STATISTICS IN THE WORKPLACE: A SURVEY OF USE BY RECENT GRADUATES WITH HIGHER DEGREES

JOHN A. HARRAWAY
Department of Mathematics and Statistics
University of Otago, New Zealand
jharraway@maths.otago.ac.nz

RICHARD J. BARKER
Department of Mathematics and Statistics
University of Otago, New Zealand
rbarker@maths.otago.ac.nz

ABSTRACT

A postal survey was conducted regarding statistical techniques, research methods and software used in the workplace by 913 graduates with PhD and Masters degrees in the biological sciences, psychology, business, economics, and statistics. The study identified gaps between topics and techniques learned at university and those used in the workplace, and points to deficiencies in statistical preparation for employment. Courses requested include multivariate statistics, generalized linear models, research design and power analysis taught with minimal emphasis on probability and mathematics. Recommendations are presented, such as expanding statistical service courses to eliminate gaps, the development of intensive workshops for postgraduate students and for workplace retraining, or involving staff from other departments to provide context for statistics teaching.

Keywords: *Statistics education research; Survey; Curriculum development; Data specialist major; Workplace needs*

1. INTRODUCTION

Like other academic topics, the teaching of statistics at the university level has to be informed, at least in part, by what graduates will have to do with their acquired statistical knowledge in their respective fields of occupation after graduation. Workplaces have unique demands and some graduates need to learn further at work as part of continued lifelong learning experience (Holmes, 1998). Yet information about workplace demands and about the actual statistical activities which graduates encounter in the workplace can be useful for planning academic curricula that are aligned with primary workplace needs. The present study was designed in light of the limited knowledge in this area.

The MEANS project (Holmes, 1998) surveyed employers and universities and sought opinion on training programmes which involved a major in statistics. However, a majority of students at the university level learn statistics as a service subject (i.e., outside departments of statistics). Indeed Holmes (1998) noted that a further survey would be needed to assess adequately the service teaching of statistics and the subsequent match with employment needs.

Statistics Education Research Journal, 4(2), 43-58, http://www.stat.auckland.ac.nz/serj

One way to characterize the statistical needs of different workplaces is by analysing the nature of statistical techniques used by researchers who operate in different fields. Harraway, Manly, Sutherland and McRae (2001) surveyed 2927 research papers published in 16 high impact journals from botany, ecology, food science, marine science and nutrition during 1999. The analysis of the results established that published research in the different subjects used a wide range of statistical procedures. The procedures used varied both between and within subject specialties. But in addition to knowing what researchers use we need to know what use is made of statistics in the workplace, especially by those who have only taken service courses in statistics at university.

An ability to manage data and use statistics software is also crucial in the workplace. Jolliffe and Rangecroft (1997) reported results from two surveys of statistical software use, one involving academic degree courses with substantial statistical content, another involving statistics consultancy organisations in the United Kingdom. The respondents to the academic survey cited the use of 21 different packages, with Minitab being by far the most popular followed by GLIM, SPSS, Genstat, SAS, S-Plus and Matlab. The respondents to the survey of consultancy firms cited SAS as the most popular package in the field, followed by Minitab, GLIM, and then equal numbers of responses for Genstat, SPSS and Statgraphics. But these surveys need to be augmented as they did not break down usage by speciality nor did they examine techniques used in the workplace.

Finally, information about the uses of statistics in the workplace can also help to design teaching in context, which may be important for success in learning statistics. The Secretary's Commission on Achieving Necessary Skills in the USA (SCANS, 1989) has concluded that an effective way of learning any skill, including quantitative skill, is "in context", placing learning objectives within the real environment rather than insisting that students learn in abstract what they will be expected to apply. We hypothesize that the real workplace environment is different for different specialties which are taught at universities, but this has so far not been studied in detail.

Given the limitations of extant knowledge of the actual usage of statistics in the workplace, this study reports on a survey of university graduates (primarily graduates of Masters and Doctoral programmes) which was designed with three goals in mind. First, we wanted to identify statistical techniques, research methods, and statistical software used in the workplace. Second, we wanted to identify deficiencies in the training received at university in these techniques, methods and software. Finally, a further goal of our study is to summarize similarities and contrasts in statistics training for several different specialties. Our expectation was that such a survey can contribute to the creation of an evidence basis which is needed for informing the design of statistical curricula. In addition, as teachers of statistics to undergraduates and as providers of consulting services in statistics to a wide range of researchers, we hoped that such a survey can also help us find ways to teach statistics in a more meaningful and motivating way.

The paper is organized in four parts. First we describe the target population and the methods employed in surveying our sample, which involved initially graduates in seven specialties subsequently reduced to five, statistics, economics and finance, marketing, the biological sciences and psychology, after deleting nutrition and food science. Next, results of the survey are reported and deficiencies with the teaching of statistics at university in the different subjects are summarized. Finally, the discussion focuses on recommendations for improved teaching in the light of our results and on suggestions for further research.

2. METHODS

2.1. SUBJECTS

The target population for our survey comprised students from the seven New Zealand universities who had completed PhD and Masters degrees in statistics, economics and finance, marketing, the biological sciences, psychology, nutrition and food science in the period 1995-2000. In the case of statistics there were few PhD or Masters graduates in New Zealand, hence it was decided to include those who had completed a Bachelor degree only as well. The period 1995 to 2000 was chosen to ensure most surveyed would have had the opportunity to be in employment while also able to recall what they had been taught at university; their relatively recent graduation should ensure accurate long term recall of topics that they had been taught. Brennan, Lyon, Schomburg and Teichler (1994) comment: "Around two years after graduation, most graduates will be settled into their first proper jobs, transitional shocks will be largely over, and graduates will be more dispassionate about their higher education but will still remember it." A number of those who had completed a Masters degree were currently enrolled for a PhD at a university in New Zealand or overseas, and these students were viewed as being in employment.

