Teaching Statistical Consulting: Enabling skills transfer and adding value

Kaye E. Marion

RMIT University, Melbourne, Vic 3000, AUSTRALIA
k.marion@rmit.edu.au

Abstract

A statistical consultant needs to understand the scientific process in order to develop models suitable for testing, have the quantitative expertise to apply the statistical test appropriately and communicate the issues involved effectively. A successful consultant will not only address the problem but will teach the client something about statistics and with this transfer of skills enable the client to make better decisions in the future.

This paper will discuss the skills that need to be acquired by a statistical consultant and cover some techniques developed to teach these skills in the Master's Degree in Statistics and Operations Research at RMIT University in the first semester of 2009. The author draws on over 40 years as a consultant developing testable models for medical researchers, business analysts, government policy makers and academic researchers with examples where ingenuity was required to handle non-standard decision problems.

Keywords and phrases: Statistical consulting, graphs, communicating statistical concepts, statistical thinking, problem solving.

1. Introduction

Students often have limited experience in making decisions where the data are limited or the problem is not well defined. Effective consultants need experience in understanding a decision problem, the ability to convince a client that they will be able to assist, be able to select the appropriate computer packages and use them effectively. But they also need to be able to carry out administrative functions since they will need to prepare a quotation or tender and produce a quality report. Interaction with the client and follow-up activities require communication and personal skills that the student needs to develop. While there is no substitute for consulting directly with a client, the Consulting and Applied Statistics (CAS) course taught at the University of Melbourne as part of the Key Centre for Statistical Sciences was designed to prepare Masters' students by teaching them to think about data in a broad context and develop the skills required [1]. The University of Melbourne withdrew from the Key Centre in 2008 and RMIT University instituted a replacement course of the same name. Students normally complete CAS in the first semester of the academic year, followed in the second semester by Consulting Practice, where they apply the skills learnt in CAS. The students' backgrounds influenced the course design as approximately half of the students came from the Business stream with the rest from the Statistics and Operations Research stream. The Business students did not have a strong training in statistics and the Statistics and OR students had not previously been required to carry out group work. The students were asked to assess themselves (out of 10) on their each of their administrative, statistical, computing and report writing skills. For the first task, an integer program was used to construct groups of no more than 3 students so that the maximum difference in total skills for each group on each skill category was minimised. Later teams were constructed to maintain the skills' distribution as evenly as possible, but also to keep the number of students working with someone that they had worked with before as small as possible.

Taplin [2] provided a useful summary of the experiences of authors who have developed resources and courses in statistical consulting, focussing on those that avoid mathematical recipes by teaching statistical thinking and describing the Introduction to Statistical Methodology undergraduate course implemented at Murdoch University. Taplin argued that encouraging students to carry out statistical work early in their studies both motivates them to study further statistics and enables them to provide statistical consultancies.

Although the RMIT students were at the master's level, there was a similarity with the Murdoch students in that they had not previously undertaken many statistical courses. The RMIT CAS course was also designed to include as many as possible of the recommended skills listed by Kenett and Thyregod [3].

2. CAS Components

2.1. The Consulting Approach

The author's experiences in consulting were distilled in the first two weeks. 'Statistical FAQs and some answers' focussed on the questions consultants are most frequently asked, beginning with "What sample size should I use?" Recognising that many consulting jobs are urgent, 'Statistical First Aid and Bridge Building' used case studies to demonstrate some of the techniques one can use to understand the problem and to communicate the results to the client so that the solution can be understood. Various cases were presented to the students to extend their thinking about statistical analysis and to develop their skills in presenting data graphically. Two examples are given below.

Case 1: A student had asked for help in designing an experiment to determine the optimal time to remove eggs from crayfish to maximize the survival of the progeny. Originally the student thought that he would have access to at least 50 crayfish and non-parametric analysis would be suitable because the data would not be normally distributed. In the event, only 7 crayfish were available and further, a large proportion of the eggs did not survive. He thought that he would fail his master's degree. Fortunately we had planned the experiment with controls (crayfish develop two rows of eggs, so eggs were only stripped from one side leaving the eggs on the other side to develop without interference). After a search of the literature, it was found that a parametric test on the data (shown below) was possible.

