FACTORS AFFECTING PERFORMANCE IN A UNIVERSITY SERVICE COURSE ON BIOSTATISTICS: AN UPDATE ®

John A. Harraway
University of Otago
New Zealand

About 800 students each year enrol in a subject Introduction to Biostatistics at the University of Otago. It is a compulsory subject for students applying to enter the health sciences professional courses. At school there are two subjects, mathematics with calculus and mathematics with statistics, with many students studying only one of these the year before university. There is debate about which one best prepares students for gaining the high marks in biostatistics necessary for entry to the competitive professional health sciences programmes. The school syllabus in mathematics with statistics is first compared with that in Introduction to Biostatistics. Results from the analysis of marks achieved in biostatistics are reported. The fitted regression models show prior knowledge of statistics from the school subject has no effect on performance in biostatistics, that there is no gender effect and that prior knowledge of calculus may be beneficial. Reasons for these results are discussed and proposals made to improve the presentation of statistics to students of the health and biological sciences.

INTRODUCTION AND COURSE CONTENT

In the final year at New Zealand high schools there is a national bursary examination. The majority of students study five subjects and often these include either mathematics with calculus or mathematics with statistics or both. It is felt that an introduction to statistics from bursary the year before will improve performance in statistics courses at university and enhance the chance of entry to health sciences programmes.

The school bursary subject mathematics with statistics is taught over the year. The syllabus (Barrett, 1997) includes (i) measures of centre and variability, graphical procedures, time series analysis using moving averages; (ii) probability including mutually exclusive and independent events, conditional probability, tree diagrams, discrete random variables, binomial, Poisson and normal distributions; (iii) sampling distributions of the mean and proportion, confidence intervals for means, proportions and differences between means; (iv) an independent investigation from the planning and sampling phases to the analysis of data and discussion of the findings; (v) mathematical topics accounting for half the subject and including combinations, the binomial theorem, introductory calculus, graphs, solutions of linear systems of equations and solution of non-linear equations by bisection and Newton-Raphson methods.

The university course on introductory biostatistics is also taught over the year. The syllabus includes (i) the research process and study designs, measures of centre, variability, disease frequency and disease association; (ii) probability laws, tree diagrams and diagnostic testing, binomial and normal distributions, sampling distributions and confidence intervals for means, proportions and differences, *p*-values, clinical importance, power and sample size; (iii) contingency tables, Simpson's paradox and the concept of a confounder, confidence intervals for relative risks and odds ratios; (iv) design and appraisal of surveys and analytic studies, internal and external validity, bias, confounding, case-control and cohort studies, randomized controlled trials; (v) simple linear and multiple regression with use of dummy variables for categorical predictors, control of confounding and appraisal of published research; (vi) one way analysis of variance, post hoc tests, block designs and two way analysis of variance.

One third of the biostatistics course covers the bursary syllabus and concepts like the confidence interval occur throughout the other two thirds. It is therefore reasonable to believe that a student with prior knowledge of statistics from school should show improved performance in the biostatistics course when compared with a student who has never studied statistics. This hypothesis is investigated in the following sections with the results leading to a proposal for improving the presentation of statistics to students in the health and biological sciences.

METHOD

In 1999, 560 out of the 800 candidates in the biostatistics class were first year at university and directly from school. Regression models were developed to predict marks in biostatistics for these 560 students from their school results the year before.

The predictors were a binary variable identifying whether a candidate had studied statistics, a binary variable identifying whether calculus had been studied, gender (value = 1 if male, 0 otherwise), and, to neutralize the overall ability of a student, the average bursary mark (Bav) for each student was introduced as a covariate. The scatter diagram in Figure 1 of biostatistics mark against bursary average shows a non-linear pattern and the square of the bursary average (Bavsq) was introduced to take account of a leveling off of the marks in the biostatistics course for high average bursary marks. Ten influential cases were deleted from the analysis leaving a sample of 550. The deletions were justified because of unusual marks probably caused by illness resulting in fewer examinations being sat. The analysis of the 550 marks for 1999 indicated that performance in the biostatistics course was not influenced by a prior study of statistics at school in 1998 (Harraway & Sharples, 2001) but some mathematical maturity did appear to result in a small, statistically significant (p-value = 0.015) improvement.

