# LESSONS WE HAVE LEARNED FROM POST-GRADUATE STUDENTS

Sue Finch and Ian Gordon
Statistical Consulting Centre, The University of Melbourne, Australia
sfinch@unimelb.edu.au

*The Statistical Consulting Centre at The University of Melbourne has been providing a post-graduate consulting service for many years. Each year, the service is used by about 200 students who come from a wide range of disciplines. Consultants can assist these highly motivated students with any part of the research cycle, from refining a research question and developing a suitable design to presenting and communicating findings. The rich and varied consulting interactions with these students provide us with an opportunity to observe applied researchers using statistical knowledge to find out about things that matter to them. We report on a survey of consulting sessions and identify important aspects of statistical thinking that commonly arise across the research cycle. These observations help us develop approaches to graduate education that are applied and contextually relevant.*

## THE STATISTICAL CONSULTING SCHEME

The Statistical Consulting Centre at The University of Melbourne has been providing a post-graduate consulting service in various guises since 1995. Between 2000 and 2009 inclusive, consultants assisted over 1100 post-graduate students. In the past few years, all students have had access to up to five hours of free consulting per year; consultation beyond the five free hours is available but must be paid for.

Recently, the service has been used by about 200 students each year. Students come from a wide range of disciplines, from medicine to traditionally non-quantitative fields like development studies. Figure 1 shows the distribution by faculty of the 210 students using the scheme in 2008 with 45% coming from medicine, dentistry and allied health professions. Most of the 210 students (65%) were first time users of the service. The scheme is open to masters and doctoral students; in 2008, about 70% were doctoral students. In that year, just over half (55%) of the students used the service for two hours or less, 27% used between two and five hours, and 18% paid for the service beyond the five free hours. The maximum amount of time a student used the service in 2008 was 29 hours.

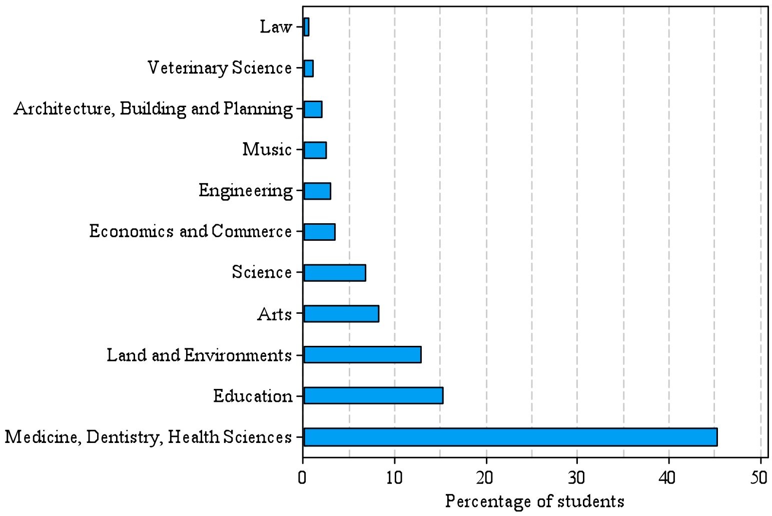


Figure 1. Percentage of students by faculty, post-graduate statistical assistance scheme, 2008

Students are highly motivated but the variation in their statistical competence of the students is large. Very few students have very good or very poor statistical knowledge; the majority have fair to good knowledge. With this variation in statistical knowledge comes great variety in the type and complexity of the questions the students bring to consulting sessions. Questions have

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php 

ranged from how to record data in a Word file (not recommended) to how to fit complex generalized linear mixed models.

Consultants help students with any *statistical* aspect of their research project. The definition of statistical we use here is very broad. It includes, for example, use of software, methods of analysis and interpretation of statistical output, and also takes in fundamental ideas about the use of data-based methods in making decisions in uncertain situations and effective methods of communicating and representing statistical inferences. Wild and Pfannkuch (1999) described a cycle of statistical enquiry that characterizes "the way one acts and what one thinks about during the course of a statistical investigation" (p. 225). The cycle of Problem, Plan, Data, Analysis and Conclusion (PPDAC) is a useful framework for considering the interactions between consultants and students; consultants typically can have an active role in any or all parts of the cycle.

