objectives.

The exploratory analysis allows you to confirm that suspected complexities in the data are actually present. As an example consider the striga count data from the experiment 'The influence of improved fallows on soil phosphorus fractions, an on-farm trial'. On three different dates striga plants were counted in each plot. We would expect such data to show more variation for dates or treatments that have higher means. The exploratory analysis shows this to be the case. Graph 1 below is a scatter plot of the standard deviation (across the 9 reps) for each treatment and date plotted against the mean (across the 9 replicates). The increasing variation with increasing mean is striking.




Graph 1. Scatter plot of the standard deviation (across the 9 reps) for each treatment and date plotted against the mean

Does it matter? The means are still valid and useful summaries of the data, revealing the changes in striga counts with date and treatment. However this pattern of changing variance will affect the formal analysis. Standard analysis of variance assumes constant variation for all treatments. The residual variance, used to estimate standard errors of treatment differences, is estimated by pooling or averaging the residual variance for each treatment. This average variance will underestimate the uncertainty in the large means and overestimate the uncertainty in small means.

The last example above illustrates the importance of understanding which aspects of the data influence different parts of the analysis.

In Session 15 we show that there is a range of options for handling problems in measurements, ranging from 'quick fixes' to comprehensive models. Here we use the general approach of finding a model that represents the data and design realistically. It builds on the 'general linear model' used in Sessions 7 and 8.

## Example 1: Screening of suitable species for three-year fallow.

This is the same example, which was analysed as a general linear model in Session 7. There we looked at treatment effects on crop yield. The data can also be analysed as a generalized linear model, Figure 1, next page, and the results would be identical to the earlier analysis.

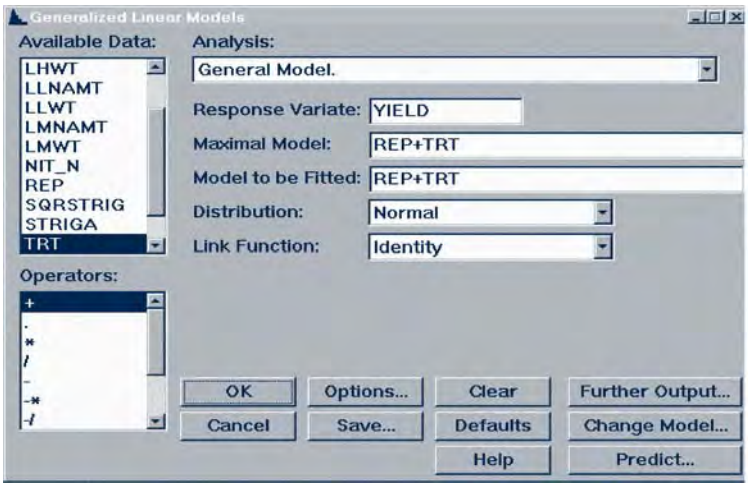



Figure 1. Generalized linear model dialogue

The above dialogue box shows that the data were analyzed, by fitting models, and looking at the effects of the different components in the model. But now the distribution of the data needs to be specified, and a transformation, or link function, used.

In the context of the model:

data = pattern + residual

the above ideas can be thought of as:

(i) Specifying the distribution of the residual variation.

(ii) Transforming the data in order to link it to the pattern associated with the different effects.

Within the generalized linear model framework there are several distributions and several link functions, and for each distribution there is a default (canonical) link. There is nowhere in the dialogue to explicitly describe the 'variance function'; that is the way in which the variance changes with the mean. In fact this is also determined by the choice of distribution. For a normal distribution the variance is constant, not changing with the mean. We showed an example in which the standard deviation seems to be proportional to the mean, or the variance proportional to mean$^2$. The gamma distribution has this property, so may be an appropriate choice for analysis of that data.

Had we fitted the above model as a general linear model, the distribution is automatically assumed to be Normal and no transformation is necessary. Here we must state that the distribution is Normal and that the link is the identity, which just means that the link between the response and model parameters is 'one-to-one', i.e. no transformation.

## Example 2: MCH (pages 325-327)

These are data from a seed germination study, looking at the effects of four different chemical concentrations and four temperature regimes on germination. The treatment structure is a factorial one, and under each combination of concentration and temperature, there were four dishes of 50 seeds which were stored for germination.

Inspection of the data suggests that different numbers, or proportions of seeds, germinate under the different conditions; and so we want to model the effects of concentration and temperature on the proportion, which germinate.

The full analysis is discussed in MCH. We will only fit a model with effects for concentration, temperature and their interaction.

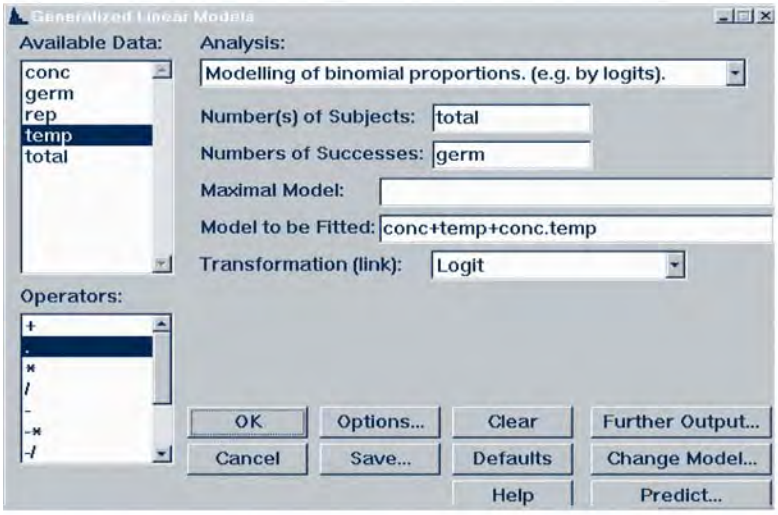



Figure 2. Dialogue for the model with effects for concentration, temperature and their interaction

The following information is needed to complete the dialogue:

(i) There are two components to our 'response'. The number of the 'successes' and the total number from which the proportion is calculated. In this case we have the number of seeds, which germinate out of the total number in the dish.

(ii) Counts of 'successes' out of some total number often follow a Binomial distribution.

(iii) With Binomial data the default link is the logit. We are now modelling our proportion data on a logit-transformed scale.

(iv) The above model produced the following output (Table 4).

```
*** Accumulated analysis of deviance ***

Change                        mean   deviance   approx
         d.f.   deviance  deviance    ratio    chi pr

+ conc      3     213.703   71.234    71.23    <.001
+ temp      3     831.993  277.331   277.33    <.001
+ conc.temp 9      92.464   10.274    10.27    <.001
Residual   48      55.641    1.159
Total      63    1193.801   18.949

.    MESSAGE: ratios are based on dispersion parameter with value 1
```

Table 4.

Compare Table 4 with the format of the analysis of variance table.

Instead of sums of squares, we now have changes in deviance, and instead of F-tests there are Chi-square tests.

The deviance tests only have a Chi-square distribution, under a certain assumption; namely that the seeds in the dishes behave independently of one another. This can be detected here by inspecting the residual deviance. If they behave independently, then the residual mean deviance should be approximately 1. Here the value of 1.159 in the table above gives no evidence to reject the assumption of independence. Hence, you see from the message in Table 4, that we have assumed the value of 1 in calculating the ratios above. This is not always the case with seed germinating data, where the residual deviance can often be greater than 1. We should then want the ratios to be calculated using the estimated residual mean deviance.

[In GenStat the **[Option]** button in the regression dialogue would allow us to do this.]

```
*** Predictions from regression model ***

Prediction
     temp      1       2       3       4
     conc
     1      0.14    0.49    0.11    0.12
     2      0.27    0.61    0.06    0.21
     3      0.48    0.80    0.16    0.33
     4      0.74    0.96    0.10    0.29
```

Table 5.

To summarize the results we use the **[Predict]** button in the regression dialogue. This gives expected proportions for the treatments as shown in Table 5 above. These are calculated using the model parameters, just as adjusted means are calculated in the case of the general linear model. So for instance the proportion germinating on temperature regime 1 with the first level of concentration (which was water) was estimated to be 0.14 or 14%.

The analysis above shows how to fit a logistic regression model and interpret the output. The approach is similar for other generalized linear models. Count data, often have skewed distribution, and can often be assumed to have a Poisson distribution. Ordered categorical data, can be analysed by methods which are just extensions of the logistic model for binomial data. For further technical information on generalized linear models, the reader should refer to MCH or some other textbook on the subject.

## Layout complexities

As we layout the trial we might be able to predict that the layout pattern will give rise to complexities. Here we concentrate on possible problems due to an 'irregular' blocking pattern.

### Why is it a problem?

Consider a simple design as represented in Table 6 below, in which the letter shows the treatment (A, B or C) applied to each plot.

| | | Blocks | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| Plot Number | 1 | A | C | B | B |
| | 2 | B | A | C | C |
| | 3 | C | B | | A |

Table 6. Treatment (A, B or C) applied to each plot

Remember that blocks are selected because we think the response in each may be different. If this is the case then we cannot simply compare means of A, B and C to evaluated treatment effects. If A turns out to have a higher mean, is that an effect of the treatments or is it because A does not occur in block 3, and that is a poor block? A similar difficulty may arise if there is more than one replicate of a treatment in some blocks.

### Spotting the problem

Some methods for spotting that the problem exists are:

1. Look at a table of counts of observations cross-classified by treatment and block (note: this is good for spotting data entry mistakes as well as layout complications).
2. GenStat's ANOVA command (and the equivalent in other software) will give a warning or fault if you try to use it. Try it on the data and interpret the message.

## What can be done?

Again we have a range of options (see Session 15). It may be that ignoring one or two observations will allow a straightforward analysis. This might be acceptable, at least as a preliminary analysis. However, here we show that using an appropriate model solves the problem.

### Example: Fertilizer, *Tithonia* and *Lantana* mulch as sources of phosphorus for maize

The experiment has 10 treatments and a blocked design with three blocks. Attempting to produce an analysis of variance of the yield produces the results in Table 7 below.

```
******** Fault (Code AN 1). Statement 1 on Line 47
Command: ANOVA grain
Design unbalanced - cannot be analysed by ANOVA
Model term treat (non-orthogonal to term block) is unbalanced, in the
block.*Units* stratum.
```

Table 7

The message is not altogether clear, but suggests there is a problem with the blocks and treatments. Table 8 showing the number of occurrences of each treatment within each block helps.

```
Nobservd
       block 1    2    3
       treat
        1    1    1    2
        2    1    1    1
        3    1    1    1
        4    0    1    1
        5    2    2    1
        6    1    1    1
        7    1    1    1
        8    1    2    2
        9    1    1    1
       10    1    1    1
```

Table 8.

While most treatments occur once in each block there are some irregularities. The means for treatments 4 and 5 cannot be 'fairly' compared, as there are two observations for 5 in block 1, but none of 4. If block 1 is a 'good' block this will bias the treatment 5 mean up compared with 4. However, there is a similar problem in determining which blocks have high or low means.

The simple solution is to remove observations that cause the irregularity. This is not ideal because:

- It is arbitrary (which of two observations of a treatment in a block will you drop?).
- It ignores useful information (why measure a plot then throw the data away?).
- It does not help when the irregularity is a 'missing', rather than an extra, treatment.

The modelling strategy is exactly that used in Session 7; fit a model with block and treatment effects, and use it to estimate 'adjusted' means. The commands and results are shown in Table 9 below:

```
MODEL grain
FIT [p=*] block
ADD [p=a] treat

***** Regression Analysis *****

*** Accumulated analysis of variance ***

Change     d.f.            s.s.          m.s.      v.r.
+ block       2        4.335E+06     2.167E+06      3.81
+ treat       9        8.979E+07     9.977E+06     17.53
Residual     22        1.252E+07     5.692E+05

Total        33        1.066E+08     3.232E+06


PREDICT treat
RPAIR  !P(treat)

*** Predictions from regression model ***


 Response variate: grain
      Prediction
      treat
       1      140
       2      510
       3      690
       4      856
       5     1678
       6     4246
       7     4432
       8     2821
       9     3953
      10     4417
```

Table 9. Commands and results for a modelling strategy

For comparison, the unadjusted means are shown (Table 10):

```
Mean
  treat
      1       237
      2       519
      3       699
      4       787
      5      1617
      6      4255
      7      4441
      8      2799
      9      3962
     10      4426
```

Table 10. Unadjusted means

In this case the differences are not great, but the adjusted means are a sounder basis for inference.

One complication of the analysis is that we can no longer give a single standard error of difference. The amount of information about different treatment comparisons is not constant, so the s.e.d.'s will not all be the same. Table 11 below shows the s.e.d.'s.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | | | |
| 2 | 578 | | | | | | | | | |
| 3 | 578 | 616 | | | | | | | | |
| 4 | 659 | 696 | 696 | | | | | | | |
| 5 | 513 | 552 | 552 | 643 | | | | | | |
| 6 | 578 | 616 | 616 | 696 | 552 | | | | | |
| 7 | 578 | 616 | 616 | 696 | 552 | 616 | | | | |
| 8 | 508 | 552 | 552 | 634 | 482 | 552 | 552 | | | |
| 9 | 578 | 616 | 616 | 696 | 552 | 616 | 616 | 552 | | |
| 10 | 578 | 616 | 616 | 696 | 552 | 616 | 616 | 552 | 616 | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Table 11. Standard errors of differences for all pairs of treatments

Options for presenting such results were discussed in Session 9. A possible method here is to give the minimum, maximum and average s.e.d. If the number of replicates of each treatment is added to the table of treatment means, then it is clear to which pairs of treatments the different s.e.d.'s apply.

There is one more complication with this example. The variance of the residuals does not appear to be constant, but is higher for treatments with higher mean. Strategies for this complication will be covered in Session 15.

### Example: *Leucaena trichandra* seed production trial

This trial was designed with 20 replicates of 20 families. However the original design used incomplete blocks of size 4. Each replicate had 5 such blocks in it. The design idea was sound; the site was thought to be variable and blocks of 20 plots, with all treatments represented equally, though may well have been too large. Thus this design has the same problem as the previous example. What are the options for analysis?

1. Ignore the small blocks. The blocks form groups, which make up complete replicates, so we could allow for these large blocks and forget about the small ones.
2. Use a model with small blocks and treatments included, as in the previous example. That is straightforward.
3. Use a model that recognises the 'between block' information in the design.

The last option introduces a new idea, which depends on the small blocks within each replicate being thought of as another random effect. The model used then is:

response = replicate + family + small block effect + plot effect + tree effect

The last 3 of these are random effects. The REML procedure in GenStat (Figure 3) fits such a model.

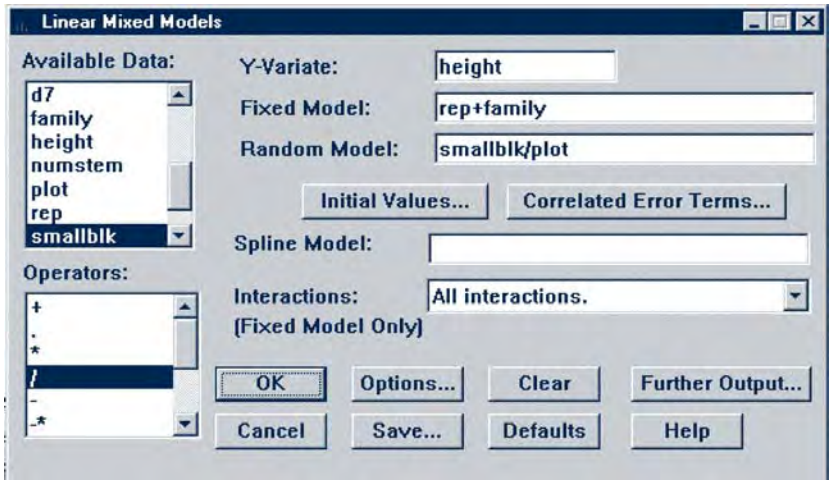



Figure 3. REML procedure in GenStat

Part of the output is shown in Table 12 below. It indicates that the block-to-block variation within replicates is actually very small, so option 1 above, of ignoring them, is a reasonable strategy. It has the advantage that the analysis and presentation is particularly simple for the balanced ANOVA analysis.

```
*** Estimated Variance Components ***

Random term        Component      S.e.

smallblk               0.00246     40.
smallblk.plot        220.          76.
*units*             2460.         101.


*** Wald tests for fixed effects ***


  Fixed termWald statistic        d.f.

  rep                  86.1        19
  family              251.7        19
```

Table 12.

The REML method will be explained further in Session 13.

## Treatment complexities

In Session 5 and 6 we saw that questions about treatments can often be answered by simple contrasts of treatment means, e.g. how does including a tree component compare with crop-only control; or by inspection of main effects and interactions when the treatments have a factorial structure. Sometimes, though, the objectives of an experiment are addressed by formulating complex questions about the treatments, which cannot be answered by these approaches. The ways in which this can arise depend on the nature of the treatment structure and the objectives. Three examples are discussed here.

### 'Near factorial' treatment structure

Some experiments have a factorial treatment structure and an added control, in which case the questions of interest lead us to compare (a) control versus treatment, and then (b) main effects and interactions of the factorial set.

Another example is where, for practical reasons, the treatments are an incomplete factorial set with one or two treatment combinations missing. For instance, in an on-farm trial looking at effects of manuring rate (none, 1 bucket, 2 buckets) and sowing rate (usual density, double the usual density) the farmers were not prepared to have a high sowing density and no manure. The five treatments under study can be viewed as either an incomplete factorial (3x2 – 1), or it could be treated as a complete factorial with an extra treatment (2x2 + 1).

Dealing with the above treatment structures is just an extension of the idea of contrasts, except now we have to set up several new factors and incorporate them into our linear model. How to do this in practice is dealt with in the practical.

A related point, which is illustrated with the experiment below is that even when the treatments are a full factorial set, the standard factorial breakdown does not necessarily answer the objectives.

'Hedge row intercropping' (HI) for soil fertility, is a system in which lines of pruned trees (hedges) are grown to produce mulch or green manure. Intercropped between the hedges are annual crops. The trees are regularly cut back and the resulting leafy biomass applied to the crops to improve soil fertility.

Experiments with HI showed variable results; on some soils and in some climates the result was higher crop production than in a crop-only control, but this was certainly not always the case. In an attempt to understand more about what was going on, and when HI might show a production advantage over a monocrop system, a series of experiments were done. Their aim was to separate the positive fertility effect of the hedges (due to the mulch being added to the crop) from the negative competition effect of the hedges (they take up space and growth resources otherwise available to the crop). The treatment set included the following four:

1. Hm Standard HI system with hedges and mulch applied to the crop
2. Ho HI from which mulch is removed (i.e. hedges present but no mulch added to crop)
3. Cm Crop with no hedges but mulch added
4. Co Crop-only with no hedge or mulch

Analysis looked at the following contrasts:

I = Hm - Co measures the overall effect of the HI system
F = Cm - Co measures the effect of the mulch, a fertility effect
C = Hm - Cm measures the effect of the hedge, a 'competition' effect

Then I = F + C. The overall effect is broken down into two components, F (expected to be positive) and C (expected to be negative). I depends on the balance of the two.

Now notice that one treatment, Ho has not been used. Was this not needed, or have we missed some useful information? There are alternative and equally useful definitions of F and C,

F*= Hm - Ho
C* = Ho - Co

We still have I = F* + C*. The difference F* - F measures how much the fertility effect is changed by the presence of the hedge. Likewise C* - C measures how much the competition is modified by the mulch.

An alternative way to view the treatments is as a 2 x 2 factorial set, with two levels of mulch (with and without) and two of hedge (with and without). The usual mulch x hedge interaction is (Hm-Ho)-(Cm-Co). This is exactly F*-F and C*-C.

## Quantitative treatments

In many experiments the treatments are defined by a quantity of something; the amount of fertilizer, the planting density, the daily feeding rate and so on. A typical example is the trial 'Fertilizer, *Tithonia* and *Lantana* mulch as sources of phosphorus for maize'. In this trial the treatments are defined as different quantities of mulch (containing organic P) or inorganic P, the full set of treatments being:

1. Control with no input
2. Inorganic 12.5 kg P $ha^{-1}$
3. Inorganic 25 kg P $ha^{-1}$
4. Inorganic 50 kg P $ha^{-1}$
5. *Lantana* mulch 5 t $ha^{-1}$ (=12.7 kg P $ha^{-1}$)
6. *Lantana* mulch 10 t $ha^{-1}$ (=25.4 kg P $ha^{-1}$)
7. *Lantana* mulch 20 t $ha^{-1}$ (=50.8 kg P $ha^{-1}$)
8. *Tithonia* mulch 5 t $ha^{-1}$ (=14.7 kg P $ha^{-1}$)
9. *Tithonia* mulch 10 t $ha^{-1}$ (=29.4 kg P $ha^{-1}$)
10. *Tithonia* mulch 20 t $ha^{-1}$ (=58.8 kg P $ha^{-1}$)

The interest is not in the actual rates of P or mulch, but in the effect of changing rates. Does increasing P increase grain yield? Is the rate of change the same for different P sources? What rates of P application might turn out to be profitable? Is there an optimal rate? Does the response level-off and if so, is that level the same for each source? And so on. The summary graph (Graph 2) below reflects this interest in responses and response curves, rather than particular treatments and differences between them. This emphasis should continue in the formal analysis. It is certainly not helpful to carry out an analysis of variance followed by comparison of means that determines whether grain yield with 12.5 kg P $ha^{-1}$ is 'significantly different' from that with 25 kg P $ha^{-1}$, then again to compare 25 kg P $ha^{-1}$ with 50 kg P $ha^{-1}$ and so on.

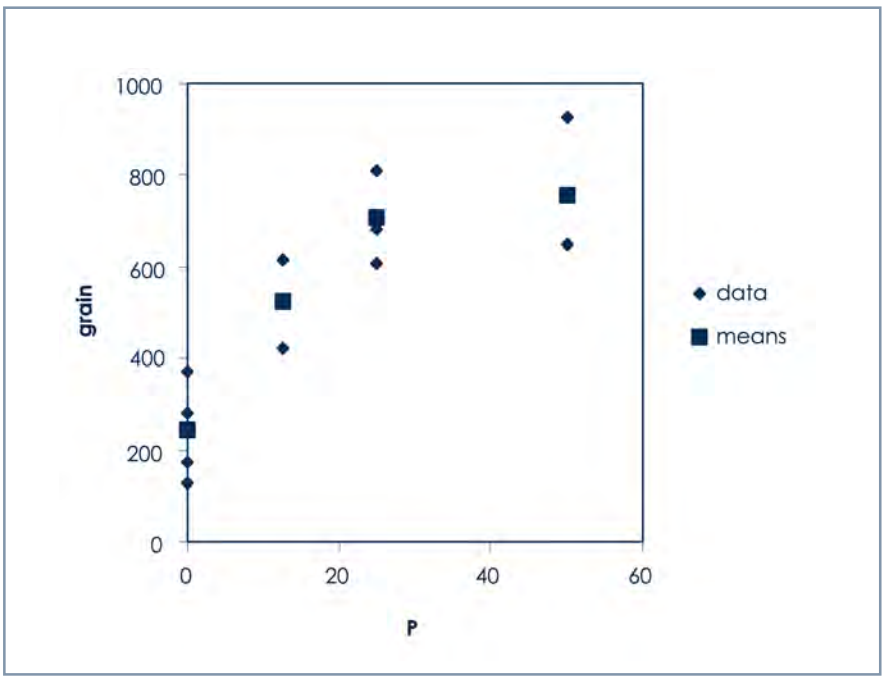
In order to illustrate the approach to analysis of this type we will restrict attention to treatments 1 to 4, the inorganic P inputs and zero input.

The usual analysis of variance (allowing for the unequal replication of treatments in each block) produces the following results in Table 13.

```
*** Accumulated analysis of variance ***

Change   d.f.         s.s.        m.s.     v.r. F pr.
+ bloc      2       94777.      47389.     3.90 0.082
+ rate      3      505187.     168396.    13.87 0.004
Residual    6       72834.      12139.

Total      11      672798.      61163.

*** Predictions from regression model ***

 Response variate: grain
   Prediction          S.e.
         rate
         0.00         244.9        55.4
        12.50         525.1        64.0
        25.00         705.0        64.0
        50.00         755.6        80.3
```

Table 13. Analysis of variance (allowing for the unequal replication of treatments in each block)

As shown in Graph 2 below, the analysis essentially estimates a mean for each of the four rates used.




Graph 2.

The model used is:

Response = (block effects) + $P_i$ + residual

Here $P_i$ is a set of 4 parameters, one for each of the 4 means. As these measure deviations from the block effects, there are actually 3 independent parameters to estimate. The model tells us nothing about what happens at other values of P input, and the ANOVA and table of means would be identical if the values of P used in the analysis were changed to any other numbers, or even just to A, B, C and D.

The graph (Graph 2) shows a clear response of increasing grain with increasing P, and this is not a straight line response. The simplest curve that might describe this is a quadratic:

Response = (block effects) + b.P+c.$P^2$ + residual

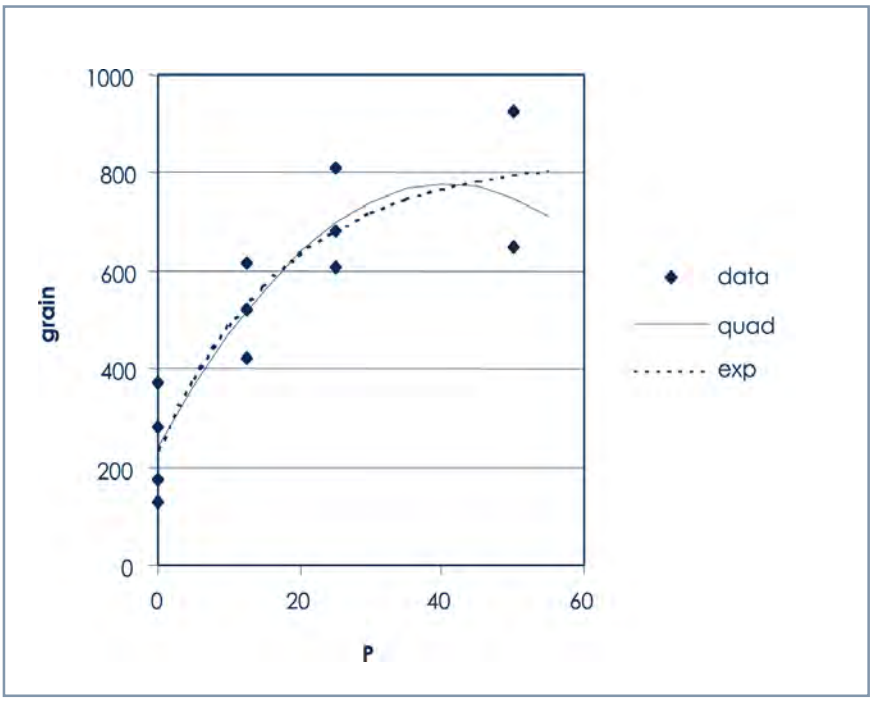
Here P is now the actual amount of P added and b and c are parameters to estimate. GenStat gives the following, in Table 14.

```
***** Regression Analysis *****

 Response variate: grain
   Fitted terms: Constant + block + P + P2

*** Accumulated analysis of variance ***

Change       d.f.        s.s.        m.s.       v.r.
+ block         2      94777.      47389.       4.55
+ P             1     411507.     411507.      39.55
+ P2            1      93678.      93678.       9.00
Residual        7      72836.      10405.

Total          11     672798.      61163.



*** Estimates of parameters ***

         estimate        s.e.
Constant    164.1         72.7
block 2     157.3         80.5
block 3      67.8         76.9
P            26.56         5.49
P2           -0.327        0.109
```

Table 14. GenStat Regression analysis

This shows that the estimate of b is 26.56 (se = 5.49) and of c is – 0.327 (se = 0.109) and the curve is represented on the graph in Graph 3 below. This is a great improvement. It allows us to make predictions about the response at levels of P input not actually used in the experiment, can be used as the basis for comparison between the different types of input, or could be used to estimate various optima.

This quadratic curve model has explained very nearly as much of the variation as the previous less useful model (residual SS of 72836 compared with 72834). Note also, that we have reduced the number of numbers needed to describe the results, from the 3 parameters of the first model to the 2 parameters of this model. This sort of 'parsimonious' modelling, using a description of the data that captures the important features while minimizing the number of things to estimate, can lead to increased precision in important quantities. For example, the b parameter represents the slope of the response curve at zero (i.e. the initial rate of increase of yield with increasing P). The quadratic model gives an estimate of this of 26.6 with an s.e. of 5.5. The best estimate of the same quantity we can get from the first model is the difference in the first two means divided by the increase in P, 12.5. This is 22.4 with an s.e. of 6.8. Not only is this estimate less precise, it is also not quite what we want, being the average slope between P = 0 and P = 12.5, not the value at P = 0.




Graph 3. Summary graph reflecting the interest in responses and response curves, rather than particular treatments and differences between them.

The curve used however is not ideal. The fitted quadratic shows a peak and then declines for high values of P. This shape cannot actually be seen in the data. Furthermore, widespread experience with crop response to P shows that a curve that levels off, rather than decreases for large P, is more usual. Finding a mathematical curve with such properties is not difficult. A suitable shape is given by:

$$\text{Response} = (\text{block effects}) + k.r^{P} + \text{residual}$$

Here k and r are parameters to estimate. The non-linear nature of the parameters means that a different method is used by the software to find them, but the principles of estimating then interpreting the curve are the same. GenStat estimates r as 0.94 and k as –553, the fitted curve also being on the graph (Graph 3) above.

The full analysis of this data could now continue by estimating similar curves for the other two sources of P and interpreting them.