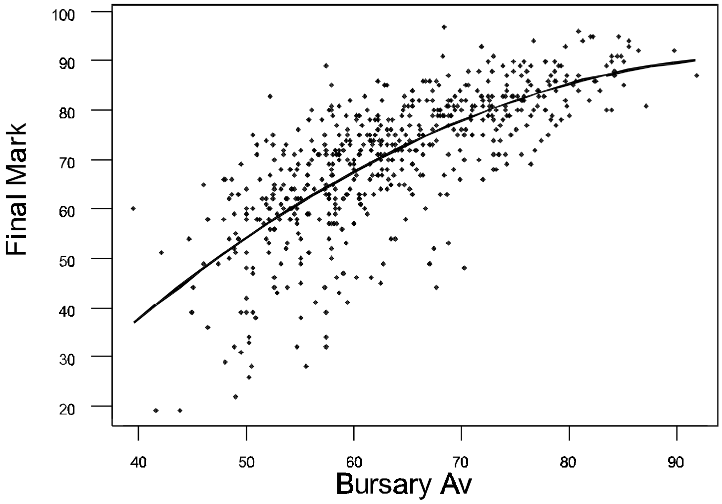

Figure 1. Mark in Biostatistics v's Bursary Average.

Because these conclusions from the regression models were unexpected it was decided to validate the results on the students who studied biostatistics in 2000. There were 490 students included in this validation sample and regression models were fitted to the marks for the combined sample of 1040 students. A dummy variable Y (taking value 1 for year 2000 and 0 for year 1999) was added to all the earlier models along with the interactions Y by bursary average and Y by bursary average squared to test whether the slope and curvature of the fitted equation for 2000 differed from the slope and curvature for the 1999 fitted equation. The extra sum of squares in Table 1 for the model involving S only shows that there is no evidence of a difference in slope and curvature coefficients for the two years. For this reason the year effect variables are dropped in the subsequent analyses and regression models reported for the combined sample. Similar results occurred with the other models involving first C and then gender.

Table 1

Extra Sum of Squares for Year Effects

SOURCE OF VARIATION	SS	DF	MS	F
Reg (S, Bav, Bavsq)	115954.928	3		
Reg (Year effects | S, Bav, Bavsq)	378.117	3	126.039	1.092(N/S)
Reg (S, Bav, Bavsq, Year effects)	116333.045	6		
Residual	119223.259	1033	116.415	

RESULTS

The final model for explaining the variation in biostatistics marks for 1040 candidates included S(value 1 if Bursary Statistics studied, 0 otherwise) and the covariates Bursary average and Bursary average squared. The results with estimated regression parameters, standard errors, 95% confidence intervals and *p*-values are in Table 2, first for a model including S only and then for a model including the covariates. There is evidence that those who had studied statistics at school scored 4.00 marks less on average than those who had not studied statistics. But at school in New Zealand there is anecdotal evidence of a tendency for the more capable students to be encouraged to take mathematics with calculus. To control for ability the average bursary mark over all school subjects and its square were introduced as covariates. The difference between the means for the two groups then reduced to 0.182 which was not significant (*p*-value = 0.805). The final model explained 49.2% of the variation in the biostatistics marks and residual plots showed no evidence of problems with the fitted quadratic. There were 725 students out of the 1040 who had studied statistics at school, such a large sample from all over New Zealand being representative of the country as a whole.

Table 2
Two Models for Biostatistic Mark if Statistics Studied

Model Variables	Estimated coefficient	Standard error	95% C.I.	*p*-value
Constant	72.848			
Bursary Statistics (S)	–3.999	1.009	–5.979 to –2.019	<0.0005
Constant	–63.932			
Bursary Statistics (S)	–0.182	0.735	–1.625 to 1.261	0.805
Average Bursary Mark	3.165	0.381	2.417 to 3.914	<0.0005
Average Squared	–0.016	0.003	–0.022 to –0.010	<0.0005

A similar analysis for mathematics with calculus produced the results in Table 3. Students who had studied calculus scored a significant (*p*-value = 0.002) two marks more on average than those who had not studied calculus. There were 620 students out of the 1040 who had studied calculus at school. The analysis for the smaller cohort of students who had studied both mathematics subjects at school showed an improvement on average similar to that for calculus.

Table 3
Model for Biostatistics Mark if Calculus Studied

Model Variables	Estimated coefficient	Standard error	95% C.I.	*p*-value
Constant	–64.317			
Bursary Calculus	2.069	0.681	0.732 to 3.405	0.002
Average Bursary Mark	3.146	0.379	2.401 to 3.890	<0.0005
Average Squared	–0.016	0.003	–0.022 to –0.010	<0.0005

Table 4
Model for Biostatistics Mark in Terms of Gender (1 = male, 0 = female)

Model Variables	Estimated coefficient	Standard error	95% C.I.	*p*-value
Constant	–63.567			
Gender	–0.749	0.691	–2.103 to 0.605	0.279
Average Bursary Mark	3.155	0.381	2.408 to 3.902	<0.0005
Average Squared	–0.016	0.003	–0.022 to 0.010	<0.0005

Further, the analysis in Table 4 indicates no evidence (*p*-value = 0.279) of a difference in marks on average for men and women. There were 385 men and 655 women in the study. These numbers, however, certainly confirm that more women are applying for the health sciences programmes in medicine, dentistry, pharmacy and physiotherapy, which is interesting in itself.

