Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 177-195.

Printed in Granada, Spain.

JOHN HARRAWAY, BRYAN MANLY, HILARY SUTHERLAND,
& ALLAN MCRAE

MEETING THE STATISTICAL NEEDS OF RESEARCHERS IN THE BIOLOGICAL AND HEALTH SCIENCES

The results of a survey on the use of statistics in research in five subject areas representative of the biological and health sciences are reported. The main component of the survey is a review of statistical methods in 2927 research papers published during 1999 in 16 high impact journals from botany, ecology, food science, marine science and nutrition. A factor analysis establishes that research papers in the different subject areas use different methods. The opinions of research staff and postgraduate students working in these areas are also reported. To support these opinions we provide details of five postgraduate studies involving advanced statistical analyses, which have either resulted in publication or should result in publication in the near future.

Discussion develops recommendations about topics important in a statistics curriculum for research students, where statistics courses should be taught, what is needed in terms of level of theory, the use of short courses and workshops, and the value of project work.

1. INTRODUCTION

Serious problems for research students in the sciences often revolve around inadequate training in study design principles and statistical methodology, both essential for planning research and analysing data, which are frequently collected in the field over long periods of time.

In many cases the research student has attended a first course in traditional statistical procedures involving data summary methods, the binomial and normal distributions, t tests, simple linear regression, correlation, and the analysis of variance. But these methods alone are insufficient for the scientists of today and frequently are taught early in an undergraduate degree with the result that many of the ideas are forgotten by the time they are needed.

This paper identifies methods currently being used by scientists through a survey of 2927 research papers published in 16 high impact journals during 1998 and 1999. The opinions of PhD students at the University of Otago are also canvassed and five of their case studies, which are leading to publication, are reported to provide support for the results of the survey.

The concluding discussion will describe how the statistical needs are presently being met and will make a number of additional recommendations about how to develop statistics training for researchers beyond the standard procedures in a first course.

2. LITERATURE REVIEW

Several studies, both general and subject specific, have identified curricula content for statistics programmes designed for researchers and graduate students.

Brogan and Kutner (1986) recommended a two-course statistics sequence with graduate students needing to learn about general linear models including multiple linear regression, analysis of variance and categorical data analysis as well as the use of statistical computer packages.

Elmore and Woehlke (1988, 1998) have focussed on the education research area by conducting statistics content analysis of 1906 papers published in three education research journals covering the period 1978 to 1997. Excluding descriptive statistics, the predominant methodology used was analysis of variance and covariance in 12% of the papers surveyed, followed by multiple regression (8%), multivariate procedures including factor analysis and cluster analysis (8%), qualitative methods (5%), correlation (5%), meta analysis (5%), nonparametric procedures (4%), *t*-tests (4%) and structural equation modelling (3%). The use of qualitative methods and graphical methods has increased substantially over the last ten years in these education journals.

Blumberg (2001) reports on the training of special education teachers in the use of statistics. In addition to the topics reported by Elmore and Woehlke for education research, Blumberg identifies the single-subject design, important because small samples are inherent in studies in special education due to the rarity of many disabilities.

Aitken et al (1990) investigated graduate training in statistics in psychology by surveying 222 psychology departments granting doctoral degrees. The majority of the departments offered courses including analysis of variance, multiple regression, and multivariate analysis of variance, which were viewed as the traditional “old standards of statistics”. “More advanced statistical considerations” including causal modelling, time series, power analysis and repeated measures using multivariate procedures were newer but used less frequently.

They also defined a concept of “in depth” which implied that students could perform an analysis in question themselves. Generally, new PhD’s were competent to handle the traditional techniques but not the newer ones. This idea will be supported by our investigation, which suggests that much of the research in the subjects we have selected can only be successfully completed by enlisting professional statistical advice.

Hammer and Buffington (1994) in a study which is not dissimilar to ours reviewed 1062 articles published during 1992 in six veterinary journals finding that 51% did not include statistical analyses or used only descriptive statistics. They made no attempt to determine whether statistical tests were used correctly and this is the case with our study. The most commonly used tests were analysis of variance and *t*-tests which along with contingency tables, nonparametric tests and simple linear regression permitted access to 90% of the veterinary literature surveyed.

Altman (1998) reports that authors writing for medical journals have poor statistical knowledge and many articles are error prone in relation to statistical content. He comments that there is now much greater use of methods such as logistic regression, proportional hazards regression for survival data, the bootstrap, Gibbs sampling, generalised additive models, classification and regression trees, generalised estimating equations, multi-level models and neural networks with this trend towards greater complexity continuing.

He emphasises the role of the journal editors and statistical referees in the review process. Curtis and Harwell (1998) make this point as well. We also believe this effect has an influence on the extent of occurrence of statistical procedures in the journals that we have reviewed. In particular it will be noted that the *Journal of Vegetation Science* has a different pattern of statistical use when compared with the *New Zealand Journal of Botany* and *Plant Physiology*.

Curtis and Harwell (1998) report the results of a survey of 27 quantitative methods programs in the field of education producing results similar to those of Elmore and Woehlke (1988, 1998). The majority of doctoral students receive training in the "old standards" analysis of variance, multiple regression, and traditional multivariate procedures as well as logistic regression and log-linear models. Nonparametric procedures are now less prevalent. In terms of recent procedures the majority of programmes train their students in meta analysis and structural equation models while some of the programmes include computer intensive methods and multilevel models.

They also report that students have "strong" skills in standard data analysis packages such as SAS and SPSS, but data base management and programming skills are more varied. Mathematical statistics training was ranked midway between "weak" and "strong". Another line of research suggested was a study of the opinions of doctoral students themselves regarding the adequacy of their own training and this issue is taken up in our paper by surveying students at the University of Otago and describing several recent studies.

3. METHODS

The 16 high impact journals for our survey were selected using *the Journal Citation Reports* (Science Edition) and were confirmed by the opinions of University of Otago staff in the five target departments who publish in these journals. Altogether, 2927 papers published in late 1998 or during 1999 were reviewed and details of statistical procedures present in each paper were recorded.

