Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 197-213.

Printed in Granada, Spain.

DAVID J. SAVILLE

A HANDS-ON, INTERACTIVE METHOD OF TEACHING STATISTICS TO AGRICULTURAL RESEARCHERS

In this paper, I focus on two topics. The first topic is several day-long workshops, which I run annually for agricultural researchers. These cover the "linear model" methods, based upon the normal distribution, which are very commonly used by agricultural researchers. These methods are analysis of variance, regression and analysis of covariance. I describe the manner in which these workshops are run, cover the content of the workshops, describe the course evaluations, speculate upon why the workshops have proven to be popular, and attempt to draw conclusions.

The second topic is a method of teaching linear model theory using N-dimensional geometry. This method has been successfully used for both second-year and graduate-level university statistics courses.

1. INTRODUCTION

I work as a consulting biometrician/statistician in a government-owned agricultural research institute, which deals with animals, pastures, crops, plants, weeds, insects, diseases and other agricultural entities. Most of my work involves helping researchers with the design of agricultural experiments, the analysis of the resulting data, and the interpretation and writing up of the results.

This work involves continual communication between me (the statistician) and the researcher – the better the communication, the better the agricultural research. On the one hand, the research work is prejudiced if its statistical foundation is shaky. On the other hand, a firm statistical foundation is of little assistance if I have misunderstood the objective of the research, and we have answered a question of little interest. The best results are obtained when we each have a basic knowledge of the other's field in addition to specialist knowledge of our own field. That is, the researcher has a basic understanding of the statistical methods being used, and I have a basic understanding of the research area and the objectives of the research. This point is also made by Bangdiwala (2001).

In prioritising my everyday work, I therefore give highest priority to the teaching of statistical ideas to my agricultural research colleagues, to learning about research projects with which I am involved, and to collaboratively working on project designs. Next priority is given to analysis of designed experiments or surveys, followed by interpretation and writing up of such work. Lowest priority is given to attempting an analysis of data from unplanned work.

I teach statistical ideas to my agricultural research colleagues in two main ways. Firstly, I always make time during my everyday consultations to explain any statistical concepts that are relevant to the topic of conversation. Secondly, I run one-day statistics workshops each winter (the period when most researchers are inside the office). These workshops are specifically designed to teach ideas rather than methods of calculation, and have enjoyed an unusual level of popularity. These workshops are the first main topic of this paper.

Turning to the second topic, my interest in the teaching of statistics led to two periods of university lecturing in 1984 and 1985. The first course in which I was involved (with Graham Wood) was a second-year applied statistics course at the University of Canterbury in Christchurch, New Zealand.

For this course, we developed a novel teaching method that involved the use of *N*-dimensional geometry. This allowed an easy integration of the mathematical basis for linear models with the practical application of the models and the necessary computing. The second course was on "design, analysis and interpretation of experiments" for graduate students in agriculture at the University of Davis, California, U.S.A. For this course, I further simplified the geometric approach.

The enthusiasm of the students for the geometric approach led to Graham and I writing two statistics textbooks (Saville & Wood, 1991; 1996). The first book was comprehensive, and included all of the topics covered in the California course. The second book was written as an easy-reading introduction to the teaching method, and was aimed at lecturers to whom the method was new, as well as a general audience of statisticians and users of statistics.

In this paper I briefly describe my everyday consultations, then focus on (1) the annual workshops to agricultural researchers and (2) the geometric method of teaching statistics.

2. EVERYDAY CONSULTATIONS

Everyday consultations are of many varying types, and cannot be adequately summarised in a short section. Some involve the design of an agricultural experiment for which the primary variable or its transform can be assumed to follow a "normal" distribution. Here the main hypothesis of interest may correspond to a particular "contrast" between the treatment means, and in this case, attention may focus on ensuring maximum precision for the estimate of this contrast.

This may involve choosing an appropriate experimental layout (e.g., a randomised complete block or a "split plot" design), then carrying out "power" calculations to determine the number of replications. Alternatively, if the primary variable follows a binomial distribution, different power calculations may be carried out. During such "design" consultations, I normally discuss the relative virtues of various competing experimental layouts, the level of replication, and points such as whether some treatments should be replicated more heavily than other treatments.

Before consultations concerning the analysis of data, the researcher would normally enter the data into columns of an Excel spreadsheet, and email it to me before visiting my office with a copy of the field plan. We then have a general discussion about how the experiment went, during which the researcher will inform me of any peculiar problems that were encountered and how this may have affected the data. We discuss the list of treatments in relation to the hypotheses of interest, and in the case of "normal" data I write down the contrasts between the treatment means, which we (hopefully) planned prior to the experiment being conducted (most researchers find this specification of contrasts very difficult).

We then discuss the field layout and how this is taken into account in the analysis of variance or covariance (or, occasionally, regression). Usually I copy the data into an appropriate statistical package and process the data. This processing includes routine assumption checking such as a histogram of the residuals from the model (to check normality) and a plot of the residuals versus the fitted values (to check for heterogeneity of variance). At this point in time unusual values are often identified, leading to checks of the data entry, possibly a transformation of the data using a logarithm or square root transform, and/or discussions about what is an appropriate course of action. Finally, we usually decide upon how to present the treatment means and indication of precision (e.g., least significant difference), whether in a table, a bar graph, a scatter plot, or other option.

Other consultations may involve the analysis of binomial data, either by simple 2 x 2 chi-squared tables, partitioned tables (c.f., contrasts), or using generalised linear models. Alternatively, the consultation may concern how to describe the statistical design and method of analysis in the research report, how to respond to comments of a statistical nature written about a manuscript by journal referees, or a multitude of other possible topics.