Female	Stripping time								#Eggs	Control
	12 hour		24 hour		36 hour		48 hour		(test)	N hatched
	N	#Alive	N	#Alive	N	#Alive	N	#Alive		
1	31	0	26	6	34	17	31	17	122	145
2	31	0	0	0	35	0	63	57	129	Eggs died
3	38	0	36	0	40	0	66	43	180	Eggs died
4	66	7	59	0	50	7	51	12	226	Eggs died
5	31	0	47	0	64	2	60	51	202	251
6	40	0	47	0	76	0	35	34	198	19
7	127	37	113	101	154	140	147	141	541	516

Figure 1 Crayfish experiment data

The appropriate analysis for binomial data is given by Rao [4] for the proportions of eggs that remain alive, $X_{ij} = \arcsin\sqrt{p_{ij}}$ where $p_{ij} = \text{Number Alive}/\text{N}$, N = number of eggs, $i = 1,2,3,4$ for each stripping time of 12, 24, 36 and 48 hours respectively and $j = 1,\cdots,7$ for each female. The general linear model (using time as a covariate) indicated a difference between females (p = 0.02) and a strong time effect (p < 0.00005), showing that later removal of the eggs greatly enhanced survival. (The model's estimated residuals were consistent with being normally distributed.)

The significance of this case is that students are reminded of the importance of the literature search.

Case 2: An experiment by a pharmaceutical company for a therapeutic skin cream in two formulations (A, and B) had been tested by 10 patients over two intervals with their preferences recorded. Increasingly the Therapeutic Goods Administration (TGA) requires the applicants to report *p*-values of statistical tests. However, in the first place, the company only provided the aggregate number of preferences. Of course, nothing could be determined from these aggregates because one did not know how the preferences changed over time. One had to have the individual observations. When these were provided, some ingenuity was required to represent the data graphically while at the same time keeping track of the individuals.

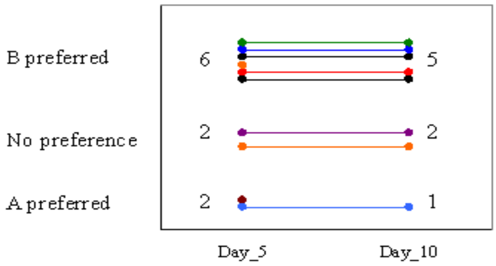

Figure 2 Therapeutic cream preferences

The total numbers of preferences at each stage are shown in the figure above. Although the sign test for the median could be applied to these data, its use would not be valid as the observations are not independent. This example introduced the concept of bootstrapping. The following scoring system was devised: +1, 0 or –1 when the subject expresses preference for B, no preference or preference to A respectively at day 5 and +2 or–2 when the subject expresses preference for B or A respectively on day 10 and 0 for no preference or a missing observation. This gives the sample set (n = 20)

+1, +1, +1, +1, +1, +1, 0, 0, –1, –1,
+2, +2, +2, +2, +2, 0, 0, 0, 0, –2

Using the bootstrap, an approximate 95% confidence interval for the mean score is (0.113, 1.09) so we can conclude that cream B is preferred to A.

The significance of this case is that although a standard technique may not be obvious, some ingenuity can devise an appropriate test. The students could also carry out some sensitivity analysis with different scoring regimes.

2.2. Aspects of the consulting process

The consulting process was examined in detail along the lines of Cabrera and McDougall [5], but presented around specific consulting jobs carried out by the author. Experience in understanding a decision problem and in choosing the right statistical test was facilitated by encouraging the student to frame questions of a potential client ("What would the data look like if there was no effect and what would it look like if there was a strong effect?"), roughing out statistical ideas as the problem was presented to show one's understanding and help the client explain their problem more clearly, using simulation to give clients a feeling for uncertainty (some sample size calculators in Excel were provided) and providing advice on how to cope if the client suggests a technique which is unfamiliar to us. (Ask the client to explain why they think the technique is appropriate and what they expect it will tell them. Often we then recognize the technique they are talking about – it has often been given a different name in another discipline.)

The students (working in groups) were required to prepare a quotation/tender for a specific consulting problem (presented as the ABC Company's problem) on which the author had worked. They were instructed, as a minimum, to define objectives clearly (to help avoid project creep later on), identify the various tasks and estimate the work needed (and time to be taken), provide clear time lines, allow for data cleaning and potential delays in provision of essential data, provide confidence that they will be able to complete the work (provide referees, establish emergency back-up) and set clear deliverables (another technique to avoid project creep). The emphasis was on the importance of good documentation, encouraging students to write clearly and develop good word processing skills.

