

Contributions of Biometrics to the Developing World

Maria Darmi and Rosie Poultney

Statistics Department, AFRC Institute of Arable Crops Research, Rothamsted Experimental Station, Harpenden, U.K.

◆INTRODUCTION ◆

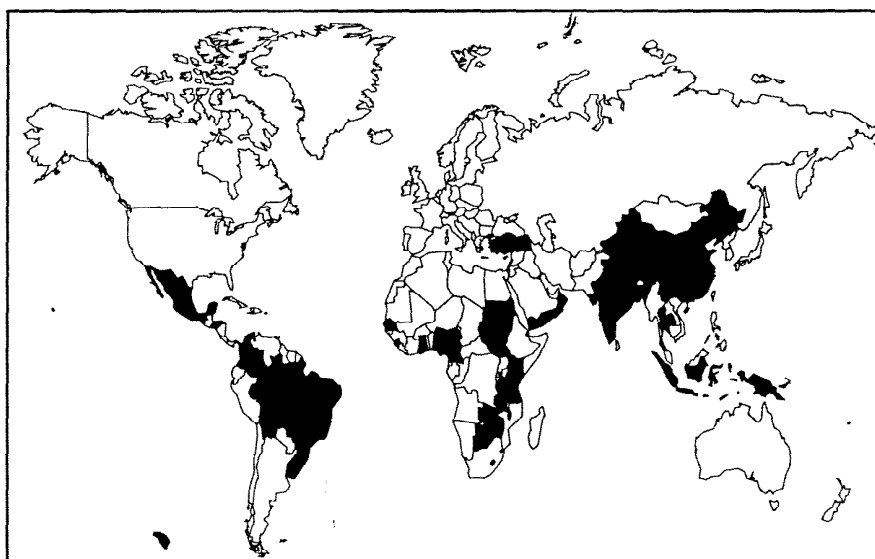
THE Institute of Arable Crops Research (IACR) is spread over three sites; Rothamsted Experimental Station in Hertfordshire, Long Ashton Research Station near Bristol and Broom's Barn Experimental Station in Suffolk. The Institute is one of a number funded by the Agricultural and Food Research Council (AFRC). There are approximately twenty-five statisticians within IACR, the majority being based at Rothamsted.

These statisticians might be called biometricians, as biometrics is the use of statistical methods to help understand biological systems. However, for historical reasons they are

known to the Institute as statisticians and we have followed this convention here. Their work includes providing statistical advice and support for the three scientific divisions of IACR; Crop and Environment Protection, Plant and Cell Biology, and Soils and Crop Science. However, the statisticians are more than simply support staff. In the fast changing area of scientific research, data are being collected which do not conform to the assumptions which underpin many statistical methods. Research into new statistical methods is an important part of the work of the Statistics Department.

Of these twenty-five statisticians, four, including ourselves, are funded separately and work solely with scientists based in developing countries.

Overseas work has long been part of the history of Rothamsted, and in the 1960's funding was provided for one person to concentrate on such work. Over the years, the funding of the Overseas Unit has increased to keep pace with the workload. In essence, the Unit's work is similar to that of the other statisticians at Rothamsted in that we provide statistical help and advice to scientists, the difference being that our scientists are based in developing countries all around the world rather than within the Institute. The map shows the countries with which we have had links over the last 5 years.



Most work is sent to the Unit by post or by Fax, but each of us spends up to three months overseas collaborating on projects in person or giving training courses on statistics and statistical computing. We each have somewhat different jobs within the Unit.

Rosie provides statistical input to a number of overseas projects, whilst Maria is spending three years carrying out much needed research into the statistical needs of one particular discipline. We have attempted below to give you a flavour of the work in which we are involved. Although the two applications seem different the final aim, that of maximising the quality of statistics within overseas projects, is the same.

◆ROSIE'S WORK ◆

I joined Rothamsted in 1987 and have worked on a variety of applications since then. My first project was to analyse the effect of different intensities of grazing

on a natural grassland pasture in the Falkland Islands. This turned out to be an excellent introduction to the work which I 'typically' receive.