3. ANNUAL WORKSHOPS

The primary aim of the annual statistics workshops is to *instill basic statistical ideas* into the minds of the researchers, not to teach them how to manually do arithmetic calculations nor how to process data using statistical packages. Most modern-day researchers are capable of feeding data into a computer package - what many are *not* capable of doing is choosing the correct menu item from the bewilderingly long list which pops up on the computer screen. My aim is to at least educate researchers to the level that they know the choice is important (and therefore ask for help if uncertain as to the correct choice).

The workshops *do* use simple exercises, which involve arithmetic calculations, but each exercise is designed to illuminate a particular statistical idea. The exercises ensure the people attending each workshop are involved in a hands-on, participatory manner. During each exercise, each person randomly selects experimental units (e.g., sheep or plots of pasture) using a set of random numbers, so that each person has their own data set to work with. In effect, each person is doing a "simulation."

To be more specific, I shall describe the first workshop in some detail (§3.1), and the remaining workshops in brief (§3.2). I then describe how the workshops were invented and how they have since developed (§3.3). I will describe the course evaluations (§3.4), give my philosophical thoughts on what I consider to be the essential ingredients of the workshops (§3.5), and conclude with a few general comments (§3.6).

3.1. WORKSHOP #1 IN DETAIL

Workshop #1 (Basic Statistics/Analysis of Variance, Day 1) starts at "zero assumed knowledge" and finishes by carrying out a simple analysis of variance. It runs for a whole day, from 9 am until 4.15 p.m. Workshop members are rested often (they need it!). Lunch lasts for one hour and there is a 20-30 minute break for refreshments halfway through both morning and afternoon sessions.

The basic equipment which people bring on the workshop is paper, pencil, eraser and battery-operated calculator. In terms of the venue for the workshop, I try to use a room which will comfortably hold enough tables for 20-30 people, allowing each person plenty of table-top space and allowing room between the tables for me to wander around and help people who are experiencing arithmetic problems.

The workshop commences with introductions. We each say our name, where we work/study, and what our area of research is. Then I put up a "plan for the day" in the form of a flow diagram on a large sheet of paper, which I attach to the wall. I go through the items on the plan and explain that we shall initially go quite slowly for the purpose of consolidating the "basics."

The first exercise is called "Why Use Statistics?" from Bishop (1980). I describe an experiment with 12 sunflowers grown in pots on a windowsill for nine weeks. There are three replicate pots for each of four liquid fertiliser treatments (A, B, C and D), with pots assigned in a completely random manner to treatments. I give the weights of the sunflowers at the end of the experiment. I then ask each class member to think of the three weights for each treatment as a sample of a larger population of weights for other sunflowers grown with the same treatment.

I ask: "*On average over this larger population*, do sunflowers grown with treatment A weigh more than sunflowers grown with treatment B?" Class members are asked to write either Yes, No or Don't know on a sheet of paper, along with reasons for their answer. They are also asked to do the same for the five other pair-wise comparisons (A versus C, A v. D, B v. C, B v. D, and C v. D). After 5-10 minutes to ponder the questions, I ask the class members to swap pieces of paper before I tabulate the responses in terms of Yes/No/Don't know (plus the nil responses). Then I ask people to give reasons for their own answers. This generates discussion of sample means, overlap or otherwise between sample values, variability within each sample in comparison with the size of the differences between sample means, and so on. It also serves to "break the ice" and set the scene of the workshop as a participatory event.

The next section of the workshop is mainly me talking. I give each person an empty workshop folder plus the first 17-page section of a booklet, which summarises the work that we cover during the rest of the day (Saville 1980). Included is a population of the gains in weight of 50 lambs during a three-month period (the data were from lambs 1-50 in an experimental group of 150 lambs in a local New Zealand experiment).

I then describe various alternative ways of measuring the centre and spread of this population of weight gains. For measures of centrality I discuss the median, mode and mean, and for measures of spread I discuss the range, the inter-quartile range, the mean absolute deviation and the mean squared deviation (or variance) plus its square root, the standard deviation. By chance, lamb 50 had a very unusual weight gain, so I discuss the influence of this one weird value on the various measures. I then explain that the methods we most commonly use are based on the mean and variance, and that this is because of the simple mathematical solutions which are possible by courtesy of Sir R.A. Fisher and Pythagoras' theorem in N-dimensional space.

This takes us through to the morning break at about 10.30 am. After the break, at about 11 am, I hand everyone a Xerox copy of some random numbers, and we embark on a sampling exercise. The idea is that we know the "truth" about our population of 50 weight gains in that we know the true mean and true variance; however, we pretend that we have only a small sample from this population, and set about "estimating the truth" by random sampling.

Firstly, this is done for a sample of size three. As in all class exercises, I run through the exercise myself on a white board, then each person does it for themselves. In the current exercise, each class member shuts their eyes, stabs their random number table with a pencil, then takes the next three random numbers in the range 1-50, and writes down the three corresponding weight gains. The sample mean and sample variance (with divisor n-1) are then calculated.

When everyone is finished I do a histogram of the sample means from the class on a "printable" white board. The class discusses how well they did in estimating the true mean. I then do a histogram of the sample variances to the right of the first histogram, and the class discusses its shape and approximate centre. I also make some points about the long-run shapes of the histograms. The class repeats the exercise using a sample of size six. While they are working on the arithmetic, I draw up the outline of the "n=6" histograms, immediately below the "n=3" histograms and using the same scales. Each person then comes up and enters their own value on each histogram. The class then discusses the effect of increasing the sample size on the spread of each histogram and I make some comments upon how the shapes change. I print off the white board (containing the four histograms), we break for lunch, and I Xerox a copy for each person's folder during the lunch hour.