Because there were insufficient graduates in our chosen specialties at any one university, and there were doubts about address accuracy that could result in lower response rate, it was decided to approach all New Zealand universities in order to increase the number of potential responders and as a consequence the sample size. Six of the seven New Zealand Universities agreed to post out a questionnaire on our behalf and at our expense to their graduates in the nominated subjects, after we approached the alumni offices at each institution for permission to access the addresses of their graduates. Final permission was in each case granted by the Registrar or Vice Chancellor at each university. Participation by the major New Zealand universities guaranteed coverage of the target population. The seventh university which did not participate has only 3% of total student roll in New Zealand.

Ethical approval for the survey was obtained from the University of Otago Ethics Committee and other universities accepted this. A condition was that all information would be confidential and recorded by identification number only. Each university wrote a covering letter explaining the purpose of our survey and its importance for the training in statistics at each institution. This we felt would improve the response rate as many graduates are loyal to their own institution. But it did complicate the organisation associated with subsequent follow-up of non-respondents as we had to rely on the commitment of the various alumni offices to fit our project around their own work.

2.2. QUESTIONNAIRE CONSTRUCTION

An eight-page questionnaire was developed. General questions on academic background, attitudes to university teaching and the nature of employment were based on the Graduate Opinion and Employers' Survey of the University of Otago. We also designed questions on 46 statistical techniques and research methods taught in statistics programmes. These were compiled from techniques and research methods listed in the papers and reports discussed in the literature cited in the introduction, together with a report by Moore (1997) on studies of the use of statistics in several fields. We also consulted four statisticians and 15 other university teachers from departments that apply statistics or require statistics as part of their major requirements. These teachers were generally happy with the structure and clarity of our questionnaire. The questionnaire was then tested on a group of 20 graduates from the target population. Some minor changes were made as a consequence to remove ambiguities and the average response time of 15 minutes noted. This information was included in a covering letter sent out with the survey document in order to encourage response.

The questionnaire comprised six sections. Sections 1 and 2 collected details of university degrees, major subject and main field of employment. The crucial section 3 aimed to establish the frequency of use in the workplace of the 46 statistical techniques and research methods as well as six computer packages, and whether these techniques, methods and the use of each package had been taught at university. Section 4 investigated the types of statistical activity prevalent in the workplace. Section 5 surveyed course experiences in relation to the statistical methods which the respondents had been taught at university and invited respondents to nominate topics for short retraining courses and workshops that would support their current work. Section 6 asked for written comments, first on the relevance of the statistics training and education received at university for the subsequent employment of each respondent, and second on how statistics training at university could be improved to better prepare graduates for the workplace.

2.3. PROCEDURE

The questionnaire was administered via a mail survey in April-May 2002. All the recommendations made by Edwards, Roberts, Clarke, DiGuiseppi, Pratap, Wentz and Kwan (2002) for increasing response rates to postal questionnaires were used except for incentives. Either national or international post paid reply envelopes were sent with the questionnaire depending on whether the survey was posted to an address within New Zealand or overseas. There were 2758 questionnaires distributed to graduates in our target group and 721 completed questionnaires were returned over the next two months. Non-respondents were followed up by further mailings of the questionnaire between October 2002 and March 2003 after progressively compiling the addresses of the non-respondents from each University. This resulted in a further 256 completed questionnaires. We were not permitted to follow-up by phone, email addresses were not available and we did not have facilities for on-line response. A total of 977 responses were received and 353 questionnaires were returned undelivered or found to have been sent to graduates not in the target population. The effective response rate was therefore 39%.

A preliminary comparison of responses from the initial and follow-up phases showed no important differences in the calculated proportions, and for this reason all 977 completed questionnaires were used initially to produce results. All 977 were manually checked for incompleteness and inconsistency. All records entered in the database were checked and mistakes corrected. Data were entered in an excel spreadsheet but SPSS was used for analysis.

3. RESULTS

Of the 977 respondents in the survey, 93% were either employed or engaged in PhD study having already completed a Masters degree within our target time frame of 1995 to 2000, 5% were unemployed having been previously employed, and 2% only were unemployed. Opinions expressed therefore represent the views of employed graduates with higher degrees in subjects that use statistics.

There were 172 PhD respondents, 759 Masters or equivalent respondents and 46 Bachelor degree respondents who had completed their degree in statistics and who were added to the statistics PhDs and Masters respondents. Nutrition (34 respondents) and food science (30 respondents) were omitted from further analysis because these small numbers were insufficient for reliable conclusions. As a consequence only 913 respondents are included in the results reported for the five remaining specialties.

3.1. TYPE OF STATISTICAL ACTIVITY IN THE WORKPLACE

The dominant statistical activities carried out in the workplace by the respondents in the five specialties are listed in Table 1. Several of the activities are performed by many of the respondents, hence the proportions of the activities performed in each specialty can sum to more than one. The most common activity across all specialties involves the carrying out of data analyses. Next in importance is reading the results of published research. Only 11% of respondents report they have no need of statistics. Review of proportions which exceed one half in each speciality in Table 1 shows that biological sciences and psychology graduates design studies, carry out data analyses and have a need to read published work. Graduates in all other specialities also carry out data analyses. But, in addition, the statisticians are involved with report writing, the economics/finance graduates carry out, as expected, financial analyses and read published work while the marketing graduates, as expected, carry out market research and write reports.