DISCUSSION

This study was undertaken originally to answer a question posed by many New Zealand high schools having time to teach only one of the two mathematics subjects. The question was "which one, calculus or statistics, should students take given that they will need high marks in their first year at university in order to gain entry to medical school?" Certainly, prior knowledge of statistical concepts should assist university study in this area leading to higher marks and greater chance of gaining entry to professional health programmes. But, on the other hand, many schools in New Zealand prefer capable students to study mathematics with calculus, which provides a deeper understanding of mathematical skills and possibly greater flexibility for future study. A second question related to whether a student with a high mark (over 75%) in statistics from school would have good understanding of statistics for future work in the health sciences professional programmes and therefore be able to be exempted biostatistics at university.

The results here provide confirmation of our earlier conclusion (Harraway & Sharples, 2001) that prior knowledge of elementary statistical methods has no effect, on average, on the mark obtained in Introduction to Biostatistics. Schools will not therefore disadvantage their students by encouraging them to study calculus rather than statistics. As far as the exemption is concerned, the fact that statistics learnt at school is not enough to impact on marks in biostatistics indicates the school subject will be insufficient for later statistical work in the health sciences professional programmes. This is not surprising given the emphasis on study design issues, bias, confounding and the appraisal of recently published research articles in Introduction to Biostatistics. An exemption should not therefore be permitted even though there is strong pressure from some health professionals to allow this.

About 60 students each year apply for an exemption. Anecdotal evidence suggests they have found statistics "boring" at school. They also have a strong dislike of independent project based learning. Several papers presented at ICOTS5 (Swanson & McKibben, 1998; Wood & Wasimi, 1998) allude to the "boring" nature of introductory service courses in statistics. These papers, and others, all suggest interesting ways of correcting the problem by using project based learning, collaborative group learning, the use of real data, the use of computers and statistical packages, and the development of statistical thinking rather than methodology. Two further comments can be made. First, there is growth in graphical procedures, probability and statistical concepts taught in the lower levels of school. This is to be encouraged but caution must be exercised later to prevent too much repetition of the concepts at higher levels and thus onset of boredom. Second, there is an urgent need to access real data in situations which appeal to students in their first year of university study. The same can be said about project work. Such data and suitable projects can be difficult to find, but the effort is worth it.

Comments made by students in the evaluations of the university course, Introduction to Biostatistics, indicate that they found the first sections of the course uninteresting. But at university we have the advantage of access to recent research in the health sciences through working with colleagues in other subjects. We are able to present information on, for example, melanoma and sun exposure, factors affecting cot death, the health and fitness of New Zealanders, dietary factors affecting iron levels in new born children, trials involving cholesterol lowering drugs, screening tests for cancers as well as many other epidemiological studies. Students find these examples very interesting and I have on a number of occasions heard the encouraging remark "This is interesting. I had no idea this was what you did in statistics." Further, Yanagawa (1998) has identified Simpson's paradox as a topic of interest. I agree, and it provides an excellent way of introducing confounding issues prior to the use of regression for adjusting confidence intervals for the effects of confounders.

The students involved in this study are high ability, seeking entry to competitive programmes. But a further analysis (not reported here) for students of more modest ability in the social and biological sciences indicates similar results. My view is that enjoyment of statistics will be enhanced if students in these groups as well can be exposed to recent research studies by scientists active in research. Some investigations which I have used involved dolphin habitat selection, sea lion bycatch, biomass of shellfish in tidal inlets, growth of shellfish in different habitats, and the best location for traps used to catch predators.

About 20 percent of the students in the courses described here for the health and biological sciences are older, new start or second career students. The additional challenge is to identify factors that will encourage the statistical learning of older students and reduce their fear of the subject. Although these students have been excluded from the analyses and results reported in the previous section, discussion with them confirmed again the central role of current data in providing motivation. These older students have either no background or distant knowledge of statistics and mathematics yet they are able to perform as well as the younger students who have come directly from school.

REFERENCES

Barrett, R. (1997). *Pass university bursaries mathematics with statistics.* Auckland (NZ): ESA Publications Ltd.

Harraway, J.A., & Sharples, K.J. (2001). A first course in biostatistics for health sciences students. *Int. J. Math. Educ. Sci. Technol*, *32*, 873-886.

Swanson, D.A., & McKibben, J.N. (1998). On teaching statistics to non-specialists: a course aimed at increasing both learning and retention. *Proceedings of the ICOTS-5, 1*, 159-165. Singapore: International Association for Statistical Education.

Wood, G.R., & Wasimi, S.A. (1998). Transforming first year university statistics teaching. *Proceedings of the ICOTS-5, 1*, 167-172. Singapore: International Association for Statistical Education.

Yanagawa, T. (1998). Teaching the Mantel-Haenszel procedure to medical students. *Proceedings of the ICOTS-5*, *1*, 295-301. Singapore: International Association for Statistical Education.