After lunch I consolidate "where we're at" by drawing bell-shaped normal curves to represent the "parent" population of 50 lamb weight gains, plus increasingly narrow bell-shaped curves to represent the two "derived" populations of sample mean weight gains for samples of size three and six respectively.

We then embark upon a further exercise concerning the variability of sample means, using the same population of 50 lamb weight gains. The idea here is to generate several sample means, then directly calculate their variance (thinking of the several means as a sample from the population of all possible sample means). Each class member uses their random numbers to select four samples of size two, and calculates the four sample means and the variance of these four means.

They then add their "estimated variance for means of samples of size two" to a histogram on a fresh page of the printable white board. They redo the exercise for four samples of size four, adding their "estimated variance for means of samples of size four" to a second histogram directly below the first on the same scale. After this, the class discusses their results, hopefully noting that the centres of the histograms differ. After the discussion has dried up, I mention the theoretical "rule" that the variance for sample means of n values is the variance of the parent population divided by n.

To see that this is reasonable, we calculate the theoretical values for n=2 and n=4 and add these to the two histograms. [As an aside, I mention also that on the original scale (kg of weight gain), the "rule" is equivalent to saying that the "standard error (=deviation)" of a mean of n values is the standard deviation of the parent population divided by the square root of n.] I print this page of the white board and Xerox it for distribution to class members.

Throughout the day, I explain that the above exercises are intended to illuminate the ideas that are required for an understanding of an "analysis of variance." We now embark upon two analyses of variance. In both, there are three experimental treatments, each with four lambs randomly assigned to them. The treatments are "untreated," "treated once with anthelmintics for worm control," and "treated twice with anthelmintics for worm control."

The first analysis of variance is a "dummy run" in that each class member draws the three sets of four lambs from the same population (above), thereby simulating an experiment in which all three treatments are equal. An "F ratio" is calculated as the ratio of the "actual" variance of the means (calculated directly from the three treatment means as above) and an "expected" variance of the means.

The latter is the variance of the means which one would "expect" if all three treatment means were identical. It is calculated by calculating the variance *within* each treatment group, averaging these three numbers to obtain the best estimate of the average variance of the three parent populations, then applying the theoretical rule by dividing by n=4. Note that if there are no treatment effects, the "actual" and "expected" estimates will both average to the same quantity, while if there are treatment effects, the actual estimate will be inflated while the expected will be unaffected; hence the F ratio is increased by the presence of treatment effects.

After completing the necessary calculations, each class member writes their actual and expected variances and the resulting F ratio on a fresh page of the printable white board, and adds their F value to a histogram of F values. We now discuss the class results. It is hopefully clear that the two variance estimates have a similar average, though one is more variable than the other. At this point I introduce the $F_{2,9}$ reference distribution, and hand out a table of F "critical" values which contains its tabulated 95 and 99 percentiles (4.26 and 8.02). The histogram of F ratios generated by the class tells us the approximate shape of the $F_{2,9}$ distribution, and it is interesting to see how many values are outside the "95% normal range" of 0 to 4.26. This page of the white board is again printed and xeroxed for distribution.

After a refreshment break, we embark upon the second analysis of variance, "the Real McCoy," meaning an experiment in which there *are* treatment effects. For this exercise, the first set of four lambs ("untreated") is randomly selected from the population used above. The second set of four lambs is randomly selected from a second population which I provide (the data were from lambs 51-100 in the same experimental group of 150 lambs as described above, except that I added 5 kg to each data value). Similarly, the third set of four lambs is randomly selected from a third population, which I provide (the data were from lambs 101-150 in the same group of 150 lambs, except that I added 10 kg to each data value).

In effect, each person is simulating an experiment in which the second and third treatment means are 5 and 10 kg greater, respectively, than the first treatment mean (these effects are very large in relation to the variability in these data). The same calculations are carried out as described above. In this case, each person finds that the "actual" variance of the mean is many times larger than the "expected" variance of the mean, with the result that the F ratio is large (usually greater than 10-20) and usually "highly significant" statistically. As above, each class member writes their actual and expected variances and the resulting F ratio on a fresh page of the printable white board, the class results are discussed, and a xerox is made for distribution.

To wrap up the day, three things remain to be done. Firstly, I write out the more usual "analysis of variance table" for the Real McCoy data set which I had worked through on the white board prior to the class doing their own simulation. The purpose here is to relate the class method to the more usual computer printouts. Here the usual "treatment" and "error" mean squares are obtained by multiplying the "actual" and "expected" variances of the mean by the sample size (n=4 here). Secondly, I draw up a 2 x 2 table with the "truth" along the top (No differences, Differences) and the experimental decision down the side (No differences, Differences). Two of the four cells constitute correct decisions, while the other two cells correspond to Type I and II errors (the class find this easy to relate to, since they have had plenty of discussion of such errors during the afternoon). Thirdly, I quickly go over the plan for workshops #2 and 3 (the two remaining workshops in the first series).

3.2. OTHER WORKSHOPS IN BRIEF

The first three workshops (*basic statistics/analysis of variance, days 1 - 3*) form a series in that the workshops build on each other and people cannot join in partway through. The fourth and fifth workshops (*simple regression and analysis of covariance*

(ANCOVA)) form an independent second series in that people can attend them without having attended workshops 1-3 (in practice, though, only a few people exercise this option). The *ANCOVA* workshop builds on the regression workshop, so the latter is a pre-requisite.