Table 1. Numbers and proportions for statistical activities in the workplace

	Subject Specialty				
Activity	Statistics (119)	Econ/Finance (85)	Marketing (82)	Bio Science (344)	Psych (283)
Own data analyses	80 (0.67)	66 (0.78)	47 (0.57)	265 (0.77)	190 (0.67)
Reading published research	38 (0.32)	45 (0.53)	34 (0.41)	213 (0.62)	199 (0.70)
Report writing	49 (0.41)	36 (0.42)	45 (0.55)	160 (0.47)	118 (0.42)
Designing Studies	24 (0.20)	23 (0.27)	27 (0.33)	183 (0.53)	144 (0.51)
Financial analysis	38 (0.32)	52 (0.61)	27 (0.33)	28 (0.08)	28 (0.10)
Understanding a consultant	13 (0.11)	8 (0.09)	6 (0.07)	80 (0.23)	32 (0.11)
Market research	19 (0.16)	11 (0.13)	54 (0.66)	12 (0.03)	22 (0.08)
Quality control	13 (0.11)	5 (0.06)	7 (0.09)	30 (0.09)	14 (0.05)
None needed	18 (0.15)	4 (0.05)	5 (0.06)	42 (0.12)	29 (0.10)

3.2. STATISTICAL TECHNIQUES USED IN THE WORKPLACE

The 46 statistical techniques and research methods nominated in the questionnaire are listed in Table 2. For the five specialties in the study, numbers and proportions of graduates who use the techniques and methods are listed. To interpret pattern in the responses we define a technique with proportional use > 0.35 as frequently used, a technique with proportional use from 0.10 to 0.35 as moderately used and a technique with proportional use < 0.10 as seldom used.

Combining responses across the five specialties identifies frequent use of graphical procedures, basic tests, analysis of variance and simple linear regression; moderate use of contrasts, multiple regression, survey design, nonlinear and logistic regression, factorial and repeated measures designs, multivariate analysis of variance, principal components and statistics theory; some of the techniques seldom used are ordinal and nominal regression, survival analysis, cross over designs, path analysis, confirmatory factor analysis, correspondence analysis, Bayesian statistics, randomisation testing, state space and lag models, adaptive sampling and bioinformatics.

Table 2. Numbers and proportions using the techniques and methods in the workplace

	Subject Specialty				
Technique or method	Statistics (119)	Econ/Finance (85)	Marketing (82)	Bio Science (344)	Psychology (283)
1. Graphs	84 (.71)	70 (.82)	49 (.60)	247 (.72)	196 (.69)
2. Basic tests, t, χ^2	50 (.42)	43 (.51)	24 (.29)	201 (.58)	145 (.51)
3. ANOVA	34 (.29)	28 (.33)	13 (.16)	178 (.52)	115 (.41)
4. Contrasts	22 (.18)	10 (.12)	15 (.18)	108 (.31)	99 (.35)
5. Simple reg.	63 (.53)	55 (.65)	28 (.34)	182 (.53)	127 (.45)
6. Multiple reg.	44 (.37)	41 (.48)	12 (.15)	78 (.23)	71 (.25)
7. Nonlinear reg.	22 (.18)	19 (.22)	5 (.06)	64 (.19)	33 (.12)
8. Nonparametric reg.	7 (.06)	11 (.13)	3 (.04)	49 (.14)	29 (.10)
9. Mixed models	14 (.12)	10 (.12)	2 (.02)	43 (.13)	19 (.07)
10. Logistic reg.	30 (.25)	13 (.15)	6 (.07)	53 (.15)	30 (.11)
11. Ordinal/Nom reg.	8 (.07)	6 (.07)	3 (.04)	17 (.05)	10 (.04)
12. Loglinear models	17 (.14)	9 (.11)	3 (.04)	37 (.11)	8 (.03)
13. Survival analysis	10 (.08)	4 (.05)	1 (.01)	18 (.05)	8 (.03)
14. Factorial designs	11 (.09)	2 (.02)	7 (.09)	80 (.23)	58 (.20)
15. Blocking	10 (.08)	0 (.00)	3 (.04)	70 (.20)	23 (.08)
16. Repeated measures	10 (.08)	0 (.00)	3 (.04)	75 (.22)	56 (.20)
17. Crossover designs	7 (.06)	1 (.01)	1 (.01)	22 (.06)	23 (.08)
18. Clinical trials	7 (.06)	2 (.02)	0 (.00)	13 (.04)	42 (.15)
19. MANOVA	6 (.05)	4 (.05)	8 (.10)	74 (.22)	52 (.18)
20. Principal compts	16 (.13)	5 (.06)	10 (.12)	72 (.21)	31 (.11)
21. Factor analysis	9 (.08)	3 (.04)	15 (.18)	16 (.05)	33 (.12)
22. Path analysis	7 (.06)	3 (.04)	6 (.07)	13 (.04)	22 (.08)
23. Confirmatory F.A.	5 (.04)	1 (.01)	6 (.07)	7 (.02)	23 (.08)
24. Cluster analysis	15 (.13)	3 (.04)	14 (.17)	63 (.18)	16 (.06)
25. Correspondence A.	3 (.03)	0 (.00)	2 (.02)	32 (.09)	3 (.01)
26. Discriminant A.	10 (.08)	0 (.00)	5 (.06)	40 (.12)	15 (.05)
27. Scaling/Ordination	3 (.03)	0 (.00)	3 (.04)	42 (.12)	7 (.02)
28. Canonical Corr.A.	3 (.03)	0 (.00)	2 (.02)	35 (.10)	3 (.01)
29. Statistics theory	27 (.23)	24 (.28)	9 (.11)	55 (.16)	34 (.12)
30. Estimation theory	21 (.18)	17 (.20)	5 (.06)	39 (.11)	8 (.03)
31. Bayesian statistics	11 (.09)	2 (.02)	3 (.04)	16 (.05)	7 (.02)
32. Jackknifing	5 (.04)	1 (.01)	1 (.01)	19 (.06)	3 (.01)
33. Simulation	26 (.22)	12 (.14)	1 (.01)	42 (.12)	0 (.00)
34. Randomisation test	5 (.04)	3 (.04)	0 (.00)	31 (.09)	7 (.02)
35. ARMA	27 (.23)	23 (.25)	5 (.06)	7 (.02)	7 (.02)
36. Forecasting	25 (.21)	28 (.33)	19 (.23)	3 (.01)	7 (.02)
37. Markov chains	7 (.06)	6 (.07)	0 (.00)	11 (.03)	0 (.00)
38. State-space models	2 (.02)	5 (.06)	0 (.00)	3 (.01)	0 (.00)
39. Lag models	1 (.01)	8 (.09)	0 (.00)	3 (.01)	3 (.01)
40. Mark-recapture	6 (.03)	0 (.00)	0 (.00)	54 (.16)	0 (.00)
41. Survey design	27 (.23)	8 (.09)	28 (.34)	95 (.28)	64 (.23)
42. Adaptive sampling	2 (.02)	0 (.00)	3 (.04)	14 (.04)	3 (.01)
43. Power analysis	13 (.11)	2 (.02)	2 (.02)	62 (.18)	34 (.12)
44. Meta analysis	5 (.04)	2 (.02)	1 (.01)	14 (.04)	27 (.10)
45. Data mining	15 (.13)	6 (.07)	8 (.10)	13 (.04)	7 (.02)
46. Bioinformatics	2 (.02)	0 (.00)	0 (.00)	20 (.06)	3 (.01)