The ABC Company problem was especially chosen in that it was one in which the company's stated goal could not be achieved. The data were deficient (there were some time series that were observationally equivalent and this was not known when the consultancy was accepted). This was to give the student confidence in handling a consulting project which could 'go wrong'. The consulting task then changed from 'solving the problem' into communicating to the client why the original goal could not be achieved and giving guidance on what was needed for the client to obtain the information or re-frame the objective that would allow a decision to be made. If the student faces a similar situation in the future, they will have had previous experience to help them cope.

While the course does not explicitly teach statistical theory or techniques, we emphasized that, in carrying out statistical analyses, the students need to be critical in preliminary data analysis and to be flexible and imaginative in analyzing the data. One must use existing research and keep up to date with statistical techniques and software. Clients are often unaware of the assumptions required for a test to be valid, so the analyst must ensure verification is done. In addition, the most successful consultants are able to bring multi-disciplinary techniques – from econometrics, time series analysis and operations research for instance to the solution of a problem.

Students often leave documentation until last, a mistake we try to correct by encouraging students to document their analysis as it is carried out. Not only is this easier when the analysis is fresh in the student's minds, it is more accurate. The other benefit is that it often provides the basis of the report, saving time later.

Other activities covered included discussions on ethics. A case study, the DEF Company, was used where the client believed that a 'For Sale' sign placed at several points of time on the building that it was renting was detrimental to the business. After the students had submitted their reports, which showed a strong correlation between a slowing in sales with placement of the sign but included a disclaimer that the correlation did not prove that the signage caused the loss of sales (the students were directed to include this), the students were asked what they would do when the client asked for the disclaimer to be removed from the report. (This had happened in the real consulting job.) The ensuing discussion also provided an opportunity to examine the responsibilities of, and the resources available to, the statistician when acting as an expert witness, including the guidelines issued by the Chief Justice [6].

2.3. Activities devised to enable the student to give confident statistical explanations

In consulting, one is often required to explain statistical concepts to non-statisticians or even to people whose quantitative skills are not strong. To give students experience in this, they were required to explain, with only a few minutes notice, a statistical concept from a list provided. The lecturer was careful not to be judgmental and not to comment (unless the explanation was misleading) in order to increase the student's confidence in speaking in class. The student was then able to select another concept from the list and provide a written answer, also meant to be succinct. In this way, students were exposed to a large range of brief statistical explanations.

Students had to answer in class, again with only short notice, a real query selected from a list gleaned from statistical websites. Class members were able to assist and, after this opportunity, the lecturer's written answers to the queries provided a basis for further discussion. This session was so successful (in terms of student feedback) that another along the same lines was run with students being given statistical computer output to explain.

A session on critiquing research papers required the students (in groups) to supply a journal paper from an area in which they were interested (not a statistical journal) which contained some statistical analysis. They were required to present their analysis to the class considering the key question being addressed, the study design, how the variables were measured and collected, the type of analysis used, the appropriateness of the design and analysis, whether the conclusions could be justified and what additional information, if any, might have been useful. In addition, some papers were chosen by the lecturer which enabled discussion on statistical concepts to which a large proportion of the class had not formerly been exposed. Fletcher's 'Uncontrolled Studies' [7] provided the basis for introducing the following concepts, in context: The Hawthorne effect, Regression to the mean, Healthy worker effect, Placebo effect and Confounding by indication. A facilitated discussion on 'Significance' (based on Sterne and Smith, [8]) gave guidelines on how to express statistical results and on what can (and cannot) be concluded from statistical tests. A small number of marks were allocated to each student's contribution to class and a log was kept. Students who had not been as actively involved as others were given further encouragement to speak as the semester progressed.

2.4. Report writing

The students produced several reports. They had been advised to keep the reports concise, not include analyses or plots just because they were carried out but to consider whether they are needed to make a coherent story. They had to make the report easy for someone to read, for instance by putting comments about graphs on the same page so the reader did not have to flip back and forth between pages. The principles defined in Tufte [9] were recommended since graphical methods can be extremely insightful and convincing. Also, they had to make it clear what the reader could and could not conclude from the graphs and analysis. They were advised to use the terminology of the client's industry and avoid statistical jargon and unnecessary acronyms.

