PRACTICE IMPROVES PRESENTATION

Flavia Jolliffe
University of Kent, United Kingdom
F.Jolliffe@kent.ac.uk

Practice in doing consultancy, and in teaching consultancy skills to statisticians, improves that teaching, regardless of the teaching method used. Teaching consultancy skills involves emphasis on communication skills, making statisticians aware of the range of problems they might meet, and showing them ways of dealing with non-standard problems. Examples based on actual consultancy sessions are discussed in this paper, and suggestions are made as to how to integrate these into a course intended to equip students for work as a practical statistician.

INTRODUCTION

Practice, by a statistical consultant, in the presentation of statistical methods and results to clients can be expected to improve the quality of the presentation. Similarly, if statisticians practise the activity of consulting this will improve the way they present themselves to the client. Another interpretation of this paper's title, and the main focus of the paper, is that practice in doing consultancy will improve the "teaching" of consultancy skills to statisticians.

Consultancy skills can be taught in many different ways (Belli, 2001). Increasingly courses with a mix of classroom sessions and an element of practical experience of consulting (see for example Jersky, 2002; Cordani, 2003) are being introduced at university level. Short intensive courses of up to 1 week designed more as continuing professional development are also often taught in this way. A mentor scheme, when an experienced statistical consultant is paired with someone who is relatively new to consulting, gives the person who is "learning" some hands-on experience as well as the benefit of instruction before a consultation and feedback afterwards. Other ways of teaching consultancy skills are through material on the internet, and through printed material. These last two have little to recommend them as the sole method of teaching. A combination of these various methods of teaching consultancy skills might well be used, of course, and often is.

In this paper *instructor* is used to denote either a teacher or mentor or an author of material designed to teach consultancy skills. A *statistician* is taken to be someone who has specialised in statistics, and could refer to those who are part or full-time statistics students as well as to those who are practising statisticians. For convenience clients who are seeking statistical help are taken to be masculine, and instructors and consultants are taken to be feminine. The problems described in the paper are based on ones which the author has met during the last three years while acting as a statistical consultant to staff and students at the Natural Resources Institute of the University of Greenwich. The Institute's research is mainly conducted in developing countries.

TEACHING CONSULTANCY SKILLS

A statistician acting as a consultant needs to have a good knowledge of a variety of statistical methods, and to have some experience in the application of statistics to real problems, including the use of statistics packages to implement methods and the use of spreadsheets and/or databases for storage of data. Skills in these areas would for the most part be taught separately from the teaching of consultancy.

Statistical consulting is essentially a practical activity and has to be learnt by practising it, in the same way that doing statistics or driving are learned through practice. The statistician can be told and shown what to do, but in the end the onus is on the statistician to acquire the necessary skills. In teaching statisticians how to do consultancy, then, the instructor's main role is guiding the statistician as to what skills are appropriate. It is important that the instructor observes the statistician in using these skills, in either a real or a mock consultancy session, and gives both feedback on the statistician's performance, and advice as to how the statistician could become a better consultant. At times, when the statistician is not sure how to answer a client, the instructor

might feel it beneficial to all concerned to discuss the situation with the statistician, perhaps with the client present.

The main skills that a statistician needs to be a successful consultant, in addition to those mentioned above, are: both spoken and written communication skills, listening and questioning skills, interpersonal skills, and pedagogic skills (Jolliffe, 2003). These are all interrelated. For face to face and telephone consultation sessions spoken and interpersonal skills are clearly important. However, consultation can involve written exchanges, increasingly by email, and a consultant will often need to write something for the client during or after a meeting - perhaps comments about a report, perhaps an explanation of a method. The latter involves pedagogic skills which are also often used during a consultation session. Part of a statistician's role as a consultant is to extend a client's knowledge of statistics, but this has to be done in an unobtrusive way, involving good interpersonal skills.

Statistics is a growing discipline and few statisticians would claim to know all of statistics. A consultant should be willing to say that she does not know how to proceed, but will try to find out. A consultant who knows something about the topics her clients are investigating, and is willing to learn more, is likely to realise what help a client needs. In addition, engaging with clients in this way would perhaps lead to fruitful collaboration between the consultant and the client (Jolliffe, 2003). Library and web research skills are thus almost essential for a consultant, and these should be covered when teaching consultancy. Other skills which it is useful for a statistician to have when acting as a consultant are data management skills, and what might be called advertising skills. Data management skills include coding and entry of data, and checking and editing of data. Teaching these is sometimes overlooked or covered in passing in statistics courses, so in teaching consultancy it is sensible to check whether the statistician has experience here. The instructor will be better at presenting these skills if she has practised them in a real application. Advertising skills are to do with being proactive in promoting statistics and encouraging potential clients to consult with statisticians. Here too presentation of material concerned with advertising will be improved by an instructor's own practice in the advertising of consultancy services.

MEETING CLIENTS

Interpersonal and communication skills are particularly important when a consultant meets with a client. Instructors with consultancy experience are in a good position to advise statisticians as to how to use these skills to develop a good working relationship with clients. They can refer to specific examples of different types of client behaviour and suggest how to react. Although written mainly for a North American readership, parts of chapters 2 and 4 of Cabrera and McDougall (2002) are helpful as regards the conduct of a consultancy session.