The workshops have been run most winters since 1978, except for the more specialised *ANCOVA* workshop, which is run about every second year. There is also a third independent series of two workshops which has been run on three occasions; this is entitled "Statistics done properly" and is based upon the *N*-dimensional geometry which is at the heart of these methods.

Successive workshops are normally run with at least one day's break between each workshop, to allow people's brains to recover. I only break this rule if my travel/accommodation costs make it necessary – on such occasions, I notice that the class members become quite tired by the end of the second day. Tuesdays and Thursdays are the days I aim to use for workshops.

I shall now briefly outline the contents of workshops 2-5. The second workshop (*basic statistics/analysis of variance, day 2*) starts by revising the work of day 1 using a second set of data.

Here the population consists of 64 plots in a pasture arranged in an 8 x 8 grid, with the data being pasture dry matter yields expressed in kg/ha (these are data made up by myself, and are not real data). The pasture is highly variable, with yields ranging from 360 to 930 kg/ha (unlike the weight gains used on day 1, which were relatively low in variability). Three treatments are hypothesised (untreated, "superphosphate fertiliser (super)" and lime), and three plots randomly assigned to each treatment by each class member.

To generate known treatment effects, 200 and 100 kg/ha are added to the yields for the second and third treatments respectively (that is, the "true" super response is 200 and the true lime response is 100 kg/ha). The analysis of variance is carried out, and most class members fail to "detect" any treatment effects using the overall *F* test. This demonstrates the low "power" of such an experiment. The exercise is then carried on to estimate super and lime responses and the significance of these responses. This also leads to the calculation of the "least significant difference (LSD)" between any two treatment means and a discussion of multiple comparison procedures (Saville, 1990). The class results are then used to show how one would combine data from a series of identical trials. Presentation of results for a report or journal article is then covered, as are sample size calculations.

The third workshop (*basic statistics/analysis of variance, day 3*) starts off by introducing the idea of a statistical model, using the day 2 design as an example. This sets us up for dealing with a randomised block design experiment.

As a class exercise involving such a design, the pasture data from day 2 is re-used in conjunction with "scores" of the pasture plots which were assigned by the experimenter before the experiment was set up (these contrived scores range from 3 to 9, and are very highly correlated with the subsequent yields).

Each class member designs their own experiment by choosing blocks of three plots with similar scores within each block, and randomly assigning treatments within each block. They then use the corresponding yields, adding 200 and 100 as in day 2, and fit a randomised block model. The corresponding analysis of variance then uses "actual" and "expected" variances of the mean as in days 1 and 2. Failing arithmetic errors, class members notice a dramatic increase in precision as compared to their day 2 results. As in day 2, responses are estimated, the LSD(5%) is calculated, and there is a discussion of power for both overall and more specific tests. The analysis of variance table is related to the workshop method, and there is a full discussion of assumption checking. Lastly, the "whys and hows of blocking" are discussed.

The first workshop in the second series (*simple regression*) again uses a data set that I invented for the purpose of the class exercise. I first set up a hypothetical straight line ($y = 80 + 10x$). Then, for each of $x = 2$, 5 and 8, I generated 20 y values from a normal distribution centred on the appropriate point on the line with a standard deviation of 20.

Each class member uses random numbers to select two y values for each of $x = 2$, 5 and 8 from the 20 which I provided, then calculates the intercept and slope of the fitted line using formulae which I provide. The class histogram of slopes gives people an idea as to the standard error of the slope. The latter is then estimated for each person's regression by calculating the variance about the line and using this in the appropriate formula. Hypotheses about the true slope are tested, and the equivalence of F, t and r (correlation) tests are discovered by the class. As time permits, 95% confidence bands for the "true line" and for predictions are calculated. The pitfalls of the method are discussed, as are Anscombe's Quartet, the influence of outliers, the dangers of extrapolation, and why a "high r value does not ensure the curve is a good fit."

The more specialised final workshop (*analysis of covariance*) deals with the fitting of parallel straight lines. The same basic data set is used from the simple regression workshop, except that course members draw their data as two "x-y" samples, and learn how to fit two parallel lines. For the first line, a single y value for each of x=2, 4, 6 and 8 is used, while for the second line, a single y value for each of x=2, 4, 4 and 8 is used. To offset the second line, each class member adds 60 to each y value for the second set of data. In the analysis, hypotheses of interest are whether the common slope is significantly different from zero, and whether one line is at a significantly different elevation to the other. The common uses of this method are also discussed in relation to real examples.

Workshops 1-4 cover the basic material which is of most interest to researchers, so are the most popular. Workshop 5 is useful to some researchers, but not others, so attendance is always lower for this workshop.

Workshops 6-7 (*statistics done properly, days 1 and 2*) explain the mathematics that is at the heart of the methods covered "heuristically" in workshops 1-5. In my opinion, these methods cannot be "properly" explained without the use of geometry.

During workshops 1-5, I mention words such as "degrees of freedom" for which I can give only an intuitive "hand waving" explanation – on these occasions, I comment that the real explanation lies in the geometry. Such comments arouse the curiosity of some of the class members, so that usually about one third to one half of them are keen enough to learn more by attending workshops 6-7. In the event, about half of those who do workshops 6-7 find them really rewarding. The workshops are largely based on simplifications of the four class exercises in Saville and Wood (1996). For reasons of space, I shall not detail these workshops here.

Over the years, I have also run other workshops on more specialised topics. I ran one on "orthogonal contrasts in analysis of variance," but found that most people could not grasp the concept of orthogonality (which is basically the idea of right angles in N-dimensional space, so hard to teach without mentioning N-space). I ran another one on "combining data from a series of trials," but concluded that this specialised topic was not one that warranted a workshop.