The journals surveyed, with numbers of papers analysed, are summarised in Table 1, which also includes the percentages of papers in each journal using some form of statistical analysis. For example, ninety five percent of the papers in the *Journal of Vegetation Science* and eighty two percent of all papers contained statistical methodology.

The methods and tests used in each paper were classified initially into 60 classes, but these were subsequently pooled resulting in the fourteen major categories listed in Table 2. A more detailed list of topics is in Appendix 2.

Several decisions were made when classifying the topics. In the analysis of variance category no attempt was made to distinguish between experimental designs like split plots which arose in the biological journals and cross over designs which were common in the nutrition journals. Prospective or retrospective, cross sectional or longitudinal, cohort or case-control studies and the principle of blinding occurred in many nutrition journals, but no attempt was made to differentiate between these and the random, transect or stratified sampling schemes in the other subjects.

Circular data were classified as spatial. Although meta analysis was developed in the social sciences, and more recently has been used in ecology, the few instances of meta analysis were classified as medical statistics on the grounds that the technique is now used quite widely in the health sciences.

Table 1: Summary of Papers Analysed

	Journal	Subject	Number of Papers	Percentage with statistics
a)	Journal of Vegetation Science (Dec 98 – Oct 99)	Botany	88	95
b)	NZ Journal of Botany (Dec 98 – Sept 99)	Botany	63	35
c)	Plant Physiology (Dec 98 – Nov 99)	Botany	446	51
d)	Oikos (Nov 98 – Oct 99)	Ecology	242	72
e)	Journal of Ecology (Dec 98 – Oct 99)	Ecology	89	89
f)	Ecology (Dec 98 – Oct 99)	Ecology	246	96
g)	Journal of Dairy Science (Nov 98 – Oct 99)	Food Science	317	86
h)	Journal of Sensory Studies (Dec 98 – Sept99)	Food Science	28	89
i)	Food Quality & Preference (Jan 99 – Sept 99)	Food Science	237	84
j)	Journal of Food Science (Nov 98 – Oct 99)	Food Science	51	92
k)	Marine Environmental Research (Feb 99 – Dec 99)	Marine Science	54	93
l)	Canadian Journal of Fisheries & Aquatic Sciences (Oct 98 – Sept 99)	Marine Science	213	96
m)	Marine Ecology Progress Series (Oct 98 – Nov 99)	Marine Science	359	87
n)	American Journal of Clinical Nutrition (Nov 98 – Oct 99)	Nutrition	241	96
o)	European Journal of Clinical Nutrition (Oct 98 – Sept 99)	Nutrition	146	95
p)	British Journal of Nutrition (Sept 98 – Nov 99)	Nutrition	107	95
	Total		2927	82

Table 2. Description of Categories

1	Analysis of Variance (anova)
2	Posthoc Tests and Contrasts (posthoc)
3	Simple Regression and Correlation (simpreg)
4	Regression and Modelling (regmodel)
5	Logistic Regression (logreg)
6	Contingency Tables and Log Linear Modelling (loglin)
7	Multivariate Methods (multivar)
8	Basic Tests and Procedures (basic)
9	Medical Statistics (medstat)
10	Population Estimation (popestim)
11	Spatial Analysis (spatial)
12	Computer Intensive Methods (compinte)
13	Stochastic Processes (stochpro)
14	Field Specific (fieldspe)

Each journal had a small number of specialised methods, which would not generally be taught in a general statistics paper, and these were classified as field specific; two examples were diversity analysis and genetics. A small number of papers involving some form of mathematical analysis were excluded on the grounds that these methods would not be taught in a statistics paper.

The category classifications made for *Plant Physiology* by one of the authors were made also by a second author with agreement being found between the two classifications thus confirming classification reliability. Counts were made of the occurrences of the 14 categories in each of the 16 journals and the counts were converted to the percentages of the total number of papers for each journal shown in Tables 3 and 4.

For example, of the 88 papers analysed in the *Journal of Vegetation Science*, 38% involved some analysis of variance procedure, 24% involved a post-hoc analysis and so on for the remaining entries in the table.

Table 3: Percentage Occurrence of Topic Category in each Journal

Journal	anova	posthoc	simpreg	regmodel	logreg	loglin	multivar
1	38	24	42	45	11	26	44
2	13	6	13	8	5	3	3
3	8	9	14	16	0	1	5
4	36	12	31	37	6	16	12
5	38	18	47	46	17	12	17
6	52	26	55	64	7	18	19
7	44	30	29	36	7	5	5
8	50	21	25	21	0	0	50
9	67	33	22	25	0	20	47
10	45	40	23	41	0	2	5
11	39	26	37	37	2	8	31
12	41	22	44	61	5	16	17
13	43	23	42	46	3	11	12
14	54	27	53	43	12	20	4
15	38	14	42	42	10	16	2
16	58	33	27	27	5	6	1

Table 4: Percentage Occurrence of Topic Category in each Journal (continuation)