Wild and Pfannkuch (1999) describe statistical knowledge, context knowledge and the information in data as the raw materials of statistical thinking. Moreover, Cobb (2007, p. 338) stated that "context is an essential part of statistical thinking, and some of the worst teaching of statistics occurs when the teacher or textbook tries to treat context as irrelevant". A post-graduate consulting session can be an ideal experiential learning opportunity, because the students have a rich understanding of the context *as well as* a strong motivation to use statistical ideas and tools to inform their understanding of their data and the conclusions it supports. At the same time, it is potentially one of the most challenging learning opportunities as the complexity of the statistical issues involved can be poorly matched with the student's level of understanding.

With this background, we carried out a survey of post-graduate consulting sessions. Our goal was to characterize the scope of statistical thinking discussed in these sessions. Together the consultant and student are jointly engaged in statistical problem solving, allowing students to 2observe and develop skills in the *art* of statistics (Pfannkuch & Wild, 2000). These observations allow us to reflect on the nature of statistical education and the important lessons post-graduates need to learn.

## THE SURVEY

Over 14 weeks towards the end of 2009, six consultants at the Statistical Consulting Centre recorded information about the interactions they had with post-graduate students. The interactions may have been in face-to-face meetings, over the phone or by email. Interactions with purely administrative content were disregarded. These three types of interactions cover almost all of the learning opportunities arising in using the service.

Data were collected in 97 sessions; the survey included multiple consulting sessions for some students. On average there were 1.7 sessions per student. Consultants recorded the type of session, the length of the session, and provided a short description of what was discussed in the session. They also listed the important statistical ideas discussed. 'Ideas' was a general term, and could include concepts, topics, models, points of detail, big ideas, and so on.

Most (75%) sessions were face to face; 18% were by email. Shorter sessions were generally by email or phone. About one quarter of the sessions went for less than half an hour, and two-thirds were between half an hour and two hours (inclusive).

Our analysis focuses on the statistical ideas discussed in these sessions. We present two levels of analysis. First, we classified the statistical ideas discussed in terms of the PPDAC cycle. Second, we identify themes that arose in the details of the descriptions of the statistical ideas referring to different stages of the enquiry cycle.

Different consultants have different styles and our aim is to characterize themes rather than to quantify rates at which ideas were discussed. Table 1 shows the proportion of sessions which included important ideas from different parts of the PPDAC cycle. Typically students come to a consulting session with some formulation of the research problem of interest. Hence consultants infrequently work with students at the phase of Problem. The consultant must, of course, gain an understanding of the steps from the Problem to Plan (and if available) to Data in order to carry out any meaningful analysis (Pfannkuch & Wild, 2000). However, in two cases students had collected data but were not able to clearly articulate the question of interest. The Plan and Data stages are closely related and we have combined them for the purposes of this paper; Data includes implementation of Plan.

In over 80% of the sessions, important statistical ideas in relation to the analysis were discussed. This arises, in part, because post-graduates more often use the consulting service once data have been collected, rather than when they are planning their work. The consultant needs to know, for example, about the design, measurement and data collection process, and these are always discussed with the student.

Table 1. Percentage of sessions in which important ideas referred to parts of the PPDAC cycle

| Problem | Plan/Data | Analysis | Conclusions |
|---|---|---|---|
| 2% | 33% | 82% | 48% |

Most of the important statistical ideas referred to Plan/Data, Analysis and Conclusions. We consider the themes arising in the each of these parts of the cycle in more detail in subsequent sections. Parts of the PPDAC cycle provide a useful structure for considering these themes; of course, ideas can be relevant to different parts of the cycle. We focus less on the procedural aspects and techniques, and more on aspects of the discussion that referred to themes in statistical thinking.