On two occasions, I set up workshops on "design of experiments" (meaning designs more advanced than the ones covered in the basic statistics workshops). The first was an informal, interactive workshop for a small group of 7 researchers all working on similar problems at a distant research station – this went well. The second was a workshop for about 16 researchers from diverse fields of research – I felt this was poorly received, since the designs used in different fields are quite different, and any particular design that I spoke about was only of interest to about one third of the people in the room, meaning each person was bored about two thirds of the time! My conclusion was that more specialised topics are more efficiently taught during one-to-one consultations, and that only the core ideas are best covered by a workshop.

3.3. INVENTING THE WORKSHOPS

In 1978, when I invented the workshops, I was unaware of any of the literature on statistics education. However, I did have the benefit of advice from a school teacher friend. She told me to focus on the stumbling blocks, which are the things which researchers find hard to understand about statistics.

These blocks seemed to be everywhere, but especially in the very basic material, so it seemed to me that the only sensible approach was to cover the material *from the beginning*, working on each stumbling block in turn, and taking time to overcome each stumbling block.

In some cases, I had to think hard about how to overcome a particular obstacle. For example, researchers take a sample of observations and estimate the mean and standard deviation; then they calculate the *standard error of the mean*; this is hard for them to understand - they have only one mean, so how can it vary? This is one reason why I set up an exercise in which the students took several samples, calculated the mean for each sample, and then calculated the variability of these means (after lunch on day 1, as detailed in §3.1).

I had various other ideas in my mind when I designed the workshops. As a pure mathematician who learnt statistical methods "on the job," I found that I needed to go through the calculations for each type of statistical analysis by "long hand." I therefore felt that it was important for each researcher to go through each example long hand. At the same time, I thought it was a good idea if everyone was in effect doing a "simulation" so that we could summarise the class results in histograms and other ways which would illustrate statistical principles and results. Lastly, in the case of analysis of variance, I worked out a method of doing the arithmetic (involving the variance of the mean) which tied in with a gradual build up of ideas via a sequence of simple class exercises.

The workshops have slowly evolved and improved over the years based on feedback from class members. Also, during the 1990s a fellow statistician and ex-secondary school teacher, Mrs Lesley Hunt, highlighted for me the value of the "Discovery" method. This led to me more consciously standing back to let the class do the talking as much as possible ("discovering" statistical truths for themselves), with me just filling in any gaps at the end. She also suggested drawing up a "plan" for each day so that class members could see where they were at each stage of each day. The course evaluation sheets were a further improvement suggested by Lesley.

The "basic statistics" series originally consisted of two workshops. However, the material included in these workshops slowly increased over the years. In addition, the Discovery method, while an improvement, meant that discussions were encouraged more actively, and lasted longer. This meant that it was difficult to cover all of the planned material in two days, and some material was covered in a rush (especially the Day 2 material). This led to some 1998 class members suggesting that the material should be spread over three days. I decided that this was long overdue, so I redesigned the series before the 1999 workshops. The same material was covered in the same manner, but I spread the old Day 2 over two days and added new material on assumption checking.

3.4. WORKSHOP EVALUATIONS

From 1978 to 1996, the workshops were run without any formal evaluations. People who attended the workshops seemed to be very pleased with what they learnt, and feedback was generally very positive (people did suggest minor improvements, so I felt that negative feedback would have been forthcoming if it had been warranted).

In general, I was always pleased that I had made the effort to run the workshops, and felt motivated to re-run them the following winter (except for an occasional year, such as the year I was overseas and the year we were restructured from a government department to a new government-owned institute). People attending the workshops recommended them to their colleagues, and there was a continuing “supply” of attendees (though fluctuating, in the range 6 - 26 per annum), so I concluded that the workshops were filling a continuing need.

For the last three years (1997 – 1999) the attendees have filled out evaluation sheets (Figure 1; note that the lines of text are evenly spaced over an A4 sheet). In general, most respondents have been very positive about the workshops, with similar responses in all three years. To indicate the sort of responses given, I shall summarise the responses for the most recent year (1999).

In response to question 1, most people expected to gain a better understanding of statistical ideas and methods, some hoped to gain increased confidence and/or increased ability in interpretation, and some had no or low expectations. The responses in terms of whether expectations were met are summarised in Table 1.

The second column of Table 1 is the total number of different people who attended one or more workshop. Only 1 out of the 29 respondents said that the workshops failed to meet their expectations and a further respondent said they only partially met their expectations. On the other hand, 69% said the workshops met expectations and 21% said that their expectations had been exceeded (replies like “Yes, and more,” “Very much so,” “Definitely,” “Yes, very well”). Amusingly, one person had two expectations. The first, “enlightenment and understanding,” was met. The second, “swamped with figures/formulas and left confused,” was not met (to my relief). Another person wrote that their expectation (from the workshop advertisements) was “not to learn anything new,” but that this expectation was not met!

Figure 1. Course Evaluation Sheets Used in 1997, 1998 and 1999.

Statistics Courses 1999
Feedback and evaluation sheet

Course name *Your name (optional)*

1. What were your *expectations* of the course? Were these met?
2. What did you *learn* from the course?
3. What did you *like* about the course?
4. What did you *dislike* about the course?
5. What could be *improved* about the course?

Use reverse of page if you run out of room

Table 1. Responses to the Question "Were These (Expectations) Met?" in 1999.