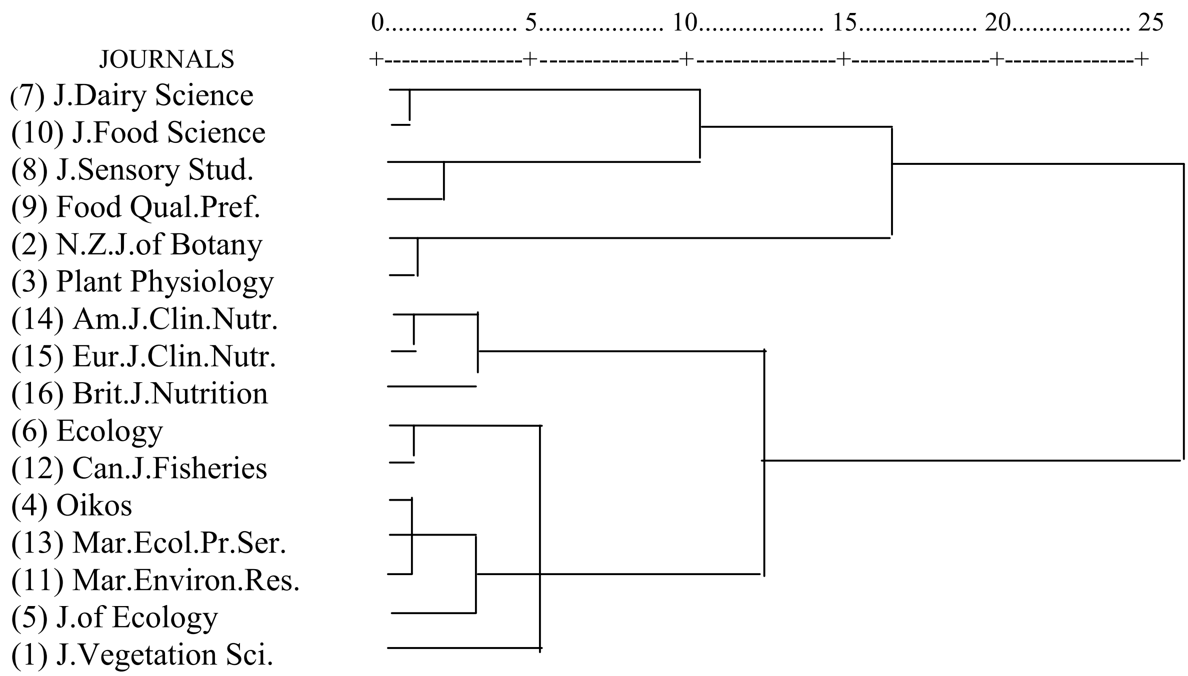
Journal	basic	medstat	popestim	spatial	compinte	stochpro	fieldspe
1	42	1	6	37	26	1	3
2	13	2	8	5	0	0	7
3	20	0	0	0	1	0	6
4	41	1	7	8	9	3	2
5	27	0	6	22	10	6	7
6	54	1	14	13	20	4	1
7	21	1	0	0	3	0	3
8	32	0	0	0	11	0	4
9	41	0	0	0	6	0	4
10	16	0	0	0	2	1	1
11	50	0	7	7	6	2	2
12	49	0	8	3	27	5	4
13	51	0	3	3	10	3	8
14	71	5	0	0	2	0	2
15	73	14	0	0	1	0	1
16	64	6	0	0	0	0	3

4. RESULTS

Cluster analysis using Ward's method identified four clusters of journals shown in the dendrogram in Figure 1.

Cluster 1 included four Food Science Journals, cluster 2 two of the Botany Journals, cluster 3 the Nutrition Journals and cluster 4 the Ecology and Marine Science Journals as well as the *Journal of Vegetation Science*. Both multidimensional scaling and correspondence analysis produced similar journal groupings, but a principal component factor analysis is chosen for further analysis since it shows clearly relationships between the journals and topic categories.

Figure 1: Dendrogram, Ward's method, for 16 Journals



The principal component analysis on the correlation matrix for the 14 categories produced four components explaining 80.74% of the total variance in the percentages. The plot of the scores of principal component one against principal component two is shown in Figure 2.

This confirms the same journal groupings identified in the dendrogram. There is a clear grouping of nutrition journals with high scores on principal component two, and a grouping of food science journals with negative scores on principal component one. The *NZ Journal of Botany* and *Plant Physiology* with low scores on both principal components one and two are similar but the *Journal of Vegetation Science* is more like the ecology and marine science journals, all having positive values on principal component one.

This, we believe, reflects the influence of at least one of the editors of the *Journal of Vegetation Science* who is known to have statistical interests.

A varimax rotation produced the four common factors with loadings greater than |0.100| listed in Table 5. Factor one, explaining 26.0% of the total variance in the percentages in Tables 3 and 4, is a measure of the extent of regression and modelling techniques, computer intensive methods and population estimation procedures.