## THEMES IN PLANNING AND DATA

Although the majority of students seen by consultants had already collected data, important issues relating to the plan and its implementation were discussed in about one third of the sessions. Basic issues concerned identifying independent measurement units, and distinguishing between true and false replication of measurements. Sometimes data files included falsely replicated observations as the problem had not been noticed when the study was planned. Considering variation is fundamental to statistical thinking (Wild and Pfannkuch, 1999); identifying measurement units and appropriate replication is an important part of understanding the structure of data and the *sources* of variability. Ideas about structure and sources of variability also arose in discussion of various types of experimental designs. We return to these aspects of statistical thinking again in discussion of themes in the stage of analysis.

Methods for making measurements, asking questions, recording responses, and obtaining valid and reliable measurements were important for those with little experience in planning research studies. A trickier issue arose in relation to making appropriate measurement in studies with variable follow-up times; here statistical thinking appeared to be constrained by knowledge of simple, fixed factor designs, as consultants needed to discuss the need to measure time-to-an-event rather than occurrence of the event *per se*. Along with the question of how to measure, *what* to measure was discussed in relation to potential confounding between measured variables or factors.

*Transnumeration* is a term that was coined by Wild and Pfannkuch (1999) to describe transformations of data representations that help our understanding of a system. Methods of describing, graphing and summarizing data are aspects of transnumeration, as is modelling. It also includes activities where measured variables are transformed or (re)coded in some way to obtain useful outcome measures or explanatory variables (Pfannkuch & Wild, 2000). This describes good statistical analysis, but it emphasizes the goal of understanding; transnumeration is not occurring, for example, when an analyst is (merely) following a prescribed sequence of analysis steps. The process of moving from measurements to variables was a strong theme in the planning phase, particularly when the measurements were made on a categorical or ordered categorical scale. Issues relating to the statistical knowledge of the student, the tractability of the analysis, and the potential ease or difficulty in communicating findings to the relevant audience influenced the ways in which outcomes and variables were defined. Scoring sets of measurements or responses in a logical way to produce one or more variables was another kind of transnumeration covered.

There were many discussions of sample size, including requests for sample size calculations (even after data were collected). These calculations were based on power considerations or precision (confidence interval width), and were for a variety of situations (e.g., comparisons of means, comparisons of proportions, correlations). Here the consultant relied on the context knowledge of the student to identify if there were practical constraints on sample size, to estimate likely variation in the primary outcomes of interest, and to determine the magnitude of a practically important effect size. If this discussion took place at the time of study planning, this may be the first time the student has been asked to think directly about the meaningful *quantification* of the effect of interest. Students need to have a *sense of number* (Wild & Pfannkuch, 1999). We return to this when discussing conclusions.

Many other perspectives on sampling were covered including key distinctions between bias and precision. The common concern with sample size may reflect an understanding of the importance of precision, although 'getting a sample size' can be seen as a simply a hurdle to obtain ethical approval for a study. Various methods of random sampling and different structures that can arise in populations were discussed in some sessions, along with the disadvantages of convenience samples. Here the consultant emphasized the importance of understanding the population from which the sampling units are taken and the process by which the sample is obtained. For example, the sampling of 'control' groups to match clinical or treatment groups was sometimes seen as a trivial problem easily solved by a convenience sample.

Appropriate data management requires both understanding of the data structure and how measurements taken should be recorded and possibly (re)coded. This theme involved discussion of the care needed in data recording, entry and coding, and in the basics of data file structures. Pfannkuch and Wild (2000) refer to the psychological idea that measured data are seemingly sacred; at the data entry stage however, due care is not always taken. In interactions about data management, consultants try to encourage students to think logically about the measurements they have taken and to adopt a meticulous and careful disposition.

## THEMES IN ANALYSIS

The largest number and variety of ideas arose in analysis. Many fundamental concepts that would be found in an introductory course were discussed. A limited selection of examples includes definitions of standard deviations and standard errors, the central limit theorem, sampling distributions, and the distinction between statistics and parameters. Models from simple two group comparisons to generalized linear mixed models were covered. Among more specialized topics were factor analysis, survival analysis and meta-analysis. It is not our intention to detail all these concepts, topics and models, but rather to identify some fundamental themes about statistical thinking.