Location	Number of different people on workshops	Number of different people responding	Number of people responding as follows			
			No and/or partially	No expectations	Yes	Stronger than yes
Lincoln	12	8	0	0	5	3
Palmerston North	36	12	0	0	9	3
Wallaceville	16	9	2	1	6	0
Total	63	29	7%	3%	69%	21%

The subsequent questions had many, varied responses, so I shall just summarise the Palmerston North responses, which were about average. To question 2 (what was learnt), replies were:

"Insights from Dave about analysis options"; "Basic trial design and how to think about design"; "Reinforced the theoretical that was learnt previously with practice"; "The fundamental idea behind statistical equations, which makes them seem less daunting and more understandable";

"(ANCOVA workshop) How to correct treatment means for covariate effects, test for significant treatment effects (very useful), interpret F, t, r and so on (very useful), understand importance and relevance of these values and how false interpretations can occur (pitfalls)";"The hows and especially the whens of regression and ANCOVA".

"Basic analysis of variance, how to understand the different terminology, how LSDs are useful, how to understand ANCOVA (needed for person's own trial which had a nuisance factor)". "LSD, how to calculate sample size, "what the numbers I have been computing for so long actually mean."

"(Geometry workshops) Greater understanding of the geometric methods behind stats". "The applications and ability to calculate regression and ANCOVA myself. In addition I now really understand where F comes from!"

To question 3 (what people liked about the courses), replies were:

"Emphasis that "simple is best." Generous provision of time for exercises, limiting the hassle (anxiety) associated with some courses where basic skills and confidence are a bit low".

"Practical working"; "Hands on. Prepared notes"; "The chance to hand work the analysis – we can become too dependent on computers doing this for us"; "Hands on approach to learning (best way to learn stats in my opinion). Time spent on answering specific questions relating to material (improves relevance to own work)".

"Good number of people"; "It was done at a pace that I could keep up with, using very understandable language"; "Good practical examples. Honesty about limitations of stats". "Worked examples made things easier to follow; "The informality. Smaller class for workshops 4-5"; "Slow pace, reinforcement of examples. Class aggregation of data to demonstrate principles"; "Dave's easy, relaxed method of teaching. It was magic to see and hear someone with a bit of humour about stats!".

To question 4 (dislikes), six people left the section blank or wrote "nothing." Other replies were:

"Attendees who ask too many questions"; "Some of the verbal explanations were a little

inadequate – but due to time constraints – understandable"; (Workshops 1-3) *"Too much of routine calculations following the "rules" versus not much discussion of ideas and concepts behind the rules".* (Workshops 4-7) *"No dislikes".*

(Regression and ANCOVA)*" Sometimes went too quickly over difficult areas";* (Workshops 1-3) *"Time spent waiting for class results".* (Workshops 4-5) *"Got lost at one point in the algebra –still can't figure out why we did it as we didn't use it later."*

(Geometry workshops) *"Last hour a bit rushed – otherwise not a lot (of dislikes)".*

To question 5 (improvements), four people left the section blank or wrote "nothing." Other replies were:

"Request for an informal, spiral-bound course text with sequential exercises worked through with explanations, followed by appendices with Dave's supporting papers, stats tables, etc". "Limit on number of people, say 20. At end of course: single handout sheet of formulas plus explanations and symbols plus meaning, to be used as a prompt for returning to the notes". "Addition of information on chi-squared tests and partial correlation coefficients".

(Workshops 1-3) *"To have more discussion on concepts and ideas".* (W 4-7) *"Nothing". "Go quickly over initial stuff and more slowly over difficult stuff". "Less people, smaller class, go faster".*

(Geometry) *"Perhaps a fully worked example of one of the more complex procedures (e.g., ANCOVA or RCBD). Additional session with computing".*(Regression and ANCOVA) *"Speed about right. More depth on 3 covariates – but appreciate that this was beyond the remit of the course".*

Further evidence of attendee satisfaction is provided by the low attrition rate. Researchers are encouraged to attend the first workshop in the "Basic Statistics" series with the assurance that there is no obligation to attend the other workshops if they do not find the first workshop useful to them (in line with this, non-AgResearch staff only pay for the workshops that they attend, with invoicing being done after the workshops rather than beforehand). In Table 2 we present the numbers of people starting and finishing the "Basic Statistics" series of workshops (2-day series, 1995-1998; 3-day series in 1999), the 1-day Regression workshop and the 1-day Analysis of Covariance (ANCOVA) workshop (which was not always offered).

Table 2: Numbers of People Starting and Finishing the Different Workshops

		Basic series numbers		Numbers attending	
Location	Year	First day	Last day	Regression	ANCOVA
Lincoln	1995	22	21	18	-
	1996	26	25	18	11
	1997	18	16	16	9
	1998	22	19	17	-
	1999	10	9	9	8
Palmerston North	1998	24	24	19	-
	1999	26	24	18	15
Wallaceville	1999	16	11	-	-

As seen from Table 2, however, most researchers complete the entire "Basic" series, and many return for the Regression workshop (though the numbers include a few people for whom this is the first workshop). Lower numbers return for the ANCOVA workshop, but this is largely because not all researchers perceive this method to be important. Note that sickness is one component of the attrition rate. Note that the high attrition rate at Wallaceville was largely because 7 of the attendees (all from one division) were called away to a hastily scheduled restructuring meeting halfway through the afternoon of Day 2, with only 3 of these people returning on Day 3.

3.5. ESSENTIAL INGREDIENTS

There are various aspects of the workshops which I feel contribute to their ongoing popularity. It is hard to know which aspects are the most important – however, I shall list the ones I feel are most crucial. Two of the most vital ingredients are:

1. Start at the *beginning*;
2. *Go slow.*

These two ingredients require nerves of steel in the presenter! No one likes to be considered stupid or uneducated, so the natural inclination is to exhibit one's intelligence and/or vast knowledge, preferably as soon into the course as possible! The worry is that the attendees will decide that they cannot possibly learn anything from a workshop which starts off dealing with such elementary material, and walk out "en masse," to the presenter's never-to-be-forgotten embarrassment. I have had an occasional individual who has made this very decision and departed one hour into the first workshop, but luckily, the contagion has never spread!