Figure 2: Plot of 16 Journals for the First two Principal Components

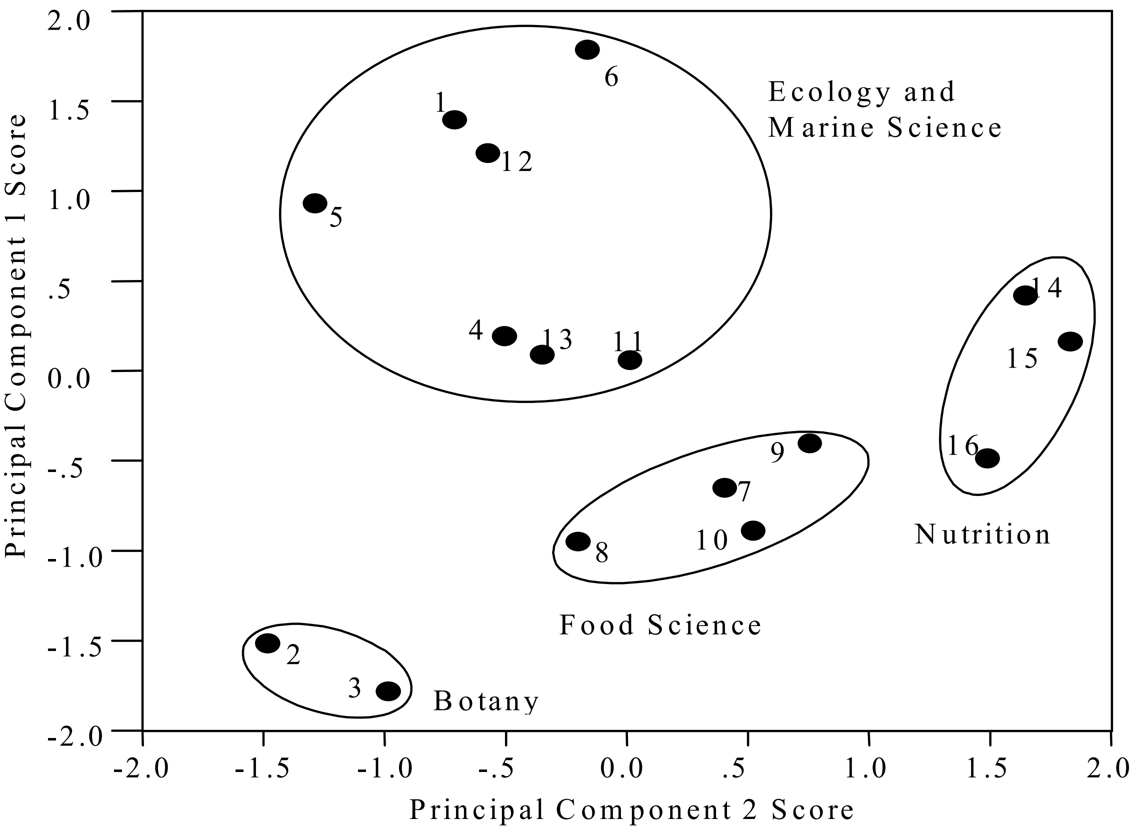


Table 5: Factor Loadings

	Factor			
SUBJECT	1	2	3	4
STOCHPRO	.934			
REGMODEL	.819	.331	.400	
POPESTIM	.756		-.248	.223
SIMPREG	.682	.595	.276	.102
COMPINTE	.674			.602
MEDSTAT	-.341	.827		-.319
BASIC		.766	.421	
LOGREG	.401	.698	-.252	
LOGLIN	.344	.626	.198	.519
POSTHOC		-.144	.895	
ANOVA		.161	.890	.182
FIELDSPE		-.375	-.596	
MULTIVAR		-.235	.239	.881
SPATIAL	.446	.205	-.290	.684

Factor two, explaining 20.7% of the total variance, reflects basic testing procedures, medical statistics, logistic regression and log linear modelling. Factor three, explaining 19.2% of the total variance, is a measure of the extent of analysis of variance and post-hoc procedures with few field specific topics while Factor four, explaining 14.9% of the total variance, reflects the extent of multivariate methods, computer intensive methods and spatial data.

The factor scores for the 16 journals are listed in Table 6. The *Journal of Vegetation Science* has an emphasis on multivariate, computer intensive and spatial analyses with the *NZ Journal of Botany* and *Plant Physiology* having an emphasis on field specific topics, population estimation and spatial analyses with lesser use of analysis of variance and post-hoc procedures.

The *Journal of Ecology* and *Ecology* have emphasis on simple regression, modelling, computer intensive techniques, population estimation and spatial analyses with a lesser emphasis on these topics in *Oikos*.

The *Journal of Dairy Science* has little use of population estimation, spatial analyses and multivariate procedures. The *Journal of Sensory Studies* and *Food Quality and Preference* are low on population estimation, and computer intensive procedures but have extensive use of multivariate methods. *The Journal of Food Science* is high on analysis of variance and post-hoc procedures.

The marine science journals show moderate use of all the techniques identified but the papers in the *Canadian Journal of Fisheries and Aquatic Sciences* has extensive use of simple regression, modelling, stochastic processes and population estimation.

The nutrition journals, in particular the American and European Journals have extensive use of medical statistics procedures, logistic regression and log linear modelling as well as the basic testing procedures.

The occurrences of the different techniques in the five subject areas provide some clues to topics that should be emphasised in a postgraduate statistics curriculum for each of the subjects.

Table 6: Factor Scores for the Sixteen Journals

Journal	Factor 1	Factor 2	Factor 3	Factor 4
1	0.11	0.85	-0.46	2.79
2	-0.69	-0.54	-2.28	-0.33
3	-0.89	-0.92	-1.59	-0.60
4	0.39	0.17	-0.55	-0.01
5	1.48	0.14	-1.14	0.14
6	1.83	0.27	0.71	0.09
7	-0.32	-0.47	0.37	-0.74
8	-1.09	-1.05	0.37	0.98
9	-1.24	-0.57	1.27	1.42
10	-0.09	-1.31	1.23	-1.10
11	0.16	-0.41	0.41	0.18
12	1.73	-0.34	0.32	-0.18
13	0.70	-0.44	0.01	-0.46
14	-0.31	1.63	0.72	-0.51
15	-0.91	2.50	-0.25	-0.82
16	-0.87	0.48	0.86	-0.86

5. THE STAFF SURVEY

The survey of research scientists and staff in the five target departments had 46 respondents. The resident department of each respondent was unfortunately not identified, and therefore the results and conclusions cannot be subject specific. The first question asked staff to nominate, from a given list, topics that they use in their research and topics that they feel their graduate students should know how to use.

The percentage use of each statistical method is presented in Appendix 1. As expected there is a significant correlation between the techniques used by supervisors and techniques seen to be important for students (p-value < 0.001). There is high use of the "old standard" methods involving t tests, ANOVA, simple linear regression, correlation, confidence intervals and multiple comparisons while the more specialised or theoretical methods including matrix methods, time series, survival analysis, canonical correlation, correspondence analysis and multidimensional scaling are less frequently used.

70% of the respondents believe that postgraduate students have inadequate statistical background before embarking on their research. To overcome this problem, 93% feel postgraduate students should attend intensive postgraduate workshops, 26% feel that completing more undergraduate papers could be worthwhile and 30% feel that relying on their own reading during their research could be beneficial. The package Excel is used by 65% of the staff with SPSS (24%), MINITAB (30%) and SAS (20%) being the other most used software.

6. THE STUDENT SURVEY AND CASE STUDIES

There were 43 responses to the postgraduate survey with five additional students being interviewed about their research to illustrate the extent of the problems that arise. The results showed that 86% of the postgraduate students had taken at least one statistics paper as part of their undergraduate degree and 49% had taken at least one mathematics paper. Those who took a mathematics paper were more likely to have taken a statistics paper as well.