There are differences in the proportions across the subject specialties, however. A hierarchical average linkage clustering based on the proportions identifies three clusters. The biological sciences and psychology form the first cluster, statistics and economics/finance form the second cluster, while marketing has its own unique pattern of statistical activity in the third cluster.

In addition to the frequently used techniques, the biological sciences and psychology graduates together display moderate use of nonlinear, nonparametric and logistic regression, factorial and repeated measures designs, multivariate analysis of variance and principal components, survey design and power analysis. The biological sciences graduates further show moderate use of mixed models, loglinear models, cluster analysis, discrimination, ordination, canonical correlation analysis, estimation theory, mark recapture and simulation while psychology graduates show moderate use of clinical trials, factor analysis and meta analysis.

The statistics and economics/finance graduates together display moderate use of mixed models, nonlinear and logistic regression, statistics theory and estimation theory, auto regressive moving averages, forecasting and simulation, while statistics graduates in addition have moderate use for principal components, cluster analysis, data mining, survey design and power analysis.

Marketing graduates have moderate use for multivariate analysis of variance, principal components, factor analysis, cluster analysis, forecasting, survey design and data mining in addition to the frequently used basic procedures.

3.3. TECHNIQUES IN USE BUT NOT TAUGHT AT UNIVERSITY

Those surveyed were asked whether statistical techniques and research methods used in their employment had been included in their training at university. Responses summarized in Table 3 give, for each subject specialty, the numbers and proportions of graduates who use each technique or method but who have received *no* instruction in that technique or method. Inspection of the numbers greater than 46 (5% of 913 responders) when summing across the five specialties indicates training deficiencies overall in multiple, nonlinear, nonparametric and logistic regression, mixed models, multivariate analysis of variance, principal component analysis, cluster analysis, discriminant analysis, simulation, survey design and power analysis.

Table 3. Numbers and proportions not taught but using techniques or methods in the workplace

	Subject Specialty				
Technique or method	Statistics (119)	Econ/ Finance (85)	Marketing (82)	Bio Science (344)	Psychology (283)
1. Graphs	3 (.03)	2 (.02)	0 (.00)	19 (.06)	6 (.02)
2. Basic tests, t, χ^2	3 (.03)	0 (.00)	0 (.00)	13 (.04)	5 (.02)
3. ANOVA	2 (.02)	0 (.00)	0 (.00)	21 (.06)	6 (.02)
4. Contrasts	2 (.02)	0 (.00)	1 (.01)	19 (.06)	8 (.03)
5. Simple regression	2 (.02)	1 (.01)	0 (.00)	18 (.05)	11 (.04)
6. Multiple reg	4 (.03)	1 (.01)	0 (.00)	29 (.08)	21 (.07)
7. Nonlinear reg	8 (.07)	3 (.04)	3 (.04)	34 (.10)	13 (.05)
8. Nonparametric reg	1 (.01)	4 (.05)	3 (.04)	24 (.07)	16 (.06)
9. Mixed models	8 (.07)	3 (.04)	2 (.02)	22 (.06)	11 (.04)
10. Logistic reg	8 (.07)	2 (.02)	3 (.04)	28 (.08)	18 (.06)
11. Ordinal/Nom reg	7 (.06)	0 (.00)	2 (.02)	11 (.03)	6 (.02)