Yet, if you lose your nerve and start anywhere except at the beginning, or go fast, you will certainly lose your class (especially if they are volunteers, as in my classes). In reality, most researchers have had bad experiences with their previous statistics courses, so approach the idea of another statistics course with fear and trepidation. They are therefore, in the main, relieved and eternally grateful that the workshop material is understandable, starts at the beginning, and goes only at their pace.

Two points that are also essential are:

3. Provide *hands-on* work with agricultural data sets;
4. Encourage *participation* and *interaction*.

The hands-on work with data is really appreciated by attendees, and is favourably commented upon in many of the evaluation sheets. I feel it is important because the class members get "the feel" of how to do statistical calculations, and this gives them a confidence in their own statistical ability. Participating in a group learning experience is also something most people enjoy and benefit from. The group as a whole quickly start to interact among themselves and with myself, ideas are bounced around, jokes made, and a nice friendly atmosphere develops. The class members develop a sense of ownership of their own data set, and mentally identify with the hapless researcher who has obtained such variable data.

Once "the scene is set" (i.e., once each class exercise is set up), class members *discover* many things for themselves, and this sinks in much better than if I told them the answer. The discussions of each set of class results can be quite wide-ranging as people raise all sorts of statistical issues which have been worrying them – some of these issues can be discussed naturally at the time they are raised, and others I need to postpone till later in the workshops and/or discuss outside the workshop time. I feel that the attendees perhaps learn as much from these impromptu discussions between a practising biometrician/statistician and a group of agricultural researchers as from the more formal parts of the workshops. What they subconsciously assimilate is "how a biometrician thinks," which tells them a great deal about statistics.

Yet two more elements that are crucial are:

5. *Experiencing the variability;*
6. Learning to *live with uncertainty.*

Variability is at the heart of statistics, and *experiencing the variability* is an important thing that the researchers get out of the workshops. They are always highly impressed by the fact that their result is *not* statistically significant while that of their neighbour *is* significant. When there are truly no differences between populations, they are "right," while when there are differences, they are "wrong." This leads to good-natured joking between the class members, and leads to a good atmosphere in the workshop, while simultaneously explaining the idea of "Type I and II errors" and the concept of "power."

An issue which worry researchers and which is always raised is "How do I know whether I've made the right decision?" My answer is that it is like being a jury in a court of law. You never really know whether you've hung an innocent person (Type I error), let a guilty person go free (Type II error), or made the correct decision – all you can do is design your experiment so that you minimise the chances of both types of error, and/or carry out a second experiment to confirm your apparent findings.

Throughout the workshops, this "uncertainty" in the results is continually discussed in relation to the various scenarios which we simulate. In some scenarios, we know there are truly no differences, yet some class members decide there are differences; in other scenarios, there are differences whose magnitude we know, yet some class members obtain estimates which are quite different from the true value. Since researchers in practice never know the truth in terms of the responses that they are estimating, they find it a useful exercise to see how good or bad the statistical methods are when they do know the truth.

Two further vital ingredients are:

7. *Confidence* building*;*
8. Building *interest in statistics*.

Many attendees come to the workshops with a dislike for statistics and a lack of confidence in their own ability in the subject. I like to think that they go away with a greatly increased level of confidence and a much higher level of interest in statistics. During the workshops people learn only the very basic facts of life "statistically speaking." They also learn during the discussion periods that there are many things which are outside the scope of the workshops – however, their basic grounding means that they are in a much better position to consult with a statistician concerning their more difficult designs or analyses. An improved appreciation of statistical ideas means that people are also more likely to seek statistical advice when they need it.

3.6. GENERAL COMMENTS

With the widespread availability of modern computers, one might wonder whether there is still a place for workshops based upon simulations done by "calculator arithmetic." Would it not be better to set up a room full of PCs so that each student can click the "simulate" button and save themselves a great deal of "mindless" arithmetic? My feeling is "no," though I would be interested in anyone's attempts to prove me wrong! My reasons are two-fold. Firstly, I feel that the mind is working whilst the arithmetic is being done, and that things are being learnt. The "no pain, no gain" phrase comes to mind – clicking a button is very easy, and perhaps as easily forgotten. Secondly, the PCs could be a distraction from the main event, which is thinking about statistics.

Another question, which is periodically raised, is: "Do the workshops need to be updated?" This is an alias for the thought: "You've run these courses for 22 years now – surely they must be out of date?" My answer is three-fold. Firstly, the methods covered by the workshops are still the ones most commonly used in agricultural research, and there is still a high demand for courses which explain the basic ideas behind these methods. Secondly, if the workshops work, why change them? After all, humans have been eating and drinking for centuries – some things take a while to go out of fashion! Thirdly, there is the automation issue, which I discussed in the first paragraph.

The value of the geometry-based workshops 6-7 is also periodically questioned, with the idea that researchers have a pragmatic view of statistics, and would not be interested in the mathematical ideas which underlie the methods (c.f., "they just need to be able to drive the tractor, not understand how it works"). I find that this is true of some researchers, just as it is true of some statisticians. However, equally, some researchers are fascinated by the elegance of the geometry and the way in which it unifies a large number of apparently unrelated, seemingly "ad hoc" statistical methods (this is also the fascination for myself). For example, one such researcher attended all of my workshops and told his colleagues that the geometry workshops were the most interesting of all. In addition, several people who have attended the geometry workshops have been interested enough to buy one or other of our books (Saville and Wood, 1991 and 1996).