The grass in the Falkland Islands is called whitegrass (*Cortaderia Pilosa*) because after an initial period of growth, the top section of the grass dies giving a white appearance to the pasture. Whitegrass grows in clumps with blades emerging from stalks called tillers. The hypothesis is that if the grass is kept short, it will remain in the growing stage and thus will provide better nutrition for the sheep which graze on the pasture. For the experiment fourteen plots of land were used: three grazed to each of 4cm, 8cm, 12cm and 16cm in height, one plot which was grazed freely, and one plot with no grazing. The experiment was handled by three scientists concerned with different aspects of the work. I received data on whitegrass yields, both the actual amount and the mineral composition, on growth through time of selected plants in each plot and, finally, on species composition of grasses on the plot.

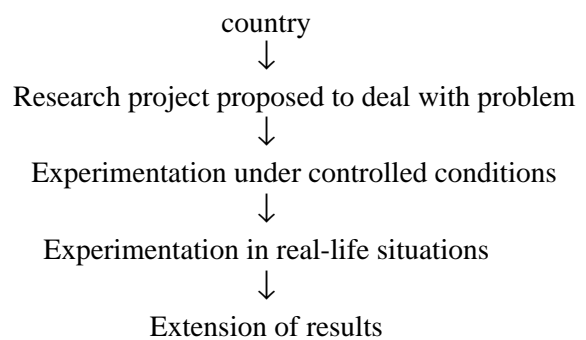
The data collected were various: continuous measurements of yields or weights, counts of the number of live leaves on a plant, proportions of each species in an area and percentages of trace elements, to name but a few. The experiment was designed to be long-running since it was thought that the effects of grazing might well take a number of years to become apparent. As frequently occurs in long-term trials, there were a number of monitored plants which died or could not be re-identified at a later measuring date. The resulting data needed a range of analyses to cope with the different types of data, the missing values, the correlations between the measurements and the repetition in time. The experiment is now in its sixth year and is showing very positive results.

I described this earlier as a 'typical' project which, given the description of the experiment, might seem an odd adjective to choose. By typical I mean an experiment which may be straightforward in its constituent parts, but rapidly becomes complicated by, for example, the need to combine results over locations or times, or to adjust for missing values in the data.

◆RESEARCH PROJECTS◆

My role in the Unit is also to provide advice and practical help on statistics during the different stages in a project's life. These stages are closely related but can be categorised into the following five steps:

The need for research into a topic is identified in



My main input would be in the three middle stages. I have included the first and last stages simply for illustration. Research must be directly applicable, such as controlling diseases in cattle, increasing maize production, or developing appropriate methods and technology for processing food. Each project is assessed before approval not only in terms of its benefits and likelihood of success, but also for its social and environmental implications.

- a) *Research Project proposed to deal with problem*
Research project proposals are examined by a number of experts to ensure that the approach itself is feasible. Projects involving complex trials or surveys are referred to the Unit for help and advice. In the ideal situation, this is where my involvement with a project begins. Contact with the scientists at this stage allows me to contribute to the design of the program of experiments to ensure the statistical validity of the research programme and the results.
- b) *Experimentation under controlled conditions*
Most initial research is carried out at institutes or universities under relatively controlled conditions. For example, at an institute involved in plant breeding, there is likely to be a number of trials comparing the performance of new varieties of a crop. The land used for these trials previously may have been part of a different experiment. Thus there will be additional information available on, for example, fertility trends in the soil or previous applications of fertiliser, which can be used to account for some of the inherent variability using analysis of covariance. The people who take the measurements and apply the treatments will have plenty of experience and will make few mistakes. These two facts make it possible to either control, or adjust for, much of the variation between individual samples in the experiment. It is possible to have more complex experiments. As the complexity increases, so does the need for good statistical advice to ensure that the design is appropriate for the situation and correct in its layout. I advise on the best way to record data to maximise accuracy and ease of data entry into a computer. It

is obviously important to be able to analyse and interpret the experimental results, and this is where I give most help and advice.

c) *Experimentation in real-life situations*

It is necessary to test results under less controlled conditions to ensure that they remain true in real-life situations. Designs for on-farm experiments, as I shall call them, are based on a different set of criteria than for on-station trials. Often the amount of land or number of animals available is small since the farmer may not wish to risk his or her livelihood on an 'untried' venture. This experimental material is also likely to be more variable than that at a research station, with little information available about the fertility of the soil or the previous history of the animals. The treatments are likely to be applied and managed by the farmer, with the scientist having less control over the experiment. Additionally, the trial may be used to demonstrate the new approach. The challenge to me is to produce a design which will allow simple experiments at a number of sites, yet will give a comprehensive set of results when the data from these sites are combined. This is not a trivial problem.