12. Loglinear models	5 (.04)	1 (.01)	1 (.01)	22 (.06)	6 (.02)
13. Survival analysis	4 (.03)	2 (.02)	0 (.00)	13 (.04)	7 (.02)
14. Factorial designs	0 (.00)	0 (.00)	2 (.02)	15 (.04)	4 (.01)
15. Blocking	0 (.00)	0 (.00)	2 (.02)	20 (.06)	1 (.00)
16. Repeated measures	2 (.02)	0 (.00)	2 (.02)	22 (.06)	6 (.02)
17. Crossover designs	3 (.03)	1 (.01)	0 (.00)	6 (.02)	3 (.01)
18. Clinical trials	5 (.04)	2 (.02)	0 (.00)	5 (.01)	8 (.03)
19. MANOVA	2 (.02)	1 (.01)	1 (.01)	41 (.12)	17 (.06)
20. Principal compts	4 (.03)	2 (.02)	1 (.01)	45 (.13)	18 (.06)
21. Factor analysis	4 (.03)	2 (.02)	1 (.01)	10 (.03)	17 (.06)
22. Path analysis	5 (.04)	2 (.02)	5 (.06)	10 (.03)	19 (.07)
23. Confirmatory F.A.	3 (.03)	1 (.01)	3 (.04)	5 (.01)	17 (.06)
24. Cluster analysis	3 (.03)	2 (.02)	2 (.02)	41 (.12)	11 (.04)
25. Correspondence A.	1 (.01)	0 (.00)	2 (.02)	22 (.06)	2 (.01)
26. Discriminant A.	3 (.03)	0 (.00)	3 (.04)	30 (.09)	10 (.04)
27. Scaling/Ordination	1 (.01)	0 (.00)	2 (.02)	29 (.08)	3 (.01)
28. Canonical Corr A.	2 (.02)	0 (.00)	2 (.02)	29 (.08)	2 (.01)
29. Statistics theory	0 (.00)	1 (.01)	1 (.01)	16 (.05)	4 (.01)
30. Estimation theory	3 (.03)	1 (.01)	2 (.02)	27 (.08)	3 (.01)
31. Bayesian statistics	3 (.03)	0 (.00)	2 (.02)	14 (.04)	4 (.01)
32. Jackknifing	2 (.02)	0 (.00)	1 (.01)	13 (.04)	2 (.01)
33. Simulation	7 (.06)	8 (.09)	1 (.01)	36 (.10)	0 (.00)
34. Randomisation test	2 (.02)	1 (.01)	1 (.01)	21 (.06)	3 (.01)
35. ARMA	1 (.01)	1 (.01)	0 (.00)	1 (.00)	3 (.01)
36. Forecasting	2 (.02)	5 (.06)	6 (.07)	4 (.01)	2 (.01)
37. Markov chains	0 (.00)	2 (.02)	0 (.00)	9 (.03)	1 (.00)
38. State-space models	0 (.00)	2 (.02)	0 (.00)	4 (.01)	1 (.00)
39. Lag models	0 (.00)	1 (.01)	0 (.00)	2 (.01)	3 (.01)
40. Mark-recapture	2 (.02)	0 (.00)	0 (.00)	29 (.08)	0 (.00)
41. Survey design	4 (.03)	7 (.08)	3 (.04)	21 (.04)	11 (.04)
42. Adaptive sampling	1 (.01)	0 (.00)	2 (.02)	6 (.02)	1 (.00)
43. Power analysis	5 (.04)	2 (.02)	2 (.02)	30 (.09)	17 (.06)
44. Meta analysis	5 (.04)	2 (.02)	1 (.01)	9 (.03)	22 (.08)
45. Data mining	14 (.12)	4 (.05)	7 (.09)	12 (.03)	5 (.02)
46. Bioinformatics	2 (.02)	0 (.00)	0 (.00)	17 (.05)	3 (.01)

Deficiencies within each specialty are also apparent. Proportions greater than 0.10 (reflecting the smaller numbers in each specialty) indicate some lack of training in the biological sciences for nonlinear regression, multivariate analysis of variance, principal component analysis, cluster analysis and simulation. There is a lack of training among statistics graduates in data mining. In addition, 0.07 to 0.09 of the graduates in the biological sciences have received no training in logistic and multiple regression, the multivariate techniques of ordination, discrimination and canonical correlations, power analysis and mark-recapture. At an equivalent level there are gaps in training in multiple regression, path analysis and meta analysis for psychology graduates; mixed models, nonlinear and logistic regression for statistics graduates; survey design and simulation for economics/finance graduates; forecasting and data mining for marketing graduates.

3.4. COMPUTING IN THE WORKPLACE

Much statistical activity involves use of computer software for both data management and statistical analyses. Table 4 lists Excel, Access, SPSS, SAS, Minitab and S+/R as packages used by over 5% of the sample. Several other packages are used to a lesser extent as follows: Statistica (34 users); Shazam (33); Systat (29); Matlab (22); Datadesk (16); Statview (16); Stata (15); Sigmastat (13); Genstat (10).

The results in Table 4 show widespread use of Excel. Access is frequently used overall but to a lesser degree by graduates in the biological sciences and psychology. SPSS is frequently (0.37) used by psychology graduates and moderately used by marketing and biological sciences graduates. SAS on the other hand is mainly used by statistics and economics/finance graduates. Minitab is seldom used in the workplace except by biological sciences graduates where usage is moderate (0.12). S+ and R are seldom used except by statistics graduates. The full implications of the free to use package, R, are only now being realised, and this is likely to result in greater use of R in the next few years.

Table 4. Numbers and proportions using statistical packages in workplace

	Subject Speciality				
Package	Statistics (119)	Econ/Finance (85)	Marketing (82)	Bio Science (344)	Psychology (283)
1. Excel	91 (0.76)	72 (0.85)	62 (0.76)	279 (0.81)	178 (0.63)
2. Access	49 (0.41)	33 (0.39)	43 (0.52)	111 (0.32)	66 (0.23)
3. SPSS	10 (0.08)	5 (0.06)	14 (0.17)	63 (0.18)	106 (0.37)
4. SAS	41 (0.34)	18 (0.21)	2 (0.02)	51 (0.15)	25 (0.09)
5. Minitab	10 (0.08)	4 (0.05)	0 (0.00)	42 (0.12)	2 (0.01)
6. S+/R	22 (0.18)	4 (0.05)	2 (0.02)	23 (0.07)	2 (0.01)

Table 5 reports the numbers and proportions of graduates using each package with no prior instruction, and points to gaps in statistical package training that have implications for needed changes in training in order to cover these gaps. Overall, 45% of those using Excel and 30% of those using Access have had no prior instruction. There are deficiencies in the teaching of SPSS to psychology and biological sciences graduates and deficiencies in the teaching of SAS to statistics and economics/finance graduates.