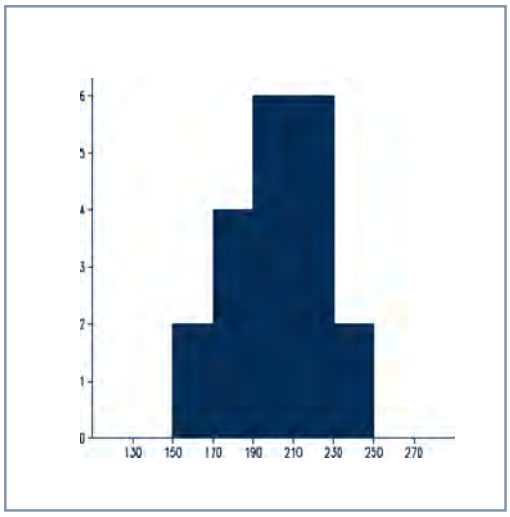
If there is more than one factor with quantitative levels, for example different amounts of several nutrients, or both within and between row spacing, then there are many more opportunities to improve the analysis by fitting realistic response surfaces and interpreting them.

## Random set of treatments

As a final example of treatment structure for which something other than simple estimation of means and contrasts may be appropriate, we look at the case in which treatments are not individually interesting because they only represent a sample from some large group. This is common in experiments concerned with breeding and genetic evaluation, but has other applications as well. The example '*Leucaena trichandra* seed production trial' is used to illustrate the ideas.

The treatment structure for this experiment is simple: there are 20 families of *Leucaena* to compare. Analyses in earlier sessions calculated means for each (perhaps adjusted for small blocks) and standard errors of differences between them. However the 20 families are a sample of those within some population. We are not so much interested in whether families 1 and 2 are different, but in how much variation there is between families and the pattern of such variation. Geneticists are able to make deductions from this about heritability and it helps in planning selection and breeding programmes.

Some simple exploratory analysis shows the nature of the variation in the 20 *Leucaena* families. The histogram (Graph 4) below is of the mean heights for the 20 families. The mean heights are spread between about 150 and 250 cm. There are no obvious clumps or outliers, but a single continuous distribution of values. The variance of these means is 562 $cm^2$.




Graph 4. Mean heights for the 20 families

The formal analysis of these data can proceed now using a similar idea as others in this session; modifying the model used so that it reflects the nature of the data and our analysis objectives more exactly. Earlier in this session we used the model:

response = replicate + family + small block effect + plot effect + tree effect

in which the last three terms are random effects, for which we wish to estimate a variance, but the family term actually represented a set of 20 parameters. The analysis above suggests that this should also be replaced by a random term for which we estimate a variance rather than 20 means. The REML results are shown in Table 15:

```
*** Estimated Variance Components ***

Random term Component      S.e.

family          517.       182.
smallblk          1.        40.
smallblk.plot   215.        75.
*units*        2463.       101.


*** Wald tests for fixed effects ***


  Fixed termWald statistic d.f.

  rep           85.4         19
```

Table 15. REML results taking family as random

The variance of the family term is estimated as 517 $cm^2$, very similar to the estimate of 562$cm^2$ calculated simply as the variance of the 20 means. The difference is due to allowing for the small blocks, and the fact that some of the variation in the family means is actually due to the variation in individual trees and plots, and REML removes that. The results show that the family variance is much smaller than the variation between individual trees within families, suggesting that selection of families may not be very effective.

Session 12

# Dealing with categorical data

*E. Allan, R.D. Stern, R. Coe and J. De Wolf*

## Categorical dependent variables

If we take a closer look at the dataset of the study ''Improved fallows and rock phosphate: farmers' experiences', we see that many responses are measured on qualitative scales. Often questions are answered with yes/no, absent/present, positive/neutral/negative, poor/good/very good. In other cases farmers are asked to give a score between 1 and 5. Analyses used so far, based on models, which assume constant variance and normal distributions, would not seem appropriate.

Session 11 described the general ideas of dealing with non-normal distributions using generalized linear models. These are our main tools for tackling this complication. The simplest situation is when the answer can only be one out of two possibilities (e.g. yes/no, absent/present, failure/success, false/true). In this case we have a binary response, which can take one of two values, conveniently coded 0 and 1.

Categorical data can also consist of more than two possibilities. Sometimes those categories do not have any particular order. For example, household type may have 3 classes: male headed (male present), male headed (male absent), female headed. If these are labelled 1, 2 and 3 the labels themselves are meaningless and the information in the data would be the same if they were changed. Household type is called a 'nominal categorical response'. In other cases, there will be some order in the various categories. Think about answers that can be 'never', 'seldom' or 'often', or that can be 'negative', 'neutral' or 'positive'. In these cases the answers are structured in a logical order that cannot be changed. Such responses are called 'ordinal' or 'ordered categorical responses'.

For simple cases of these categorical response variables, useful descriptive summaries of these can be constructed as tables of counts or percentages, with the response on one margin and a classifying variable (such as treatment) on the other. Examples were used in Session 4. It is harder to find good summaries when interested in the relationship between a categorical variable and continuous 'explanatory' variables.

A cross tabulation of counts is sometimes referred to as a contingency table. For two-way contingency tables a simple confirmatory test exists, the Chi-square test, but the limits of the applicability of this method will be soon reached, as we will see. For more general problems, a modelling approach is used. The logistic regression model introduced in Session 11, is explored further in this session as it is widely useful and the basis for other developments.

## Contingency tables

We said in Session 4 that cross-tabulations, or contingency tables, often give sufficient information for our report, without the need for formal analysis. Sometimes though, a more formal test is needed, and we consider this here.

As an example look at the most important benefit of improved fallows, reported by men and women: A tabulation of the data, either using Excel or GenStat, would produce the following summary in Table 1.

Observed values:

| | crop | seed | soil | weed | wood | total |
|---|---|---|---|---|---|---|
| F | 24 | 4 | 26 | 20 | 16 | 90 |
| M | 14 | 5 | 13 | 3 | 15 | 50 |
| | 38 | 9 | 39 | 23 | 31 | 140 |

Table 1.

These data are relevant to address the question: 'Do the genders differ in their perception of the most important benefit of improved fallows?' The patterns in the table are easier to see if the data are presented as percentages, as in Table 2.

| | crop | seed | soil | weed | wood | total |
|---|---|---|---|---|---|---|
| F | 27 | 4 | 29 | 22 | 18 | 90 |
| M | 28 | 10 | 26 | 6 | 30 | 50 |

Table 2.

It appears, for example, that women value the benefits of weeding more than men, maybe to be expected as they do much of the weeding. On the other hand more men value the wood produced in the fallow. Now a confirmatory test can be done, to check whether these patterns could just be due to sampling variation. A Chi-squared test is appropriate.

This Chi-square test calculates, for each cell in the table, the frequency that is 'expected' for each category, assuming no difference between males and females. A comparison is then made between the observed frequencies and the expected frequencies; the further these two are apart, the more convincing evidence there is to reject the hypothesis.

Totals are included in Tables 1 and 2 above, to show how expected values can be calculated. The table of expected values is given in Table 3 below, calculated as follows: If there is no gender effect then the proportion who say 'crop' is estimated to be 38/140. Hence we would expect that, of the 90 females in the study, 90 x 38/140 - i.e. 24.43, would respond with 'crop'. And so on for the rest of the cells in the table.

Expected values:

| | crop | seed | soil | weed | wood | |
|---|---|---|---|---|---|---|
| F | 24.43 | 5.79 | 25.07 | 14.79 | 19.93 | 90 |
| m | 13.57 | 3.21 | 13.93 | 8.21 | 11.07 | 50 |
| | 38 | 9 | 39 | 23 | 31 | 140 |

Table 3.

The test then compares the observed and expected values (Table 4), a large discrepancy being evidence against the hypothesis used to calculate expected values.

```
*Warning: some cells in the table of expected values have less than 5

Pearson chi-square value is 8.98 with 4 df.
Probability level (under null hypothesis) p = 0.062

Likelihood chi-square value is 9.71 with 4 df.
Probability level (under null hypothesis) p = 0.046
```

Table 4. GenStat's Chi-square test output

Many researchers will be familiar with the Pearson's Chi-square statistic, which is the commonly used form of the Chi-square test. What is less well known, though, is the other form, which can be used, the Maximum Likelihood Statistic. Both tests focus on the comparison of the observed and expected data under an assumption of no difference between males and females. It is therefore not surprising that they have similar chi-square values, at around 9, and similar p-values, close to the 5% significance level. There is no advantage in using one as opposed to the other. The Pearson statistic is the one used traditionally, because the hand calculation is easy; but most software packages now give both.

The GenStat warning, points out one limitation of the chi-square test, that the expected value of a particular cell in the contingency table should not be too small. As a rule, the test will be valid provided that fewer than 20% of cells have an expected count below 5, and none are below 1. The above warning is highlighting the low expected count for the 'seed' benefit for men, when no difference between men and women is assumed. This is one cell out of 10 (10%), and all other cells have an expected counts greater than 5, so the test is acceptable. This restriction about expected cell counts, implies that the Chi-square test is only useful for large data sets or for crude categorizations of data.

The low p-value(s) indicate that the hypothesis of no difference between men and women is difficult to sustain, i.e. there is evidence to suggest that there is a difference between the genders. From the contingency table we see that this is caused by the difference between the genders when it comes to both weeds and wood. For both of these benefits, the observed data are quite far from what one would expect if there were no difference between men and women. The chi-square statistic and the accompanying p-value make only a general statement concerning all categories. This situation is similar to when we used ANOVA and its F-test. For ANOVA a more detailed analysis using s.e.d.'s had to be used to look at particular differences. In the case of contingency tables this will have to be done by analysing sub-tables, for instance one that leaves out the category 'weed'.



The problem with these Chi-square tests is that they are only applicable for simple two-way tables. Unfortunately, this rarely corresponds to our analysis objectives. In Session 4 we saw that we needed a multi-way table to relate 'use of rock phosphorous' to 'the sex of the farmer' and 'their prior experience with the use of fertilisers'. More generally we might want to relate our response to a continuous explanatory characteristic such as farm size or amount of fertilizer used.

In the remaining sections we see how to analyse categorical dependent variables, when the structure of the data is more than just a simple comparison of groups. We begin with binary responses first and then consider how to deal with responses where there are more than two categories to choose from.

## Modelling binary responses

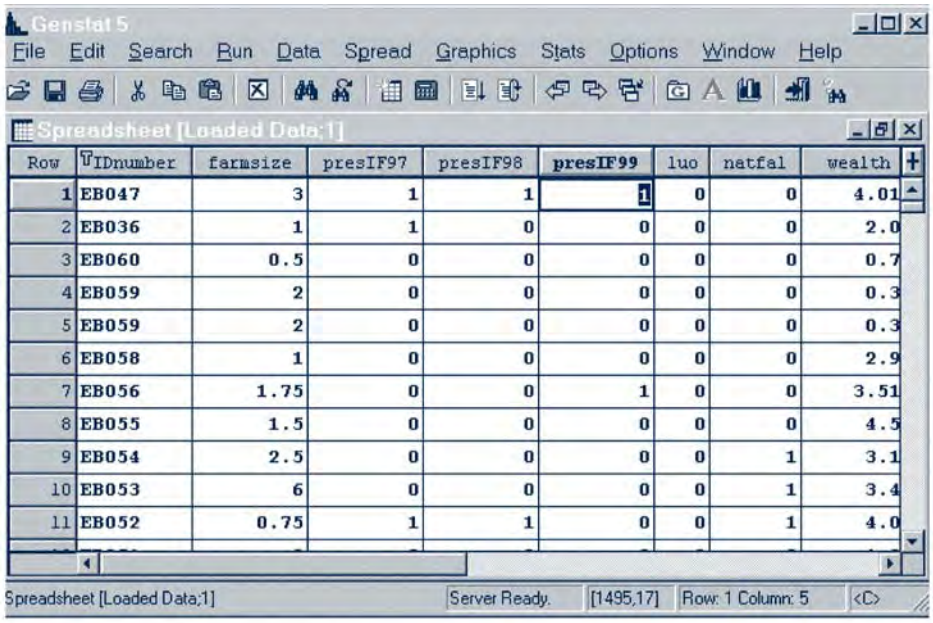
Many hypotheses or research questions concern a relationship between a binary response variable , such as 'whether or not improved fallows are used', and how this relates to (a set of) continuous or discrete explanatory variables. For example in our type 3 trial, we might want to investigate how large the plot under improved fallow can become (acreage) before it starts to

give the farmers problems with labour. This could be analysed by relating the indication of the labour problem (yes/no) to the plot size.

Another common situation is when we want to evaluate factors associated with a certain behaviour, for example 'How do users of improved fallow differ from non-users?' This would involve relating several explanatory variables to the user/non-user response.

To do this we move to fitting models to data, bearing in mind that our dependent variable is no longer continuous. The general principles of model fitting still apply. With continuous data, we used means and standard errors to summarize the pattern in our responses; we then went back to the raw data to fit a model, which assumed an underlying normal distribution. Here we use multi-way contingency tables to summarize the patterns in our binary responses, but go back to the raw data to model the effects of various features on this response.

Let us use the example of 'the determinants for having improved fallow' to illustrate the analysis of binary data. The raw data are depicted in Figure 1 below, each row representing a different farm.




| Row | IDnumber | farmsize | presIF97 | presIF98 | presIF99 | luo | natfal | wealth |
|---|---|---|---|---|---|---|---|---|
| 1 | EB047 | 3 | 1 | 1 | 1 | 0 | 0 | 4.01 |
| 2 | EB036 | 1 | 1 | 0 | 0 | 0 | 0 | 2.0 |
| 3 | EB060 | 0.5 | 0 | 0 | 0 | 0 | 0 | 0.7 |
| 4 | EB059 | 2 | 0 | 0 | 0 | 0 | 0 | 0.3 |
| 5 | EB059 | 2 | 0 | 0 | 0 | 0 | 0 | 0.3 |
| 6 | EB058 | 1 | 0 | 0 | 0 | 0 | 0 | 2.9 |
| 7 | EB056 | 1.75 | 0 | 0 | 1 | 0 | 0 | 3.51 |
| 8 | EB055 | 1.5 | 0 | 0 | 0 | 0 | 0 | 4.5 |
| 9 | EB054 | 2.5 | 0 | 0 | 0 | 0 | 1 | 3.1 |
| 10 | EB053 | 6 | 0 | 0 | 0 | 0 | 1 | 3.4 |
| 11 | EB052 | 0.75 | 1 | 1 | 0 | 0 | 1 | 4.0 |

Figure 1. Example of binary data

The variable 'presIF99' is our binary response; it takes the value of 1 if improved fallow is used in 1999, and 0 if it was not used.

The variables 'luo' records the ethnic group (either 0 or 1, indicating the Luo or Luhya ethnic group), and 'natfal' the use of natural fallow (either 0 or 1, indicating the absence or presence of traditional fallow). The continuous variable 'farmsize' gives the size of the farm in hectares.

Consider the simple question 'Is the use of improved fallows related to ethnicity?', which can be addressed by the contingency table (Table 5):

| | No | Yes | *Total* |
|---|---|---|---|
| Luo | 768 (84%) | 150 (16%) | 918 |
| Luyha | 429 (74%) | 148 (26%) | 577 |
| *Total* | 1197 | 298 | 1495 |

Table 5.

Instead of being seen as codes to indicate the use/non-use of improved fallows, the raw data can be regarded as a count of a particular event, i.e. whether or not improved fallow was used on a plot. And this is a count out of a total of 1 plot. For example, at farm EB047 improved fallow was used on 1 plot out of 1, whereas at farm EB036 improved fallow was used on 0 plots out of 1. Viewed like this, the data have a format similar to the seed germination data of Session 11 (number germinating out of the total on a dish), and can therefore be modelled using logistic regression.

In the dialogue box below (Figure 2), 'Number(s) of Subjects' receives the number of observations per unit, in this case 1. The box for 'Numbers of Successes' indicates the response variable, in this case 'presIF99'. The 'Model' boxes contain, as always, the description of our pattern, in this case the ethnicity of the farmer.

The link function is set, by default, to Logit. We will not discuss the effect of this transformation here except to say that it constrains the model to ensure that any estimated proportion cannot fall outside of the range 0 to 100%.

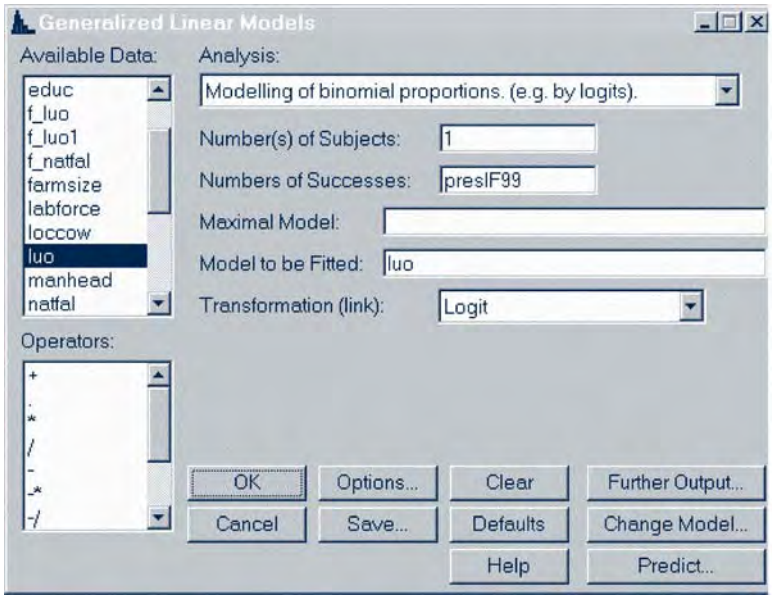



Figure 2. GenStat's Generalized Linear Model dialogue

As before, the output includes an analysis of deviance table and the model deviance (Regression deviance Table 6 below) is used to assess the significance of effects. What is different with binary data, compared to the binomial seed germination data, is that we cannot now interpret the residual deviance in any sensible way.

```
**** Regression Analysis *****

 Response variate: presIF99
 Binomial totals: 1
  Distribution: Binomial
 Link function: Logit
  Fitted terms: Constant, luo


*** Summary of analysis ***

             mean  deviance      approx
             d.f.  deviance    deviance   ratio chi pr
Regression      1       19.     18.8786    18.88 <.001
Residual     1493     1475.      0.9876
Total        1494     1493.      0.9996


* MESSAGE: ratios are based on dispersion parameter with value 1

Dispersion parameter is fixed at 1.00
```

Table 6. Regression analysis

The 'Regression' deviance is the deviance associated with our pattern, namely the ethnicity of the farmer. It follows a Chi-square distribution, and its value of 18.88 on 1 d.f. suggests a highly significant difference between the two ethnic groups ($p<0.001$).

In the same way as the regression modelling of Session 7 gives us mean values (often 'adjusted') for our effects of interest, the logistic model gives us expected, or predicted, proportions, or probabilities. Standard errors are approximate and provide a measure of uncertainty about these figures (Table 7).

```
*** Predictions from regression model ***

Response variate: presIF99
          Prediction   S.e.
   luo
  0.0000   0.1634     0.0122
  1.0000   0.2565     0.0181

* MESSAGE: S.e.s are approximate, since model is not linear.
* MESSAGE: S.e.s are based on dispersion parameter with value 1
```

Table 7.

The model estimates that 16% of Luo and 26% of Luhya farmers used improved fallows. [Another way of interpreting these predictions is in terms of probabilities, the probability of a Luo farmer using improved fallows is 0.16 whereas for a Luhya it is 0.26.]

Had we treated the data as a 2-way contingency table (Table 8), we would have got:

```
Pearson chi-square value is 19.24 with 1 df.
Probability level (under null hypothesis) p < 0.001

Likelihood chi-square value is 18.88 with 1 df.
Probability level (under null hypothesis) p < 0.001
```

Table 8.

Comparing this with our modelling approach, we see that the Chi-square deviance test for the ethnicity differences in our logistic regression is exactly the same as the likelihood ratio chi-square for the two-way contingency table. The two tests are identical, but modelling allows us to investigate more complex structures in the data.

To show this, let us investigate how the use of improved fallow is related to both ethnicity and the use of natural fallow. An exploratory analysis would require a three-way cross-tabulation of improved fallow use by ethnicity, and natural fallow use. From this one could see the effects of ethnic group and natural fallow use on the use of improved fallows, and also how each of these influences the other. The formal analysis is similar; we fit a model to the data that described the pattern in terms of the ethnicity and natural fallow main effects and their interaction.

The summary analysis of deviance table is no longer informative when there are several explanatory variables in the model, since it gives no information on the individual components of the pattern. By adding the model terms one by one, we can display the equivalent of the 'accumulated analysis of variance', namely the 'accumulated analysis of deviance'. Predictions are now presented in two-way tables (Table 9).

```
[Note: 9 farms did not have their use of natural fallow recorded. Hence the
analysis is based on only 1486 farms.]

*** Summary of analysis ***

mean deviance approx
             d.f.       deviance    deviance     ratio      chi pr
Regression      3            29.      9.7442      9.74      <.001
Residual     1482          1457.      0.9834
Total        1485          1487.      1.0011

Change         -1           -10.      9.5347      9.53      0.002
* MESSAGE: ratios are based on dispersion parameter with value 1

Dispersion parameter is fixed at 1.00


*** Accumulated analysis of deviance ***

Change                                  mean      deviance   approx
             d.f.       deviance    deviance      ratio     chi pr
+ luo           1        18.5728     18.5728      18.57     <.001
+ natfal        1         1.1250      1.1250       1.13     0.289
+ luo.natfal    1         9.5347      9.5347       9.53     0.002
Residual     1482      1457.4088      0.9834

Total        1485      1486.6413      1.0011
* MESSAGE: ratios are based on dispersion parameter with value 1
```

Table 9.

We have already discussed the benefit of fitting models to data to explore their structure properly. Here, with our binary data, we again see the benefit. By looking at not just the simple effect of ethnic group, but also whether farmers do or do not use natural fallow we learn more about the use of improved fallows. Not only is there a significant effect of ethnic group, but there is also strong significant ethnic group x natural fallow interaction ($p<0.001$ and $p=0.002$ respectively).

```
*** Predictions from regression model ***

 Response variate: presIF99
   natfal   0.0000       1.0000
         Prediction     S.e.    Prediction    S.e.
luo

0.0000     0.1373       0.0142     0.2132     0.0229
1.0000     0.2820       0.0257     0.2279     0.0254


*  MESSAGE:  S.e.s  are  approximate,  since  model  is  not  linear.
*  MESSAGE:  S.e.s  are  based  on  dispersion  parameter  with  value  1
```

Table 10.

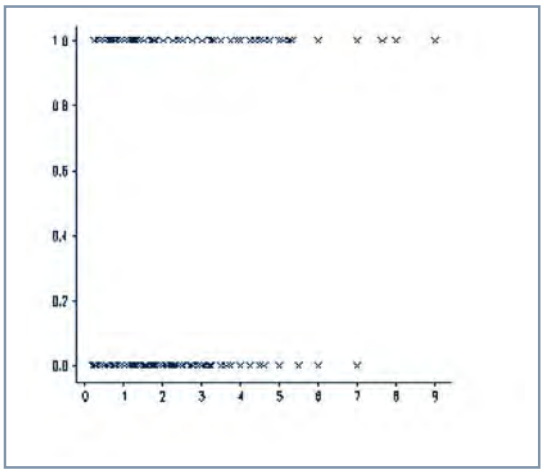
From the table of predicted proportions (Table 10) we see that the use of improved fallows in the two groups is influenced by whether or not the farmer uses natural fallows. Behaviour of the two ethnic groups with respect to use of improved fallows is much the same for those farmers that already use natural fallows. However for farmers who do not already use natural fallows, Luo are more likely to try improved fallow.

## Logistic models with continuous explanatory variables

Above we have looked at models containing categorical explanatory variables, but there is no reason to restrict the explanatory variables to just categorical ones; continuous variables can also be included. To show this, let us return to the question about relating farm size (continuous variable 'farmsize') to the use of improved fallow.

Few farms are very big (up to 31 acres), and the majority of the farms are well below 5 acres. Provided no error occurred, these large farms are certainly odd among the rest of the farms, and so for the purpose of this example we restrict the farm size data to the ones larger than 0 and less than 10 acres.

A point plot of the raw data is not informative, as demonstrated in Graph 1 below. The points come on two horizontal lines, but there is too much overlap to get a good impression of any relationship, or lack of relationship.

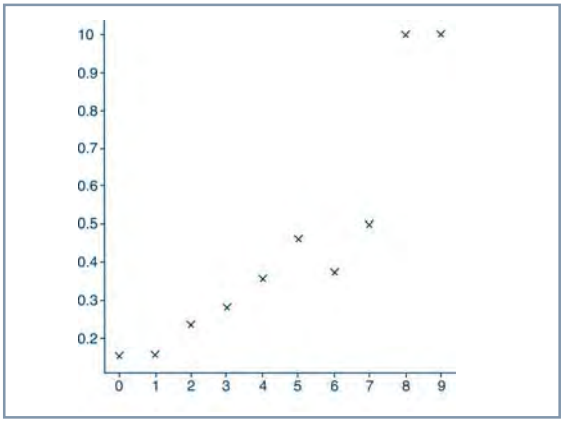



Graph 1. Point plot of presence of fallow on farms according to farm size

A better form of exploratory analysis is to group the observations for classes of farm sizes, say groups of farms with less than 1 acre of land, farms with between 2 and 3 acres of land, and so on. Within each class, proportions of farms with improved fallows can be calculated and plotted against the farm size. Graph 2 suggests that there is some positive relationship between use of improved fallows and farm size.

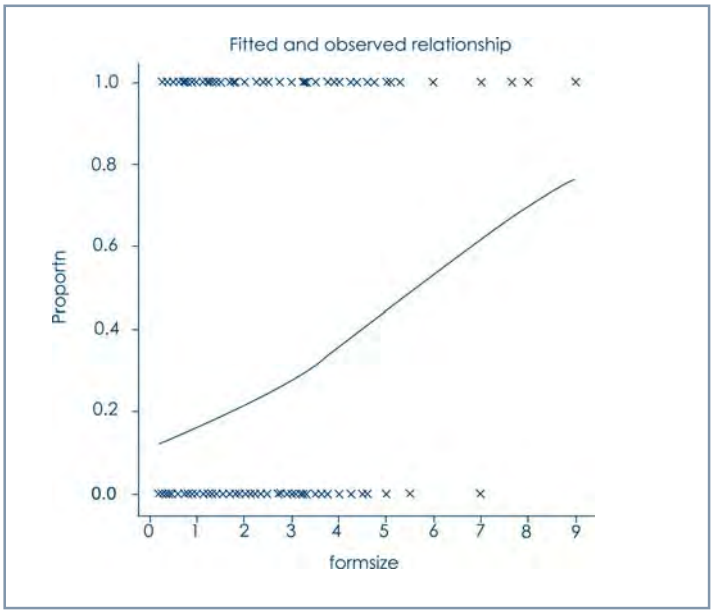
How can we model this to draw some firm conclusions? Our raw data are binary and so simple linear regression is clearly inappropriate. A straight line would not fit the points of the graph of the raw data at all. Moreover, we do not want the model to give values that are negative or are greater than 1, which would be meaningless in this context. The model should be such that for very small farms the probability goes asymptotically towards 0, whereas for very large farms it reaches asymptotically 1, in other words a flattened S-shaped curve, confined between 0 and 1. The logit transformation used in the logistic regression model produces this kind of function.

The specification of the model is no different from the ethnicity example, we declare the continuous variable 'farmsize' as our 'pattern'.




Graph 2. Proportion of farms with fallow according to farm size class

The fitted model can be displayed in a graph (via the **Further output…Fitted Model** buttons in GenStat, and filling in 'farmsize' as 'Explanatory Variable'.) Graph 3 shows the observations, either at 0 or 1, with an S-shaped curve between these two asymptotes. The fitted curve seems similar to the grouped data summary of Graph 2.




Graph 3. Logistic regression of presIF99 on farmsize

The output shows the analysis table of deviance for the regression (Table 11).

```
***** Regression Analysis *****

 Response variate: presIF99
 Binomial totals: 1
  Distribution: Binomial
 Link function: Logit
  Fitted terms: Constant, farmsize


*** Summary of analysis ***

                                mean    deviance   approx
              d.f.  deviance deviance     ratio     chi pr
Regression       1       51.   51.0346     51.03     <.001
Residual      1468     1412.    0.9617
Total         1469     1463.    0.9958
*  MESSAGE:  ratios  are  based  on  dispersion  parameter  with  value  1

Dispersion parameter is fixed at 1.00

* MESSAGE: The following units have high leverage:
Unit  Response Leverage
       30       1.00     0.0047
       33       1.00     0.0194
       90       0.00     0.0070
        …
     1320       1.00     0.0070
     1347       1.00     0.0070
     1382       1.00     0.0070
     1428       0.00     0.0116
```

Table 11. Analysis of deviance table for the regression

The message about 'high leverage' farms tells us that there are some farms, which have a strong influence on the fitted model. They are real to our experiment and tell us something about the use of improved fallows. However, if we were to explore the relationship with farm size further we might want to fit the model with and without these farms in order to see the extent of their influence.

The analysis above indicates that farm size is a factor that is related to having improved fallow in 1997.

As with ordinary linear regression modelling, the results include parameter estimates. The parameter estimate for 'farmsize', highlighted below in Table 12, is positive and therefore tells us that the probability of having an improved fallow on a particular farm **increases** with the farm size. This result is intuitive, as a farmer has to have enough land available to allow some to be taken out of production during a cropping season.

```
*** Estimates of parameters ***

                                                    antilog of
             estimate      s.e.     t(*)    t pr.    estimate
Constant     -2.043      0.117  -17.44    <.001      0.1296
farmsize      0.3603     0.0503   7.16    <.001      1.434
*  MESSAGE:  s.e.s  are  based  on  dispersion  parameter  with  value  1
```

Table 12. Parameter estimate for 'farmsize'

We can also use the model to predict the probability of having improved fallow for a certain farm size, just as in the earlier analysis of ethnicity, etc., and just as one would in simple linear regression. The model that has been fitted to p, the chance of using an improved fallow, is

$$\log(p/1-p) = -2.043 + 0.3603 \times farmsize$$

or

$$p = \frac{\exp(-2.043 + 0.3603 \times farmsize)}{1 + \exp(-2.043 + 0.3603 \times farmsize)}$$

This is the curve plotted in Graph 3.