The final stage of any project is the extension of the benefits of the research into the community. Possible approaches include the use of demonstration plots and training local farmers.

◆ TRAINING ◆

A component of my time is spent in teaching both statistics and the use of statistical software packages. The actual amount varies from year to year and is dependent on the demand. I might spend two days training one person in the use of a statistical software package or three weeks lecturing in advanced statistics to a group of fifteen people. The former would take place at Rothamsted, whereas the latter would be held overseas in the host country. Training fits easily into the project framework described above. The need for training is identified and a course proposed. After the course, the participants are expected to apply what they have learnt to their work and to teach others. Thus, the extension of information is built into the project.

In the last couple of years, I have been involved in a biometrics training programme in Bangladesh. The Bangladesh Agricultural Research Council (BARC) is somewhat similar to the AFRC here in UK. Both coordinate agricultural research across a number of institutes specializing in different areas. To work effectively, biometricians need knowledge of a wide

range of techniques, good facilities, and the ability and opportunity to work in tandem with researchers. The training courses for these biometricians were designed after careful assessment of their statistical knowledge, facilities, and the needs of their scientists. Many of the institutes had micro-computers, others were soon to have them delivered. The first course, which we later repeated, centred around the analysis and interpretation of experimental results and included a number of techniques which were new to the participants. These included multivariate analysis and the analysis of repeated measures, both of which are impractical to use without good computing facilities. The statistical package Genstat 5 was donated to each institute. This package was developed at Rothamsted to deal with the needs of AFRC statisticians and is equally useful to the biometricians in Bangladesh. A second course, dealing with the use of Genstat 5, was held in early 1992. As a result of these courses, the participants, and their colleagues, have the ability to analyse their data accurately and efficiently using methods which until recently were not available to them.

• MARIA'S WORK •

I am solely involved in a three-year research project studying the statistical properties of fish-farming systems. These systems are associated with a large scientific field called aquaculture. Aquaculture studies the living members, animals and plants, of aquatic environments, freshwater or sea water. One of the main activities of aquaculture is to look for and study ways with which to transform aquatic environments into sustainable food resources. This is particularly useful for countries where aquatic products, and in particular fish, make up a large part of the human diet, as is the situation with many countries in the developing world. Furthermore, aquaculture studies ways of protecting the aquatic environment from any adverse consequences of human evolution.

In the early history of aquaculture, and in particular fish-farming, information was obtained by farmers through trial and error. Even today, given their day-to-





A typical small-scale fish farming system in Thailand.

day experiences, farmers are still able, through trial and error and intuition, to arrive at solutions; that process, though, can be very slow. Problems can be resolved faster and more effectively if this trial and error process is organised and systematised. It is then that this process becomes scientific research. In scientific research, the scientist, in our case the aquaculturalist:

- (1) observes the problem and where it occurs,
- (2) formulates an hypothesis (a tentative explanation or solution).
- (3) plans an experiment to test the hypothesis.

Experimental trials are then set up, either to provide information for use in further research, or to answer questions and arrive at solutions to problems. When experiments in aquaculture are performed to acquire information for use in further research, trials are set up in controlled conditions and factors that may distort the experimental results are closely regulated. However, when experiments are performed to solve problems already encountered by farmers or to provide better farming techniques for the farmers' available resources, experimental trials are set up to resemble real aquaculture practices; these have all the advantages and disadvantages of a real situation, but are done in a controlled manner.

Aquaculture is a relatively new scientific field and therefore there has been very little statistical involvement. Statistics in aquaculture is vital in organising investigations, in setting up well planned experimental trials to answer specific scientific inquiries, and in interpreting relationships or trends in the observed measurements.

In order to study the statistical properties of fish-farming systems, I had to acquire datasets from experiments designed to answer real world problems with regard to fish farming practices. Because of the role of our Unit in the overseas aid program, the great majority of the datasets that I acquired came from experimental trials performed in developing

countries, for example Thailand and Malawi. The aim of this study was to understand the statistical relationships between various components in fish-farming systems, and to make this knowledge available to scientists. They will then be in a much better position to search for improved fish-farming techniques and help small-scale farmers in the developing world to improve their livelihood.

◆EXPERIMENTAL PROBLEMS ◆

The datasets I am studying come from experiments concerned with fish growth and yield under different experimental treatments, for example feed, population density and culture practices. The data are mainly measurements of individual fish length, snout to tail, and of individual fish weight. Many of the experiments are conducted in earthen ponds. The reason for this is that small-scale farmers in the developing world quite frequently dig a pond within the farm land they utilise, thus acquiring an economic resource in addition to crop farming or animal breeding.