Courses or workshops were attended by only 35% as part of their postgraduate study, and 23% did not believe a workshop would be helpful. Of those who would like to attend a workshop/course, modelling techniques and multivariate statistics were the most common topics of interest. 44% of students had training in the use of computer packages before commencing postgraduate study. Excel (60%) was the most common package used in research followed by SAS (19%) and SPSS (26%). About one quarter only of students felt that their knowledge of statistical procedures was adequate for their research.

Several recent studies, which typify the problems confronting postgraduate students in the subjects being investigated in this paper, are described. Each study has involved substantial data analysis assistance from professional statisticians. In each case the analysis has resulted in or is leading to publication.

Case Studies 1 and 2: Marine Science

A biology student with background in algebra, calculus and some introductory probability completed papers for a Master of Marine Science degree, which included a course on statistics for marine scientists, before commencing field work over two summers on the behaviour of Hector's dolphins in a bay in the South of New Zealand. His research led to the publication of two research articles.

The first (Bejder et al 1998) investigated the association pattern of Hector's dolphins using the Half-Weight Index as a measure of association of two animals. A randomisation test was used to test the extent to which the observed association index values differed from those of a randomly associating population. Association patterns were also investigated using cluster analysis, multidimensional scaling and minimum spanning trees.

The second article (Bejder et al, 1999) reported results from tracking by theodolite the path of a dolphin pod and a tour boat over a period of several hours on 24 days. The direction of movement relative to the boat during the time the boat was in the bay was recorded at 40 second intervals.

In Table 7 the 853 bearings relative to the boat are classified according to the time the boat was present. The proportions of boat directed movements at different times were modelled using logistic regressions. Deviance differences and the Akaike Information Criterion (AIC) established the best model was quadratic in time. This showed the dolphins were attracted to the boat for about 50 minutes after which they lost interest. This conclusion has been used by the Department of Conservation in New Zealand to limit the times tour boats may be in the vicinity of the dolphins.

A second marine science doctoral student with prior training only in basic statistical methods collected data on the presence and absence of Hector's dolphins at 980 locations around the South Island of New Zealand. The sea surface temperature, water depth, water clarity and season of the year were measured at each location sampled.

Table 7: *Bearings Towards Boat Classified by Time*

Time into encounter (min)	Total (n)	Frequency towards boat
0-10	94	32
10-20	133	48
20-30	168	50
30-40	133	52
40-50	111	38
50-60	98	27
60-70	59	16
70-80	34	3
80-90	23	3

The purpose of the study was to establish the preferred environment for dolphins. This entailed fitting habitat selection models using AIC and deviance differences.

A serious problem, which arose with this study, was the result of an inadequate sampling protocol. Initial consultation with a statistician had not taken place. Fortunately, it has been possible to refine the sample and the results are now likely to lead to publication. Although the two students were excellent biologists with good data base management skills, deficiencies in advanced statistical methods, in particular randomisation procedures, multivariate procedures, advanced regression modelling and principles of sampling design would not have allowed completion of their work without consultation with statisticians.

Case Study 3: Food Science

A doctoral student who is a chemistry graduate with first year general mathematics, some advanced calculus but no statistics is working in the area of sensory science involving assessment of taste and smell.

Her research involves training panelists to evaluate mixtures of compounds, with each compound at low, medium and high levels. There are several odour attributes associated with each compound, and the panelists evaluate the mixtures and "score" each of the attributes on unstructured line scales.

Response surfaces for each of these attributes are produced and these indicated how the particular attribute varies with changes in concentration of mixture components. The panel evaluates some intermediate points not used to generate the response surfaces, and checks that these lie on the response surfaces, the idea being that if they do, the response surfaces could be used predictively to establish the attributes for any mixture of these given compounds.

A second study is looking at the perception of trained panelists versus untrained people and the effect on the task they are asked to do – having them evaluate mixtures according to a list of attributes is different to what we do naturally when we consume a food or smell a smell. The trained panel's results from this attribute-based assessment are compared with their results from a task, which involves judging only similarity of mixtures.

An untrained panel perform the similarity task and compare their results with the trained panel. The actual task of the panelists is to select the most similar pair from a set of three samples – so for example in the first experiment there are two compounds either absent, or at low, medium and high levels, giving nine samples. There are 84 possible combinations of these nine samples, and all the panelists see these 84 combinations. These data are analysed using multidimensional scaling.

In interview this student felt she would have benefited from taught work on study design to help formulate the questions to ask a statistician, a rundown on available methods with their strengths and pitfalls, a rundown on appropriate statistical packages, and an opportunity to bring actual data along to any workshop aimed at teaching particular aspects of statistical methodology.

Case Study 4: Nutrition

The nutrition doctoral student involved in this study had completed only an introductory business statistics paper before commencing her research.

Her problem (Taylor et al, 1998) was to assess the relative ability of body mass index (BMI), waist girth and waist hip ratio (WHP) to correctly discriminate between subjects with low and high trunk fat mass or central fat mass as measured using dual-energy X-ray absorptiometry (DEXA).

The subjects consisted of 96 adult women. The DEXA measurements of trunk fat mass and central fat mass were the reference measurements and the anthropometric techniques were the screening tests. Studentized residuals were constructed for all variables. The 75^{th} percentile for each age-adjusted DEXA measurement defined a subject as truly positive. Sensitivity and specificity of 19 percentile cut-offs were calculated for each screening measure and receiver operating characteristic curves (ROC's) were constructed. A bootstrapping procedure in STATA was used to calculate the areas under the curves, the 95% confidence intervals and the differences between the areas.

In a second study (Taylor et al, 2000) a similar analysis was conducted except the conicity index instead of BMI was evaluated in a large group (n = 580) of children and adolescents. Internal z scores were created for each variable. A z score $\geq + 1$ identified a subject as truly +ve. Positive and negative likelihood ratios were also calculated in this paper.