All of the above shows us how a continuous explanatory variable can be included in logistic regression modelling, and how the output can be interpreted. That, combined with the previous section, means that we can build quite complex models for binary data. Modeling is based on the raw binary response data and the models can incorporate any mix of factors, variates and interactions as explanatory variables. Analysis of deviance tables can be used to assess the significance of effects. Multi-way tables of predicted proportions can be used to summarize the results for the categorical explanatory variables. Parameter estimates can be used to quantify linear relationships.

## Modelling responses with more than two categories

We saw in the section on contingency tables that we could deal with two-way tables and that the Chi-square test provided a way of making some inference from these tables. Whilst this is suitable for a response variable with any number of categories, a limitation to the method was the rather limited applicability. We could only deal with one single independent variable.

We saw in the next section that logistic models can deal with more complex models containing several independent variables and their interactions, and they can describe both categorical and continuous explanatory variables. A drawback of the logistic model in the overall discussion of analysing categorical response data, is that the model is only suitable for responses that contain only two categories.

A simple, and often satisfactory, way of dealing with several categories of response is to reorganize the categories or ignore some of them temporarily to reduce it to a binary problem. This is often a logical way of dealing with the data when there is some hierarchy in the responses, e.g. 'poor/good/very good' could be reorganised to 'poor/(very)good' or the five household types could be reduced to the types 'male-headed household' and 'female-headed household'. Lumping or ignoring categories could also be a satisfactory solution when very few cases have been recorded for some categories. Sometimes it could also be informative to look at only the two extreme categories, e.g. in understanding an answer that can be either 'negative', 'neutral' or 'positive', most informative cases will be the one that answered 'negative' or 'positive'.

Another strategy is to divide a response variable with several categories into a series of binary categories. For example, the response to a question 'Which type of trees do you intend to plant?' might be 'none', 'fodder trees', 'fruit trees' or 'both fruit and fodder trees' . This could be broken down as variables that indicate intention to plant any tree, intention to plant fodder trees, and intention to plant fruit trees. Analyses of these will not be independent but if the variables correspond to analysis objectives this does not matter.

However, if we want to use all the categories of response at once, which is quite likely when there is no ordering to them (e.g. as in the example of benefits of improved fallows), then logistic models are not appropriate.

In the particular case of all independent variables being categorical, so that we can build a multi-way frequency table with the data, a solution is to use log-linear modelling. This is also part of the larger family of generalized linear models, just like the logistic model, but the logic behind it is different from the way we deal with our other models. We do not discuss the details

of the approach here, other than to say (i) that it can be used to model patterns in a multi-way contingency table, provided that the data are not sparse, and (ii) only categorical explanatory variables are possible in the model.

This last point is important, since explanatory variables can often be continuous (farm size, amount of fertilizer, etc.). In these circumstances there is no one 'correct' approach to the analysis of responses with more than two categories. If many of explanatory variables are continuous, one option is 'categorizing' these and fit a log-linear model. Alternatively, the data can be recoded as binary data and logistic models fitted to them, using the explanatory variables in their continuous form.

The answer to which approach to take, will depend on the questions of interest, the amount of data and the number of continuous explanatory variables in each particular circumstance. The log-linear approach will provide information about all the categories of response, whereas with the logistic regression approach some information is lost. On the other hand the logistic approach also allows us to quantify any linear relationship with the explanatory variables (if they exist), whereas the log-linear approach cannot. Also if there are several continuous explanatory variables to be 'categorized', it is not always possible to fit all the appropriate main effects and interactions using the log-linear approach, unless the dataset is quite large. This is much less of a problem with the logistic regression approach.

Session 13

# Getting more out of on-farm trials and multilevel problems

*R. Coe, R.D. Stern, E. Allan*

## Introduction

In Sessions 4, 5 and 6 the analysis of data involved calculating summaries of 'patterns', often treatment means, and then assessing the uncertainty in the means. The uncertainty depends on the 'unexplained' variation in the data. In simple experiments such as a randomized block design this is interpreted as the plot-to-plot variability within blocks. In Session 8 the idea was extended to split-plot designs. There are now two types of plot (main-plots and split-plots) and hence two different variabilities - between main-plots and between split plots. These show up as two error lines in the analysis of variance table, and are important in making inferences.

The split-plot design is an example in which 'information' occurs at more than one level. The treatments are applied to two different levels in the layout hierarchy. The measurements are taken at one level (the split-plot), but important information occurs at two levels. There are many other cases in which information and perhaps measurements and treatments occur at more than one level in a design. When the design is as 'neat' as a split-plot experiment, then the analysis poses no problems. However, in the general case the analysis can become complex, and in Session 11 we considered the types of complexities that might occur.

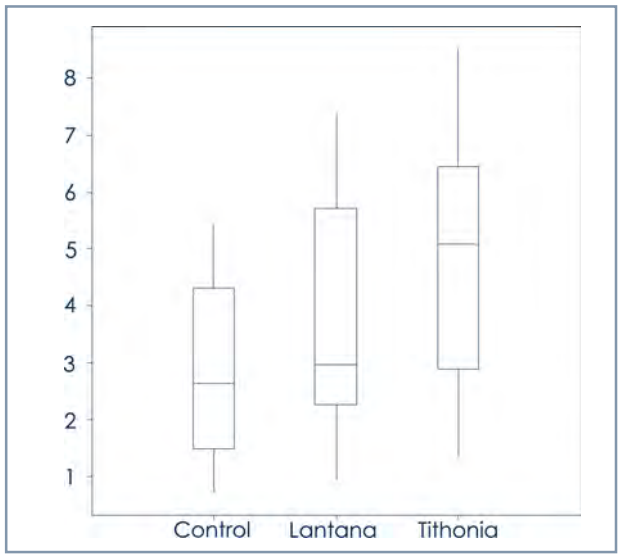
In Session 11 we looked at each type of complexity separately, namely either complexities in the treatment structure, the layout factors, or in the types of measurement being taken. Here we look in more detail at problems when there are measurements at multiple levels in the context of on-farm trials. Many on-farm trials give rise to multi-level information, but it is not unique to them. We explain how to recognize when multi-level information may cause problems in the analysis, introduce some simple methods of handling this that are useful in most cases, and show that more general methods are available.

The example used in this session is the on-farm trial 'Effect of *Tithonia diversifolia* and *Lantana camara* mulches on crop yields in farmers' fields'. The motivation for the analysis is the need to extract more information from the trial, to move beyond the simple comparison of treatment means, hence the session title 'Getting more out of on-farm trials'.

## What additional information is there?

We will start by looking at the data for the Central district. The trial has 16 farmers in this district, each comparing grain yield from three treatments, *Tithonia* mulch, *Lantana* mulch

and a no-mulch control. A boxplot (Graph 1) suggests that there is an increase of yield with both *Lantana* and *Tithonia,* the increase being larger with *Tithonia*. However there is a lot of variation about the mean patterns. The observations come from 16 different farms. We can separate the farm and treatment variation using ANOVA or fitting a simple model. The results are shown in Table 1 below.




Graph 1. Boxplot showing yields of *Lantana* and *Tithonia*

```
***** Analysis of variance *****
Variate: grain

Source of variation d.f.      s.s.         m.s.      v.r.      F pr.

farmer stratum       15    154.4270      10.2951    14.31

farmer.*Units* stratum
treat                 2     37.2980      18.6490    25.92      <.001
Residual             30     21.5869       0.7196

Total 47  213.3118


* MESSAGE: the following units have large residuals.

farmer 11     *units* 1      1.66      s.e. 0.67
farmer 11     *units* 2     -1.86      s.e. 0.67


***** Tables of means *****

Variate: grain

Grand mean 3.80

  treat     Control     Lantana     Tithonia
            2.76        3.74        4.91


*** Standard errors of differences of means ***

Table      treat
rep.       16
d.f.       30
s.e.d.      0.300
```

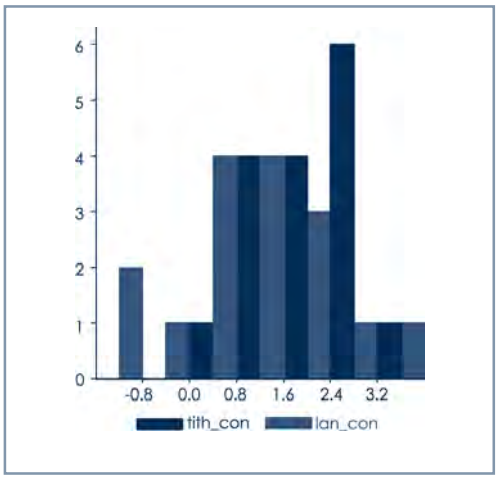
Table 1. Analysis of variance

These results are clear but they do not answer all the objectives or use the data very well, and suggest further questions. For example, the residual mean square of 0.72 gives a residual standard deviation of 0.84. Thus individual plots may be expected to vary by ±1.7 in addition to the effect of the treatment within a single farm, which is quite a lot. Can we explain any more of the variation?

The residual line can also be interpreted as the treatment x farmer interaction. It measures the extent to which treatment effects seem to vary between farms. This is worth looking at directly and can be seen by calculating a treatment comparison for each farm and looking at its variation. In the example here, there are three treatments, therefore more than one possible difference between treatments to consider. Two are calculated: *Lantana*-control and *Tithonia*-control. Summary statistics (Table 2) and histograms (Graph 2) are given for the grain yield value of the contrasts.

| | *Lantana*-control | *Tithonia*-control |
|---|---|---|
| Mean | 0.98 | 2.16 |
| Variance | 1.67 | 0.77 |
| | | |
| Minimum | -1.08 | 0.64 |
| Lower quartile | 0.12 | 1.36 |
| Median | 1.18 | 2.26 |
| Upper quartile | 1.89 | 2.93 |
| Maximum | 3.39 | 3.62 |

Table 2. Summary statistics for treatment differences




Graph 2. Histogram of treatment differences

The mean differences are the same as the difference in treatment means. What we can now see though is how the size of the treatment effect varies between farms. *Lantana* gives a mean grain yield increase of 0.98 t $ha^{-1}$. However individual farmers experience anything from a decrease of 1.08 t $ha^{-1}$ to an increase of 3.39 t $ha^{-1}$. Understanding what distinguishes a –1 farmer from a +3 farmer is clearly important. Several variables have been measured that might explain some of this (in fact they were measured because, as the experiments progressed, researchers noted what was happening on the farms and guessed, or hypothesized, that they would be important). There were three management variables (manure, weeding and incorporation of the mulch) and three weed/disease variables (striga, streak and couch). To start with we will look at streak, (Table 3) recorded as two levels (high or low).

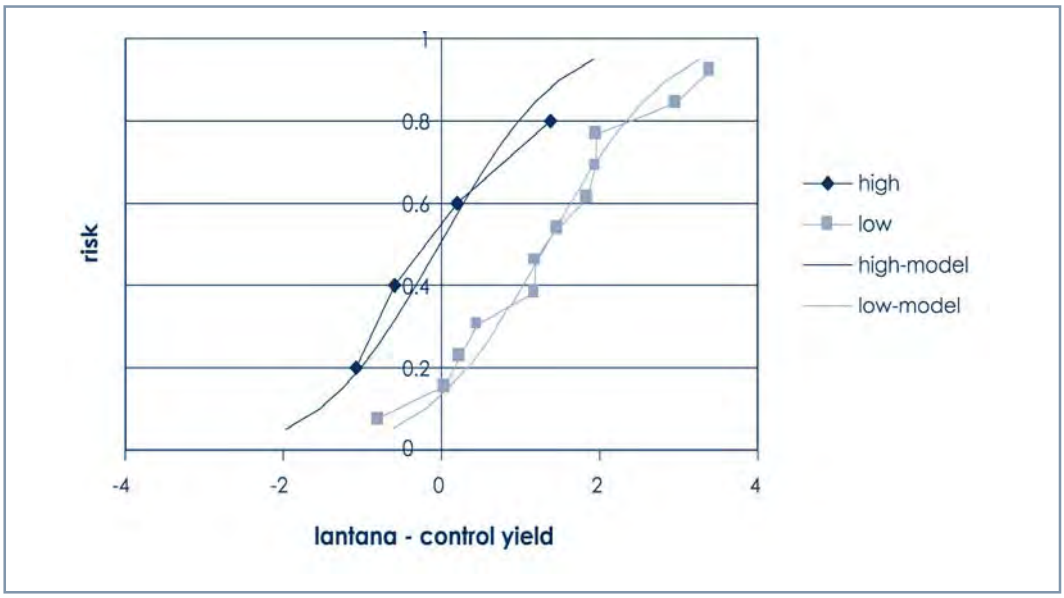
```
*****  Regression  Analysis  *****

 Response variate: lan_con
   Fitted terms: Constant, streak1


*** Summary of analysis ***

                d.f.        s.s.          m.s.   v.r. F pr.
Regression         1        5.36         5.360   3.82 0.071
Residual          14       19.65         1.403
Total             15       25.01         1.667

Percentage variance accounted for 15.8
Standard error of observations is estimated to be 1.18


*** Estimates of parameters ***

            estimate        s.e.    t(14) t pr.
Constant      -0.022       0.592    -0.04 0.970
streak1 L      1.337       0.684     1.95 0.071


Response  variate:  lan_con
                   Prediction         s.e.
            streak1
                  H     -0.022        0.592
                  L      1.314        0.342
SED  =  0.68
```

Table 3. Regression analysis of streak

There is some indication that there is an effect of streak. With low streak the average effect of *Tithonia* is to increase yield by 1.3 t ha$^{-1}$. With high streak the average increase is zero, but the se of the difference between these is 0.7 t ha$^{-1}$.

The residual variance is still large, showing there is a large variation between farms in the effect of *Lantana*, even after allowing for the fact that some farms have streak and others do not. Before looking at ways to explain more of this variation we show how the model can be interpreted more informatively than simply looking at means.

The model tells us that the farms fall into two groups defined by high and low streak. The variation between farms in each group is an indication of the risk a farmer faces when choosing to use the *Lantana* technology. For example, of the 12 farms with low streak, only 2 observe a yield increase of more than 2 t ha$^{-1}$. We could estimate the risk of a farmer getting less than 2 ha$^{-1}$ as 10/12 = 0.83. In fact a better estimate is 10/(12+1) = 0.77. To see why this is true, think what happens if we ask for the chance of less than 3.5 t ha$^{-1}$. This is larger than the biggest observation, so we would estimate the chance as 12/12=1. Yet a probability of 1 means certainty, and although we would say it is likely that the yield increase is less than 3.5, it would be unwise to claim that a yield increase of greater than that is impossible. These estimates have been plotted in Graph 3 below for both groups of farmers.




Graph 3.

The problem with this analysis is that there is very little data on which to base conclusions for the high streak group; just 4 farms. If we put confidence intervals around the risk estimates, they would be very wide. In doing the modelling above we made some strong assumptions about the data; that the two groups have different means but the same variance between farms, and that the variation follows a normal distribution. If we are happy with these assumptions then we can use the model to produce risk estimates. For example, the probability that a farm with low streak produces a yield increase of less than 2 t $ha^{-1}$ is the probability that an observation from a normal distribution with mean 1.314 and variance 1.667 is less than 2. This can be calculated or found from tables and is 0.702. It is calculated as the probability that an observation from a standard normal distribution is less than (2.000-1.314)/ √1.667.

In Excel, for example, this would be found as

=NORMINV((2.000-1.314)/√1.667)

and in GenStat as

CLNORMAL(2.000;1.314;1.667).

These estimates are also plotted on the graph above (Graph 3) and seem to give a good representation of the data.

Now we can attempt to explain more of the variation by adding other variables to the model. This is nothing new: simply incorporate variables for which there is reason to believe that they may be important and note which do indeed explain more of the variation. The results below (Table 3) show that adding another disease variable (striga) and two management variables (number of weedings and whether the mulch was incorporated) make little difference.

There is no strong evidence that any of these are responsible for the farm-to-farm variation in the size of the effect of *Lantana* on grain yield.

```
*** Accumulated analysis of variance ***

Change          d.f.          s.s.          m.s.      v.r. F pr.
+ streak1          1         5.360         5.360      3.57 0.088
+ striga1          1         0.525         0.525      0.35 0.568
+ weeding1         2         2.716         1.358      0.90 0.436
+ incorp2          1         1.374         1.374      0.91 0.362
Residual          10        15.033         1.503

Total             15        25.007         1.667
```

Table 4

## Reviewing the strategy

Looking back at the analysis in the previous section we can see that the strategy might be characterized as follows.

1. The data look like the following:

| Farmer | Streak | Treatment | Yield |
|---|---|---|---|
| 7 | L | *Tithonia* | 6.78 |
| 7 | L | *Lantana* | 6.02 |
| 7 | L | Control | 4.08 |
| 8 | H | *Tithonia* | 2.91 |
| 8 | H | *Lantana* | 2.14 |
| 8 | H | Control | 0.76 |
| 9 | H | *Tithonia* | 4.96 |
| 9 | H | *Lantana* | 2.38 |
| 9 | H | Control | 2.97 |
| 10 | L | *Tithonia* | 5.66 |

Table 5.

The level of streak varies between farms. The treatment and grain yield vary within farms.

2. The data are rearranged and the treatment effects (*Lantana* – control) and (*Tithonia* – control) are calculated as shown in Table 6:

| Farmer | Streak | *Tithonia* yield | *Lantana* yield | Control yield | *Tithonia*-control | *Lantana*-control |
|---|---|---|---|---|---|---|
| 7 | L | 6.78 | 6.02 | 4.08 | 2.70 | 1.94 |
| 8 | H | 2.91 | 2.14 | 0.76 | 2.15 | 1.38 |
| 9 | H | 4.96 | 2.38 | 2.97 | 1.99 | -0.59 |
| 10 | L | 5.66 | … | | | |

Table 6.

3. Now the two levels of variation in the data (between farm and between plot) have been reduced to a single level. For any treatment contrast such as *Lantana*-control the variation between farms can be explored, described and understood.

The advantages of this method include:

- It is intuitive. If the aim is to find out how a treatment effect varies across farms, then it is natural to calculate it on each farm and look at the variation.

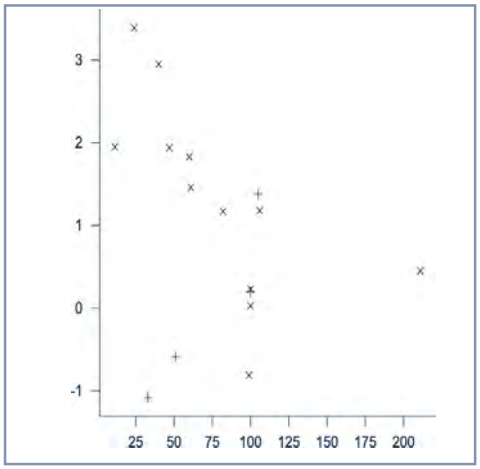
- It can be 'tuned' to the objectives. The effect that is studied, such as *Lantana* – control, can be chosen to be one, which is directly related to the objectives.

- In the example above, each farm has each treatment. The method works equally well if some farms do not have all treatments. We simply use those farms which did have each treatment needed to estimate the contrast.

- The relevant variation can be described and summarized simply. The histograms and boxplots shown earlier in this section are examples.

- The statistical modelling of the effects is straightforward.

Note that the regression models used can include factors (such as the disease and management variables used above) or continuous variables. An example of a continuous variable is the amount of mulch applied. The scatter plot (Graph 4) shows that if there is an effect of amount of mulch on the value of the *Lantana* – control contrast then it depends on the level of streak. Further more, there seems to be a negative effect of amount of mulch on low streak farms. This sort of counter intuitive result needs careful investigation before conclusions are reached.




Graph 4. Lantana -control contrast v mulch (+high streak) (x low streak)

While the approach described above is powerful and flexible, it has some disadvantages.

## Many possible contrasts

In the example used so far in this session, there were just 3 treatments. Normally this would suggest it is sensible to look at two contrasts, and we analysed both *Lantana*-control and *Tithonia*–control. Another sensible pair of contrasts would be mulch-control and *Tithonia* mulch–*Lantana* mulch. It might even be of interest to look at all 4 of these contrasts if each corresponds to a particular question of objective. However we would be repeating information, as they are not 'orthogonal', implying that we are just presenting the same information in a different format.

Now think what would happen if there were 10 treatments, a situation not at all uncommon, for example when farmers are evaluating different species for timber production. There would be at least 9 contrasts, maybe many more, if we are happy repeating the same information in different formats. Each of these would require a different analysis, exploring factors associated with the variation. This possibly large number of separate analyses is unsatisfactory as we might expect some common patterns amongst them. The method gives no obvious way to identify and describe these common patterns. A better approach would be to model all the data together, rather than each contrast separately.

## Incomplete designs

The case for an overall approach becomes even stronger if we look at incomplete designs. Imagine the *Tithonia* – *Lantana* experiment carried out with the same three treatments but with only 2 on each farm. Three of the farms might have the following (Table 7):

| Farm | *Lantana* | *Tithonia* | control | contrast | value |
|---|---|---|---|---|---|
| 1 | ✓ | | ✓ | l – c | $Y_1$ |
| 2 | | ✓ | ✓ | t – c | $Y_2$ |
| 3 | ✓ | ✓ | | l - t | $Y_3$ |

Table 7.

If we want to look at the *Lantana*–control contrast then the method described so far will only use information from the first farm, the value $Y_1$. However the other two farms also tell us something about the contrast as $Y_3 - Y_2$ also measures the size of *Lantana*–control. The method described so far ignores this information. If there are just a few farms without the complete set of treatments this lost information will not be important. However some experiments are deliberately planned to have this feature. For example, we might want to compare 10 species but there is only room for three on each farm. The method described so far will then miss a substantial part of the relevant information. The result will be s.e.d.'s of treatment effects that could be much larger than those from an analysis that uses all the information.

## Missing main effects

The analysis described so far, has concentrated on identifying the interactions between treatments and other factors, for example the level of streak virus. Once we have identified that streak does alter the effect of the treatments we may want to step back and find the overall effect, i.e. the main effect, of streak. The analysis presented so far does not show this. The main effect of streak, that is, the average yield on high streak compared with low streak farms, could of course be found but it requires a separate analysis of the data. We could take the average yield across the three treatments on each farm, then model the effect of streak, as shown in the output in Table 8 below, in which there is no sign of an overall effect of streak on the farms in the Central district.

```
***** Regression Analysis *****
*** Accumulated analysis of variance ***
Change          d.f.          s.s.          m.s.          v.r.
+ streak1          1         0.105         0.105          0.03
Residual          14        51.371         3.669
Total             15        51.476         3.432

*** Estimates of parameters ***
            estimate          s.e.          t(14)        t pr.
Constant       3.661         0.958          3.82         0.002
streak1 L       0.19          1.11          0.17         0.868
```

Table 8.

This analysis is correct and informative, but it raises some questions. For example, how would we deal with farms that had only 2 of the 3 treatments present? The average yield for such a farm is not comparable with the average for farms that have all three treatments.

## Mixed levels of variation

The analysis presented so far has depended on a feature of the data, which is unlikely to be found in many studies. We formed the contrast, say *Lantana* – control, for each farm and then related the value of the contrast to the level of streak for the farm. Using multiple regression we were able to look simultaneously at the effects of streak, striga, weeding and incorporation of the mulch. However we did not include the level of couch infestation or of manure use. Looking at the data shows that these two variables have a different structure from the first four. They are not constant within a farm. Each farm has either high or low striga, so we know the value to attach to the contrast when doing the analysis. However, some farmers used manure on one plot but not on another. Likewise some farms have high couch in one plot but not in others. Thus we have a value of 'couch' for each plot, not for each farm.

Again, there are simple ways to overcome the problem, but which are not altogether satisfactory. We could omit from the analysis those farms, which do not have a constant value for couch. In fact in this data set there is only one of them, so that would not have much impact

on the results. But there are an additional 6 farms that do not have constant levels of manure, so we would end up omitting 7 farms. In other studies an even greater proportion of the whole data set may have to be dropped.

Each of the problems outlined above suggest the need for an approach, which uses the available data from variation both within and between farms in an integrated way, and is not dependent on particular regular or simple patterns that emerge. Such a method is introduced in the next section.

## Assumptions behind the analyses

When introducing different analyses earlier, we pointed out that any analysis makes certain assumptions about the data. It is essential that the analyst understands what these assumptions are, and is able to check that they are reasonable. The analysis presented so far appears to be a standard regression analysis, so the usual assumptions made (observations are independent with constant variance and roughly normal distribution) still apply.

Looking carefully at the problems, shows that there are other assumptions made which may or may not be reasonable. For example:

- When we find that every treatment does not occur on every farm, there are a number of reasons why that may be the case. The trial may have been designed that way (for example there was not room for all treatments) with the researcher allocating treatments to farms. The farmers may have selected the treatments they were interested in, or some treatments may have failed, with no data collected. In these latter cases there is information in the 'missing data'. This information needs to be included in the overall analysis. Then the further results are 'conditional on data being collected'. That is, a mean treatment difference is an estimate of the treatment difference that would occur on the types of farm that can or choose to use the treatments, not an estimate for all farms.
- How were farms selected? Some of the interpretation of the results requires that the farms were a random selection of those in the district. This is unlikely to be the case, but we can check that they behave like a random selection. For example, we could check that they are not all farms of richer people, or of those that consistently get higher than average crop yields.
- Similarly, we need to know something about how plots were chosen within each farm and how treatments were allocated to them. If the farmers are selecting the plots and treatment allocation, then they may deliberately do something different from the 'as

random' selection we assume. For example, when farmers know mulch is expected to improve fertility and crop performance they may deliberately put the mulch treatments on an infertile plot and the no mulch control on a fertile plot. Their aim may then be to see if the mulch can improve an otherwise poor plot. If this was the method of choosing plots and allocating treatments then a zero treatment difference is actually impressive!

- When regression models are used to estimate the mean effects of treatments of other effects, the analysis is robust to the assumption of normal distributions. That is, the residuals can have a distribution, which is distinctly non-normal and yet the results are still valid. The same is not true when we start interpreting the models in terms of risks for individual farmers. The estimates then depend crucially on the normal distribution assumption being realistic.

## Modelling between and within farm variation

When analyzing simple experiments, such as randomized block designs, we started using GenStat's analysis of variance. The methods worked well until we were faced with data sets that did not have the regular patterns (such as each treatment occurring exactly once in each block) required. The solution was to write down a model that describes the pattern and residual, and then fit the model using methods, which do not require the data to have a particular structure. We use the same strategy here.

The key to finding a suitable model is understanding the structure of the data. In the current example the critical features are:

- observations of the response (grain yield) are made at the plot level,
- treatments are applied to plots,
- covariates are measured at plot or farm level,
- there will be variation within farms and additional variation between farms.

We therefore need to find a model that can describe these features, in particular the variation and measurements made at the two levels, something the standard 'regression' methods do not allow for.

The general form of the model is still:

(plot level data) = (pattern) + (residual)

The pattern will be both the treatments and any other explanatory variables (streak or manure, for example). The residual is more complex than in earlier models. We know that there is random variation contributing to the residual, both between plots, within farms and between farms. We therefore break the residual down into these two components and during model fitting, estimate the variance of both.

(plot level data) = (pattern) + (between farm variation) + (within farm variation)

The model is numerically complex to fit, and many statistics packages will not cope. The output may also look non-standard, but is often easy to interpret. In GenStat the REML commands complete the analysis. The dialogue below in Figure 1 is suitable for analysis of the streak effect and its interaction with treatment, for comparison with earlier results. The specification of the model is self-explanatory once it is remembered that GenStat automatically adds the plot level residual variation term to the model specified, called **units** in the output below.

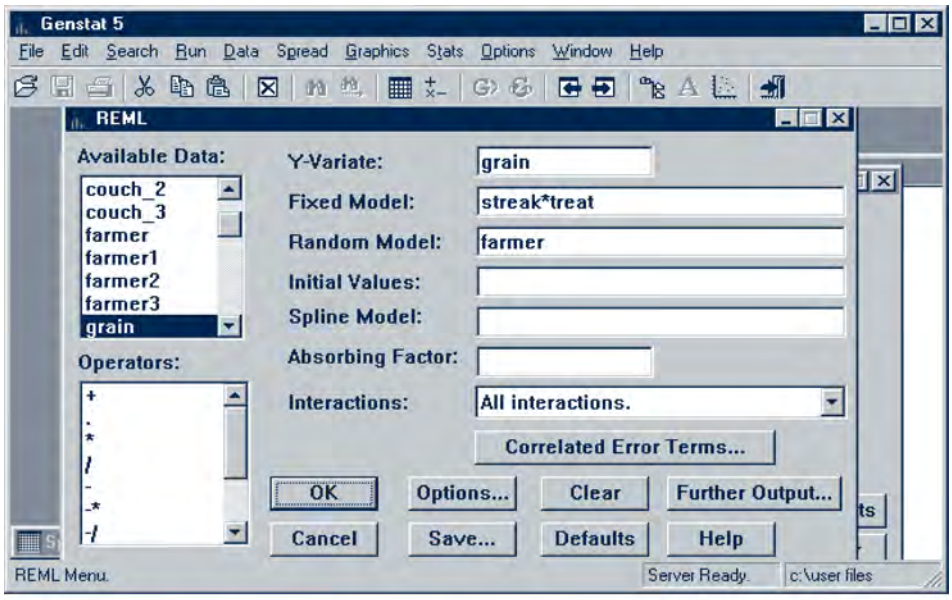



Figure 1. GenStat's REML dialogue box

Part of the output is shown below in Table 9.

```
*** Estimated Variance Components ***

Random term     Component          s.e.

farmer             3.4767        1.3878


*** Residual variance model ***

TermFactor   Model(order)     Parameter    Estimate         s.e.

Residual         Identity        Sigma2       0.578       0.1545


*** Wald tests for fixed effects ***

  Fixed termWald statisticd.f. Wald/d.f.Chi-sq prob

* Sequentially adding terms to fixed model

  streak             0.03             1        0.03        0.866
  treat             64.53             2       32.26       <0.001
  streak.treat       9.35             2        4.67        0.009


* Message: chi-square distribution for Wald tests is an asymptotic
           approximation
 (i.e. for large samples) and underestimates the probabilities in other
cases.


*** Table of predicted means for treat ***

 treat            Control       Lantana    Tithonia
                    2.779         3.425       5.060

Standard error of differences:   0.3104



*** Table of predicted means for streak.treat ***

 treat         Control           Lantana    Tithonia
streak
    H               2.825         2.803       5.355
    L               2.732         4.047       4.764

Standard error of differences: Average      0.8671
Maximum             1.163
Minimum             0.3104

Average variance of differences:0.8880

Standard error of differences for same level of factor:

streak          treat
Average             0.4240        1.163
Maximum             0.5376        1.163
Minimum             0.3104        1.163
Average  variance  of  differences:
                    0.1927       1.352
```

Table 9.