Table 5. Numbers and proportions not taught but using packages in workplace

	Subject Speciality				
Package	Statistics (119)	Econ/Finance (85)	Marketing (82)	Bio Science (344)	Psychology (283)
1. Excel	50 (0.42)	40 (0.47)	22 (0.27)	161 (0.47)	133 (0.47)
2. Access	43 (0.36)	29 (0.34)	33 (0.40)	105 (0.31)	63 (0.22)
3. SPSS	8 (0.07)	4 (0.05)	1 (0.01)	57 (0.17)	56 (0.20)
4. SAS	15 (0.13)	16 (0.19)	0 (0.00)	32 (0.09)	16 (0.06)
5. Minitab	1 (0.01)	4 (0.05)	0 (0.00)	15 (0.04)	2 (0.01)
6. S+/R	4 (0.03)	4 (0.05)	1 (0.01)	16 (0.05)	2 (0.01)

3.5. RETRAINING

Respondents were invited to nominate courses or intensive workshops they thought would help reinforce the statistics techniques and research methods needed for their employment. Some respondents expressed an interest in attending retraining courses but did not list topics. But for the 396 respondents who listed courses, Table 6 summarises the preferred workshop requests in order of most mentions. Some respondents made multiple requests. The three most requested topics are multivariate methods, generalized linear models and survey design including power analysis.

Table 6. Preferred workshops with numbers and proportions of requests

Workshop Topic	Number of respondent requests	Proportion of respondent requests
Multivariate methods	99	0.25
Regression/generalized linear models	88	0.22
Survey design and power analysis	84	0.21
Statistical software developments	75	0.19
Introductory statistical methods	75	0.19
Stochastic processes	39	0.10
Psychology topics/clinical trials	35	0.09
Forecasting/Time Series	32	0.08
Theory including Bayesian methods	20	0.05
Experimental designs	19	0.05
Marketing related topics	16	0.04
New methods e.g. meta analysis, data mining, bioinformatics	12	0.03
Computer intensive statistics	8	0.02

Table 7 summarizes the three most cited topics for retraining, by specialty. The table also records the number of times each course was nominated in that subject area. Many of the respondents noted that these topics could be taught alternatively in the undergraduate programmes at university as part of the requirements for majors.

Table 7. Preferred courses requested by specialty

Statistics (119)	Regression/Generalized linear models (14 nominations) Forecasting/Time Series (12) Statistical software developments (8)
Econ/Finance (85)	Forecasting/Time Series (20) Regression/Generalized linear models (13) Multivariate methods (7)
Marketing (82)	Marketing related topics (10) Regression/Generalized linear models (7) Multivariate methods (6)
Bio Science (344)	Survey design and power analysis (55) Multivariate methods (48) Regression/Generalized linear models (31)
Psychology (283)	Multivariate methods (31) Statistical software developments (28) Psychology topics/clinical trials (27)

The course nominations match moderately or frequently used techniques as listed in Table 2 or the training deficiencies noted from Table 3. For example, for the graduates in the biological sciences, training deficiencies were identified initially in aspects of regression, multivariate methods, data mining, power analysis and mark recapture; these topics bear a close resemblance to the three most frequently nominated requests, survey design and power analysis (55), multivariate methods (48) and regression/generalized linear models (31).

3.6. COURSE APPRECIATION

Respondents were invited to provide free-form comments on the relevance of their statistical training for their employment and to make recommendations for improving university teaching in statistics to better prepare them for their workplaces. These responses have not been analysed quantitatively, instead being used to provide anecdotal examples to help gain deeper appreciation for the linkage (or lack thereof) between studying statistics and research methods at university and the actual demands of the workplace with consequential training gaps.

One theme raised in the written comments about the work environment cited the importance of statistics which had in many cases not been realised while students were studying at university. One graduate wrote about statistics:

> *I would be impotent without it. Keep statistics compulsory – I never would have done it otherwise – and it is now the most incisive tool in my arsenal.*

A zoology and ecology MSc graduate made the following comment along similar lines:

> *I am not currently employed in the field for which I studied but hope to be eventually. There is little requirement for stats understanding in my local government position but I know my knowledge of statistics will be insufficient if/when I begin an ecological career, which was the focus of my study. I did not understand the importance of stats in ecology when I began my BSc so only took one first year course, as that was all that was compulsory. Biological Science students should be forced to do at least second year stats courses. When I came to do honours research in my 4th year I found my statistical understanding to be poor and relied heavily on outside help and did not really understand the computations behind the statistical software I was using. I introduced myself to SPSS software and quoted results of the analyses in a paper that was eventually published but would have benefited greatly from more guidance on the results' interpretation.*

It was clear from the comments that only after finishing their study do many students come to realise that they need more statistics. A further common view expressed was that the disciplines, especially those in science, should promote statistics more. Many respondents thought that more statistics at each level of university study would have been beneficial. By the time they begin their research a lot of what they have learnt has been forgotten. A few said they had found it hard to fit extra statistics courses into their major study and therefore statistics courses should have been more a part of their major requirements. One graduate commented:

> *Stats teaching needs to be more integrated with the discipline in which it is being applied so that students think of it as another important tool within that discipline rather than perceiving it as scary numerical stuff outside the discipline they're studying and thus able to be avoided.*

Another advocated teaching in context by suggesting that:

> *Courses targeting analyses commonly used in a specific discipline (e.g. psychology, marketing) would increase student interest.*

Respondents also suggested that there should be a more co-ordinated approach between departments with greater collaboration between major research disciplines and departments of mathematics and statistics in order to provide courses that suit. Many psychology students in particular wanted statistics subjects to be relevant to their work. Some thought that their statistics training was too general and recommended that statistics should be taught within their discipline so that topics would be applicable. One graduate maintained that courses on statistics taught within psychology departments were "*excellent*". Yet on the other hand, another two graduates who obviously had had different experiences supported the view that statistics departments should teach statistics by stating:

> *Psychology and education departments should be forced to have the statistics department teach their students – as psychology and education departments do a bad job.*
>
> *Statisticians teaching stats, not biologists teaching stats.*

These diverging views need further research to ascertain the best location for the teaching of statistics and to identify how to incorporate statistics in each of the disciplines which use statistics.