2.5. Interaction with the client: Adding value

Although the students would not have practical experience in dealing with clients until the second semester, class discussion assessment tasks gave them experience in expressing and substantiating opinions. In particular, the ABC Company's problem highlighted the necessity to keep the client up to date with progress reports and to advise the client immediately when problems arise so that a final negative report does not come as a surprise.

The concept of adding value to the consulting project is that the consultant does more than solve the current problem. The client will be able to make better decisions in the future if the consultant has enabled the client to understand the statistical thinking that has led to the statistical technique(s) being applied.

2.6. Project wrap-up

A strategy that the author has found extremely useful when the project is complete is to ask the client if they would be prepared to act as a referee with potential clients in the future. If they would not, then ask what would need to be done for them to change their mind. On the rare occasions when this happened, only a slight modification was needed to improve the consultancy work and satisfy clients who might otherwise not have articulated their dissatisfaction.

2.7. Student feedback and future directions

Students reported that although providing statistical explanations of concepts without advance notice in class was terrifying, they found the information gained extremely helpful. Anecdotal evidence from previous years' students ("CAS was a waste of time" from four students independently) was not supported by the 2009 class. A good teaching scale score of 76.2% and an overall satisfaction index of 85.7% was achieved (questionnaire response rate = 64%). The students appreciated being exposed to actual local consulting projects, with several commenting that these were the best part of the course.

The students requested some help in presentation skills and networking strategies. These will be incorporated in future courses. More case studies will be written up to give the students some choices in their submissions, and to provide cases which can be used in role-playing. Class role plays will be designed to give the students practice in listening to the client and asking questions to help determine what analysis the client's problem really requires.

3. Further case studies

Space does not permit full details of the cases covered, but a subset, and the reasons for their inclusion are described below.

(a) A Lymphoedema Compression Garment Program exemplified difficulties in analysing questionnaire data when the questionnaire had not been well designed. For instance, there seemed to be a mismatch between the proportion of patients reporting that they were receiving funding (26%) whereas 72% of therapists reported their patients were receiving assistance. However, the therapists were not asked what proportion of their patients were receiving assistance, but only if they had *any* patients receiving assistance. This case study was also used to demonstrate how one can handle text-heavy responses.

(b) Film Blowing Design of Experiments (DoE) enabled comparison between the emphasis appropriate in developing models 'in practice' with "teaching" examples where the focus is usually on detecting significance. A pilot analysis on film blowing was conducted where the input variables included Blower speed and Temperature. These together determine Freeze Line Height (FLH), the variable used in the literature as an input. The output variables were Haze and Tear strength, Stress, Gloss and Shrink in the Machine and Transverse Directions. The objective was to find the settings of the input variables that would provide the best values of the output variables depending on the use to which the film would be put. The client was initially resistant to the proposition that the input variables needed to be blower speed and temperature set by the experimental design, rather than FLH (which can be achieved by a range of blower speed-temperature combinations that would be uncontrolled and unrepeatable being arbitrarily set by the operator at each trial).

The client was convinced on this point eventually. He was, in contrast, very comfortable with the remainder of the analysis where response surfaces could be used to determine the optimal machine settings facilitated by the computer package ECHIP [10].

(c) A pharmaceutical experiment provided a case where statistics could not resolve problems with the experiment. Therapeutic cream was applied to the right forearm, the competitor's cream to the left, and the difference in moisture was measured using a corneometer at the start and after 1, 3, 5 and 24 hours, with the results as shown in Figure 3. The hypothesis that the mean (or median) difference was zero was not rejected.

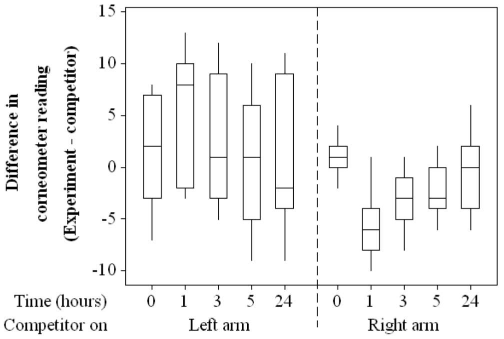

Figure 3 Corneometer reading differences (n = 6)

The experimental cream was also applied to the left arm, the competitor's to the right. But the variation in differences was noticeably greater than when the treatments were on the reverse arms. The experimenter could give no explanation for the variances differing between arms. No elementary statistical technique can compensate for such discrepancies in experiments.