The output looks a little different from that produced by regression or ANOVA commands, but the main components remain. First come estimates of the variance components. These are the variances of both the farms and plots within farms. In a regression model there is only one variance, estimated by the residual mean square. REML provides estimates of both the between farm and between plot variances. REML provides estimates of both these. These are analogous to the main-plot residual and split-plot residual variances produced by an ANOVA of a split-plot experiment. As expected, the farm-to-farm variance is much larger than the plot-to-plot variation within a farm.

Then there are 'Wald Tests' for the fixed effects. In this case the fixed effects are the treatment effect and covariate effects. These are similar to the lines for treatment effects in an ANOVA table. The statistic given is large if the effect is large, small otherwise. Formally the statistics can be compared with a Chi-squared distribution with the appropriate degrees of freedom to test the hypothesis of no effect.

Then follow tables of estimated means and s.e.d.'s. The size of the contrasts, are identical to those in the earlier analysis. The s.e.d.'s differ slightly because this analysis is based on the whole data, not just the information on one contrast. On page 138 we found that the mean size of the *Lantana*-control difference was –0.022 (s.e.d 0.592) on high streak farms and 1.314 (s.e.d. 0.342) on low streak farms. In REML the difference for low streak farms is 2.803-2.825 = -0.022 and for high streak is 4.047-2.732 = 1.315. REML reports the s.e.d. as lying in the range 0.3104 to 0.5376.

Notice then the s.e.d. for the streak main effect (low – high) is 1.106, much larger than the s.e.d. for treatment main effect of 0.3104. This is to be expected. Comparison of two treatments is a 'within farm' comparison as each treatment occurs on each farm, so the s.e.d. depends on the within-farm variance. Streak varies between farms, so the streak s.e.d. depends on the farm-to-farm variation that is much larger. Notice that REML has correctly used the right variance for each s.e.d. The user has not had to specify that the between-farm variation should be used when comparing the two streak means.

The strategy for modelling with more than one source of random variation is much the same as that for simpler regression models, where there is a single random term. For example, having fitted the model above, we can check that the assumptions are reasonable and we can investigate whether additional terms in the model are justified. As an example (Table 10) we consider whether the application of manure is a useful addition to the model.

```
***** REML Variance Components Analysis *****

Response Variate : grain

Fixed model: Constant+streak+manure+treat+streak.treat+manure.treat
Random model    : farmer

Number of units : 48

* Residual term has been added to model

*** Estimated Variance Components ***

Random term    Component          S.e.

farmer            2.9150        1.2076


*** Residual variance model ***

Term  Factor Model(order) ParameterEstimate S.e.

Residual        Identity        Sigma2      0.542     0.1546


*** Wald tests for fixed effects ***


Fixed term  Wald statistic   d.f.  Wald/d.f.  Chi-sq prob

* Sequentially adding terms to fixed model

  streak            0.03            1       0.03      0.854
  manure            0.49            1       0.49      0.483
  treat             1.72            2      35.86     <0.001
  streak.treat      7.69            2       3.85      0.021
  manure.treat      6.33            2       3.17      0.042

* Dropping individual terms from full fixed model

  manure.treat      6.33            2       3.17      0.042
  streak.treat      5.76            2       2.88      0.056

* Message: chi-square distribution for Wald tests is an asymptotic approximation
 (i.e. for large samples) and underestimates the probabilities in other cases.


*** Table of predicted means for manure ***

 manure               N            Y
               3.618        4.773

Standard error of differences:  0.7033


*** Table of predicted means for manure.treat ***

                 treat       Control    Lantana   Tithonia
                manure
                     N         2.763      3.368      4.723
                     Y         2.840      4.737      6.740

Standard error of differences: Average     0.7701
                Maximum         1.162
                Minimum        0.3255

Average variance of differences:0.6690

Standard error of differences for same level of factor:

                manure         treat
Average         0.6126         0.8715
Maximum         0.9960         1.094
Minimum         0.3255         0.6796
Average  variance  of  differences:
                0.4653         0.7885
```

Table 10.

Notice the analysis uses the data on manure from both within and between farms without splitting the analysis up. Reading the table of s.e.d.'s for the treatment x manure, means we see that the s.e.d.'s for comparison of treatments within levels of manure are lower than those for comparing manure means within treatments. This is because most of the information for comparison of treatments for a fixed level of manure application, comes from within farms. Most farms do have a constant level of manure application for the 3 treatments. The model correctly uses both within and between-farm variation to estimate both the effects of interest and their standard errors.

Finally we consider what might have happened if we tried to look at the effect of streak without fully understanding the structure of the data and using an appropriate model. A first attempt might be to use the regression tools introduced and widely understood and available in statistical software. In the example below (Figure 2 and Table 11) we try to look at the effects of streak and the treatments on grain yield in an attempt to reproduce earlier results. The model is the natural one to attempt to use, having terms for treatment, streak and their interaction, plus a term for farmer. The dialogue is similar to that used earlier with REML, but there is no place for the random Farmer term so we include it as a fixed term in the model. The results are not encouraging.



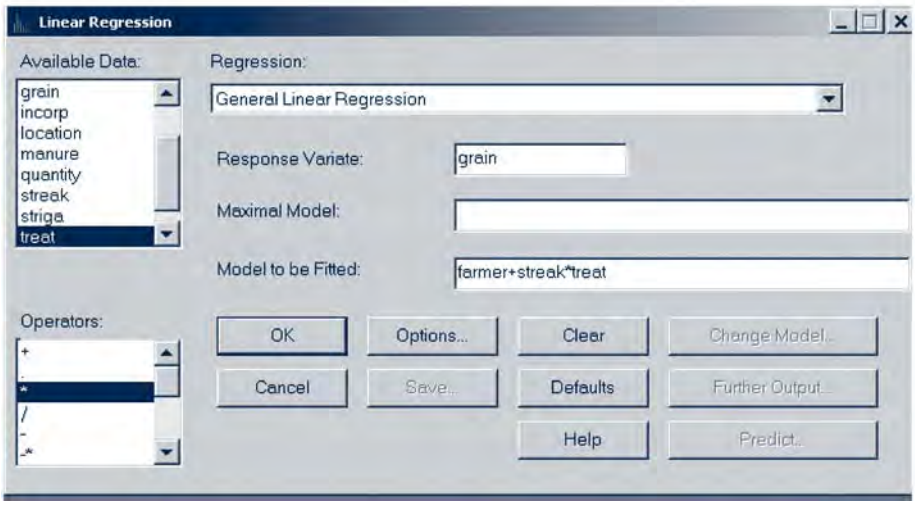



Figure 2. Linear regression dialogue box

```
*  MESSAGE:  Term  streak  cannot  be  included  in  the  model
 because it is aliased with terms already in the model

 (streak L) = 1.0000 - (farmer 8) - (farmer 9) - (farmer 11) -
  (farmer 12)

***** Regression Analysis *****

*** Accumulated analysis of variance ***

Change          d.f.         s.s.       m.s.      v.r.    F pr.
+ farmer          15     154.4270    10.2951     17.81    <.001
+ streak           0       0.0000          *
+ treat            2      37.2980    18.6490     32.26    <.001
+ streak.treat     2       5.4022     2.7011      4.67    0.018
Residual          28      16.1846     0.5780

Total             47     213.3118     4.5385
```

Table 11.

First there is a warning message, which is difficult to interpret but a sign that something has gone wrong. Secondly we see 0 d.f. for the streak term, which is not as expected. Remember the ANOVA here describes the variation attributable to each term allowing for those fitted before. We know streak varies between farms, yet all the farm-to-farm variation is accounted for by the farmer effect. Thus after fitting 'farmer' it is not possible to say anything about streak. Both the error message and the 0 d.f. are due to this.

We could try changing the order of the terms farmer and streak (Table 12):

```
***** Regression Analysis *****

*** Accumulated analysis of variance ***

Change            d.f.         s.s.         m.s.         v.r.
+ streak             1       0.3145       0.3145       0.54
+ farmer            14     154.1124      11.0080      19.04
+ treat              2      37.2980      18.6490      32.26
+ streak.treat       2       5.4022       2.7011       4.67
Residual            28      16.1846       0.5780

Total               47     213.3118       4.5385
```

Table 12. Changing the order of the terms farmer and streak

This gives an ANOVA, but we should still be suspicious. For example, the s.s. for streak is compared with the residual (within farm) variation. Yet we know that streak varies only between farms, so we should be comparing it with some between farm variation. If we then form predicted means and s.e.d.'s for the streak term we will have the same problem.

## Conclusions

As in the last session, the complexity (in this case caused by multiple levels of variation) requires the following steps.

1. Understanding that the problem might actually exist. In the example you are alerted to the problem by noticing that there are at least two important types of variation in the design, between and within farm.

2. Spotting if it really is a problem. This might be by suitably chosen diagnostic tables, or perhaps from unexpected error messages that the analysis software produces.

3. Finding an analysis that either avoids the problem or allows for it. Problems due to multiple levels of variation may be avoided by 'moving the analysis to a single level', or allowed for by using a model that includes several variation terms.

Session 14

# Complications in agroforestry trials

*R. Coe*

## Introduction

The first part of the course introduced the principles and steps of analysing experiments, which apply to many situations. In the second part of the course we have focused on the complications, which make the analysis of real experiments seem harder than the theory suggests. These are complications due to the design (Sessions 11 and 12) and due to surprising patterns in the data (Session 14). In this session we look at complications that arise because the trials are agroforestry experiments. There are no complications unique to agroforestry, but some, which commonly occur.

In this session we look at two of these, and briefly at a third. Agroforestry trials contrast with many agronomy of crop breeding trials, in two important ways. They are typically long term, as trees take more than a season to grow, reach harvest time and have significant effects on environment. Secondly, it is typical for different parts of a single plot to be of specific interest. This contrasts with crop-only plots in which all parts of a plot are considered to be equivalent. Both these situations produce 'repeated measures'; multiple observations on a plot either in space or time (or both). In many agroforestry trails, multiple components (trees and crops) are measured and analysis focuses on their interaction, not just on each separately.

For each of these three cases, we have to follow the same three steps; understand the nature of the problem, determine if it exists in our data and for our objectives, and find an appropriate strategy for coping with it.

## The problem: repeated measures

One message from earlier sessions is that the analysis of data from an experiment has to use the 'right' level variation to assess treatment effects. 'Right' means variation between the units to which the treatments were randomized. We saw that when measurements were taken at a lower level (for example individual trees within a plot were measured, or individual soil samples taken and measured), the appropriate analysis averages or sums these to the plot level. Even if the calculation is hidden with GenStat's ANOVA command, the result is the same as summarizing the multiple observations per plot to one number per plot before analysis.

However it is not obvious that this works in all cases. Two examples are:

- Observations are taken in each plot at different time points. The reasons might be to measure growth rates or to determine when flowering starts and peaks. The repeated observations are taken because we want to know something about changes over time, and averaging or summing over time looses the information.
- Observations are taken at different positions in a plot, but these positions are not all equivalent. For example they might be different distances from a tree line, or different depths in the soil or canopy. These measurements are taken because we want to compare positions (distances or depths), and so they appear in objectives a bit like treatments. However they have not been randomly allocated to the experimental units; 50 cm deep is always 50 cm deep, and cannot be randomly allocated to a depth of 100 cm. Again averaging or totaling these observations may loose the information we really want.

## Spotting the problem

The existence of repeated measures is detected by paying close attention to the match between the layout and measurements made.

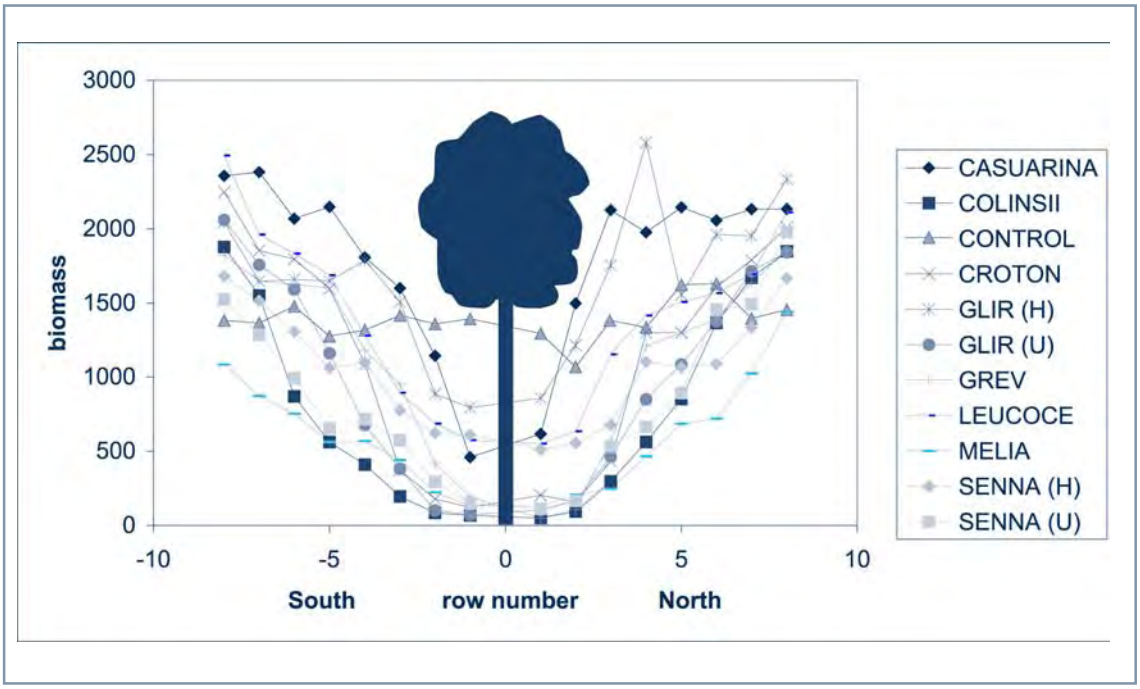
Remember that a 'plot' is that unit of land (or other experimental material) that, under the randomization scheme is certain to be of the same treatment, but any two different plots could have different treatments. Now how many observations do you have? If there is more than one per plot you have repeated measures! We have already seen this in the *leucaena* trial in which 4 trees per plot were measured; repeated measures in space. In that example we were not interested in those particular trees. There was, for example, no reason to compare tree 1 with tree 2.

Contrast that with the example that follows.

### Example: Roots and Competition (RAC) trial

The trial was laid out as a randomized block design with 4 replicates and 11 treatments, a total of 44 plots. In each season there are 704 observations of crop biomass and yield. The crop is measured row by row for each of two sides (north and south) of the tree row.

The interest is the effect of trees on the crop yield and how this varies between tree species and management. Exploratory analysis shows there is a rich set of patterns to evaluate. The graph below (Graph 1) shows the mean biomass (over replicates) for each treatment, plotted against position relative to the tree line.



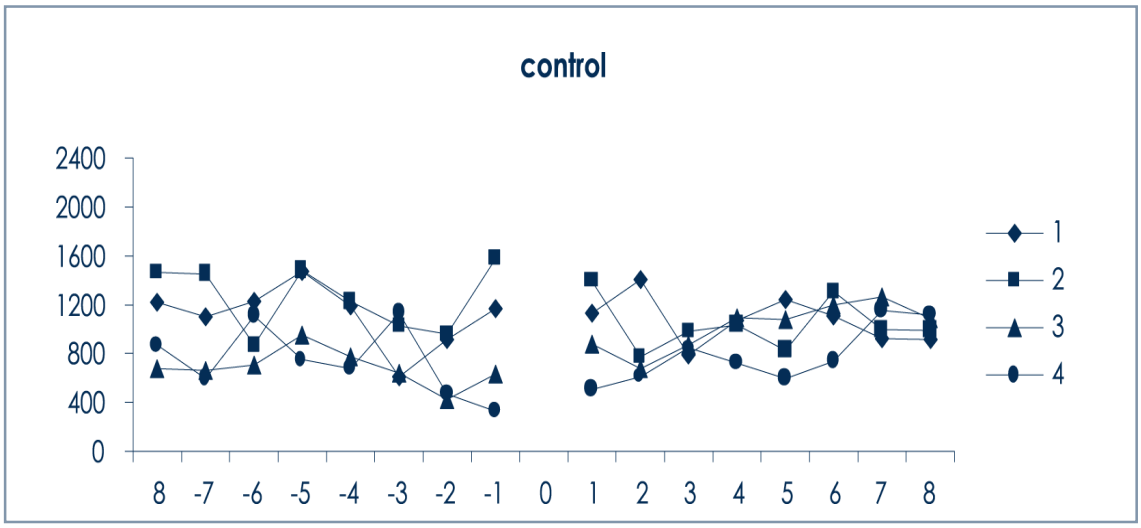

Graph 1. Mean biomass (over replicates) for each treatment, plotted against position relative to the tree line

## Exploring the data

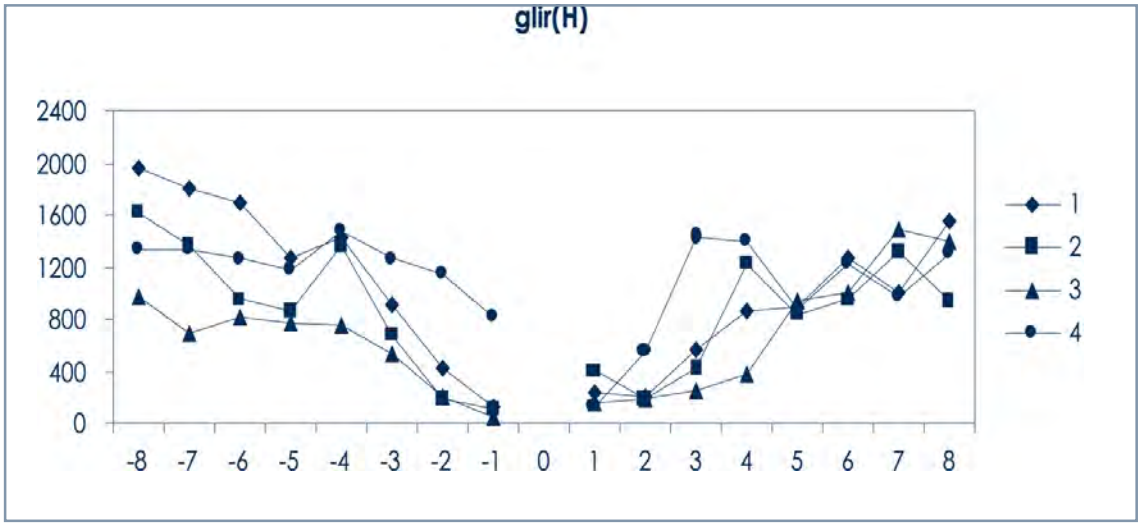
The data exploration is as important with repeated measures data as with any other data structure; possibly more so, as the potential patterns and problems are more complex. As with other data, the aims are twofold; to make summaries that help meet the specific objectives and identify odd or informative patterns.

In the current example the regular structure to the data makes Excel's pivot table command very useful for constructing summaries. For example, the Graph 1 above was drawn from averages over replicates and summarizes the effects of treatments together with distance from the trees on both the north and south sides. We can see apparently clear effect of trees, which vary with species, and give distinctive patterns of grain yields with increasing distance from the tree line. There is no obvious difference between responses on the north and south sides of the trees.

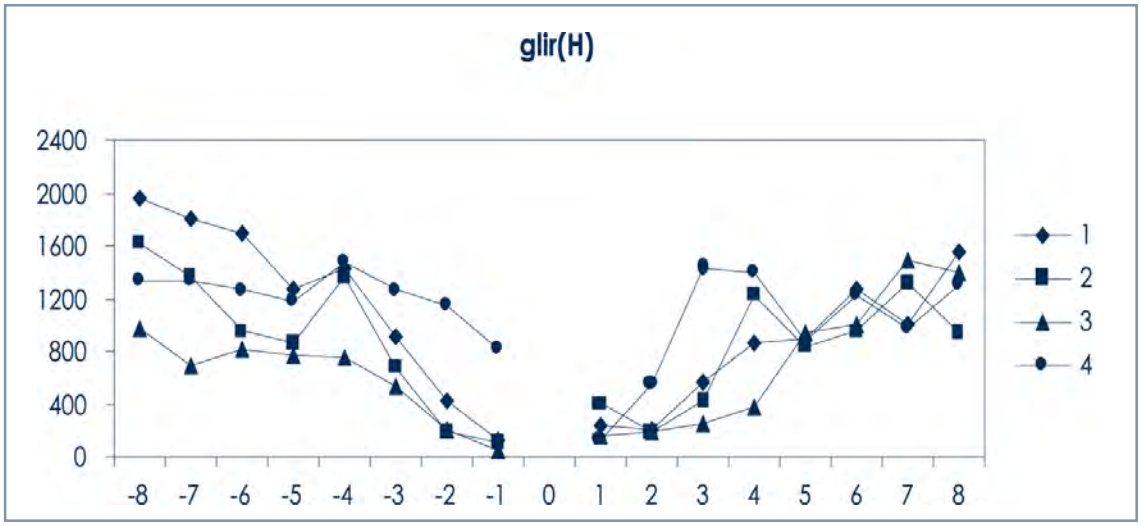
The graph above (Graph 1) involves averaging. The pivot table tool in Excel also makes it easy to look at 'slices' of the raw data. For example, we may want to reveal how the responses above vary between replicates, after all, it is consistent patterns across the replicates that will lead us to conclusions of repeatable treatment effects. A graph with the data for all treatments is too cluttered, but plots for three critical treatments are shown in Graphs 2, 3 & 4 below.




Graph 2. Variation in yields across control plots




Graph 3. Variation in yields across *Gliricidia* hedge plots




Graph 4. Variation in yields across *Gliricidia* upperstorey tree plots

These graphs contain a lot of information. For example, from the control plots (Graph 2) we see no consistent pattern across rows in the field, as we would expect. However there are consistent (across rows) differences between plots, which will need to be allowed for in any analysis. The graph for the *Gliricidia* hedge plots (Graph 3) shows that in only one plot was it possible to get a reasonable grain yield in row 1. Look for any distinctive features of this plot (for example the hedge not growing well) either in the data or in records and observations of field staff. The graph for the *Gliricidia* upperstorey trees (Graph 4) shows how the variation between replicates is smaller for the smaller means (rows close to the trees). This may be obvious but will need to be allowed for in an analysis.

## Approaches to analysing repeated measures

### 1. Analysis of separate components

The objectives may well require that each repeated measure be examined separately. For example, if we measure the yield every season, we may wish to analyse each season separately because we are interested in the treatment effect each season. However the analysis does not focus on changes or trends over time. It also cannot make sense when the observations are taken arbitrarily close together. If the observations are close enough together in time then we cannot be learning anything new by each separate analysis of each time.

Do not automatically dismiss analysis of each repeated measurement; check whether it will meet the objectives.

In the RAC example, there is a curious pattern. Yield in row 4 on the north side looks very high for one treatment. An analysis of just that row gives the following (Table 1):

```
Variate: BIOMASS

Source of variation       d.f.   s.s.       m.s.      v.r.    F pr.

BLOCK stratum               3  1693102.    564367.   1.83

BLOCK.*Units* stratum
SPECIES                    10 15772219.   1577222.   5.12     <.001
Residual                   30  9243825.    308128.

Total                      43 26709147.


* MESSAGE: the following units have large residuals.

BLOCK 2  *units* 5     1580.   s.e. 458.
BLOCK 4  *units* 10    1064.   s.e. 458.
```

Table 1. Analysis of Yield in row 4

The analysis of residuals suggests there is an odd observation for that treatment in Block 2. Going back to the original data and comparing grain yield and biomass, reveals this biomass observation to be in error so it is omitted from now on.

## 2. Analysis of plot averages or totals

Again, check if this meets the objectives. It may give all the information you need.

In the RAC example we are interested in the yield loss (relative to control) that is incurred by introducing trees. This is simply the total for control plot - total for a tree plot. Hence the appropriate analysis (Table 2) is based on the total for each plot.

```
***** Analysis of variance *****
Variate: TOTAL

Source of variation    d.f.        s.s.        m.s.    v.r. F pr.

BLOCK stratum             3   4.554E+08   1.518E+08        6.60

BLOCK.*Units* stratum

SPECIES                  10   1.379E+09   1.379E+08    6.00 <.001
Residual                 30   6.899E+08   2.300E+07

Total43 2.524E+09

***** Tables of means *****

Variate: TOTAL

Grand mean 18369.

 SPECIES   CASUARINA   COLINSII   CONTROL    CROTON    GLIR (H)    GLIR (U)    GREV
             28643.     12353.    22169.    18046.     26177.      15394.   17539.

 SPECIES     LEUCOCE       MELIA   SENNA (H)     SENNA (U)
              22038.       9557.      16671.        13475.

 *** Standard errors of differences of means ***

TableSPECIES
rep.           4
d.f.          30
s.e.d.      3391.0
```

Table 2.

## 3. Informative summary statistics

The steps are:

(a) Plot the data against time or distance, to highlight the types of patterns in the data.

(b) Choose one or more summary statistics that describe key features of the pattern, which relate to your objectives. These may be something simple (e.g. the average yield in the three rows closest to the trees) or more complex (the proportion of the total soil nitrate that occurs within the crop rooting zone). They may even involve fitting a model. For example, for many species, the tree root length density (rld) decreases exponentially with horizontal and vertical distance from the tree. The pattern can often be modelled

using rld=Aexp(-k $\sqrt{}$ ($x^2$+c$y^2$)). The variables x and y are horizontal and vertical distances and c measures the relative importance of vertical and horizontal. A is the maximum density and k the rate of decrease. Each of these parameters can be interpreted.

(c) Calculate the summaries for each plot,

(d) Subject the summaries to analysis of variance to show how they differ between treatments.

All the above three methods actually reduce to the same thing; we choose variables, perhaps constructed from the original variables in a complex way, which have two properties: (1) they measure something which we need to know about, to meet the objectives, and (2) they result in a single number per plot.

Looking back at Table 2 for the RAC trial, suggests several more important patterns. Of course we should focus on those specified in the objectives, but as those are often rather vague it does no harm to follow up 'interesting' observations.

1. There seems to be considerable variation in yield between treatments immediately adjacent to the trees. A summary that describes this is just the average yield in the four rows 1 and 2 on north and south sides. Analysis of these (Table 3) produces:

```
***** Analysis of variance *****

Variate: row12

Source of variation  d.f.          s.s.          m.s.      v.r. F pr.

BLOCK stratum           3       738048.       246016.     4.11

BLOCK.*Units* stratum
SPECIES                10      6970739.       697074.    11.65 <.001
Residual               30      1795564.        59852.

Total                  43      9504351.


***** Tables of means *****

Variate: row12

Grand  mean  475.

SPECIES    CASUARINA COLINSII   CONTROL  CROTON  GLIR (H) GLIR (U) GREV
              930.       74.      1279.     166.      937.    110.   191.

SPECIES    LEUCOCE       MELIA   SENNA(H)     SENNA (U)
             611.         173.    575.          181.


*** Standard errors of differences of means ***

TableSPECIES
rep.          4
d.f.         30
s.e.d.    173.0
```

Table 3.

There are certainly substantial differences between treatments. Notice that the unpruned *Senna* and *Gliricidia* have very similar crop biomass adjacent to the tree, but when pruned, the *Gliricidia* gives a much higher crop growth than the *Senna*. This type of result needs explaining in terms of light, water or nutrients.

2. The rate change in crop biomass as we move away from trees, seems to vary between species. The summary statistic chosen to describe this is the slope of the regression line of biomass on row number, for rows 3 to 6. This is calculated for each plot and the resulting slopes subjected again to analysis of variance (Table 4).

```
***** Analysis of variance *****

Variate: slope

Source of variation d.f.(m.v.)      s.s.      m.s.      v.r.      F pr.

BLOCK stratum            3           85910.    28637.      4.48

BLOCK.*Units* stratum
SPECIES                 10          535688.   53569.      8.37     <.001
Residual                29(1)       185575.    6399.

Total                   42(1)       801981.


***** Tables of means *****

Variate: slope

Grand mean 204.

SPECIES CASUARINA  COLINSII CONTROL  CROTON    GLIR (H)  GLIR (U) GREV
          85.      283.      59.      415.      93.       354.     250.

SPECIES  LEUCOCE     MELIA    SENNA(H)    SENNA (U)
          228.       128.      138.         209.


*** Standard errors of differences of means ***

Table  SPECIES
rep.         4
d.f.        29
s.e.d.    56.6
```

Table 4.

Again there are clear differences between tree species. Remember the units are now (biomass per increase in row number). What does the mean for 'Control' in the above analysis represent? Think about what you expect it to be and how you might test if the data are consistent with this expectation.

The above two examples are both simple summaries. One that might be a little harder to define, is the distance from the trees at which the biomass becomes equal to that of the control. Think about how such a summary could be calculated.