Generally in her research this student had used log transformation, t-tests, one-way ANOVA, simple factorial ANOVA, multiple regression analysis, curve estimation, calculation of sensitivity and specificity and construction of receiver operating characteristic curves, calculation of internal and external z scores, positive and negative predictive values, cross-tabulations and chi-squared tests. The SPSS package had been used for most of the calculations. In interview the student commented that it is crucial to increase the statistics knowledge of those involved in research. Although it is possible to learn from trial and error it is a very steep learning curve.

Case Study 5: Botany

This doctoral student had taken no mathematics papers in his undergraduate degree and one introductory statistical methods paper covering the "old standards" which was subsequently supplemented by a further paper on similar material taught within the Botany Department. All the statistical analyses were performed using the programs Teddybear, Statistix and Glim.

The PhD topic was "The comparative ecology of rare and common *Aceana* and *Chionochloa* species". The work comprised glasshouse and garden experiments, where species were grown from each genus together, subjecting them to various treatments relating to intrinsic growth and reproductive traits, and taking account of responses to stress factors and competition.

Experimental responses were analysed using analysis of variance, and correlated response means with the geographic range sizes of the species, separately for each genus. Looking at the patterns within each genus reduces phylogenetic dependence in the data, since many phylogenetically correlated traits 'cancel out' when comparing congeners. Some species were represented by two populations, and because these were non-independent the mean value for each of these population pairs was used in the correlations. Consistent differences between rare and common species were being looked for. Regression methods were used frequently.

A typical experiment comprised replicated randomised blocks of species being treated with one or more controlled factors. For one experiment where the treatments were difficult to apply at the level of single plants, a split-plot design was used. The number of controlled factors and their levels varied. Examples of some of the designs used for *Aceana* are presented in Table 8.

A variety of problems arose with the analyses. There was almost always a small amount of mortality (unrelated to treatments) among the experimental plants, raising complications when it came to statistical analysis, as it was always desirable to extract a block effect. Either the block effect was included in the error term, with a reduction in power, or "missing values" were estimated. In other experiments, the mortality response where it differed between treatments was analysed. Because mortality is an all-or-nothing event, the data had a binomial error distribution. In these cases Glim was used to perform generalised linear modelling, specifying a binomial error structure.

In interview the student suggested the best way to approach the presentation of statistics to botany research students is to give examples of how poor experimental design can invalidate results. He saw two important issues to be pseudoreplication (in the field and with growth cabinets), leading to non-independence of data, and inadequate randomisation, opening up the possibility of biased treatments. The usefulness of randomised blocks in reducing noise (e.g. if plants vary in size, associate difference in size with the different blocks) should be stressed.

Table 8: *Examples of Experimental Designs Used*

Experiment	Growth rate	Reproduction	Nutrients	Shade	Competition
Design	completely randomised (in growth chamber)	randomised block	fully factorial, randomised block	split-plot	fully factorial, randomised block
Treatments	13 species	13 species	13 species	12 species (subplot)	13 species
	10 replicates	4 replicates	N(3 levels)	Shade (5 levels) (plot)	2 fertility levels
			P(3 levels)	4 replicates (plot)	2 competition levels
			5 replicates		6 replicates

7. MEETING THE STATISTICAL NEEDS

At the present time two papers outside the usual undergraduate statistics programme are provided at the University of Otago for postgraduate researchers in the biological and health sciences.

The first paper, for marine and environmental scientists, has as prerequisite an introductory methods paper including the “old standards”. It has 72 teaching contact hours and covers sampling and estimation (3 weeks), testing hypotheses (2 weeks), regression methods (2 weeks), analysis of variance (2 weeks), BACI designs (2 weeks), power analysis and sample size determination (2 weeks), time series (2 weeks), environmental monitoring (2 weeks), impact assessment (2 weeks), spatial data (2 weeks), bioequivalence (2 weeks) and censored data (2 weeks).

The purpose of the paper is to introduce students to the theory and methods of sampling and modelling in the context of environmental science. The emphasis is on correct and justifiable procedures rather than statistical theory. Project work is used to promote “learning by doing”.

For example, one project involves estimating the biomass of a shell fish in a tidal inlet. The students design a stratified sampling scheme, carry out the sampling and find an estimate of the biomass from the data. An appropriate text for the course is available (Manly, 2001).

The second paper is for postgraduate students who have knowledge of the “old standards” but who need the more advanced analysis methods routinely used in the health sciences. It has 52 teaching contact hours with less emphasis on statistical theory but more emphasis on the relationship between statistical methods and the scientific inference.

From a statistical viewpoint the paper covers regression methods including the generalised linear model (linear, logistic, Poisson regression) and the Cox regression model. From the scientific aspect the course covers definition of outcome measures, transformations, robustness of linearity assumptions, scientific aspects of model selection, and model checking. The paper relies on several data sets providing examples from a range of health science fields. The class is attended by clinical researchers and statisticians training to be biostatisticians. Bangdiwala (2001) points out that such interaction between these groups is a valuable learning experience for all involved. Aspects of the INCLEN programme described by Bangdiwala are also included in this paper.

8. DISCUSSION

It is apparent from the surveys presented above that a great deal of statistics is used by researchers in the areas of science covered, but the important techniques vary to some extent between, and even within disciplines. Given the time constraints on these researchers, it is apparent that although basic statistics (the ‘old standards’) can be taught in generic courses, it is highly desirable that advanced learning is through approaches that are targeted to the needs in different areas.

For example, we see that a botany student interested in studying the ecology of plant species in a quantitative manner, in the style of papers in the *Journal of Vegetation Science*, needs to have a good knowledge of multivariate and spatial analyses and computer intensive methods in particular. This might have to be provided by giving less emphasis to analysis of variance, regression methods and other topics. On the other hand, a nutrition student mainly needs a good knowledge of analysis of variance and regression methods.

There are three basic approaches that can be used to provide starting researchers with the statistical skills that they need:

a) They can learn informally from their own reading, with assistance from established researchers;
b) They can attend specialist short courses and workshops provided by statisticians or by established researchers in their disciplines; or
c) They can attend formal lecture courses, again either provided by statisticians or established researchers.