4. DISCUSSION

This study was initiated because we sought ways to improve training in statistics to better prepare graduates for employment. We surveyed graduates, currently employed, who had completed Masters and PhD degrees mainly in a range of specialties that have statistics pre-requisites at universities. Holmes (1998) only surveyed employers and universities and concentrated on majors in statistics. We have placed the emphasis on statistics teaching in service courses.

Our survey has identified statistical activity in the workplace and shortcomings in the statistical training received for this work. The respondents in each specialty surveyed have nominated key courses for retraining or, equivalently, key statistical techniques and research methods that should be included in coursework at university.

The SCANS report (1989) identified skills that those who have completed either high school or university should master if they are to be effective in the workplace. Table 1 summarises the main statistical activities in the workplace and this shows that carrying out data analyses, reading published research and report writing are important for 71%, 58% and 45% respectively of those surveyed. These skills relate to understanding mathematics, reading and writing which are emphasised in the SCANS report.

When they surveyed psychology departments, Aitken, West, Sechrest and Reno (1990) discovered that PhD graduates were able to handle traditional statistics techniques but were deficient in newer procedures. We show in Table 2 that the elementary techniques and topics are used frequently in the workplace and are well covered in the training of psychology students according to the numbers in Table 3. But our study has, in addition, identified three more advanced retraining workshops important for psychology graduates in the workplace. This is in agreement with the opinions of Howard, Pion, Gottfredson, Flattau, Oskamp, Pfafflin, Bray and Burstein (1986) who claim that many psychology graduates obtain employment but are poorly prepared for some aspects of their work.

Higgins (1999) argues for a concentration on the service aspects of statistics teaching. He suggests that planning and management of scientific studies and communication skills are important for graduates. This is consistent with the evidence in Table 1 that designing studies, report writing and reading published work are statistical activities occurring frequently in the workplace. What emerges is a need for an approach where service courses on the above areas and on statistical methodology are taught in a way that reduces dependence on mathematics and probability. The focused workshops nominated in our survey are consistent with the spirit of this approach.

Both Cobb (in Higgins, 1999) who describes a liberal arts major in statistics at Mt Holyoke College, and Moore (2001), also support this approach. Moore remarks:

> While the discipline of statistics is healthy, its place in academe is not. Our future there depends strongly on achieving a more prominent place in undergraduate education beyond the first methods course. To this end we must offer undergraduate programs that are popular with students … the primary intent of such programs cannot be to prepare students for graduate study in statistics, but to equip them for employment with a bachelor's degree or for further study in a wide variety of areas. Finally, success requires greater co-operation between statistics and other disciplines, ….

4.1. STUDY LIMITATIONS

This survey, involving seven subject specialties, was ambitious yet not wholly successful. The overall response rate was 39% which is reasonable for a mail survey but not as good as one would hope. There were small response numbers in some of the categories in our target population. This resulted in omitting both nutrition and food science with 34 and 30 respondents respectively from our analyses.

The sample sizes in the biological sciences and psychology on the other hand were satisfactory adding strength to conclusions reached for these subjects. The sample sizes in the business areas were not as good but we felt it was reasonable to keep these subjects in the analyses.

There were few PhD and Masters graduates in statistics at the six universities surveyed. For this reason we also mailed our questionnaire to graduates who had completed Bachelor level degrees only in statistics. At one university 38 students who had completed a Bachelor degree had also completed a second Bachelor degree in economics/finance. For our analyses these graduates were not viewed as economics/finance graduates because they only had Bachelor degrees in these subjects. One consequence of this policy on classification of respondents is that the techniques used by the statistics graduates in the workplace will be biased towards the statistics techniques used by Masters or PhD graduates in economics/finance. Care is therefore needed when generalizing the results for statistics graduates. It is interesting that the cluster analysis reported in Section 3.2 showed similarity between statistics and economics/finance.

There is a question about the generalizing of our results to other countries. Even though 100 of our respondents (about 10%) were found to be working overseas at the time of our survey, the academic education system, including in statistics service courses, is not the same in all countries and therefore conclusions should be generalized with caution to countries whose academic education system differs from that in New Zealand. It would be interesting to carry out a similar survey in other countries.

It is also possible that our survey time frame 1995 to 2000 has missed recent developments in computing like the expanding popularity of the free-to-use statistical package R. Similarly, statistical procedures and research methods used in recent advances in the area of genetics will have been missed in our investigation.

4.2. RECOMMENDATIONS

About 89% (815 respondents) of the graduates surveyed in this study used some aspect of statistics in their employment. To overcome the deficiencies in statistical training identified for almost half of them, we make five recommendations. These relate to the content of advanced university courses at either the undergraduate or the postgraduate level, instructional style, and the content of retraining statistics modules for lifelong learning while in employment.