### 4. Other methods

Textbooks, mathematicians and computer software sometimes suggest another approach. This is to model all the original observations, using a model, which accounts for the multiple levels of variation. This is consistent with the approaches used in other parts of this course. For other types of complexity in the data, we were able to either find a simple approach to analysis that met the objectives, perhaps in a rather cumbersome way, or find a comprehensive statistical model. In the case of repeated measures the modelling is not straightforward and generally beyond the scope of this course. The difficulties are:

1. There is no single model which will always work. The simpler methods are usually simplistic, giving answers that are unlikely to be valid.
2. The more realistic models are difficult to formulate.
3. The analysis is not 'tuned' to the objectives in the way that methods described above are.

If we look at the RAC example then we can see that the 'pattern' part of the model has to include the smooth response of grain, with distance from the trees, possibly of a different shape for each tree species. The 'residual' or random part of the model has to include random variation between blocks, plots sides and rows. We would expect some sort of correlation between adjacent rows, so that has to be included, as well as the way the variance changes with distance from the tree row. All this is possible but not straightforward.

## Multiple components

Strategies for handling the multiple components measured in a trial are briefly mentioned:

### 1. Analyse separately

Look at the objectives of the experiment. It may well be that many of these are met by simple analysis of each component. For example, during this course we have analysed the trial 'Upperstorey/understorey tree management', and met the trial objectives, without considering the tree and crop yields at the same time.

### 2. Form indices

Occasionally there are simple sums of components that make sense. The total biomass production over a number of years, for example, may be a good combined index to assess the biological advantage of one system over another. In applied trials we are more likely to be interested in some notion of total value, and indices which describe this, can be constructed that describe the financial value of the productivity of each plot (sum of price x yield). This type of analysis can be made more realistic and complete by (a) incorporating input costs as well as output prices, and (b) by combining outputs across seasons, perhaps discounted in an appropriate way.

It is beyond the scope of this course to give details of these financial calculations. However the statistical analysis of such data can be straightforward. If the only quantified source of uncertainty in the data comes from the plot-to-plot variation in yields, the same procedure as above can be used. Simply calculate the financial value for each plot, and carry out an analysis of these plot-wise values to compare treatments.

## 3. Ideas from intercropping

There has been much research on intercropping annual crops, and ideas for analysis of these experiments developed. Some are relevant to certain agroforestry trials. For example, if the objective is to assess the production advantage (or otherwise) of mixing trees and crops, compared with having sole-tree and sole-crop plots, we might calculate the land equivalent ration (LER). This is defined as the total area of sole-crop and sole-tree needed to produce the same as unit area of the mixture. However, pay attention to objectives again. LER is not always going to be useful!

Session 15

# Complications in data

*R. D. Stern, R. Coe, E. Allan*

## Introduction

In this session we look at problems such as missing values, odd observations and unexpected zeros that arise once data are collected. The main part of this handout examines the alternative strategies for coping with such complications. This sets the scene for the work, in Session 15, where scientists look at different ways of proceeding with the analyses of their own data.

We anticipate that there may be surprise complications in any of our three major components, namely the treatments, layout and measurements, of the trial. We may misapply a treatment, missing values may complicate a simple layout, or some of our measurements may appear strange.

Surprises are common! Like a birthday surprise, we know there will be something interesting and unpredictable about the trial, but we do not know what, nor where it will occur.

Some surprises, like a drought, may unfortunately destroy an experiment completely. In this situation there is nothing we can do. But many surprises affect only part of the experiment. For example, runoff from excessive rainfall may destroy the crop on a few plots, or bird damage may affect all plots, but to a limited extent, so we still feel that some of our objectives can be realized.

One of the main reasons we strongly encourage the use of a powerful statistics package, such as GenStat, for data analysis, is that we must be able to cope with these surprises and still be able to analyse the data. Until recently, the powerful statistical software has been too difficult for scientists to use effectively; hence they have been forced to use simpler packages. However the failure of these simpler packages to be able to analyse data effectively, when there are surprises, has been a major reason that data have not been fully analysed in the past.

The strategies for coping with surprises include ignoring the surprise, because it does not affect the objective. Sometimes we may have to modify or omit an objective; sometimes we may profit from a complication and add a new objective. For example, insect damage may enable detection of a variety that is resistant.

Often we will find that the trial has become unbalanced and hence the analysis is more complex (for the statistics package). However, we have already seen, in Sessions 8 and 11 that analysing data from unbalanced experiments is no longer a problem. The presence of zeros in the data may be simple to handle (e.g. note which varieties did not flower and then omit them), or we may wish to analyse first the binary (zero or not zero) data and then look in more detail at the plots with non-zero values. We saw in Sessions 11 and 12 that it is now easy, in principle, to analyse data from a range of distributions, including where the response is a yes/no variable.

## Common problems

Table 1 below lists some of the complications that commonly arise, most of which have been seen on the examples used in this workshop.

| Type | Complication | Example |
|---|---|---|
| Treatment | Levels modified | Farmers applied 'about' 100 kg of mulch. |
| | Applied to different units | Seed shortage of one variety, so control applied to multiple plots in reps 3 and 4. |
| Layout | Post-hoc blocking useful | 'Relay planting on *Sesbania sesban* and maize trial' indicated that blocking should have been in other direction. |
| | Missing values | a) Some farmers left trial. b) Stover yields missing on 3 out of 36 plots in mulching trial. c) Some plants missing within the row. |
| Measurement | Zero values | Some trees did not survive. |
| | Strange values | Observations for one farmer had different pattern to the others. |
| | Censored values | All trees counted, circumference only measured on large trees. |
| | Different variability | Soil nitrogen values showed more plant-to-plant variation in some treatments. |
| | Affect treatments | a) Some plots heavily infected with striga. b) Some farms have streak virus. |

Table 1. Some of the complications that commonly arise

In Table 2, we have categorized the complications as those affecting primarily the treatments, the layout and the measurements. They differ in the point in the trial when the complication can be detected. We now consider these categories in turn.

## Complications concerning treatments

Sometimes treatments are misapplied, either in error, or because there is a shortage of materials for some of the treatments, so others are substituted. This substitution has complicated the allocation of treatments to plots; it usually makes the design non-orthogonal.

In the on-farm trial, farmers were instructed to apply 100 kg of mulch per plot. In interviews afterwards it was clear that the amount applied varied from farm to farm, from as little as 11 kg to over 200 kg. The quantities applied were recorded. These differences imply that the treatments were different from farm to farm.

A similar complication is where one treatment is called 'farmer's practice', and this treatment differs from farmer to farmer.

This type of complication can normally be detected when the treatments are applied. Sometimes, as with the on-farm trial, the situation only becomes clear at a later stage. It may then be necessary to suggest further measurements, in this case the reasons for the variation in the amount applied, to decide on the appropriate action. Two possible reasons are the labour time required for the treatment and the fact that 100 kg may not be needed on some types of soil.

## Complications concerning layout

One complication is where observations in the field indicate a blocking factor, that was not incorporated when the trial was designed, would have been useful. For example 2 reps of a millet trial with 48 treatments and a 4 by 3 by 2 by 2 factorial structure was laid out in a randomized block design in a rectangular field with 12 plots by 8 plots. The neighbouring area on one of the longer sides was left fallow. During the season there was bird damage that was progressively less severe, the further the plots were from the fallow area. To what extent could the analysis incorporate this extra blocking factor?

This idea is called post-hoc blocking and seems rare in the analysis of experimental data. It is common in the analysis of survey data, because surveys often know of potential blocking factors (e.g. farm size), where the allocation to one 'block' or the other is only possible as the data are being collected.

What is common in the analysis of experimental data is the use of the analysis of co-variance. In the above example, if we were to assume a linear change in damage effect related to distance from the fallow, we could code the plots as 1 to 8 and then use that analysis of covariance.

Hence the idea of post-hoc blocking is simply a little more general, and uses a 'co-factor', rather than a 'co-variate'.

The most common complication is that of missing values. When the trial is at multiple levels, then missing values can be at any of the levels. For example, some farmers may drop out of a trial. A farmer in a trial may lose the control plot because of theft. Labels were lost for bags of maize from 6 rows in different plots.

Missing values are normally detected when data are collected, or when the data processing begins.

### Complications concerning measurements

The analysis of the on-farm trial indicated that one farmer had a different pattern of yields to the others. Omitting him from the analysis changed the conclusions. What should be done in the final reporting? Odd values are often clear from the exploratory stage of the data analysis. This complication was detected following the initial analysis of variance.

In the trial on fallow management, the measurements of soil nitrogen were found to be much more variable for some treatments than for others. This was detected using boxplots during the exploratory stage of the data analysis.

In the same trial, some plots were found to be heavily infested with striga. This was detected during field observations and the striga counts were therefore recorded on each plot.

In the *Leucaena* trial the number of stems per tree was counted and their diameter recorded. There were however, a lot of small stems, so the decision was made to record details only of the stems with a diameter greater than 10 mm.

There are many reasons for having zeros in the data and the appropriate strategy for the analysis depends on the reason. In the *Leucaena* trial some plants were small and had no stems greater than 10 mm. The calculated volume of wood was therefore given as zero for these plants.

## Strategies for coping with problems

In this section we describe the general strategies for coping with complications. We will see that they range from ignoring the problem (which is sometimes the correct strategy!), to a 'quick fix', to a small research project. It is important to be able to assess the possible impact of

the problems in your experiment and hence to suggest the type of remedial action that is needed. Ignoring a problem might mean that all conclusions are invalid. At the other extreme, a lot of work may go into solving a problem which turns out not to affect any conclusions.

## Ignoring the problem

The treatment problem, in the on-farm experiment with the varying amounts of mulch applied, has been ignored so far. The results on the effects of the different treatments are clear, and the amount of mulch applied is usually close to 100 kg. Hence ignoring this problem is a possible strategy. Also, for some objectives, and depending on the reasons, the farmers choose not to apply 100 kg; this may not be a problem.

Similarly, in the trial where the diameter of the small stems has been omitted, the aim is to estimate the total volume of wood on each plot. Very little of the volume comes from these small stems. Perhaps this problem can be ignored.

## Modifying the objectives or the definition of a treatment or measurement

If there are major problems you may find that some of the original objectives cannot be met. This does not mean that the analysis should be abandoned, because others may still be attainable. In some cases you may find that new objectives can be studied. Thus, you turn the problem to your advantage.

For example with the on-farm mulch trial, perhaps the fact that farmers did not all put 100 kg of mulch provides an opportunity to study the relationship between yield and the quantity of mulch applied.

This is attractive in principle, but often it turns out to be a 'happy dream', and not a new objective to which much time should be devoted. This case is typical, in that the experiment would have been designed differently if that objective had been proposed at the design stage.

What is common, is that major problems are of great assistance in the effective design of a future trial on the same topic. For example the treatment problem in the on-farm trial (problem 1), might lead to consideration of a design for the following year where an extra treatment was the application of *Tithonia* at two levels, namely 50 kg and 100 kg. This would increase the number of plots per farm from 3 to 4, but would also be consistent with further objectives.

An extreme case of redefinition of objectives is to define the new objective to be to use this trial as a 'pilot study' in the effective design of the trial in the next year!

A redefinition of the treatments is sometimes consistent with the original objectives. For example in the on-farm trial we could redefine the treatment to be 'about' 100 kg, with 'about' defined to be between 80 and 120. This is similar to ignoring the problem except that it might indicate that some observations (either farmers or plots) should be omitted, because the treatment falls outside the stated range.

A redefinition of the measurements may be consistent with the original objectives. For example in problem 3 the measurement could be redefined as the total volume of wood per plot from large stems. This is then consistent with what was done, though not with the original protocol.



## Do a 'quick fix'

Sometimes a 'quick fix' is built into the software. One common problem is that of some missing values in a simple ANOVA. The data are then non-orthogonal and the standard ANOVA cannot be used. The 'quick-fix' is to estimate the missing values and then to proceed as though the data are orthogonal. This is usually all that is required if just a few values are missing.

Quick fixes that are built into the software, are not risk-free! For example GenStat's ANOVA system estimates missing values automatically for any number of missing values. However, the results are only approximate. The benefit of GenStat's ANOVA for analysis, is particularly the clarity of the display of the results. So, when there are more than a few missing values, say more than 3 values or 3% (whichever is the smaller) compare the results with the exact analysis from regression or REML.

A 'quick fix' for the problem of the small stems in the *Leucaena* trial might be to assume they are all 5 mm in diameter.

A common type of 'quick-fix' is to assume a linear response to a quantitative factor and hence make simple adjustments to the yields. For the on-farm trial this might proceed as follows. Take farmer 12 in the West as an example. He had a yield of 8.42 ton from the *Tithonia* plot, which was 5.77 ton more than the control. He applied 125 kg of mulch. Perhaps we might estimate that he would only have achieved 4/5 of this increase had he applied the 100 kg, so the estimated yield for his plot is then 7.27 ton. We adjust the other yields similarly and then analyse the adjusted yields.

We caution against the overuse of this type of adjustment. It is often applied in cases such as the above, where the response is unlikely to be linear and can give very misleading results.

## Use a more flexible approach to the modelling

As an example, we consider the question of missing values again. We saw in the modelling session that the regression facilities can be used for an analysis of experimental data, and the design does not then have to be orthogonal. So, if a number of values are missing, we could simply use the regression facilities (or REML) and do the correct analysis.

The problems of misallocation of treatments and that of post-hoc blocking can be handled similarly. The complication has merely affected the orthogonality of the experiment and the regression commands, or REML are therefore used, instead of the simple ANOVA, for the analysis.

Many of the common complications concerning the measurements indicate that a more complicated model is required. One problem is when the assumptions of the analysis of variance do not hold. The standard ANOVA assumes that the observations in the whole trial are equally variable. Where this is not the case there are two common solutions. The first is to split the data into 2 or more parts. For example, if the data in the on-farm trial had different variability in the two districts, then a separate analysis could be done in each district. This is a simple solution to the problem, but the model has become slightly more complex in that there are now two separate variances, one in each district.

Where variances are obviously different for the different treatments or regions, then a different strategy is to transform the data, for example to model the square root of the yields, rather than the yields themselves. While there is nothing wrong with the use of transformations of the data, we feel that the analysis of the data in subsets is often simpler and more effective.

Another alternative to taking transformations is to use a 'generalized linear model' as was described in Session 11. One example is given by MCH. In Chapter 8 the data are described and a transformation is proposed. The data are counts and are then analysed assuming they are from a Poisson distribution (see Chapter 14). These analyses can also be followed using GWIM, pages 94 to 97 with a transformation, and pages 102 to 103 as a generalized linear model.

One special case is when some of the treatments give a zero yield. In this case we do not recommend transforming the data, because it normally hides an important component of the results. Instead the zeros against non-zeros are considered first. This may be a trivial analysis, for example to note that two varieties gave no yield. Or it may involve a model, for binary data, such as was described in Session 11. Following this stage the non-zero data are analysed.

A common difficulty is that the results from a standard analysis are so variable that it is difficult to detect any treatment differences. This is often summarized by a report that the 'cv was too high'. We find that the cv is an overused diagnostic. Sometimes it is reported for variables where it is clearly inappropriate, such as disease rating or measurements that could be negative. Even when it is legitimately used we find it is usually easier to look at its two components directly.

The cv is given by the 100 x residual standard deviation/overall mean, i.e. by 100s/xbar. So, if it is large, then either the residual standard deviation is surprisingly large, i.e. there is a lot of unexplained variation, or the overall mean is surprisingly low, or both.

One simple reason for a surprisingly low overall mean is that zeros remain in the data. We have discussed methods of coping with zeros above.

Usually the problem is of a surprisingly large residual standard deviation. One common reason that is considered below is that some odd observations remain in the data and should be examined separately. A second possibility is that there is something that was measured that should be included in the model. Perhaps a covariate will help to explain or the inclusion of an additional blocking factor (post-hoc blocking).

Sometimes that is life, and the large residual variation has to be accepted. It is still however, useful to have an explanation, because it can help in the design of future trials. For example an examination of the plots on a farmers field may show that there is large variation even **within** individual plots, due to local soil heterogeneity. In such cases the plot-to-plot variation will clearly also be large.

## Do a 'sensitivity' analysis

A common reason for a large residual variation is the presence of 'odd' observations. In such cases it is sometimes clear why a result is different, perhaps closer inspection reveals an old termite mound. In this case it is legitimate to omit the observation.

Where there is no obvious reason for the odd value, one strategy is to try the analysis both with the offending observation, and without it. If the results (in relation to the objectives) are unaltered, it is unimportant which analysis is used. If the results change in an important way then attention must be focussed on the observation, because it is driving the conclusions. Sometimes both sets of results are reported, with details of the odd observation, to allow readers to draw their own conclusions.

This same 'with and without' strategy can also be used to try transformations for the data. If there is no difference in the conclusions, then the report usually uses the untransformed results, because they are easier to interpret, with a comment that a transformation was considered.

### Build a model to try to solve the problem

A simple example would be to build a 'one-off' model for stem diameters of *Leucaena* trees. Often the construction of this type of model is not part of the original objectives, and may not be needed for those objectives. (For them the quick-fix is sufficient.) But it may be interesting and of general use in other similar experiments later. Thus adds a methodological objective that is of general use.

Typically this type of modelling work proceeds in parallel with the initial analyses that use a quick-fix solution. In such cases it is useful if all the steps in the analysis are saved (i.e. a log file is kept). Then the revised tables and graphs can quickly be produced once the model is in operation.

## Conclusions

It is rare for an experiment to have no complications. Thankfully, some of the solutions to the common problems are relatively simple. It is however important to have access to statistical software that is sufficiently powerful to make it easy to investigate and then be able to cope with complications.

With access to modern statistics packages, the analysis of datasets without complications is usually a quick process and one that scientists can increasingly do unaided. The time taken and advice needed is often largely to handle the previously unforeseen problems. It is therefore important to clarify the time scale and effort proposed for a given analysis.

Often the most time-consuming complication, once the data are available, is the lack of a good data-management strategy. This area was covered in the companion course.

Finally we caution against expecting too much from the analysis. The key to a good experiment, i.e. one that satisfies the stated objectives, is of good design and data collection. For example variability cannot be explained unless measurements were taken that enable an explanation.

This is an obvious example, but others may only become apparent once the data processing begins. For example, a large trial where 200 farmers choose the varieties they will plant, may include many farms where there is striga and also many farms where an important variety is grown. But suppose only 3 farmers with striga chose to plant the particular variety. It is then a trial that is far too small to be useful, for the particular objective of assessing the affect of striga on this variety. This complication has no solution given these data and this is sometimes a surprise to scientists, because the trial as a whole is very large.

# Session 16 Data analysis

*R. D. Stern, E. Allan, R. Coe*

## Introduction

We describe alternative ways in which data from participants can be used within the training course. The duration of the course and the number of resource persons are major determining factors of the time that can be devoted to the individual analyses.

It is important that the individual analyses are viewed in relation to the sessions of the course. If many participants have data that needs analyses that are not covered, it is likely that an alternative, or additional session needs to be provided. However, one odd set of data should not be allowed to distract either resource staff or a participant from the major objectives of the course.

Occasionally a participant brings a very large volume of data and hope that there will be time for everything to be analysed. Objectives need to be specified carefully for what is possible within the time-constraints of the course, and given the other training objectives. The data analysis is intended to assist and complement the materials in the course. They must not be allowed to distract either a participant or the resource team.

This is not to deny the importance of the sessions spent on the analysis of the participants' own data. In some workshops this will be a major objective itself and in other instances participants may arrive early, or stay after the end of the formal training, to proceed with their own analyses and reports.

## Organizing the data

Early in the course, participants should see the data sets and protocols that have been provided. They need to have equivalent materials. These include the raw data and information so that the objectives, treatments and so on are clear. Where this information is not available, it is often preferable for the participant to work with partners who have their own data, than to try to make do.

Organizing the data into the appropriate form can take time and should not be allowed to be a distraction from the course for a participant. Usually one session will be allowed early in the course for this task. Where more time is needed, then either resource persons can help, or the participant can work in the evenings or at weekends. Usually a good solution is for the resource staff to provide some help, at least to put a sample of the data into an appropriate form.

Sometimes participants will have data well organized but in a different format to the examples that we have provided. There is no need to insist on a consistent format. We have aimed for consistency to simplify the process of using different datasets on the course. The required format should simply satisfy the criterion, that the data can easily be transferred to a statistics package for analysis. Sometimes this transfer to the statistics package may be part of the session on data organization.

The aim of this part of the session is that participants can then concentrate on the analysis of their data. Many of the practical sessions suggest that analyses be tried on the participants' own data. This is not possible if the resource persons have continually to find the file and then re-import the data.



## Using data sets within the course

The data organization stage also provides an opportunity for resource staff and participants to see which examples can be integrated with the course materials. Often one or more of the resource staff may look ahead to future sessions and try the participant's data. Sometimes participants can make data available to resource people before the course starts, so that these examples can be incorporated into lectures and notes.

When data from a participant reflect a common problem, the use of these data can bring the course to life. But there are risks. The first is that the example chosen is sometimes a major preoccupation of just one participant and may be of less interest than the standard examples to the rest of the group.

The second risk is that scientific aspects may distract participants from the objectives of data analysis that are being covered in the session. For example:

'I really don't see why you are applying nitrogen on this type of soil, because …'

'Why are you still doing trials on hybrids, when poor farmers …?'

These may be most valuable points scientifically, but the time for their discussion must be severely rationed.

Many of the practical exercises assume that the analyses will start with an example we have provided, and then can continue with the participants' data. Often the first exercise is to consolidate the material covered in the lecture. This is a useful order, so a participant who is weaker on the computer, or who works more slowly, can concentrate on the basic materials. Others can proceed with the greater challenge of trying the same type of analyses on their own data.

## Further analyses and presentations

Exercises 1, 2, 4 and 9 all suggest simple analyses of participants own data as part of the practical work associated with the session. They also involve the possibility of short presentations that describe the protocol, the data and the basic analyses. We strongly suggest this route be followed, so that these analyses are integrated into the training course.

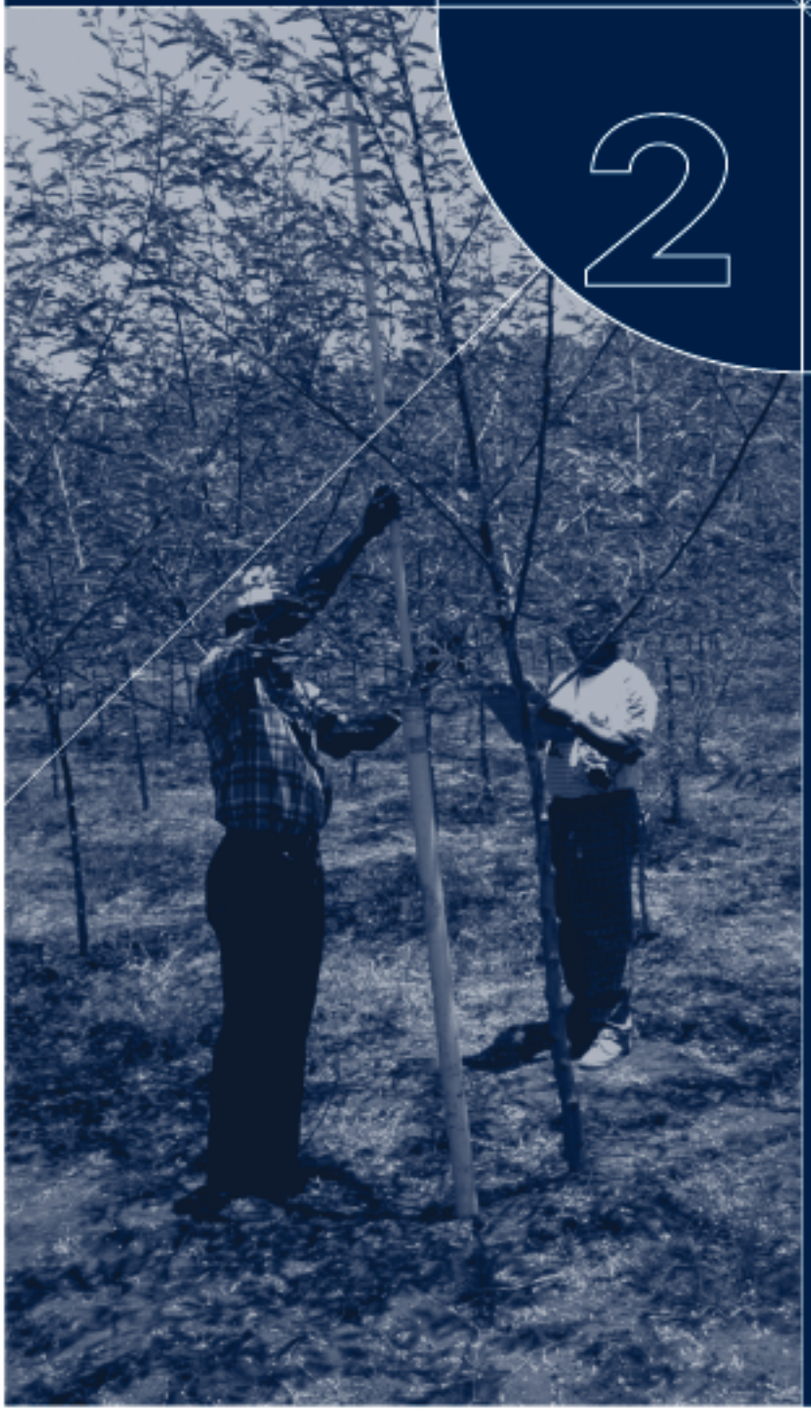
A longer report, but one that again uses just simple methods of data processing can follow this initial work. In this way these analyses will proceed in parallel with the course materials.

It is possible that materials from later sessions need to be introduced early, to enable further work to be undertaken. In particular Session 15, on common complications may need to be introduced.

Ideally a session or day, towards the end of the training period would then be devoted to the further analysis of these data and to presentations of the results. The presentations should again be restricted in length.

ISBN 92 9059 145 5

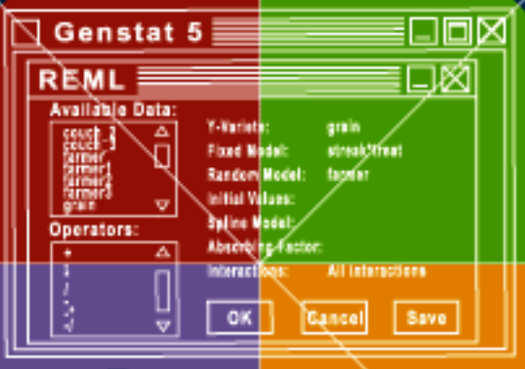
2



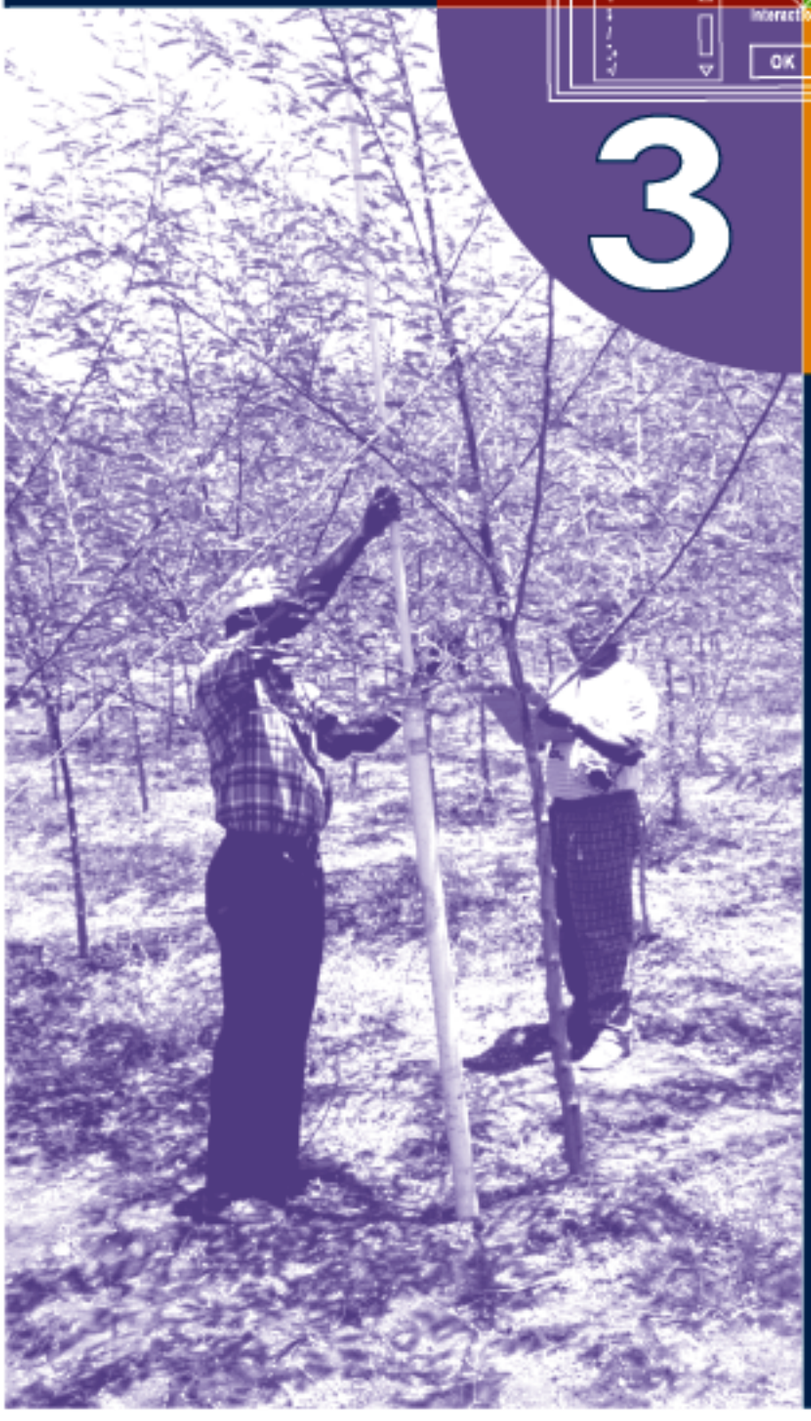
World Agroforestry Centre

TRANSFORMING LIVES AND LANDSCAPES

# Data analysis of agroforestry experiments




3



## Exercise guidelines

World Agroforestry Centre
TRANSFORMING LIVES AND LANDSCAPES

# Data analysis of agroforestry experiments

3

## Exercise guidelines

*by*

*Richard Coe*

*Roger Stern*

*Eleanore Allan*

*edited by*

*Jan Beniest*

*Janet Awimbo*

World Agroforestry Centre

TRANSFORMING LIVES AND LANDSCAPES

The World Agroforestry Centre (ICRAF) is the international leader in Agroforestry – the science and practice of integrating 'working trees' on smallholder farms and in rural landscapes. Agroforestry is an effective and innovative means to reduce poverty, create food security, and improve the environment. The Centre and its many partners provide improved, high quality tree seeds and seedlings, and the knowledge needed to use them effectively. We combine excellence in scientific research and development to address poverty, hunger and environmental needs through collaborative programs and partnerships that transform lives and landscapes, both locally and globally.