At present (a) seems to be the most commonly used approach. However, we believe that this is not altogether satisfactory because it may result in the researchers having gaps in their knowledge that they are not even aware of.

The best approaches seem to be (b) and (c), or perhaps a combination of these. If these are used then we believe that it is absolutely crucial that there is the involvement of statisticians who are very familiar with applications of statistics in the subject area. Established researchers may be well informed about current uses of statistics, but they will rarely be able to see how these relate to developments of statistical methods in general.

The danger with just using established researchers is therefore that important new developments in statistical methodology will be ignored. The situation is no better if statisticians who are not familiar with the subject area do the teaching. In that case they will not be able to judge very well what topics are important, and probably will bore their audience with irrelevant examples.

At the University of Otago, approach (c) has been used with success with courses for biological and health science students. We believe that it has been successful largely because the statisticians involved are very experienced at consulting and research in the subject areas.

Statistics staff have also provided a number of short courses on specialised areas of statistics for researchers. These have been intended primarily for scientists outside the university, but research students and university staff have been able to attend. These courses seem to fill a real need for the scientific community by providing an update on the latest developments in a restricted area of statistics. We see such short courses as complementing rather than replacing other modes of instruction.

Although there have been successes at the University of Otago, this does not mean that the situation is satisfactory. Of the staff in our survey, 70% thought that students were inadequately prepared when they started research. Of the students, 75% thought that they were inadequately prepared. One reason for this is that most of the students have not taken more than an introductory course in statistics, although more advanced courses that would be valuable for them are available.

One question that has only been considered rather indirectly in our surveys concerns how well the existing methods available for researchers to learn statistics are actually working in the end. Students may be inadequately prepared when they embark on research, but perhaps this is rectified by the time that they are ready to publish their results. Put another way: is the research that is appearing in scientific journals usually based on correct and appropriate statistical analyses?

It is difficult to know for sure what the answer to this question is because it often depends on the nature of data that are not provided in papers, it is not often possible to decide exactly how analyses were done, and in some cases the meaning of 'correct and appropriate' is debatable. However, there is no question that some very experienced consultant statisticians think that there are problems. Altman (1998) expressed this view with regard to medical researchers. Maindonald (1999) puts it quite forcefully:

> *"My reading of published papers persuades me that serious problems with the design of data collection and with data analysis are common. A cursory overview of papers in major international journals may be sufficient to reveal examples of serious statistical misinterpretation,…, there is a case for treating all published analyses as preliminary, pending scrutiny by researchers with relevant statistical skills!"* (Maindonald, 1999)

If these views are correct, which we suspect is the case, then there certainly are real problems at the moment. So how do we meet the needs of researchers in the health and medical sciences?

We suggest that the only really satisfactory approach is for statisticians with relevant consulting experience to develop structured programmes in collaboration with established researchers. These programmes must recognise the way that statistics is currently being used in the discipline concerned, but must also take into account broader developments in statistical methods so that researchers are able to make use of new approaches where this is desirable.

APPENDIX 1

Percentage Use of Statistical Method

Method	Percent Staff Use	Percent Student Use
t test	91	91
ANOVA	91	93
Simple linear regression	91	93
Correlation	91	89
Confidence intervals	87	87
Experimental design	83	74
Sampling/Power	72	72
Graphical procedures	67	63
Nonparametrics	67	74
Multiple comparisons	67	89
Transformations	63	74
Multiple regression	63	72
Survey design	63	61
Goodness of fit	61	63
ANCOVA	59	67
MANOVA	50	59
Contingency tables	46	50
Principal components	43	39
Cluster analysis	33	35
Contrasts	30	30
Bootstrap/Randomisation	30	39
Discriminant analysis	28	24
Factor analysis	28	24
Fishers exact test	26	39
Population estimation	22	24
Time Series	22	15
Logistic regression	22	39
Non Linear regression	20	35
Simulation	17	26
Correspondence analysis	17	15
Maximum likelihood	15	22
Multidimensional scaling	15	15
Loglinear modelling	13	28
Matrix methods	13	9
Canonical correlation	11	9
Jackknifing	11	20
Survival analysis	9	15

APPENDIX 2: CATEGORY DESCRIPTIONS IN DETAIL

1. *Analysis of variance and experimental designs:* Factorial designs. Blocking. Repeated measures. Split plots. Cross over designs. Random and Mixed Models. Approximate *F*-test. Levene's Test. Kruskal-Wallis and Friedman tests. Ante-dependent repeated measures analysis. Q_w test for within group homogeneity.

2. *Posthoc tests and contrasts:* Multiple comparisons. Student-Newman-Keuls. (Fisher's) least significant difference. Tukey. Duncan. Scheffe. Dunn. Ryan. Dunnett. Least square means.

3. *Simple regression and correlation:* Allometry. Fisher's *z* transformation. Durbin Watson. Pearson. Spearman.

4. *Regression and modelling:* Multiple regression. Response Surface Analysis. ANCOVA. Path Analysis. Non-linear regression. Growth curves. Nonparametric regression. Lowess curves. Ridge regression. Classification and Regression Trees (CART) GLM. GAM. Bayesian inference methods. Partial least squares regression. Structural equations. Polyserial correlation coefficients. Kernal density estimators. Linear standard curves. Split line regression. Spearman Karber 3 D plane regression procedure. Longitudinal regression models. Maximum likelihood and REML.

5. *Logistic regression:* Odds ratios. Survival analysis. Proportional hazards models. Cox regression model. GLM. GAM (for this type of analysis).

6. *Contingency tables and log linear modelling:* Fisher's exact test. Goodness of fit; χ^2. Cochran's C test. Mantel test for trend. Mantel-Haenszel method. McNemar's test. Categorical modelling; CATMOD. GLM. GAM (for this type of analysis).