One of the things that I investigate is the extent to which experimental treatments and other non-experimental factors affect the observed measurements. In analysing the data I take great care to study the existence of unexpected behaviour that is attributed, for example, to inherent differences between ponds or to irregularities within a pond population.

Sometimes ponds show different behaviour even when they are treated identically. For example, three ponds were stocked with the same number of fish and an equal amount of organic feed (mixture of 50% cassava and 50% risebran) was added to each pond.



Tilapia fish.

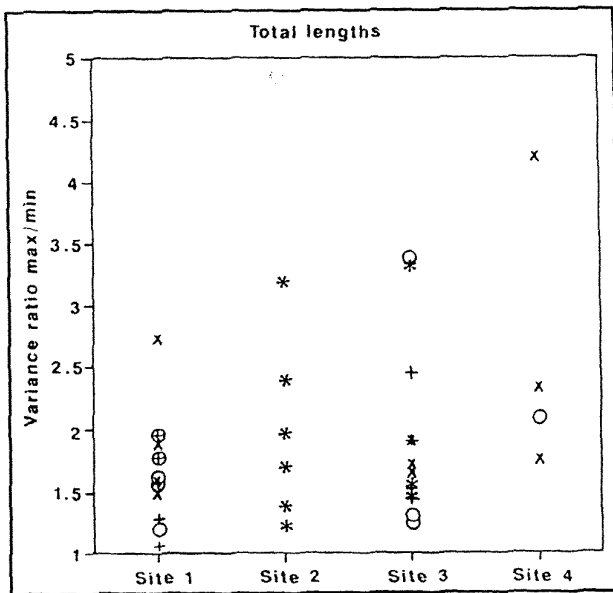


Fig. 1a.

After four months the fish were harvested. In the first pond the majority of fish had died and the rest were small and weak, in the second the fish had bred and the population consisted of a large number of small fish, whilst in the third the population number was roughly the same as initially and all fish had grown to a desirable size. These differences (variabilities) in the behaviour of identically treated ponds (replicates) were examined by studying the variances of the pond populations (Figure 1). Large variabilities were not only observed in one experimental site and for one group of ponds but in many sites and for most of the ponds considered. These variabilities appear

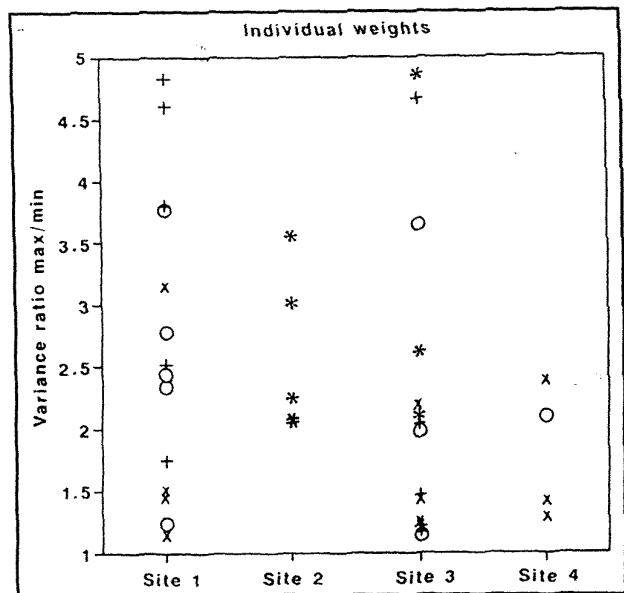


Fig. 1b.

more uniform for length measurements than for weight measurements, which suggests that fish weight is more sensitive to variability sources than fish length.

Histograms (Figure 2) show that distributions of the fish populations are often not normal; they may be bimodal exhibiting two distinct humps. The larger hump represents the main experimental population upon which investigations will be based. The smaller hump might represent a number of extreme observations, or it might represent a second, much younger, fish population that comprises part of the experimental population and is important in the investigation.

The x-axes represent experimental sites; sites are identified according to the types of treatments applied. The y-axes represent the ratio of the maximum to the minimum variance observed in identically treated ponds. Different symbols in one column represent different experiments.