ISBN 92 9059 145 5

ICRAF
The World Agroforestry Centre
United Nations Avenue
PO Box 30677
Nairobi, Kenya
Tel: + 254 2 524 000
Fax: + 254 2 524 001
Contact via the USA
Tel: + 1 650 833 6645
Fax: + 1 650 833 6646
E-mail: icraf@cgiar.org
Internet: www.worldagroforestrycentre.org

Design: Mariska Koornneef
Printed by: Kul Graphics Ltd, Nairobi, Kenya

# Table of contents

Session 1

# Review of experimental design

## Exercise 1 - Sample protocol

The task is to study the protocol for an experiment to make sure you understand and are able to describe the design used. Your group will be allocated one of the following trials:

a) Influence of improved fallows on soil P fractions, on-farm trial
b) Effect of *Tithonia diversifolia* and *Lantana camara* mulches on crop yields in farmers fields
c) Screening of suitable species for three-year fallow
d) Upperstorey/understorey tree management trial

You can find the protocols for these trials in Part 4 (Experiments portfolio) of the **Data analysis of agroforestry experiments** workshop materials.

Start by looking at the **objectives.** Are these clearly stated? Can you understand them? Do they seem to be complete?

Now look at each of the following:

- treatments
- measurements
- layout

For each of them, try to describe clearly and exactly what was done. If it is not clear, explain why you find it difficult - what further information would help?

Choose a group member to be the rapporteur. Prepare a brief (maximum 4 minutes!) verbal report on the experiment you have been looking at. Highlight the four points of objectives, treatment, measurements and layout. Do NOT spend time explaining the background or context of the trial. You may critique the trial and suggest improvements but the main focus of the report should be a simple description of the design used.

## Exercise 2 - Participants protocol

Now repeat the analysis carried out in Part 1 but looking at the protocol for a trial provided by one of your group members.

Again you are looking to see whether you can identify the objectives, treatments, measurements and layout.

You may be critical in your analysis, but be constructive as well!

Choose a new rapporteur. Again prepare a short verbal report. You do not need to explain all the details of the experiment. Instead concentrate on the aspects that made it hard to understand. You do NOT have to present a revised protocol, but positive suggestions for improvement are welcome.

# Session 2 Objectives and steps in data analysis

## Exercise 1 - Sample protocol

Working in your groups and using the protocol from Part 4 (Upperstorey/understorey tree management trial) in the Experiments portfolio that you looked at in Session 1:

- Work out what the objectives of the analysis should be, knowing the trial objectives and the data available.
- Describe the summaries (means, rates, etc.) that are needed to meet the analysis objectives.
- Design the tables and graphs, which will present the experimental results. Sketch the table or graph frame, add labels but don't put in experimental data — use rough estimates of expected data to illustrate what the graph or table might look like.
- Define exactly which data will be used. For example, if yield was measured every season, will there be separate presentations for each season? From the average across all seasons? From some **discounted** cumulative total?
- Prepare a short report (one or two overhead transparencies) that displays the planned tables and graphs and discusses any difficulties encountered.

## Exercise 2 - Participants protocol

Now repeat the task of Exercise 1, but using a trial protocol provided by a participant in your group. Prepare a brief report which displays the tables and graphs planned, but also explain any difficulties you had completing the assignment.

You can use the results of this exercise (and Exercise 2 in Session 1) to start a file for the analysis of your own data.


# Software familiarization

## Exercise 1 — Introduction to GenStat

Follow the Introductory Tutorial on GenStat. You will find it in Part 1 of the GenStat for Windows Introductory Guide (GWIM). We estimate that it will take you up to one hour.

Do not spend too long on aspects that are unclear, but make a note of any points for discussion. We will have a review at the end of this exercise where you can ask the resource person to answer any questions you have.

## Exercise 2

1. In this workshop most data are stored in Excel spreadsheets and need to be imported into GenStat for analysis. Import the data from the trial '*Tithonia/Lantana* on-farm mulch'. Check that the importing has worked as expected. Use the data to find the mean yield across all the farms. Also draw a boxplot of grain yield for each treatment.
2. We have also seen how to use RESTRICT in GenStat to examine only a subset of your data. Go to the Explanatory Analysis section of GWIM. You should find it in Part 1 of the manual in the 'Factors, data exploration and analysis of variance' section. Start on the second page of this section where it begins to use data from the '*Tithonia/Lantana* on-farm mulch' trial. In this example it restricts the data to look at only one of the districts (Central). Follow the example and produce a boxplot of grain yield for each treatment, but only for the Central district.

If you have time, use any of our other example data sets to practice the method of restricting data. You could try to restrict by rows or by column values.

## Exercise 3 — Datasets for analysis

In the initial session on Data Management we looked critically at how data can be most accurately and efficiently stored. At the end of that session's practical exercises you should have reached the stage where your own files are 'neat and tidy'. Now try to import your data into GenStat. If it works first time, with all the data looking as expected, excellent! If not, try to identify the problems, correct them in the spreadsheet and try again.

### GenStat hint:

An easy way of transferring data into GenStat from Excel is to 'name' the required section of your data first. Within Excel highlight the data you wish to transfer, then go to the box in the top-left of the spreadsheet, at the moment it should read the name of the first cell you have highlighted. Within this box you can now name your highlighted section e.g. type 'data' into the box. When you go into GenStat and ask to import your Excel file, the item R:'data' will now be shown in the list. By selecting this range your data will automatically be read into GenStat, using the correct cell range and labelling the columns as they were labelled in Excel.

Session 4

# Descriptive analysis and data exploration

## Exercise 1 — Sample experiments

In this practical you are required to carry out some preliminary analysis of the data collected in one of the following studies:

a) Influence of improved fallows on soil P fractions, on-farm trial
b) Effect of *Tithonia diversifolia* and *Lantana camara* mulches on crop yields in farmers fields
c) Screening of suitable species for three-year fallow
d) Upperstorey/understorey tree management trial

Choose the trial that you have discussed in your previous group discussion sessions.

Your specific tasks are as follows:

i) In Session 2 you were asked to identify at least one skeleton table or graph to use as a sensible way of summarizing your results, which would address the objectives of the trial. Here you are now required to extract the data to produce this (or these) table(s) or graph(s).

ii) Consider what other questions you would like to ask about the data and identify whether they can be addressed by a table or graph. Produce the appropriate summary, and comment on the results.

e.g. in the 'Effect of *Tithonia diversifolia* and *Lantana camara* mulches trial' which some of you have looked at, examples of questions might be 'How many farmers in each district achieved the target biomass application of 100 kg/plot?' or 'Are the treatment differences influenced by the incidence of striga and/or streak on the plots?'

iii) Use appropriate methods to consider the variation in the data. Are there any strange values? What can you, or should you, do about them?

You are expected to use boxplots, scatterplots, tables of sensible summary statistics (such as the mean) and cross-tabulations. In some cases you will want to do this for the whole dataset, and in other cases you might want to look at subsets of the data (for instance in the on-farm trial you might wish to look at the patterns in the two districts separately).

## Exercise 2 — Participant experiments

*[This exercise may be moved to an additional session, after Session 9 where participants can attempt a complete analysis of their data using what they have learnt in Sessions 1 - 8].*

Carry out some preliminary analyses on the dataset that you used in the second practical in Session 2 where you looked at the analysis in relation to the objectives. This might be your own, or someone else's data. In that earlier practical you should have identified tables or graphs you wanted to produce. Start there, produce them and discuss the patterns that emerge.

Then carry out further explorations using the ideas and techniques you have just learnt in Exercise 1 in this session.

### GenStat hints:

*From the Stats pull-down menu*

From Summary Statistics
Choose Summaries of Groups : for tables

*At the Spread pull-down menu*

Choose Restrictions
And Using factor levels… : to work with a subset of the data
Unrestrict Unselected Levels : to remove the restriction

*From the Graphics pull-down menu*

Boxplots… : for graphical summaries
Pointplots …

If you are using graphics and you want to keep several graphics windows open at once, then when you have the GenStat Graphics window open, click on Multiple Windows at the Options pull-down menu.

Refer to the GWIM for more detail on the commands you need.


# Analysis of variance as a descriptive tool

## Exercise 1 — Understanding ANOVA

The first exercise aims at reviewing your understanding of the role of the ANalyses Of VAriance (ANOVA) table and to practice producing an ANOVA for a range of simple designs. GWIM is used for this purpose. This should not take more than 30 minutes.

1. Read into GenStat the data from the 'Screening of suitable species for three-year fallow' trial ('Fallow N.xls'). Restrict the data to include only the 1992 season of data (you should have already learnt how to restrict your data sets). Note that this trial has a Randomized Complete Block Design with 10 treatments and 4 replicates. Do an ANOVA with grain yield as your response, setting the options so that F-probabilities and standard errors are not given. Check that you understand all of the output.

2. In the lecture we described the analysis of the 2-factor 'Upperstorey/understorey tree management' trial. Repeat the analysis (again without F-probabilities and standard errors). Check that you understand all of the output. (N.B. Read into GenStat the data from the sheet R:fact – this contains the altered structure data)

3. In the data from the example above, change some of the values to simulate possible entry errors. For example density = 0 and Under = No; change Block 1 observation to 22.5, and Block 3 to 22.1. Why is the odd residual pointed out by GenStat not one of the ones that were changed?

4. With the original data restored, give the plot of the treatment factors, as we did in the lecture. Also give the breakdown of the treatment term into the polynomial effects as shown in the lecture.

(Hint: type or use the <Contrasts> button to change the treatment term from density*under to Pol(density; 1)*under.)

## Exercise 2 — Participant experiments

Take one of the datasets that you used for the descriptive statistics in the last session and conduct an ANOVA on this set. Does it add anything of value to the presentation that you did then? In your discussion include at least the three following points:

1. Are there any odd or curious values? If so, then does it affect the analysis much if they are omitted? If not, then what would you conclude? If so, then what would you do?

2. Does the analysis help in identifying which tables of means to investigate in more detail? Also, does it tell you roughly how much importance to attach to the different levels of each treatment factor?

3. Are you satisfied with the amount of residual variation in your trial? Is it small enough that the objectives of the trial can be realized?

4. If the residual variation is large, can you think of ways that it can be reduced? Remember that the basic idea is that your **data = pattern + residual**. So can you think of any part of the residual that can be explained by some effects that are not included yet? Then they could become part of the pattern and the residual would be reduced.

# Session 6 Ideas of simple inference

## Exercise 1

In the first part of this practical we review and extend the analysis that was covered in the lecture. Use the same set of data, namely the Kenyan on-farm experiment 'Effect of *Tithonia diversifolia* and *Lantana camara* mulches on crop yields in farmers fields'. Start by restricting attention to the second district, namely West.

Repeat the ANOVA and check you understand all the output. How do the results compare with those given in the lecture for the Central district?

In the lecture, two observations in the West district seemed odd. Try the analysis without these observations. How do the results change?

Do the results seem valid? GenStat offers a range of residual plots to investigate this aspect.

For the objectives of this trial it would be sensible to look at two treatment 'contrasts', i.e. *Lantana* vs Control, and *Tithonia* vs Control. The lecture you have heard for Session 6 has explained the theory behind these comparisons. Now you should try to work out the contrasts using GenStat. In GWIM Part 2 - Analysis of Variance - Further Topics, you should find a section entitled 'User Defined Contrasts' which considers the use of contrasts for a similar problem that also has 3 treatments. Do you find it easy to use contrasts? Does it clarify the interpretation of the results? What do you conclude in this case?

Repeat the analysis, including the contrasts on the data from the two districts together. Start by trying to reproduce the tables shown in Part 2 - Lecture notes, page 57, Table 2. Then add the contrasts to the analysis. Are the results clearer for the overall analysis with the 2 districts, or is it simpler to analyse the data from each district separately? Justify your approach.

## Exercise 2

Choose a different example, either from those provided, or use your own data if it is suitable and conduct an Analysis of Variance.

You should now be able to understand all elements of the general summary that is provided as a routine. This consists of the analysis of variance table, the table(s) of means and the corresponding standard errors. Check that everything is understood and give an overall summary of the results.

Now return to the protocol and assess the extent to which this overall analysis provides answers to the objectives of the study. You are likely to find that it provides partial answers. Reasons that the answers are not complete include:

- Some different variables need to be analysed.
- Some additional variables need to be included in the analysis that you have done.
- There is too much unexplained variability.
- There are some odd observations.
- The analysis is too general; it helps, but does not answer the specific objectives.

What is the situation in your example?

What are the next steps in your analysis? Specify these and attempt them.

# Session 7 An introduction to statistical modelling

## Part 1

Review the fitting and interpretation of simple linear regression models by working through the GWIM example on pages 61-64.

Now look again at the data from the trial 'Screening of suitable species for three-year fallow'. Investigate the relationship between soil inorganic N and maize yield in the second experiment (1992). Is there any indication of a relationship? Is it a straight-line relationship? If so, obtain estimates of the slope and intercept of the line. Assess the quality of the fit of the line and interpret the model.

## Part 2

Use the regression commands in GenStat to produce an analysis of variance table, treatment mean estimates and s.e.d.'s for an example that you can also analyse with the ANOVA commands. A good example would be the second experiment in 'Screening of suitable species for three-year fallow', which you've used in Part 1. Confirm that you get the same results using the two approaches.

Now carry out an analysis to identify treatment effects for an experiment that can **not** be analysed simply using the ANOVA commands, because block or treatment factors, are non-orthogonal or unbalanced. A good example is the 'Fertilizer, *Tithonia* and *Lantana* mulch as sources of phosphorus for maize' trial (look in Part 4 of the course booklets to find the protocol), or you may have one of your own.

Carry out an exploratory analysis then carry out an ANOVA that reveals the importance of block and treatment effects. Finally estimate treatment means that are adjusted for any possible block effects and interpret them.

## Part 3

Now return to the example 'Screening of suitable species for three-year fallow'. Try to answer the following questions:

1. We saw that in Experiment 1 there was still considerable difference in yield between treatments after allowing for effects of inorganic N. Can these differences be attributed to either the aerobic mineralizable N (AEROBIC) or the striga count (STRIGA)?

2. Are the effects of inorganic N on yield the same for both Experiments 1 and 2?

# Session 8 An introduction to multiple levels

## Exercise 1

In this practical, working in pairs, repeat the analysis of the split-plot example, 'Influence of improved fallows on soil P fractions' on-farm trial, which was discussed in the lecture.

The objective here is to understand how to request the correct analysis from the software as well as to understand the output.

## Exercise 2

(a) Inspect the data from the 'Effect of *Tithonia diversifolia* and *Lantana camara* mulches on crop yields in farmers fields' trial to assess its structure in terms of how many different levels there are and what measurements were made at each level. Produce some tabulations of these factors and identify, but do not carry out, some simple analyses which could be performed.

A short discussion of this will follow.

(b) Work through the exercise on 'multiple observations per experimental unit', which is in GWIM Part 2 – Analysis of Variance – Further Topics. The objective is to learn the necessary computing steps needed to be able to carry out a satisfactory analysis when there are multiple observations per experimental unit.

(c) Use the '*Leucaena trichandra* seed production trial' data, which was discussed in the lecture to try out some of the different approaches we can use to deal with multiple levels. You should also analyse the data as (i) a randomized complete block design with replicate and family effects and (ii) as an incomplete block design with 100 blocks of 4 plots.

A short discussion of both of these will follow, summarizing the main points, which emerged in (c).

Session 9

# Writing up and presenting results

## The Challenge

During the course you have been analysing data from a number of different studies, provided by the course organizers, brought by you or brought by another participant.

Now we challenge you to:

- Take one objective from a study you have been working on.
- Carry out the statistical analysis to meet the objective.
- Present the results using a **single graph or table** and a **single paragraph of text**.

Prepare the output in two ways:

1. A single printed page, which can be given to various people to comment on.
2. A single transparency with the table or graph together with a 100 word (maximum) commentary that you can present to the whole group.

The aim of this challenge is to force you to think about the most effective and concise way of describing the results. The reports have to be complete, presenting all the information necessary to meet the objective but cannot include anything unnecessary. It is not necessary to include any of the background or methods used in the study.

Ask for help, from resource persons, when selecting a study and objective for this challenge.

Session 10

# Where are we now?- Review of basic statistics

## Review and presentation

You have been put into a group and the group has been given a topic from the course so far.

Now you are asked to prepare a brief review of that topic. The review should summarize the key points on the topic and the importance of it to your work.

The review should be in the form of a presentation taking no more than 5 minutes and using less than 5 slides.

Each group should also provide a discussant, who adds a 2 minute discussion to the review. The discussant should concentrate on the problems they faced with the topic, the extent to which these have been overcome and the confidence they now feel in the area.

# Session 11 Design and analysis complexity

## Exercise 1

Working in pairs:

a) Carry out an analysis of non-normal data using the example in GWIM pages 154 - 156.

b) Investigate the idea of exploring treatment structure further using the example in GWIM, pages 122 - 125.

## Exercise 2

Work in groups looking at either one of:

a) Explore the ideas of complex treatment structure further by analysing the data from 'Upperstorey/understorey tree management trial'.

b) Carry out an analysis of the proportion of trees surviving in October in the trial 'Fruit trees survival'.

Try to:

i) Discuss the complexity of the problem and decide how it can be addressed.

ii) Attempt an appropriate analysis and interpret the findings.

You will probably need help using GenStat. Before asking for that help try to work out exactly what you are trying to achieve with the analysis. Try to use the resource people as 'GenStat experts who know no statistics'!

Prepare a short presentation that covers:

- the nature of any complexity
- the strategy for coping with it
- the analysis and interpretation.

Session 12

# Dealing with categorical data

## Exercise

1. In the lecture, data from 'Improved fallows and rock phosphate: farmers' experiences' were analysed, focusing on the response variable PresIF99, which records whether an improved fallow was present in 1999. Repeat the analyses using the 1997 and 1998 variables (PresIF97 and PresIF98) and determine whether the influence of ethnicity, practicing of natural fallow and farm size are much the same in the earlier years.

2. In the lecture separate analyses were carried out to look at the effects of categorical explanatory variables (ethnicity and natural fallow) and the continuous variable farmsize. Try to complete and interpret an analysis that considers all three in the same model.

3. Look carefully at the protocol 'Fruit trees survival' and the associated data set. A number of trees of several species were planted on each farm. The planting niche was chosen as either shallow or deep soil and chicken manure was applied to some of the trees. The objectives of the analysis are to determine how chicken manure and planting affect survival of these species. You will have to think carefully about if and how to allow for the different farms and species in the study. You might also experience some new technical problems in fitting models. Try to understand their source and think of solutions.

Session 13

# Getting more out of on-farm trials and multilevel problems

## Exercise

Use the data from the Malawi on-farm trial, 'On farm cropping with *Sesbania* and *Gliricidia*'. Consider the following objectives, all of which refer to the '98 season, when the rainfall was reasonable and it is expected that trees will have had time to become well established.

1. How much does inclusion of trees (either *Gliricidia* or *Sesbania*) increase yield compared to a crop-only control, and how much does this vary across farms?

2. It is expected that the trees will be most effective when soil P is high (the trees add N to the system but not P). Is this the case?

3. The steeper sloping fields tend to have shallow soil and are subject to erosion. We would not expect the trees to have much effect there. Is there evidence for that?

4. We might also expect that the steeper fields, being generally less productive, are often left unweeded. What is the evidence for that?

For each of these:

a) Carry out an exploratory or descriptive analysis that answers the question.

b) Determine the level (within or between farms) at which the relevant information occurs.

c) Carry out an inferential analysis if you are able to.

d) If you are unable, describe the analysis you would like to carry out and why it is difficult.

Session 14

# Complications in agroforestry trials

## Exercise

Choose to work on either a problem with repeated measures in space or in time (you will not have time for both!).

If interested in the spatial problem, work on the RAC trial. Try to repeat the analyses done in the lecture, but using grain yield rather than biomass. Then look carefully at the summary graphs and decide if there are other summary statistics it would be informative to analyse. Carry out such an analysis.

If interested in repeated measures in time have a look at 'Prototype hedgerow intercropping systems'. Look at the objectives and try to meet them. You will need to start by producing simple graphs of the response of treatments through time, and in relation to rainfall. The choose useful summaries and analyse them.

In both cases produce a brief presentation that describes not just the results, but any difficulties you had producing them or understanding them.

Session 15

# Complications in data

## Exercise

This practical is best done with sets of data brought by participants. Where this is not possible, the agroforestry datasets can be used instead.

Participants, in small groups should take one or more problems from the table given in the lecture that corresponds with a complication in the dataset they have. They should then investigate possible solutions to that problem, which are to be described in the discussion session.

The groups should check that they are each investigating different problems.

For guidance on the types of solution, reference should be made to supporting documentation. This includes:

- The lecture note for this session.
- The GWIM guide, particularly the section entitled Analysis of Variance - Further Topics.
- MCH, particularly Chapter 8.

Where possible participants should try different solutions to the complication, so that these can be compared.

Each group should prepare a short presentation. These presentations are to describe the method(s) of resolving the complication to the whole group. They are not to discuss the data analysis itself.

The presentations should also discuss reference material that helped in the solution.

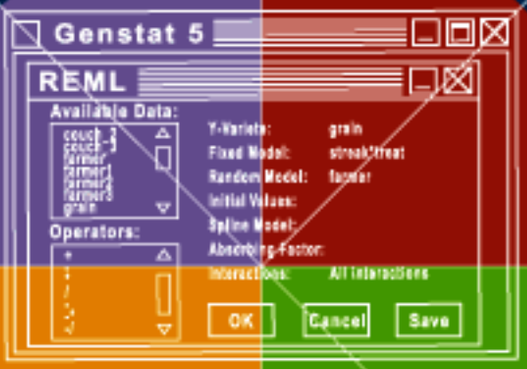
ISBN 92 9059 145 5

3

World Agroforestry Centre

TRANSFORMING LIVES AND LANDSCAPES

# Data analysis of agroforestry experiments




4

## Experiments portfolio

World Agroforestry Centre
TRANSFORMING LIVES AND LANDSCAPES

# Data analysis of agroforestry experiments

4

## Experiments portfolio

*by*

*Richard Coe, ICRAF, Kenya*

*Roger Stern, SSC, UK*

*Eleanor Allan, SSC, UK*

*edited by*

*Jan Beniest, ICRAF, Kenya*

*Janet Awimbo, ICRAF, Kenya*



The World Agroforestry Centre (ICRAF) is the international leader in Agroforestry - the science and practice of integrating 'working trees' on smallholder farms and in rural landscapes.

Agroforestry is an effective and innovative means to reduce poverty, create food security, and improve the environment. The Centre and its many partners provide improved, high quality tree seeds and seedlings, and the knowledge needed to use them effectively. We combine excellence in scientific research and development to address poverty, hunger and environmental needs through collaborative programs and partnerships that transform lives and landscapes, both locally and globally.

Founded in 1983, the Statistical Services Centre (SSC) is a not-for-profit body within the School of Applied Statistics at The University of Reading, UK. The SSC provides training and consultancy in both statistics and data management in the international arena. We aim to encourage good statistical practice, and the use of modern statistical methods in applied problems.

The SSC currently has nine statisticians, plus computing professionals and administrative staff.

ICRAF
The World Agroforestry Centre
United Nations Avenue
PO Box 30677
Nairobi, Kenya
Tel: + 254 2 524 000
Fax: + 254 2 524 001
Contact via the USA
Tel: + 1 650 833 6645
Fax: + 1 650 833 6646
E-mail: icraf@cgiar.org
Internet: www.worldagroforestrycentre.org

Statistical Services Centre
The University of Reading
Harry Pitt Building Whiteknights Road
P.O.Box 240
Reading RG6 6FN, UK
Tel: +44 (0) 118 378 8025
Fax: +44 (0) 118 975 3169
E-mail: statistics@rdg.ac.uk
Internet: www.rdg.ac.uk/ssc/


ISBN 92 9059 145 5

Design: Mariska Koornneef
Printed by: Kul Graphics Ltd, Nairobi, Kenya

# The Training Materials

These Training Materials were developed to help us present a series of courses on the analysis of data from agroforestry experiments. They are published here to assist others give similar training in the future.

The course is very practical and built around the analysis of real data sets. Concepts are explained largely without using mathematics. The computer software takes care of calculations and hence formulae are not used. Instead the course emphasises understanding of the analyses it is sensible to use, and the interpretation of results. We distinguish between learning to use the statistical software (buttons to press or commands to use) and understanding the statistical concepts, models and methods.

The course was designed initially to help with analysis of agroforestry experiments, and the examples given are from agroforestry trials. However both the statistical and teaching ideas can be applied to trials from agriculture, forestry and other application areas. Only one out of 17 sessions is dedicated to peculiarities of agroforestry research, and it should be easy to substitute other examples when using the materials. The materials refer to both on-station and on-farm trials. Emphasising the distinction between on-station and on-farm experiments is not necessary or helpful for this course. The approaches and methods for the analysis of a trial depend on its objectives, treatments, layout and measurements, not on where it was carried out.

The materials are presented in four printed parts together with a computer CD.

**Part 1** contains an overview of the course and teaching approaches, with suggestions on how the materials may be used and adapted. It also contains a summary of each of 17 teaching sessions.

**Part 2** contains the lecture notes, one for each of the sessions. They form a useful and readable resource in their own right and hence are presented as a separate document.

**Part 3** contains suggested exercises for each session. These are presented as a separate document as they are most likely to be adapted and modified to use local examples.

**Part 4** contains a protocol describing each of 16 experiments, the data from which are used in examples.

The CD contains

- a data file (in Microsoft Excel format) for each of the 16 example experiments
- files (in pdf format) for each of the 4 parts, so that further copies can be printed
- the original word processor files of all the text (in Microsoft Word format), so users may modify and adapt the text
- some additional documents (in pdf format) that are referred to in the materials

We encourage the copying and modification of these materials as long as the original source is acknowledged, and resulting products are not sold without our permission. We would appreciate being informed of any use and developments of these materials.

The materials were produced through a long term collaboration between the World Agroforestry Centre (ICRAF) in Nairobi, Kenya and the Statistical Services Centre of the University of Reading, UK.

# Table of contents

# Introduction

Following are brief protocols for each of the examples referred to in the training course notes and practicals. Each protocol refers to the Excel data file that contains the data . The Excel data files are included on the CD.

The protocols have been provided solely for the purpose of describing the data, so that a sensible statistical analysis can be done. They are not a complete record of the trial, and only describe the details that are needed for the analysis of the data provided.

The data files have been provided to give realistic examples for use in training. They are not complete sets of data from the trials. In a few cases some details may have been altered in order to make a training point. Therefore the results of the analysis should be used to understand statistical methods, but not to reach conclusions about agroforestry practices.

# Relay planting of *Sesbania sesban* and maize

## Data file name

Relay planting.xls

## Trial Location

Makoka, Malawi

## Principal investigator

Prof. J. Maghembe, ICRAF-Malawi

## Starting date

12/1988

## Justification

Two *Sesbania sesban* provenances, *S. sesban* (ex Jamhuri, Ngong, Kenya) and *S. sesban* (ex Kakamega, Kenya) grow fast and have shown high biomass production in the unimodal upland plateau ecozone of Southern Africa. In the first year of growth, these *Sesbania sesban* provenances have attained 4 m in height in a 3-4 month time period, and showed remarkable production of phytomass in the same period, of 4-3 $tonha^{-1}$, dry weight. In addition, they are highly nodulated by local Rhizobia strains and they can easily be grown from nursery seedlings and by direct seeding. These characteristics of *S. sesban* make them ideal candidates for use in the improvement of infertile soils.

Within the Southern Africa Miombo ecozone, *S. sesban* has been used to enhance soil fertility in improved fallows, and alley cropping at Chipata and Chalimbana in Zambia. Initial results from these experiments are presented in the 1989 progress reports for Chipata and Chalimbana (Kwesiga, 1989; Kamara, 1989). In general, the prospect of developing viable technologies for improving soil conditions using these benefits will depend on existing land use constraints and the biology of the trees.

The small land holdings, 0.1 - 0.2 ha/family (Minae and Msuku, 1988), in the Shire highlands and the Lilongwe land use systems in Malawi, preclude the wide use of fallows. The small farm sizes also limit the utilization of agroforestry technologies in which trees occupy a substantial portion of the land.

We propose to test an agroforestry arrangement, which strives to maintain the recommended population of maize plants/ha while utilizing the soil improving capability of the sesbanias.

## Objectives

1. Determine the soil improving potential of *S. sesban* grown in relay with maize and left to occupy the field for the dry season after maize harvesting.
2. Determine how the yield of maize is modified by relay planting with *S. sesban*.
3. Determine a good time for planting *S. sesban* in relay with maize.
4. Determine the interaction between *Sesbania* and fertilizer applied to the maize.

## Treatments

Two factors will be used in a 4 x 3 factorial design. The factors constitute the following treatments:

1. Four *Sesbania sesban* planting times;
   - at maize sowing, P1,
   - at the period of rapid maize stem elongation, i.e. at node formation, P2,
   - during tasselling of maize, P3,
   - at maize growth maturity, P4.