7. *Multivariate methods :*MANOVA; MANCOVA. Box's M. Mauchly; Pillai's trace. Test of sphericity. Hotelling-Lawley. Principal Component Analysis. Factor Analysis; Procrustes analysis. Canonical correlations analysis. Redundancy analysis. Correspondence analysis. Multidimensional scaling. Principal Co-ordinate Analysis. Methods of ordination. Canonical detrended analysis. Cluster analysis. Distance measures and analysis. Bray Curtis. Mantel. Jaccard's coefficient of similarity. Sorensen's index. TWINSPAN.

8. *Basic tests and procedures:* Descriptive statistics. Exploratory data analyses. Coefficient of variation. Skewness, Kurtosis. Profile analysis. Kolmogorov-Smirnov. Anderson Darling. Lilliefors. Simple confidence intervals and *t*-test. Power. Nonparametric tests. Kruskal-Wallis. Kendall. Freidman. Spearman. Rayleigh. Jonckheere's ordered alternatives test. Dixon's test. Likelihood ratio test. G-test. Bartlett. Shapiro-Wilks. Box's M. Mauchley. D'Agestino-Pearson K^2 test. Gini coefficient. Geary's C coefficient. Wald test statistic. Exact significance probability test. Binomial probability test.

9. *Medical statistics:* Epidemiological methods. Methods of Comparison Procedures; Bland and Altman. Measures of Agreement; Kappa; Cronbach's α. Diagnostic testing; specificity and sensitivity. Repeated measurements; AUC (Area under curve) and summary measures method. Relative risk. Odds ratios. Intention to treat analysis. Meta analysis (which developed in the Health Sciences to a large extent).

10. *Population estimation:* Mark recapture. Transect sampling. Stratified sampling.

11. *Spatial analysis:* Ripley's K Method. Diggle's G and F functions. Spatial pattern analysis. Lattice wombling. Morisita's index. Spatial variograms. Semivariogram. Kriging. Spatial autocorrelation. Moran's I autocorrelation. Mantel correlogram. Circular data involving the

William V-test and the Rayleigh test.

12. *Computer intensive methods:* Bootstrapping. Jackknifing. Simulation. Cross validation analysis. Gibb's sampling. ANOSIM. Hardy-Weinberg proportions. Neural networks. Superposed epoch analysis.

13. *Stochastic processes:* Transition matrices. Random walk. Markov processes. Time series. Periodogram. Cross correlation. Random intervention analysis.

14. *Field specific:* For example, Shannon's index, Shannon-Weaver, Shannon-Wiener and Simpson (measures of diversity). Also area of Genetics involving Phylogenetic Analysis. Parsimony analysis. Consensus tree (evolutionary trees in biology). AMOVA – analysis of molecular variance Linkage analysis; Genotype analysis. Sister-taxon comparison.

REFERENCES

Aitken, L .S., West, S. G., Sechrest, L., & Reno, R. R. (1990). Graduate training in statistics, methodology and measurement in psychology. A survey of PhD programs in North America. *American Psychologist*, 45, 721-734.

Altman, D. G. (1998), Statistical reviewing for medical journals. *Statististics in Medicine,* 17, 2661-2674.

Bangdiwala, S. I., & Muñoz, S. (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.

Bejder, L., Fletcher, D., & Bräger, S. (1998). A method of testing association patterns of social animals. *Animal Behavior,* 56, 719-725.

Bejder, L., Dawson, S. M., & Harraway, J. A. (1999). Responses by Hector's dolphins to boats and swimmers in Porpoise Bay, New Zealand. *Marine Mammal Science,*15 (3), 738-750.

Blumberg, C. J. (2001), Training regular education and special education teachers in the use of research methodology and statistics. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 231-244). Granada: International Association for Statistical Education and International Statistical Institute.

Brogan, D. & Kutner, M. H. (1986). Graduate statistics service courses. *The American Statistician*, 40 , 252-254.

Curtis, D. A., & Harwell, M. (1998). Training doctoral students in educational statistics in the United States: A national survey. *Journal of Statistics Education*, 6(1), http://www.amstat.org/publications/jse/v6n1/curtis.html

Elmore, P. B., & Woehlke, P. L. (1988). Research methods employed in *American Educational Research Journal, Educational Researcher,* and *Review of Educational Research* from 1978 to 1987. Paper presented at the *Annual Meeting of the American Educational Research Association*, New Orleans.

Elmore, P. B., & Woehlke, P. L. (1998). Twenty years of research methods employed in *American Educational Research Journal, Educational Researcher* and *Review of Educational Research*. Paper presented at the *Annual Meeting of the American Educational Research Association*, San Diego.

Hammer, A. S., & Buffington, C. A. (1994). Survey of statistical methods used in the veterinary medical literature. *Journal of the American Veterinary Medical Association,* 205 (2), 344-345.

Maindonald, J. H. (1999). New approaches to using scientific data - statistics, data mining and related technologies in research and research training. Included in Section 2 (*Research Standards: Data Collection and Analysis Standards*) of the discussion document on the web at http://www.anu.edu.au/graduate/papers/gs98_2.html

Manly, B. F. J. (2001). *Statistics for environmental science and management*. London: CRC Press.

Taylor, R. W., Keil, D., Gold, E. J., Williams, S. M., & Goulding, A. (1998). Body mass index, waist girth, and waist-to-hip ratio as indexes of total and regional adiposity in women: Evaluation using receiver operating characteristic curves. *American Journal of Clinical Nutrition,* 67, 44-49.

Taylor, R. W., Jones, I. E., Williams, S. M., & Goulding, A. (In press), Evaluation of waist circumference, waist-to-hip ratio, and the conicity index as screening tools for high trunk fat mass, as measured by dual-energy x-ray absorptiometry, in children aged 3-19 years. *American Journal of Clinical Nutrition.*

John Harraway, Bryan Manly, Hilary Sutherland, & Allan McRae
University of Otago, Department of Mathematics and Statistics
P.O. Box 56, Dunedin, New Zealand
Email: jharraway@maths.otago.ac.nz

Bryan F.J. Manly, Biological & Environmental Statistics
Western EcoSystem Technology
Inc. 217 South 1st Street, Suite 5 Laramie, Wyoming 82070, USA
bmanly@compuserve.com