Recommendation 1: expand statistical methods taught in the biological sciences to include advanced topics in regression and generalized linear modelling, multivariate methods, power analysis, mark-recapture and data mining; expand the statistical content of psychology to include regression and generalized linear modelling, survey design, multivariate methods especially factor analysis, the design of clinical trials and meta analysis; expand the statistical content of economics, finance and marketing to include regression modelling and multivariate methods. These advanced topics should follow from introductory first year statistical methods courses but should be taught without mathematics and probability prerequisites in order to make them accessible to the graduates in the specialties surveyed.

Recommendation 2: university statistics departments should develop short courses or intensive workshops for postgraduate students. Such courses will also serve as retraining modules for those in employment. Recently we had success with a 3-day workshop on multivariate statistics techniques for 25 ecologists enrolled for PhD or Masters degrees. Data sets relevant to the work of the participants were used, some of these being generated by those attending. The workshop was therefore taught in the context of ecology. It was successful and requests have been received to mount a second workshop on regression and generalized linear modelling as well as repeating the first.

Recommendation 3: seek support from staff in departments teaching the specialties we have surveyed. They should be asked to encourage their senior undergraduate students to enrol for more undergraduate statistics courses. The staff in these departments could be used as guest or visiting lecturers. They could also provide interesting data sets which would place the statistics in the context of the specialty. It could even be possible to have statisticians located in the other departments. A consequence of these approaches could be the inclusion of critical appraisal of some recently published research articles and the development of group project work using unstructured real data generated from consulting or work placement. These instructional styles should equip students with the essential skills for the workplace identified by the SCANS report (1989).

Recommendation 4: university statistics or mathematics departments should investigate the development of a data specialist undergraduate major along the lines of Higgins (1999). This could be studied in conjunction with a major in another specialty and employment prospects for graduates with this type of qualification would be enhanced.

Recommendation 5: check carefully the statistical software that is used when teaching students in the different specialties surveyed in order to meet the deficiencies noted in Section 3.4.

These five recommendations go a long way towards meeting the views expressed by many respondents who advocated that statistics must be part of their major requirement. This study provides compelling evidence in support of teaching more statistics while paying close attention to workplace needs. The support is coming not from statisticians but from recent PhD and Masters graduates, now in the workplace, who have majored in disciplines other than statistics. Conveying these attitudes to students studying at university can only encourage them to include as much statistics in their courses as possible.

4.3. FUTURE RESEARCH

The limitations caused by low response rates in nutrition and food science could be addressed by raising response rates in these areas. The use of on-line replies and phone interviews could increase sample size. There would also be benefit in focusing on statistics graduates in a further survey. The list of techniques and research methods used for the specialties in our survey could be adapted for a statistics graduate survey by including more specialised statistical procedures.

Another survey could assess the relative importance of techniques and the difficulties students encounter when they have to cope with statistics in the workplace.

Reasons why students involved in research at universities allow gaps to develop in their statistics knowledge should be investigated given the overwhelming need for extended training which we have identified. We propose therefore a survey of research students at university together with an analysis of study design, research methods and statistical procedures, if any, used in their theses. The thesis results could be cross referenced to the student opinion. We are currently attempting this exercise ourselves. Use of on-line response is giving fuller information than that obtained on written response forms. We are attempting to establish at what stage statistics help was sought, for instance at the early planning stage of the study or when the data were to be analysed.

ACKNOWLEDGEMENTS

This work was supported by a Research into University Teaching Grant from the University of Otago and a grant from the Campbell bequest to the New Zealand Statistical Association. We are grateful for this support. We also thank the four anonymous referees whose extensive comments have significantly improved this paper.

REFERENCES

Aiken, L.S., West, S. G., Sechrest, L., & Reno, R. R. (1990). Graduate Training in Statistics, Methodology and Measurement in Psychology. *American Psychologist, 45*, 721-734.

Brennan, J., Lyon, E. S., Schomburg, H., & Teichler, U. (1994). The experiences and views of graduates messages from recent surveys. *Higher Education Management*, *6*, 275-304.

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., & Kwan, I. (2002). Increasing response rates to postal questionnaires: systematic review. *British Medical Journal*, *324*, 1183-1185.

Harraway, J. A., Manly, B. F. J., Sutherland, H., & McRae, A. (2001). Meeting the statistical needs of researchers in the biological and health sciences. In C. Batanero (Ed.), *Training Researchers in The Use of Statistics* (pp. 177-195). Voorburg, The

Netherlands: International Association for Statistical Education and International Statistical Institute.

Higgins, J. J. (1999). Nonmathematical statistics: a new direction for the undergraduate discipline. *The American Statistician*, *53*(1), 1-6.

Holmes, P. (1998) Matching education, assessment and employment needs in statistics. *Report on MEANS project.* [Online: http://science.ntu.ac.uk/rsscse/means/means.html]

Howard, A., Pion, G. M., Gottfredson, G. D., Flattau, P. E., Oskamp, S., Pfafflin, S. M., Bray, D. W., & Burstein, A.G. (1986). The changing face of American Psychology: A report from the committee on employment and human resources. *American Psychologist*, *41*, 1311-1327.

Jolliffe, F., & Rangecroft, M. (1997). Statistical software use by student and consultant statisticians in the UK-a comparative study. [Online: http://www.stats.gla.ac.uk/cti/activities/reviews/97_08/software_su]

Moore, D. S. (1997). Bayes for beginners? Some reasons to hesitate. *The American Statistician*, *51*, 254-261.

Moore, D. S. (2001). Undergraduate programs and the future of academic statistics. *The American Statistician*, *55*(1), 1-6.

Secretary’s Commission on Achieving Necessary Skills (SCANS) (1989). US Department of Labor, Employment and Training Administration [Online: http://www.tvi.cc.nm.us/html/scans.html]

J. A. HARRAWAY
Department of Mathematics and Statistics
University of Otago
PO Box 56, Dunedin
New Zealand