2. Three N and P fertilizer levels;
   - no fertilizer applied, F0,
   - 50% of the recommended dose of Nitrogen and Phosphorus fertilizer, F50,
   - 100% of the recommended dose of Nitrogen and Phosphorus fertilizer, F100.

In addition there will be three control treatments of the maize without trees but with each of the three fertilizer levels.

A total of 15 treatments.

## Replicates

Three or six. The original design included a third factor (*Sesbania* provenance) which has now been dropped.

## Field layout

There were 27 plots in each of three blocks. The blocks coincided with bench terraces. Blocks 1 and 2 had all 27 plots laid out in a single line, the position indicated in the data file. In Block 3 the plots were arranged in two lines with a break between them; the positions again indicated in the data file.

## Measurements

Many tree and crop growth and soil variables were measured for a number of seasons. Here just maize grain yield for one 'typical' season in which the rainfall was good is presented.

## Field Layout:

| Distance from left-hand side | 0 | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Terrace Boundary | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Rep 1 (Terrace) | | P4 F50 | P4 F0 | P3 F50 | Control F100 | P4 F100 | P3 F0 | P2 F0 | P4 F50 | P4 F0 | Control F0 | P3 F100 | P3 F100 | P2 F50 | P1 F50 | P2 F100 | P4 F100 | P2 F50 | P2 F0 | P1 F100 | P1 F0 | P2 F100 | P1 F0 | P1 F50 | Control F50 | P3 F0 | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Rep 2 | | | | P3 F100 | P2 F50 | P2 F50 | P2 F0 | P4 F50 | P3 F50 | P4 F0 | P1 F0 | P2 F100 | P1 F100 | P3 F50 | P4 F100 | P3 F100 | P3 F0 | P1 F100 | P2 F0 | P1 F50 | P4 F0 | P1 F0 | P4 F100 | Control F50 | P3 F0 | P2 F100 | P4 F50 | Control F100 | P1 F50 | Control F0 | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Rep 3 | | | | | | | | | | | | | | | P2 F100 | P1 F50 | P3 F50 | P4 F50 | P3 F0 | P1 F0 | P2 F0 | P4 F0 | P1 F0 | P2 F50 | Control F100 | P3 F100 | P1 F100 | P1 F100 | P4 F50 | P3 F50 | P4 F0 | P3 F100 |
| | Control F50 | P1 F50 | P2 F50 | P2 F0 | P4 F100 | P3 F0 | Control F0 | P2 F100 | P4 F100 | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

***Sesbania* Planting time:** P1 = maize sowing P3 = tasselling
P2 = node formation P4 = maize maturity
Control = no *Sesbania*

**Fertilizer (% of recommended):** F0 = No fertilizer
F50 = 50% of recommended
F100 = 100% of recommended

# Effect of *Tithonia diversifolia* and *Lantana camara* mulches on crop yields in farmers fields

## Data file name

Onfarm tithonia and lantana mulches.xls

## Location

West and Central Bunyore, Vihiga District, Kenya

## Start Date

01/09/95

## Justification

Traditional hedges in internal and external boundaries for demarcation, protection and production functions constitute one of the most popular agroforestry practices found in most land use systems in the Eastern and Central Africa highlands (Hoekstra, 1988). Those hedges comprised generally *Lantana camara*, *Tithonia diversifolia* and other species and produce a large quantity of biomass, which is not fully utilized. Results from trials established in September 1994, comparing the effect of 6 different mulch species from existing hedges on maize fields have shown very promising results with *Lantana camara* and *Tithonia diversifolia*. These results were confirmed with another trial established in March 1995. Yield of maize is increased considerably if 5 t $ha^{-1}$ of *Lantana* or *Tithonia* are applied. It is therefore important to find out how these two mulch species perform on farmers' fields and under farmers' management conditions.

## Objectives

The overall objective of this trial is to find out if the good results obtained on-station with *T. diversifolia* and *L. camara* will be confirmed in farmers' fields and conditions.

In order to do this we need to:

1. Determine the effect of *Tithonia* and *Lantana* mulch on maize yield.
2. Investigate any problems associated with biomass transfer technology.
3. Measure the effect of different management practices, disease and weed problems on the performance of the mulches.
4. Assess the farmers views, opinions on the process, on implementing the technology and the effect of mulch application on crops, weed etc.

## Treatments

It was proposed that each farmer could have two or three treatments replicated once, with control (farmers usual practice) being one of the treatments in each farm. The other treatments comprise of:

+ *T. diversifolia*
+ *L. camara*

Each farmer was allowed to apply any quantity of the mulches, provided (s)he records the quantity applied and tries to use the same amount of both *Lantana* and *Tithonia*. It was recommended to use 100 kg per plot.

If a farmer uses animal manure, he was asked to use the same amount in all the treatments. Mulches can be applied before sowing or on existing crop, and the following application methods could be used:

- Spreading all over the plot.
- Point placement.
- Any other, provided the farmer states the method used.

## Farmer selection

Collaborating farmers are those who have shown interest during field visits organized in Ebukanga and Ochinga. Meeting of farmers and researchers were organized. During the meeting, the researchers discussed with the farmers the proposed research and questions arising about species, treatments, research protocols and modality of operations. Responsibilities of each of the partners were clearly explained and answered. It was emphasized to the farmers that the research was wholly under their management.

It was agreed that if the control performed better than the other two treatments, the farmers would be compensated for the difference. Harvesting the required amount of biomass, the application and incorporation are the responsibility of the farmers.

# Screening of suitable species for three-year fallow

## Data file name

Fallow N.xls

## Location

Chipata, Zambia

## Investigators

Dr F. Kwesiga, Dr E. Barrios, ICRAF

## Start Date

01/12/91

## Justification

Farmers in Eastern province complain that crop yields, especially maize, continue to decline year after year if inorganic fertilizers are not applied. This was confirmed during micro D and D in Chipata and Katete Districts (Ngugi, 1988). The problem is largely attributed to the low fertility status of the soil. To increase crop yields, farmers have adopted several strategies including use of chemical fertilizers, crop rotation with legumes and leaving land to rest (fallow) for a few years. The fertilizer strategy has become less reliable because the government subsidy has been removed. The economic price of these fertilizers is very high and most of the small-scale farmers cannot afford to buy them. In addition, the majority of these small-scale farmers are not eligible to bank loans which are offered over a short period at very high interest rates, presently at 45% per annum.

The cheaper options still available to these farmers are; (i) to continue using legumes in crop rotation, (ii) to apply scientific principles to improve the efficiency of traditional fallows.

It is now well established that legumes (groundnuts, soya beans or pigeon peas) can contribute at least 30kg N $ha^{-1}$/year and this is about 25% of the total N requirement of maize in Zambia. Besides being low, it is not known whether this N is in the available form for crop utilization at peak-time.

As for the fallows, there are still challenges which have to be overcome before this strategy can be of wide use.

1. Increasing population pressure on the land (the present growth rate in Zambia is 3.5% per annum) means that long fallow periods are not possible in most of the settled areas. In Eastern province, fallow periods of 1-5 years are very common on the plateau and are still declining (Kwesiga and Chisumpa, 1989).

2. There are studies in which some tree and shrub species have been shown to improve fallow period and increase subsequent crop yields (Kwesiga and Coe 1991, Adejuwon and Adesina1990, NCSU 1990, Palm et al., 1988, Saleen and Otsyina 1986 ). Such species improve the physical and chemical conditions of the soil in a short period when planted as improved fallows compared to natural vegetation which takes longer to reach the peak of biological productivity. Obviously, there are many fast growing trees and shrubs, which have not yet been evaluated in fallow improvement.

In order to improve on the efficiency of the natural fallow, there is need to screen appropriate trees that may have relevance in soil fertility regeneration within permissible fallow periods in Eastern Zambia.

Understanding the mechanism of improved fallow effects on crop yields in relation to rooting depth of tree and crop species, and the interaction with moisture uptake (including effects of litter on the enhancement or inhibition of inorganic decomposition and nutrient availability) is essential.



## Objectives

1. To screen species that might be suitable for 3-year fallows.
2. To study the impact of such fallows on soil changes and crop yields.
3. To evaluate the consequences of improved fallow on nutrient and water uptake of the crop.

## Treatments

T1 *Sesbania sesban* three-year fallow followed by maize
T2 *Gliricidia sepium* three-year fallow followed by maize
T3 *Leucaena leucocephala* three-year fallow followed by maize
T4 *Flemingia congesta* three-year fallow followed by maize
T5 *Cassia siamea* three-year fallow followed by maize

T6 Groundnuts in one-year rotation with maize

T7 Natural fallow for 3 years, followed by maize

T8 Continuous maize without fertilizer

T9 Continuous maize grown with 120 Kg N $ha^{-1}$

T10 *Calliandra calothyrusus* three -ear fallow followed by maize

## Statistical Design

Randomized complete block, 4 replicates.

The whole experiment was repeated starting one year later, on an adjacent site.

## Measurements

These include soil nitrogen analyses at the end of the fallow period, and before the start of the first crop, together with the crop yield that year.

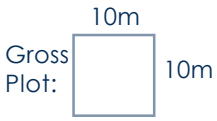
## Publication

Barrios E, Kwesiga, F, Buresh RJ, Sprent JI and Coe R (1998). Relating preseason soil nitrogen to maize yield in tree legume-maize rotations. Soil Science Society of America Journal 62: 1604-1609.

## Field Layout

Experiment 1(91)

| | | | | | |
|---|---|---|---|---|---|
| Block 1 | T3 | T7 | T4 | T10 | T1 |
| | T8 | T5 | T9 | T6 | T2 |
| Block 2 | T4 | T10 | T2 | T5 | T8 |
| | T6 | T9 | T1 | T7 | T3 |
| Block 3 | T8 | T3 | T2 | T9 | T5 |
| | T6 | T4 | T10 | T7 | T1 |
| Block 4 | T10 | T3 | T9 | T5 | T7 |
| | T8 | T4 | T1 | T6 | T2 |




Experiment 2 (92)

| | | | | | |
|---|---|---|---|---|---|
| Block 1 | T7 | T8 | T3 | T9 | T2 |
| | T6 | T1 | T5 | T4 | T10 |
| Block 2 | T6 | T3 | T9 | T1 | T8 |
| | T4 | T10 | T5 | T2 | T7 |
| Block 3 | T1 | T5 | T10 | T9 | T4 |
| | T6 | T2 | T7 | T8 | T3 |
| Block 4 | T3 | T6 | T9 | T2 | T4 |
| | T1 | T5 | T1 | T8 | T10 |

# Upperstorey/understorey tree management trial

## Data file name

Upper under storey.xls

## Location

Kabanyolo, Uganda

## Investigators

Nelson Wajja-Musukwe, NARO and Don Pedon, ICRAF

## Starting date

11/88

## Justification

While numerous upperstorey tree species and provenances are being evaluated for their suitability in agroforestry in the bimodal highlands of East and Central Africa, some were already well known. Two of these, *Grevillea robusta* and *Casuarina equisetifolia*, were chosen for management studies intended to evaluate less known management factors that might influence their performance. Two factors identified for study are optimal intra-row spacing of upperstorey trees on farmland, and the compatibility of these upperstorey trees when interplanted with understorey species such as *Pennisetum purpureum* (Napier, elephant grass) and *Calliandra calothyrsus*. Interplanting with understorey species is of great importance especially in sloping land where they may serve to control soil erosion and run-off. Other potential benefits from upperstorey / understorey combinations include production of fuel wood, green mulch, fodder, stakes for climbing beans, and stabilization of the terrace structures.

## Objectives

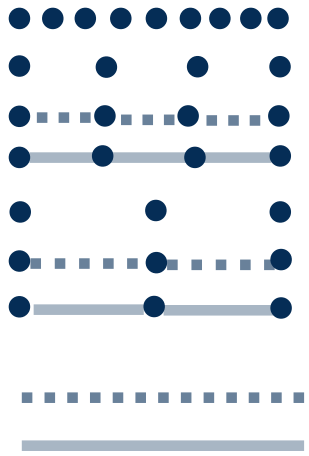
The objectives of the management trial are:

1. To determine the production potential of the upperstorey trees grown at different intra-row spacing and in association with Napier grass and *Calliandra* as understorey hedgerows.
2. To determine the effect of the upperstorey trees and understorey species on adjacent crops.
3. To determine the effect of the upperstorey and understorey species on each other.

## Treatments

Combinations of upperstorey tree spacing and type of understorey used in management trial at Kabanyolo.

| Intra-row spacing of upperstorey trees (m) | Understorey type |
|---|---|
| 1 | None |
| 3 | None |
| 3 | Napier |
| 3 | *Calliandra* |
| 5 | None |
| 5 | Napier |
| 5 | *Calliandra* |
| C | None |
| C | Napier |
| C | *Calliandra* |



'C' indicates infinite intra-row spacing (i.e. Control)

*Grevillea* Tree

Napier strip

*Calliandra* strip

## Statistical design

Randomized complete block design, 3 replicates.

## Field layout

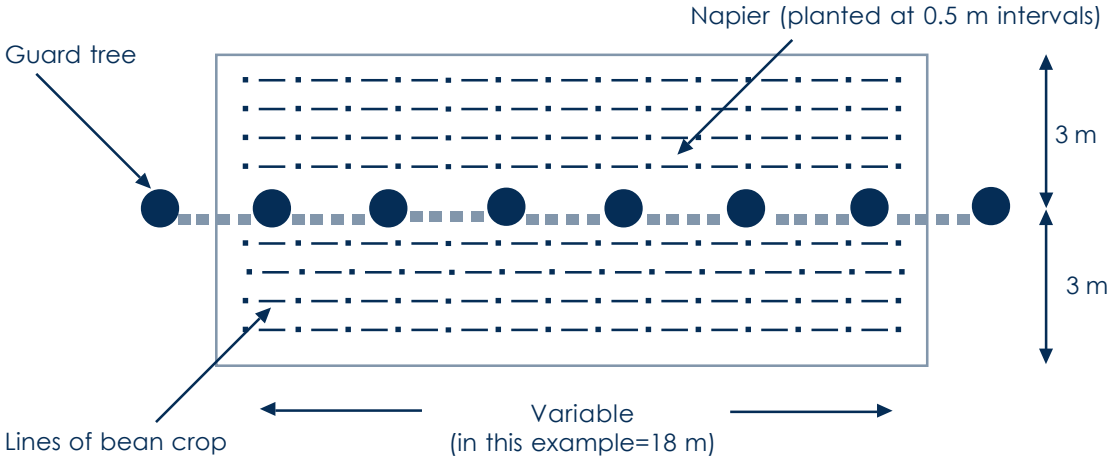
A plot consists of either a control without upperstorey trees or a single row of trees spaced at 1, 3 or 5 m intervals. Plot lengths vary according to the spacing of the trees, with 6 trees in each plot. The control (no tree) plot was 15 m long. Width of each plot is 6 m, with 3 m on either side of the tree row. Adjacent to the tree rows, on either side, a crop is raised. One guard tree is provided at each end of the plot. The understories, Napier grass and *Calliandra calothyrsus*, were planted in the same row as the trees at intervals of 0.5 m.

## Measurements

The upperstorey trees, understorey shrubs and adjacent crops were measured at regular intervals. In the data file we present crop yield for one season and several measures of upper storey tree size measured at the same time.

## Plot Design

Example Plot (Tree spacing 3metres, Napier hedgerow)




## Field Layout

Rep 1

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 6 | 4 | 8 | 2 | 9 | 10 | 1 | 3 | 5 | 7 |

Rep 2

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 10 | 3 | 2 | 6 | 7 | 4 | 8 | 9 | 1 | 5 |

Rep 3

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 4 | 7 | 8 | 2 | 3 | 10 | 5 | 9 | 1 | 6 |

# *Leucaena trichandra* seed production trial

## Data file name

Leucaena family trial.xls

## Location

Muguga, Kenya

## Investigators

James Were, Tony Simons, ICRAF

## Start Date

01/05/96

## Justification

Timely availability of adequate quantities of high-quality tree seed is vital for the success of any tree planting activity, agroforestry included. *Leucaena* spp have been used in a wide variety of agroforestry activities, although the germplasm used has been mainly of the fast growing *Leucaena leucocephala*. With the recent *Leucaena* psyllid problem in many parts of the world, there is need to diversify and use other leucaenas in place of the susceptible *L. leucocephala*. There is a need to produce seed of suitable provenances and optimize methods for this production.

## Objectives

1. To determine intra-provenance variation in survival, establishment and early growth of *Leucaena trichandra* (syn. *L. diversifolia ssp stenocarpa*).
2. To develop protocols for *Leucaena* spp. seed stand planting and management.
3. To produce seed of a superior provenance of *Leucaena trichandra.*

## Treatments

1. Twenty families representing a single provenance of *Leucaena trichandra*.

   Two other factors will be introduced at a later date:
2. Coppicing vs no coppicing.
3. Systematic vs selective thinning.

## Statistical Design

Incomplete block design. The 20 plots in one row (a complete replicate) made up 5 small blocks (each with 4 treatments). 20 replicates. Line plots of 4 trees each, no guards. 1m spacing with rows, 4 m between rows.

## Field Layout: arrangement of plots in field

| Replicate | Block 1 | | | | Block 2 | | | | Block 3 | | | | Block 4 | | | | Block 5 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 20 | 18 | 9 | 10 | 2 | 16 | 13 | 5 | 4 | 6 | 3 | 19 | 12 | 15 | 8 | 14 | 1 | 11 | 7 | 17 |
| **2** | 12 | 1 | 13 | 6 | 10 | 3 | 8 | 5 | 2 | 11 | 20 | 4 | 7 | 14 | 18 | 16 | 17 | 9 | 15 | 19 |
| **3** | 5 | 18 | 1 | 19 | 11 | 16 | 15 | 3 | 12 | 4 | 10 | 7 | 2 | 6 | 14 | 9 | 20 | 17 | 8 | 13 |
| **4** | 10 | 16 | 17 | 6 | 18 | 15 | 4 | 13 | 5 | 11 | 12 | 9 | 14 | 20 | 3 | 1 | 7 | 8 | 19 | 2 |
| **5** | 20 | 10 | 9 | 18 | 7 | 11 | 17 | 1 | 8 | 12 | 15 | 14 | 5 | 13 | 2 | 16 | 4 | 3 | 6 | 19 |
| **6** | 11 | 8 | 18 | 6 | 1 | 10 | 15 | 2 | 14 | 5 | 17 | 4 | 19 | 20 | 12 | 16 | 3 | 7 | 9 | 13 |
| **7** | 18 | 13 | 4 | 15 | 7 | 8 | 2 | 19 | 9 | 5 | 12 | 11 | 6 | 10 | 17 | 16 | 20 | 1 | 14 | 3 |
| **8** | 8 | 17 | 13 | 20 | 3 | 11 | 16 | 15 | 7 | 10 | 4 | 12 | 14 | 9 | 6 | 2 | 1 | 19 | 5 | 18 |
| **9** | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … |
| **10** | | | | | | | | | | | | | | | | | | | | |
| **11** | | | | | | | | | | | | | | | | | | | | |
| **12** | | | | | | | | | | | | | | | | | | | | |
| **13** | | | | | | | | | | | | | | | | | | | | |
| **14** | | | | | | | | | | | | | | | | | | | | |
| **15** | | | | | | | | | | | | | | | | | | | | |
| **16** | | | | | | | | | | | | | | | | | | | | |
| **17** | | | | | | | | | | | | | | | | | | | | |
| **18** | | | | | | | | | | | | | | | | | | | | |
| **19** | | | | | | | | | | | | | | | | | | | | |
| **20** | | | | | | | | | | | | | | | | | | | | |

## Arrangement of trees (X) within plots

# Fruit tree survival

## Data file name

Fruit tree survival.xls

## Location

Yucatan, Mexico

## Scientist responsible

Jeremy Haggar, ICRAF-Mexico

## Justification

Farmers in the region are keen to increase the range and number of fruit trees on their farms. However they have problems establishing the trees.

## Objective

To determine the extent to which soil depth affects early survival of fruit trees, and whether this can be improved with use of manure.

## Design

On-farm trial involving five farmers. Potential planting niches of shallow and deep soil, were identified and trees planted in June 1997. The trees were of four species. In August 1997 chicken manure was applied to some of the trees.

## Measurements

The number of trees planted in June 1997 was recorded. The number surviving in August 97, March 98 and October 98 was recorded for each niche-type, manure treatment, species and farmer.

# On farm cropping with Sesbania and Gliricidia

## Data file name

Onfarm gliricidia and sesbania.xls

## Location

Makoka area, Malawi

## Investigators

Prof J. Maghembe, G. Kooi, ICRAF

## Start Date

01/11/94

## Justification

After several years of on-station experimentation, relay cropping of maize with *Sesbania sesban* and mixed intercropping of maize with *Gliricidia sepium* shows promising results. These results should be verified under farmers' conditions.

## Objectives

1. To collect biophysical data to assess the effect of relay cropped *Sesbania sesban* and intercropped *Gliricidia sepium* on maize yield, and determine their potential in increasing maize yields.
2. To determine farmers assessments of these technologies.

## Treatments

1. Relay cropping maize with *Sesbania sesban*.
2. Intercropping maize and *Gliricidia sepium*.
3. Maize only control.

## Statistical Design

One replicate per farm. Some farmers have only one of treatments 2 and 3. A total of 42 farms, plus two farmer training centres.

# Roots and Competition (RAC)

## Data file name

Roots and competition.xls

## Location

Machakos, Kenya

## Start Date

10/1/93

## Justification

When crops and trees are grown together under semi-arid conditions, there is increasing evidence that competition for soil moisture is mostly responsible for the observed reduction in crop yields (Ong et al, 1991; ICRAF, 1992). Measurements of root distribution have shown that dense and superficial tree roots, such as exhibited by *Leucaena leucocephala*, can explain intense competition with crops. However, we still lack a clear relationship between tree root structure and competition with crops, Therefore we need to check the hypothesis that root density and distribution of a tree determines competitiveness with crops. In addition, it is not well known how tree management (e.g. pruning) can modify the tree root system and hence influence competition with crops.

In order to answer the above questions, this experiment uses a range of tree species with expected different rooting patterns, as well as contrasted management for some species. In order to correlate tree-crop competition with root distribution only, it is important to uncouple different possible interference between trees and crops. This is achieved by the canopy manipulation of the trees (pruning) according to the monitoring of transpiration rates and light interception, so that different tree species have similar water uptake and above-ground interactions with crops. Competition for nutrients is not expected since the site was previously under fallow and a cover crop assessment before the experiment did not show any sign of nutrient deficiency. Both N-fixing and non N-fixing species are nevertheless included in the experiment, since Nitrogen fixation may be a major determinant of competition with crops. This experiment is ICRAF's flagship initiative on root studies in water limiting conditions and should provide us with a better understanding of the relationships between tree root architecture and the consequences on competition for water with crops. Such information is critical both for the selection of multi-purpose trees and for the design of agroforestry technologies which minimize competition.

## Objectives

1. To measure the competition (in terms of reduced crop yield) due to differing tree species and pruning patterns.
2. To characterize important difference in patterns of competition.
3. To test the hypothesis that root structure of a tree determines the characteristics of competitiveness with crops.
4. To test the hypothesis that tree pruning induces changes in root structure, which in turn modify competition with crops.

## Treatments

T1. *Senna* spectabilis*, upperstorey tree
T2. *Senna spectabilis*, hedge
T3. *Croton megalocarpus*, upperstorey tree
T4. *Gliricidia sepium*, upperstorey tree
T5. *Gliricidia sepium*, hedge
T6. *Grevillea robusta*, upperstorey tree
T7. *Leucaena collinsii*, upperstorey tree
T8. *Melia volkensii*, upperstorey tree
T9. *Leucaena leucocephala*, upperstorey tree
T10. *Casuarina equisetifolia*, upperstorey tree
T11. Control (crop only).

(* formerly *Cassia*)

The associated cropping pattern is a maize-bean rotation.

## Statistical Design

Randomized complete block design. 4 Replicates

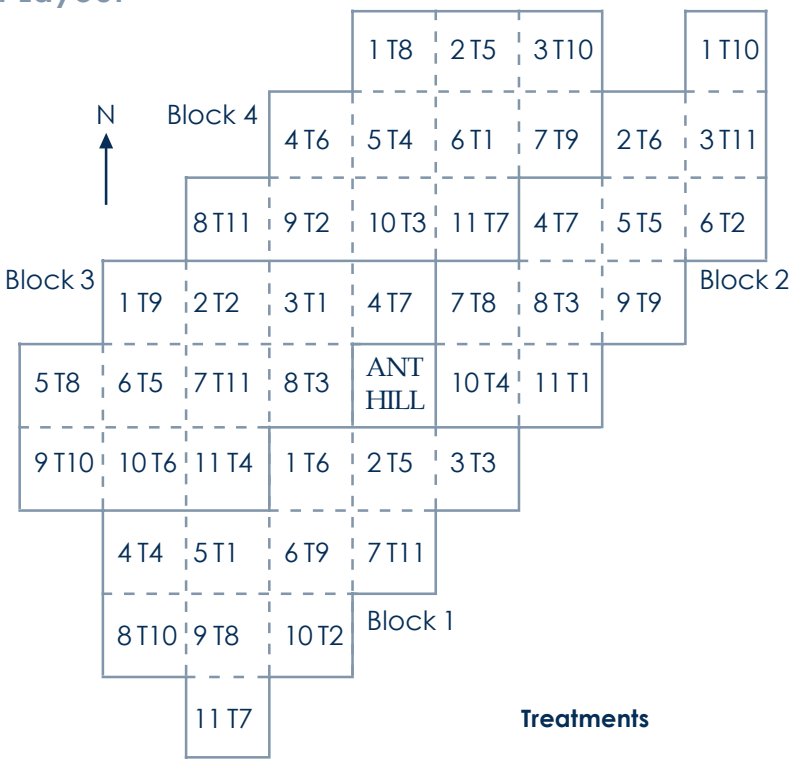
## Layout

A plot consists of a single line of trees (1 m spacing) with parallel rows of crops on either side. Trees are grown for pole (upperstorey) or mulch (hedge) production, as a single row per plot, with crops on both sides; a simple arrangement which is expected to yield clear results on competition at the tree-crop interface. Blocking was done according to a visual assessment of the land and soil. A maize cover crop grown prior to the experiment showed very uniform yields, with the exception of one ant hill which was avoided. Blocks 1 and 2 are on the upper flat zone, while blocks 3 and 4 are on the slightly sloping area to the West, where the soil is not as dark as

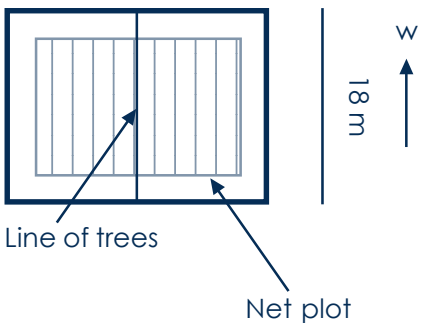
in the upper part. Tree rows in plots are East-West oriented, in order to minimize light interference with crops. Plots are sufficiently large (18 m x 18 m) to accommodate repetitive environmental and root measurements as well as possible superimposed studies. Trenches are seasonally dug and refilled in a 2 m-wide border area around all plots to check any root interferences across treatments.

Ong CK, Corlett JE, Singh RP, Black CR. (1991). Above and below ground interactions in agroforestry systems. *Forest Ecology and Management 45: 45-57*.

## Field Layout




### Plot Layout




**Treatments**

T1 = Senna spectabilis (Upper-storey)

T2 = Senna spectabilis (Hedge)

T3 = Croton megalocarpus (Upper-storey)

T4 = Gliricidia sepium (Upper-storey)

T5 = Gliricidia sepium (Hedge)

T6 = Grevillea robusta (Upper-storey)

T7 = Leucaena collinsii (Upper-storey)

T8 = Melia volkensii (Upper-storey)

T9 = Leucaena leucocephala (Upper-storey)

T10 = Casuarina equisetifolia (Upper-storey)

T11 = Control (Crop only)

# Prototype hedgerow intercropping systems

## Data file name

Prototype HI.xls

## Location

Machakos, Kenya

## Investigators

M.R.Rao, M.N.Mathuva, ICRAF

## Starting date

11/1989

## Justification

The potential of hedgerow intercropping for sustained crop production has been demonstrated by IITA in the humid to sub-humid lowlands of Africa. But its prospects in semi-arid environments as a means of improving soil fertility and maintaining crop yields, has not been proven conclusively. In these areas the woody perennial may compete with annual crops for soil moisture and the system may not show similar benefits as in high rainfall areas. Nevertheless, on sloping areas the hedges act as barriers and reduce soil erosion, and to that extent the hedgerow intercropping is beneficial even in semi-arid tropics. Hedgerow intercropping demands extra labour for establishment and management of hedges. To what extent the system is economically viable in limited labour areas has not been examined so far.

An alternative approach to the *in situ* incorporation of hedge pruning is to feed the biomass to livestock and return manure to the field for improving the soil fertility. This approach fulfils both soil fertility improvement and livestock production.

An earlier experiment at the Field Station, failed to establish the potential of hedgerow intercropping because of the absence of control treatment and fertilization of annual crops.

## Objectives

1. To evaluate the potential of hedgerow intercropping in maintaining crop yields in semi-arid conditions.

2. To compare the effects of *in situ* incorporation of hedgerow prunings with recycling of prunings through livestock as manure.
3. To determine the extent to which seasonal rainfall modifies the performance of the systems.

## Treatments

1. Annual crop, no fertilizer and crop residues removed.
2. Hedgerow intercropping, no fertilizer and biomass incorporated in alleys.
3. Hedgerow intercropping, no fertilizer, and biomass fed to oxen and manure returned to crops.
4. Annual crop with recommended fertilizer rates and residues taken out.

## Statistical design

Randomized block design, 3 replicates. Plot sizes varied in order to fit them into the site. However all were over 500 $m^2$.