Although the majority of students had collected data, an important theme in sessions on analysis was data types and the implication of data types for representation and analysis. One aspect of a *sense of number* often discussed was the arbitrary nature of assigning numerical values to (ordered) categories. This was linked with transnumeration from measures to variables described in discussion of themes in planning above.

Consultants modelled transnumeration in many analysis sessions through exploratory data analysis and often by using simple graphical tools such as bar charts, boxplots and dotplots. This kind of investigation of representations of data lays the groundwork for understanding and interpreting final models of the data.

Thinking about sources of variation was a dominant theme in analysis. For example, surprisingly many sessions discussed and distinguished two simple models–matched pairs and independent two group designs. An important distinction here was the relevant source of variation in assessing group differences. In the context of more complex designs involving measurements over time, discussions centered around the use of change scores and questions of whether to model all or part of the study design.

Discussions of more complex modelling, like aspects of the plan, included explanations of units of analysis, levels of variation, fixed and random effects, and replication. These ideas were discussed in the context of clarifying possible models for the data. Again differences between factors measured on the same units and factors measured on different units arose as a fundamental theme.

Over-simplified approaches to modelling adopted by students can reflect a failure to understand sources of variation. Examples include averaging over experimental conditions in a multifactor design, and comparing measurements from different brain areas in the same animals as a between factor design. Some sources of variation were perhaps more salient than others, as they reflected the primary variable or factor of interest; sources of variability that should have been modelled but were not of primary interest may have been less salient. Again statistical thinking may be constrained by simple, familiar ideas based on comparisons of groups.

Problem solving in analysis involved dealing with outliers, considering transformations and assumption checking. Consultants showed *how to model variation*, and the important idea of finding a model for data rather than fiddling data to fit a model. This was a view of modelling that needed explication. Principles of inference touched on included the relationship between hypothesis testing and confidence intervals, and relationships between power, sample size and effect size. Attempts to manipulate hypothesis testing procedures to find statistically significant results (by swapping between one-tailed and two-tailed tests, ignoring unmet assumptions, trying alternative possible tests and 'fishing') also reflected the primacy of familiar statistical models and a lack of understanding of principles of modelling.

Consultants needed to work with a student in a way that provides the most parsimonious model(s) without going beyond the student's competence in terms of the requirements for interpretation and communication. In this context, it is useful to briefly comment on themes in the kinds of analysis used. As mentioned above, sessions covered discussion of a vast range of techniques. However, multiple linear regression, and binary logistic regression models were often used. A number of students asked about methods of assessing agreement.

## THEMES IN DRAWING CONCLUSIONS

Wild and Pfannkuch (1999) describe a recognition of the need for data as fundamental to developing statistical thinking. In general, the post-graduates we saw appeared to understand this *need* but some were less clear about the *role* of data in situations of uncertainty. One theme arising in relation to drawing conclusions was the fundamental notion of making inferences from sample data. For example, one student thought it was sufficient to make claims about differences simply based on examining the sample means. Another wished to argue that measured explanatory variables could be deemed irrelevant without statistical inference. In some senses, uncertainty was not recognized here; the sample data had primacy. Although a general claim may be made, the distinction between sample and population was blurred and the importance of the sample for what it told us about the population was not recognized.

The misuse and misinterpretation of hypothesis tests is recognized a problem in many disciplines. An overreliance on the *p*-value as a binary indicator of the statistical significance or otherwise of a result has been observed and criticized (e.g., Cohen, 1994; Hoekstra, Finch, Kiers & Johnson, 2006; Rosnow & Rosenthal, 1989). The meaning and interpretation of *p*-values was a strong theme in discussions in the conclusion phase of PPDAC. Students were encouraged, for example, to report estimates and confidence intervals as well as *p*-values. One consultant's survey indicated she had discussed: "0.057 is not *bad*, < 0.05 is not *good*". The meaning and interpretation of confidence intervals also arose here. Principles of establishing "no differences" between treatment conditions, for example, were discussed in terms of the practical interpretation of the point estimate and the width of a confidence interval.