(d) Call centre: The client had heard about 'six-sigma' and 'lean manufacturing' and wanted to apply the techniques to his company's call centre. Their major key performance indicator (KPI) was the 'Grade of Service', the proportion of calls that were answered within 20 seconds of the caller making the last button choice. The first step was to represent the data visually with the results in one centre as shown in Figure 4 where it was clear that conditions had changed over the period of interest. Only then could the client be convinced that he couldn't apply control charts without further analysis.

4. Conclusions

A review for the students at the end of the course was reinforced with a list of the qualities needed by a statistical consultant and a list of the computer packages readily available at RMIT. Each quality was illustrated by examples from the students' work submitted during the course or from the author's experience.

The time expended on creating the teams was extremely worthwhile in giving students a wide experience in working with others and at the same time creating a cohesive class where students benefited from their colleagues' skills and gained confidence with their contributions to their groups. While nothing can replace actual consulting, the case studies, class role-plays and presentations provided individual student experiences. The practice students gained in explaining statistical concepts, interpreting computer output and critiquing journal articles in real time will enable them to pass on this knowledge to future clients and add value to their consulting.

References

[1] I. Gordon, http://www.scc.ms.unimelb.edu.au/ian.html Viewed 20 June 2009

[2] R.H. Taplin, "Teaching Statistical Consulting Before Statistical Methodology", *Aust. N.Z.J.Stat.* 45(2), pp141-152, 2003.

[3] R. Kenett, P. Thyregod, "Aspects of statistical consulting not taught by academia", *Statistica Neerlandica* Vol. 60, pp 396-411, 2006.

[4] Rao, C. R., *Linear statistical inference and its applications*, 2nd ed., Wiley, New York, 1973.

[5] J. Cabrera, A. McDougall, *Statistical Consulting*. Springer, 2003.

[6] M. E. J. Black, Practice Direction: Guidelines for Expert Witnesses in Proceedings in the Federal Court of Australia http://www.fedcourt.gov.au/how/prac_direction.html 2008 Viewed 20 May 2009.

[7] J. Fletcher, Uncontrolled Studies, *BMJ* 2009;338:b1164.

[8] J.A.C. Sterne, G.D. Smith, "Sifting the evidence — what's wrong with significance tests?" *BMJ* 2001;322:226–31.

[9] E. Tufte, *The Visual Display of Quantitative Information*, Cheshire. Conn., Graphics Press, 1983.

[10] ECHIP 7.0 http://velocitypointe.com/VP1/ECHIP.htm

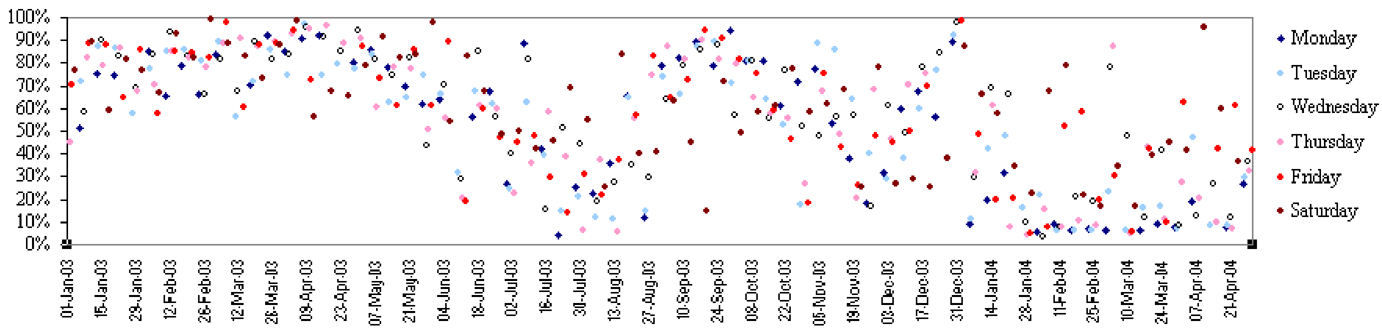

Figure 4 Grade of service at the call centre had been satisfactory until July 2003