## Measurements

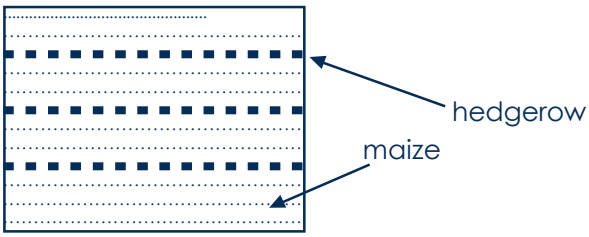
These included crop growth and yields, hedge growth and production, nutrient contents of all components, labour inputs and others. Here we present crop yield data for all seasons of the trial.

## Publication

Mathuva MN, Rao MR, Smithson PC, Coe R (1998). Improving maize yields in semi-arid highlands of Kenya: agroforestry or inorganic fertilizers? Field Crops Research 55: 57-72.

## Field layout


Plots of Hedgerow treatments (2 and 3)




(area is approximately 550 $m^2$)

# Fertilizer, *Tithonia* and *Lantana* mulch as sources of phosphorus for maize

## Data file name

Organic-inorganic P.xls

## Location

Maseno (Ochinga farm), Kenya

## Starting date

03/95

## Justification

Phosphorus is the limiting nutrient for most of the soils in Western Kenya. Phosphorus application is inevitably required to increase crop production on these soils. Small-scale farmers have insufficient capital to purchase inorganic fertilizers. Consequently there is a great potential in use of organic materials to supply nutrients for crop production. One possible source of this organic materials could be the traditional hedges, used in internal and external farm boundaries for demarcation and production function, which are one of the most popular agroforestry systems found in most of land use systems in Eastern and Central Africa highlands (Hoekstra, 1988). These hedges are comprised generally of *Lantana camara* and *Tithonia diversifolia*. They produce a large quantity of biomass which is not fully utilized and which can be harvested and used as mulch in order to provide nutrients particularly P to the crops.

## Objectives

1. To compare effects of *Tithonia diversifolia* and *Lantana camara* mulches and inorganic P at comparable P levels on maize and soil properties, in particular to measure any advantage of using organic sources.
2. To determine and compare the response of maize to different rates of inorganic and organic P inputs.

## Treatments

1. Control with no input
2. Inorganic 12.5 kg $Pha^{-1}$ (TSP)
3. Inorganic 25 kg $Pha^{-1}$ (TSP)

4. Inorganic 50 kg $Pha^{-1}$ (TSP)
5. *Lantana* mulch 5 $Tha^{-1}$ (D.M)
6. *Lantana* mulch 10 $Tha^{-1}$ (D.M)
7. *Lantana* mulch 20 $Tha^{-1}$ (D.M)
8. *Tithonia* mulch 5 $Tha^{-1}$ (D.M)
9. *Tithonia* mulch 10 $Tha^{-1}$ (D.M)
10. *Tithonia* mulch 20 $Tha^{-1}$ (D.M)

The quantity of dry matter to apply is based on the estimated P content of *Lantana* and *Tithonia* mulches (0.2% to 0.3%). 5, 10 and 20 tonnes of dry matter will correspond to 10-15, 20-30 and 40-60 kg $Pha^{-1}$ which is equivalent to the quantities of inorganic P applied.

All the fertilizers will be incorporated one day before sowing maize. The fertilizers and mulch will be applied only during the long rainy season and the residual effect will be assessed during the short rainy season.

## Statistical design

Blocked design with unequal replication.



## Layout

Rep 1

| | | | | | |
|---|---|---|---|---|---|
| 8 | 2 | 6 | 5 | 3 | 1 |
| 4 | 1 | 10 | 7 | 9 | 5 |

Rep 2

| | | | | | |
|---|---|---|---|---|---|
| 8 | 10 | 2 | 6 | 5 | 4 |
| 5 | 9 | 8 | 1 | 3 | 7 |

Rep 3

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 9 | 3 | | | | | | | | |
| 7 | 4 | 6 | 2 | 8 | 1 | 5 | 8 | 1 | 10 |

# *Calliandra* feeding trial

## Data file name

Calliandra feeding trial.xls

## Location

Embu, Kenya

## Justification

Preliminary work has shown that fresh leaves of *Calliandra calothyrsus* provide a fodder supplement for dairy cattle. Nutritional analysis suggests that 3 kg of *Calliandra* should be nutritionally equivalent to 1 kg of concentrate. This trial aims to test that.

## Objectives

Determine whether *Calliandra* will substitute (in terms of milk yield) for dairy concentrate at the rate of 3 to 1 when added to the basal diet of cows in Embu.

## Treatments

1. Basal diet alone
2. Basal + 3 kg *Calliandra*
3. Basal + 1 kg concentrate
4. Basal + 3 kg *Calliandra* + 1 kg concentrate

## Design

Twelve Farmers, each with a lactating cow, are used in the study. Each cow is allocated to a sequence of treatments in which they are on one treatment for a period, and then change to the next. A treatment period is 14 days with milk yield assessed during the second seven only.

# Effects of organic and inorganic sources of nutrients on striga, weeds and maize

## Data file name

Organic-inorganic-weeds.xls

## Location

Maseno, Kenya

## Start Date

01/09/95

## Justification

Low available soil phosphorus, striga and weeds are the major factors responsible for the low maize yields, which hardly exceed 1 $tha^{-1}$, in much of western Kenya. Fertilizer trials on farmers fields frequently showed large maize yield responses to phosphorus. However, observations in an on-going trial in Ochinga farm indicate that response to P in striga-infested fields is dependent on how well the striga is controlled.

Many resource-poor farmers cannot afford to purchase inorganic P fertilizers. An alternative seems to be the use of organic materials already available on the farm. Trials are currently being conducted in Western Kenya to evaluate the organic materials such as *Tithonia diversifolia, Lantana camara* and *Senna* spp. as sources of nutrients. Preliminary results from these on-going trials suggest that the organic materials, besides increasing maize yield, also reduce the detrimental effect of striga on maize growth.

Organic residues are important mostly for the supply of N and P. It is very well known that high N fertility reduces the striga infestation. It raises an important question whether the reduced detrimental effect of striga due to the application of organic residues, was something to do with their capacity to supply N or some other factors. In high rainfall and light soil conditions as prevailing in Western Kenya, N from inorganic fertilizers may quickly leach beyond 30 cm soil depth, where striga is most active. On the contrary, organic residues may release N slowly, depending on the characteristics of the materials and thus may maintain high N status in the top soil over a prolonged period so as to reduce the striga infestation.

A number of research issues emanate from the above background regarding the utilization of organic materials such as: a) the need to evaluate a large number of organic materials to determine their effect on striga, b) the need to determine the effect of organic vs inorganic sources of P and their integrated use (as integration becomes necessary because the amount of material needed to apply adequate quantity of P through organic sources is too large to obtain on-farm), and c) the potential benefits of organic on annual weeds and soil fauna.

## Objectives

1. To determine the effects of organic and inorganic sources of P on maize yield in the presence of striga infestation.
2. To assess the effects of different organic materials on striga incidence and in turn its effect on maize yield.
3. To observe the effect of organic materials on changes in soil fauna, particularly plant parasitic nematodes.
4. To examine the effect of organic and inorganic nutrient sources on weed dynamics (species composition and biomass).

## Hypotheses

a) Organic P from tree residues of specific characteristics is more efficient in improving maize yield than the inorganic P.

b) Organic sources of N decreases striga infestation, better than inorganic N; the magnitude of the effect being dependent on the decomposition and mineralization rates of the material.

c) *In situ* decomposition of organic residues reduces the detrimental effect of striga; the magnitude of effect being dependent on the secondary compounds present in the material.

d) Organic residues reduce the population of plant parasitic nematodes.

## Treatments

### 1. Fertility treatments applied to main plots

| Treatment | Source | Rate of biomass (t $ha^{-1}$) | Nutrient supplied(kg $ha^{-1}$) N | P | K |
|---|---|---|---|---|---|
| 1 | inorganic | | 0 | 150 | 100 |
| 2 | inorganic | | 120 | 0 | 100 |
| 3 | inorganic | | 120 | 10 | 100 |
| 4 | inorganic | | 120 | 25 | 100 |
| 5 | inorganic | | 120 | 50 | 100 |
| 6 | inorganic | | 120 | 150 | 100 |
| 7 | *Tithonia diversifolia* | 5 | * | * | 100 |
| 8 | *Lantana camara* | 5 | * | * | 100 |
| 9 | *Calliandra calothyrsus* | 5 | * | * | 100 |
| 10 | *Senna spectabilis* | 5 | * | * | 100 |
| 11 | *Sesbania sesban* | 5 | * | * | 100 |
| 12 | *Croton megalocarpus* | 5 | * | * | 100 |

* The amount of N and P added by the organics will depend on the chemical composition, which will be determined every season at the application time.

### 2. Split-plot treatment of removing striga as it emerges or not

All the selected organic materials contain fairly high N and P but differ in respect of tannins, polyphenols, etc. A 5 $tha^{-1}$ rate was chosen so that all the materials provide a minimum of P and >120 kg $Nha^{-1}$ and their effects are related to characteristics of the material without being confounded by different rates of N.

## Statistical Design

Main -plot treatments: organic and inorganic sources of nutrients.

Sub-plot treatments: presence or absence of striga.

Four replicates in complete blocks.

# The influence of improved fallows on soil phosphorus fractions; an on-farm trial

## Data file name

Phosphorus fallow.xls

## Location

Maseno, Western Kenya

## Starting date

10/1998

## Principal investigator

Roland Buresh, ICRAF-Nairobi

## Justification

Improved fallows are a promising intervention for increasing the availability of N to subsequent crops. Improved fallows however, do not recycle sufficient P to overcome P constraints to subsequent maize crops on the severely P-deficient soils, common in Western Kenya. Past work with *Sesbania sesban*, nonetheless, suggests that fallows might at least slightly increase the availability of P to crops by increasing the quantities of P in labile soil pools.

Most past research has been conducted with *Sesbania sesban*, but two other promising species for improved fallows are *Crotalaria grahamiana* and *Tephrosia vogelii*. Most work with fallows in Western Kenya has been on relatively clayey soils with little or no water deficit during the growing season. This research therefore, involves the relatively little studied crotalaria and tephrosia.

## Objectives

1. To determine the effect of fallows on soil P fractions.
2. To assess spatial variability associated with measurement of soil P fractions.
3. To measure the N and P contributions of the fallows to growth of a subsequent maize crop.

## Experimental design

Initially, a randomized complete block design with four fallow treatments, is replicated on 9 farms. Each farm contains one plot of each of the fallow treatments. After the growth and harvest of the fallow treatments, each plot will be split into two. One half of each plot will receive N fertilizer, the other half will not. The N fertilizer will be applied in the form of a split application of Urea (100 kg $Nha^{-1}$), 30 kg $Nha^{-1}$ will be broadcast and incorporated immediately before planting, and 70 kg $Nha^{-1}$ will be side dressed.

Final design is therefore split-plot with fallow treatment as main-plot factor, and N application as sub-plot factor.

Treatment combinations are as follows:

Continuous maize ± N
*Tithonia* fallow ± N
*Crotolaria* fallow ± N
*Tephrosia* fallow ± N



## Rationale for sub-plots

Both N and P could limit maize growth in the –N sub-plots, whereas N will not limit maize growth in the +N sub-plots. The comparison of +N and –N sub-plots within a main-plot will assess whether the fallows have eliminated N deficiency for maize.

Differences in maize yield among treatments for the +N sub-plot will result from differences in P plus 'fallow benefits' to maize. Differences in maize yield among treatments for the -N sub-plot will result from differences in N plus P plus 'fallow benefits' to maize.

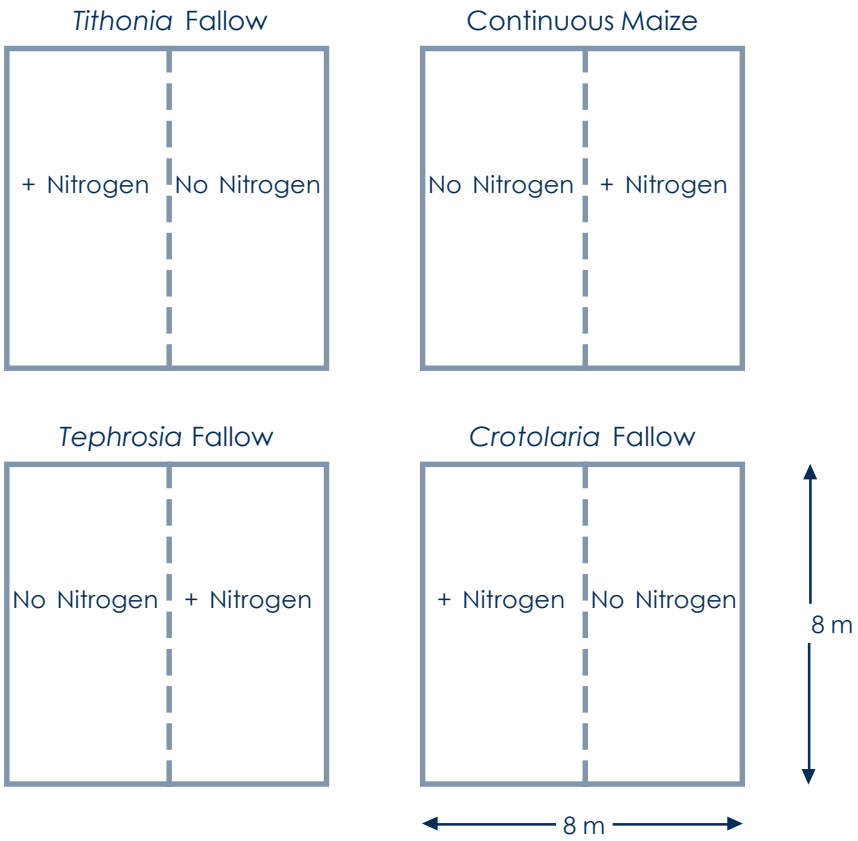
## Measurements

Among other measurements:

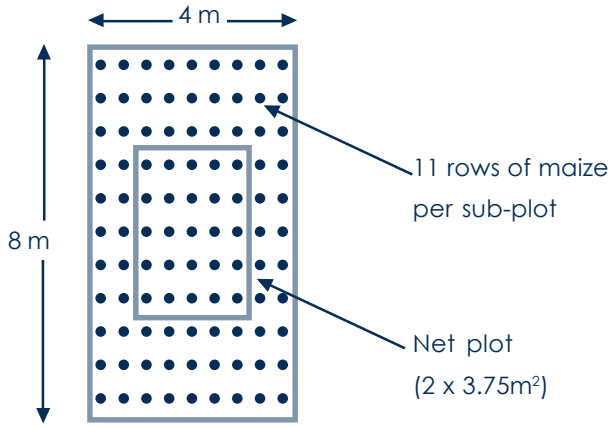
- Striga counts recorded 4 times, beginning after the 1998 short rains and ending before the 1999 long rains.
- Maize grain yield in 1999 long rains.

(Note the soil P data is not provided)

## Field Layout (example plan for one farm)




## Sub-plot Layout



# Improved fallows and rock phosphate: farmers' experiences

## Data file name

IF farmers experience - census.xls, IF farmers experience - sample.xls

## Location

Vihiga and Siaya District, Western Kenya

## Starting date

1/1997

## Justification

Fallowing by allowing development of spontaneous vegetation on a field, has always been part of the farming system. Its role is to build up organic matter and nutrients that were depleted during a period of cropping and to interrupt the life cycles of pests associated with the crops. In many places, pressure on land has shortened the period for which fields can be left fallow. Natural fallows can no longer restore soil fertility to the level needed at the beginning of the cropping cycle.

A modification, called 'improved fallows' or 'IF', is to plant trees or shrubs at a high density, instead of letting the spontaneous vegetation develop freely. Desirable characteristics of the planted species include ease of propagation, rapid biomass production, efficient nitrogen fixation, deep rooting and ease of removal. Fallows of this kind should be able to restore soil fertility more rapidly than the traditional ones and hence allow a shortened fallow period. This makes improved fallow an attractive soil management tool in areas with relatively high land pressure.

Experiences with improved fallows in other regions of Africa and in researcher-managed trials in Western Kenya, indicate a possible role of the technology to increase food production of the area. The aims of the project on soil fertility replenishment are to disseminate the technology and then to study how adoption occurs and what impact it has on the livelihoods of the farmer community.

The target area for the project is known to be deficient in phosphorus. Fallowing does not increase P levels in the soil; additional P sources are needed. Researcher controlled trials have shown that the relatively cheap rock phosphate may be effective when used in conjunction with improved fallows. The combination of both technologies might increase their adoption and it is therefore crucial to test both technologies at the same time.

Improved fallows are mainly grown for soil fertility improvement. However, additional benefits are possible. Fuel wood production and the suppression of weeds and pests are some of the known benefits. More unknown benefits could be present.

Not all improved fallow species are equally good in performing these various roles, and their suitability depends on the farming system on a farm. Although some general characteristics of these species are known, some still have to be confirmed and some might still be discovered.

This study builds further on activities that are ongoing in the area. Researcher-controlled experiments with improved fallows created interest of farmers in the neighbourhood. Seed of improved fallow species was made available and many farmers started testing the technology on their own farms. The main objects of investigation are these spontaneously testing farmers.

Primarily this study investigates whether the potential advantages of improved fallows can be realized under farmers' conditions; both the establishment of the species grown as improved fallow and the effect of the fallow on crops, may be dependent on the skills and level of management of the farmer. Secondly, it is to document the experiences of farmers with the technology; especially comparing the different species used to improve the fallows. Thirdly, it aims at establishing whether this spontaneous testing will lead to adoption and which factors might facilitate or constrain this adoption.



## Objectives

1. To understand the decision making process on planting fallows (why), the choice of species (which) and niche for fallows (where).
2. To assess the additional labour needed for improved fallow (planting and cutting) and how this fits in the normal farming calendar.
3. To assess the experiences of farmers with the planting and managing of improved fallows.
4. To assess farmers' perception on establishment, growth and biomass production of the various species, as well as their problems of pests, diseases and management.
5. To assess the experiences of farmers with the application of rock phosphate.
6. To assess the experiences of farmers related to crop responses after improved fallow, with or without rock phosphate.
7. To assess preferences of farmers among improved fallow species.
8. To assess major additional benefits of improved fallows, besides soil fertility improvement.
9. To assess farmers' future plans with improved fallows and rock phosphate.
10. To assess if the above topics depend on gender, ethnicity and resources (land, labour, cash).

11. To assess the actual adoption of improved fallows over time and define constraining and facilitating circumstances for this adoption.

## Parts of the study

Farmers that were spontaneously testing improved fallow in 1997 were offered the use of rock phosphate, at a time when farmers should have had the chance of completing a rotation cycle fallow-crop; data will be collected on the farmers' reactions.

The information collection consists of three parts:

(1) A general census that covers the entire population of the pilot project area.

(2) A questionnaire with a limited sample (121 farmers) that had experience with improved fallows.

(3) A group discussion with knowledgeable farmers on topics where the questionnaire failed to yield adequate information to be held in 4 areas.

## Census

The entire population is covered by a short questionnaire to gather baseline information about household composition, farming system, farm assets and available resources. This will take place once only.

Every year a new inventory is to be conducted to establish which farms have improved fallows, what species are used and how large the area is under improved fallow.

## Questionnaire

For more detailed information about experiences with improved fallow and rock phosphate and to compare the benefits of the various fallow species, a study is carried out on the group of farmers that has experience with the technology.

A questionnaire is used to collect the information. It covered all objectives listed, was developed and tested with 5 households that were not part of the sample. It consists of two parts. One contains questions on information at farm level. A second part contains questions about each improved fallow plot that was established on that farm. Additional questions are asked of farmers that had chosen to test more than one species. Bao-games are used to rank species and their benefits.

## Group interview

Knowledgeable farmers are invited to a group discussion to cover some aspects in more detail. Farmers are chosen on the basis of interest (either positive or negative) during the individual interviews. Topics to be covered are those needing more clarification after the questionnaire, but certainly include the following;

- Advantages/disadvantages of individual species
- Methods of establishment
- Seed production
- Use of improved fallow wood as firewood
- Methods of cutting and incorporating biomass
- Additional labour requirements
- Application method of rock phosphate
- Negative effects of rock phosphate on crops
- Choice of crop after improved fallow
- Crop performance after improved fallow
- Niche of improved fallow

One session of this group discussion was organized for each of the areas, described in the stratification below.

## Stratification and Sample

The entire population in 17 villages, i.e. 2035 households (HH) is monitored for the presence of improved fallow and the basic set of census questions. The villages are grouped in four areas. The Luhya people occupy three of these areas, the Luo people the fourth.

In November 1997, 473 families had improved fallow. This group of families was used as the population to sample the farms visited for the questionnaire. The overview of the sampling frame is given in Table 1. Within each area, households were grouped according to the fallow species they planted to ensure the presence of less-common species in the sample. A random selection of the households took place within these groups. A total of 121 households were interviewed.

| Area | Ethnic group | Total number | Total number of HH with improved fallows | Number of HH in sample |
|---|---|---|---|---|
| Bar Sauri | Luo | 747 | 136 | 46 |
| Ebukanga | Luhya | 546 | 193 | 24 |
| Emmbali | Luhya | 293 | 21 | 20 |
| Essaba | Luhya | 467 | 87 | 31 |
| Total | | 2053 | 437 | 121 |

Table 1

Note the number of records in data files may be less than the above due to incomplete information.

# On-farm trial with improved fallow and inorganic fertilizer

## Data file name

Onfarm IF and P fertilizer.xls

## Location

Vihiga and Siaya District, Western Kenya

## Starting date

6/1997

## End date

12/1998

## Justification

Fallowing by allowing development of spontaneous vegetation on a field has always been part of the farming system. Its role is to build up organic matter and nutrients that were depleted after a period of cropping and to interrupt the life cycles of pests associated with the crops. However, pressure on land has in most cases shortened the period of a field being fallow. Natural fallows can no longer restore soil fertility to the level needed at the beginning of the cropping cycle.

A modification, called 'improved fallows' or 'IF' is to plant trees or shrub species at a high density, instead of letting the spontaneous vegetation develop freely. Desirable characteristics of the planted species include ease of propagation, rapid biomass production, efficient nitrogen fixation, deep rooting and ease of removal. Fallows of this kind should be able to restore soil fertility more rapidly than the traditional ones and hence allow a shortened fallow period. This makes improved fallow an attractive soil management tool in areas with relatively high land pressure.

The combination of inorganic fertilizer with improved fallows might bring out their full potential. The combination of both technologies might increase their adoption and it is therefore crucial to test both technologies at the same time.

## General objectives

The agronomic potential of improved fallows is to be tested in farmer-managed on-farm trials.

The main aim is to compare improved fallows with traditional fallow and continuous maize cropping. The fallow species used in these trials are mainly *Crotalaria grahamiana* and *Tephrosia vogelii*.

These cropping systems are or are not combined with phosphorus fertilization. Because of the high prevalence of phosphorus deficiencies, which cannot be overcome with the fallow on its own, and the presumption that it would even jeopardize the effectiveness of improved fallows, it was assumed that phosphorus fertilizer is a necessity. The phosphorus is either applied as the standard TSP (Triple Super Phosphate) or as the cheaper, but bulkier and less easily to manipulate rock phosphate. The rate given is either a small dose of 50 kg $Pha^{-1}$ (supposed to be repeated yearly) or a large dose of 250 $kgha^{-1}$ (supposed to build up phosphate levels sufficient for five years). Varying the source and rate of the phosphorus fertilizer is to provide information on efficiency of less costly alternatives (rock phosphate at low dose) compared to what is assumed to be the optimal treatment (TSP at 250 $kgha^{-1}$). The application of 100 kg $Kha^{-1}$ is tested at a high rate of P (150$kgha^{-1}$).



## Experimental design

The experiments were conducted in 1997 and 1998 in 17 villages of the project area. The fallow is a one-season short rains fallow.

The experiments take place on-farm, superimposed on improved fallow fields that were planted on farmers' own initiative. Researchers select plots on the basis of a minimum size (to accommodate at least two plots of 8 by 8 m) and a dense (total ground cover) homogenous shrub stand. The owners are proposed to participate in on-farm trials and invited to planning meetings. During the meetings farmers can choose between several sets of simple comparisons (with a maximum of three treatments per farm) or decide not to participate. Farmers linked to one trial then meet to agree on a common set of non-experimental management principles (e.g. planting distances, monocrop – intercrop, varieties, weeding intensity, pest treatment).

Except for the fertilizer, the farmers provide all external inputs and labour, with no compensation promised. Research assistants are present during the cutting-down of the fallow stand (to assess the fallow biomass), casual checks, and the crop harvest (to measure the yields and take samples). Crops are an intercrop of maize and beans. Otherwise farmers are left alone

to manage their crops. Commonly, farmers use local varieties of beans and maize, and do not use chemicals to overcome pests and diseases.

Besides yield data, the time needed to carry out certain activities in the IF system, is monitored on several farms. In conjunction with local market prices for the produce and the inputs, these data are used to calculate economic costs and benefits of the IF system. The costs and benefits are combined over four seasons; the cropped season when the fallow is planted, the un-cropped season when the fallow is developing, and two cropping seasons after the fallow. The benefits are based on the maize and bean yields for all these seasons.

After the completing of the cropping cycles, group discussions with the participating farmers are held to elicit experiences.

## Experimental treatments

The experimental treatments are the result of a factorial combination between the improved fallow species and the fertilizer treatments.

The improved fallow species is the fallow present on the participating farms. Additional fields will be required to provide the control treatments of 'natural weed fallow' and 'continuous cropping'.

The sets of fertilizer comparison that can be chosen from are:

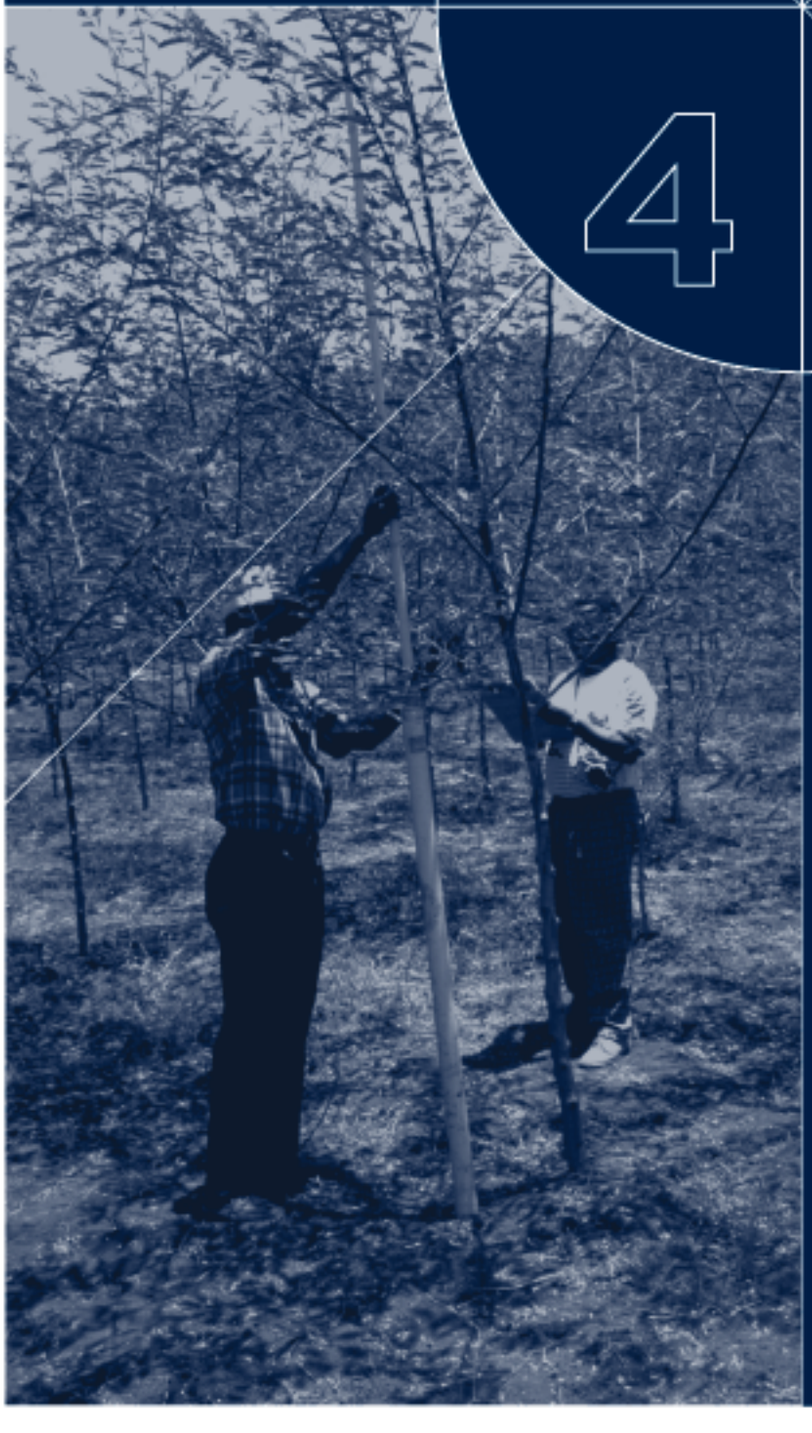
1. Rates of TSP:
   - 50 kg $Pha^{-1}$
   - 250 kg $Pha^{-1}$

2. Comparison of P sources:
   - no fertilizer
   - 250 kg $Pha^{-1}$ as TSP
   - 250 kg $Pha^{-1}$ as RP

3. Effect of potassium:
   - no potassium + 150 kg $Pha^{-1}$ as TSP
   - 100 kg K $ha^{-1}$+ 150 kg $Pha^{-1}$ as TSP

ISBN 92 9059 145 5

4



World Agroforestry Centre

TRANSFORMING LIVES AND LANDSCAPES