A related theme was distinguishing between statistical significance and clinical or practical significance. Consultants had an important role in assisting in mapping the statistical results back to the applied context. This involved interpreting point estimates and confidence intervals, and asking students to think about acceptable levels of agreement, clinically meaningful differences, and practical criteria and so on. It removes a comfortable reliance on statistical significance that perhaps satisfies a need for certainty. For example, students asked: "What is the criterion for a good (or bad) correlation?" in the hope of a simple numerical answer. Consultants explained the important of context in judging the magnitude of correlations. Statistical thinking here involves integration and synthesis of statistical and contextual aspects, and the sometimes challenging task of mapping of a clinical or practical judgment onto the measurement scale.

Interpretation of parameter estimates for some particular types of models arose commonly in the conclusions phase. Consultants often assisted with the interpretation of linear regression coefficients (binary or continuous explanatory variables), and the coefficients of covariates in general linear models. Explanation of the interpretation of parameter estimates based on means arose less often; however interactions in ANOVA and general linear model needed explanation. As mentioned earlier, consultants were assisting a number of students with analysis using binary logistic regression and, to a lesser extent, ordinal logistic regression; in this context, the interpretation of odds and odds ratios was discussed.

As discussed earlier, *transnumeration* includes methods of describing, graphing and summarizing data and modelling. Some parameter estimates, such as means, have natural and familiar analogues in exploratory and descriptive methods. Others, like many of those mentioned in the paragraph above, have less direct correspondence to the summary statistics and graphical displays that students routinely use. For example, exploratory analysis of a binary outcome in a treatment and control group might involve bar charts and comparisons of proportions, but inference may be based on an odds ratio.

Presentation and representation of results was an important aspect of some sessions. This usually focused on deciding the aspects of analysis that are important to understanding the findings and conclusions, and on appropriate graphical representation of inferences. Findings are to be reported in theses, academic journals and conferences, so there is a need to distill information from exploratory, descriptive and inferential analysis in such a way that it can be understood by academic audiences. In this process of transnumeration, there is sometimes tension between disciplinary norms and the consultant's preferred approach. This can be further complicated by the differences in the standards, conventions and requirements that can apply to presentations, articles and theses.

The final theme relates more broadly to the validity and generality of the conclusions. In this part of the cycle aspects of design and analysis are reconsidered in relation to inference; here thinking needs to move beyond the inferences obtained from the statistical analysis to consider the practical constraints on interpretation of the statistical model. Examples include considering the limitations of observational studies compared with randomized controlled trials in establishing causality, assessing the implications of bias in recruiting subjects, and considering the implications of carrying observations forward to estimate an endpoint treatment effect in a clinical trial.

## CONCLUSION

We have identified some themes in the kind of statistical thinking arising in interactions between post-graduates and statistical consultants. Consistent themes emerged across the parts of the PPDAC cycle that covered aspects of the important statistical ideas discussed. This featured several aspects of transnumeration, including how to move from measures made to meaningful variables for analysis, methods of exploring and representing data to support interpretation of models, and how to best communicate and represent inferences. We suggested interpretation of some (less familiar) parameter estimates might be helped by developing transnumeration practices in exploratory analysis. Underlying developing transnumeration practices is having a good sense of number - understanding data types and making mappings from statistical outcomes to practically important results. Thinking about sources of variability emerged as a theme across plan and analysis phases of the cycle; here graduate education could focus on teaching about complexity in design and data structure while leaving the complexities of actual analysis aside.

## REFERENCES

Cobb, G. (2007). One possible frame for thinking about experiential learning. *International Statistical Review*, *75*(3), 336-347.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49,* 997–1003.

Hoekstra, R., Finch, S., Kiers, H. A. L., & Johnson, A. (2006). Probability as certainty: Dichotomous thinking and the misuse of *p*-values, *Psychonomic Bulletin and Review, 13,* 1033-1037.

Pfannkuch, M., & Wild, C. (2000). Statistical Thinking and Statistical Practice: Themes Gleaned from Professional Statisticians. *Statistical Science, 15(2*), 132-152.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44*, 1276-1284.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry (with discussion). *International Statistical Review, 67*(3), 223-265.