In summary, workshops 1-5 for agricultural researchers are suitable as a "heuristic" way of introducing the statistical ideas which underpin the most commonly used "frequentist" methods. Workshops 6-7 and/or our two textbooks constitute an alternative, more unified approach in which theory, practice and computing are all viewed from the same perspective.

4. THE GEOMETRIC APPROACH

As mentioned above, Graham Wood and I have pioneered a new method of teaching statistics geometrically to university undergraduate statistics students and agricultural science postgraduate students. This has been the subject of two books (Saville & Wood, 1991, 1996) and a review paper (Saville & Wood, 1986). The first of these books served as the textbook for a second year applied statistics course at the University of Canterbury, Christchurch, New Zealand until the end of 1998.

This course was jointly taught by Graham and myself for the first year (1984), then by Graham on his own for several years, then by Dr Murray Smith for several years, and finally by Dr Peter Heffernan for several years. All of these lecturers were enthusiastic about the geometric approach, and the course was well liked by the students throughout the 15 years during which it was taught (sadly, the course was discontinued when the statistics courses were re-arranged in 1999). In general, the students found the material relatively easy to understand, and they found it satisfying because they were learning something relatively concrete (how to design and analyse experiments) following a method which unified theory, practice and computing.

As evidence of its popularity, I shall present the course and teacher evaluations for the final year that the course was run (Table 3). The first survey rated the course, while the second survey rated the teaching. The number enrolled in the course was 59, and the numbers responding to the two surveys were 33 and 35 respectively. The 1 – 5 rating scale is defined at the top of the table.

Table 3. Percentages of Responses To Two Surveys Carried out by the University of Canterbury Education Research and Advisory Unit at the End of the 1998 Course.

Student Rating of Course (1-5 points scale; 1 strongly disagree; 5 strongly agree)				
Standard Questions	Mean	<3	=3	>3
This was a well-organised course	4.3	0	3	97
This course helped to further stimulate my interest in the course area	4.0	6	15	79
The overall workload in this course was reasonable 1 = too light / 3 = reasonable / 5 = too heavy	3.2	9	64	27
The level of difficulty of this course was reasonable 1 = too easy / 3 = reasonable / 5 = too hard	3.2	9	70	21
Overall, this was a good quality course	4.2	0	24	76
Supplementary Questions				
The lectures were a valuable aid to my learning	4.1	3	16	81
The tutorials were a valuable aid to my learning	4.2	3	9	88
Student Rating of Teaching				
Standard Questions				
The classes were well organised	4.3	0	11	89
The lecturer was able to communicate ideas and information clearly	4.3	3	9	89
The lecturer stimulated my interest in the subject	4.1	6	20	74
The lecturer's attitude towards students has been good	4.8	0	0	100
Overall, the lecturer is an effective teacher	4.5	0	14	86
Supplementary Questions				
I found that the teaching methods used in this course were effective in helping me to learn	4.3	6	11	83

The ratings were excellent. For the questions for which 5 was the most favourable rating, the mean ratings were all between 4 and 5. For the two questions for which the most favourable rating was a 3 (workload and level of difficulty), the mean ratings were both close to this on 3.2. Of course, in such ratings the ability of the teacher is paramount in that a wonderful method of teaching will prove a dismal failure if the teacher is not equally wonderful! Nevertheless, some credit at least should go to the geometric method of teaching this material. The success of this course serves to contradict the opinion that students would not be able to cope with a geometric approach (such an opinion was expressed in several journal reviews of our 1991 book).

The geometric method of teaching statistics is especially enlightening for, though not restricted to, university students who have taken first-year courses in linear algebra (also called vector geometry) and statistics. The linear model statistical methods are seen as a simple application of linear algebra, and the linear algebra is made less abstract by being applied to a concrete problem. However, the linear algebra requirements of the method are quite minor, and could be easily learnt by Californian graduate students in agriculture during a one hour lecture.

Our second book (a primer) on the geometric approach also includes material on the way in which R. A. Fisher came to invent the linear model methods. This is material on the *p* value which is included in an appendix (D).

ACKNOWLEDGEMENTS

I would like to thank the attendees and other colleagues who have contributed to the development of my workshops over the years and AgResearch for sponsoring my attendance at the IASE Round Table conference in Tokyo, Japan, in August, 2000.

REFERENCES

Bangdiwala, S. I. (2001). Training of statisticians worldwide to collaborate as co-investigators within country clinical epidemiology units: The experience of the International Clinical Epidemiology Network (INCLEN). In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 265-275). Granada: International Association for Statistical Education and International Statistical Institute.

Bishop, O. N. (1980). *Statistics for Biology: A practical guide for the experimental biologist, 3rd edition.* Harlow, Essex, U.K: Longman.

Saville, D. J. (1980). Analysis of variance in agriculture (revised). *Ministry of Agriculture and Fisheries Biometrics Bulletin* 1, 1-49.

Saville, D. J. (1990). Multiple comparison procedures: The practical solution. *The American Statistician* 44, 174-180.

Saville, D. J., & Wood, G. R. (1986). A method for teaching statistics using n-dimensional geometry. *The American Statistician* 40, 205-214.

Saville, D. J., & Wood, G. R. (1991). *Statistical methods: The geometric approach*. New York: Springer.

Saville, D. J., & Wood, G. R. (1996). *Statistical methods: A geometric primer*. New York: Springer.

David J. Saville
AgResearch, Statistics Group
Canterbury Agriculture and Science Centre,
P.O. Box 60, Lincoln, New Zealand
E-mail: Dave.Saville@AgResearch.CO.NZ