Identical symbols in one column represent different treatments from the same experiment. For example, in the plot for weights, at the top of the first column, the symbol (+) means that for the particular experimental trial a pond is nearly 5 times more variable in behaviour than an identically treated pond.

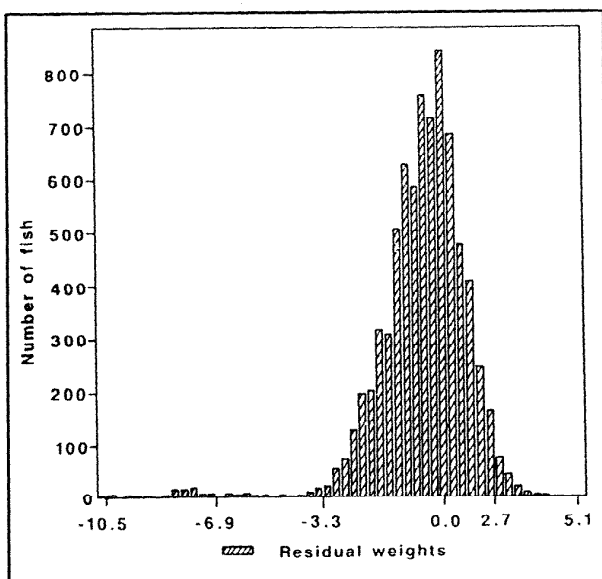


Fig. 2. Histogram of length data from fish treated with hormones. Note the small but distinct number of observations to the left of the main population.

◆ LIKELY CAUSES OF ◆
EXPERIMENTAL PROBLEMS

As a result of the above findings, questions arise about how and why such erratic behaviour develops in fish cultures. Apparently, the reasons are many and interrelated.

Variability between ponds is primarily related to pond

history. Ponds tend to accumulate inorganic as well as organic matter in the bottom mud from previous trials, and wastes of fish previously stocked. Even simple usage of the pond as a water reservoir can affect a pond dramatically. The material that is stored in the mud becomes available to the new stock of fish. Because treatments are assigned randomly to ponds, in every experiment the amounts and types of nutrients in the sediments vary greatly from site to site and also within sites from pond to pond, which explains the behaviour in Figure 1.

Variability within a pond is caused by the diversity of fish population in that pond mainly due to breeding. Erratic breeding can result from the effect of hormones that may be contained in the mud from a previous trial, or from the non-homogeneity in the age, and therefore the maturity stage, of the stocked population.

Fish-farming is very sensitive to changes induced by applied treatments. It is similarly sensitive to changes due, for example, to rain, wind currents, leaves falling in the pond or animals having access to the pond. These non-treatment effects are very difficult to control when the experiment is not properly organised and performed. By monitoring however, these changes can be identified and taken into consideration at the analysis stage and in the organisation of future experiments.

◆FUTURE DIRECTIONS ◆

The majority of the difficulties that I have encountered can be eliminated by the careful organisation of the experiment, and the careful selection of the data and of their appropriate analysis.

In order to investigate the factors that might affect the experimental results, new experimental structures might be needed. Ponds, for example, need to be monitored for the types of treatments they have received over a period of time. In this way future experimental investigations can be organised to take into consideration the pond history, for example by grouping ponds of similar history together. Another valuable investigation would be to monitor the causes of irregularities in the population distribution and study how these affect the overall behaviour of the fish population within a pond.

Furthermore, there is a need for investigations to continue, in order to establish whether these problems occur locally or concern aquaculture experiments on a global scale. To achieve this, many more datasets from different experiments, coming from different parts of the world, need to be carefully studied.

◆CONCLUSION◆

These are some examples of the type of work we do. The problems we have to face and solve are never repetitive, mainly because the projects we undertake are demand-led and focus on the needs of people from different communities with variable social and economic infrastructure. For most of the problems we face, textbook statistics is not immediately applicable and therefore, on many occasions we have to use extended techniques, or even make new developments as the need arises. Our rewards come from working on challenging problems in an environment where we have constant support from colleagues within IACR. Our jobs sometimes require us to travel overseas, usually to acquire better understanding of the nature and needs of a project, or to give training courses. Alternatively, we can be attached as a short-term consultant to overseas projects to provide intensive biometrical support. However, the most fulfilling aspect of our work is the fact that we use our knowledge and expertise to support scientists in the developing world to help their host countries toward technological, scientific and economic advancement.

Acknowledgment

The authors are funded by the UK Overseas Development Administration.