Ideal clients are ones who come well prepared, with, on paper, a summary of their research and specific questions that they want answered. Where relevant they would also come with data in electronic form and perhaps also the results of some analysis. The ideal client will, of course, have consulted both before and during the research and will have followed the consultant's advice. Although possibly having his own ideas as regards statistical aspects of the research he would be open to suggestions for improvement. It would be a two-way partnership between the consultant and client - a true collaboration (Jolliffe, 2003).

There are many ways in which clients deviate from the ideal. There are some who give far too much detail of the project on which they are working, much of which is not relevant to the consultation, for example full details of the funding and staffing of the project, or of past projects with which they have been involved. This wastes consultation time and the consultant is likely to lose concentration before the pertinent facts have been reached. With experience one learns to steer clients towards the reason why they have come. One also learns which clients are likely to be of this type and can manage subsequent sessions with them more efficiently. They might, for example, be encouraged to put background and other information in writing for the consultant to refer to if needed.

In complete contrast some clients bring almost nothing to a consultation, perhaps just some data or some output from an analysis. A lot of questioning is then needed to find out what they have done and what the aims of the research are. A further appointment needs to be made

fairly soon afterwards after making clear to them what they should bring with them. This situation should not really occur if the appointment has been made in advance but could if, for example, the appointment has been made by email and the client does not pick up later messages explaining what is expected. In presenting scenarios like this the instructor could suggest to statisticians that having a sheet explaining what to bring (and what not to bring) to a consultation is a good idea. However, clients do not always take any notice of guidance as to how to prepare for a consultation.

Then there are clients whose analysis is too ambitious either for the data or for the client, or is inappropriate. In cases such as this much tact is needed. The client has to be let down gently and the consultant needs to explain why the analysis is not suitable, perhaps giving examples of when it would be, and to suggest an alternative method, in other words the consultant has to teach the client, but without making it too obvious that she is doing so. As one example, the client was interested in the relationship between various variables and the level of water in a large lake in an African country. He had daily data on the level of the lake, rainfall, and the maximum and minimum temperatures for a period of eleven years, obtained from published sources, and was planning to fit some regression models. However, the rainfall data were for areas some way from the lake, and one does not need much expertise to realise that rainfall and temperature are not the only factors affecting water levels. The values in themselves were averages. This whole project was a non-starter, but to be fair, the client was a student who had not consulted with his supervisor as to its suitability, and who had only a minimum knowledge of statistics.

As another example, in a final report of a well-designed project involving several researchers, logistic models had been fitted to some survey data by a member of the team who was overseas. This researcher appeared to understand the method and had even commented on the fact that some of the variables were qualitative. However, on looking at the results it was clear that, in those models where a qualitative variable taking more than two values was involved, the variable had been treated as if it were a continuous variable. The problem was explained to the client who was present at the consultancy session and who wanted to know whether or not the statistical issues had been addressed adequately in the report.

This last example could well have come from the kind of client who thinks he is an expert in statistics and who presents a completed analysis, possibly expecting approval and praise, even if he is modest in manner. It might even form part of a paper that is to be presented at a conference or submitted to a journal. Often the client has used a sophisticated statistics package and the analysis appears plausible at first sight. There is, as with any client, a need to probe, and sometimes detective work is needed. Use of an independent two-sample *t*-test instead of a paired *t*-test is usually easy to spot once the details of the experiment are known, as is a *t*-test done on data where a chi-square test is appropriate, the researcher having been led astray by the lay-out of the data. These are common mistakes which anyone who has ever taught statistics at an introductory level will be familiar with, but ones that we might not expect experienced researchers to make. In teaching consultancy the instructor needs to warn that nothing should be assumed as to the statistical expertise of a client. Less easy to spot are errors where the results of an advanced analysis are presented and the client asks for, say, help in writing up an interpretation. As an example of this, a client had performed a probit analysis and had entered the number of trials as 100 in every case, having converted results to percentages.

SOME NON-STANDARD PROBLEMS

Clients sometimes come with seemingly easy questions that are not amenable to statistical analysis. In some cases it is the design that is at fault, in others it is more that there are no obvious methods to use, or that the methods are too complex to be suitable as many researchers are subject to time constraints, and in some cases advanced statistical analysis would be taking them well beyond their current statistical knowledge. Such problems throw up challenges for the consultant. However, being a general statistician rather than a specialist in one application area is an advantage, and lateral thinking often helps. Methods usually associated with medical statistics, for example, might well be applicable to data collected by natural scientists. Statisticians who are taught consultancy skills from an instructor who uses examples that she has met will be prepared for the challenge of acting as a consultant.

An example of an experiment where the data were of little use is as follows. Pairs of insects, one male and one female, were placed in cages, one pair to a cage. The researcher was interested in the insects' sexual activity and how this related to other activities such as sleeping and eating. An observer spent five days observing, at the end of each half hour period, except during the night, what the male insect in each cage was doing. The researcher wanted to estimate the proportion of times the insects spent copulating. Had the observer been asked to record activities and the times at which these changed this question could have been addressed. The consultant can do little in a situation like this except stress the importance of discussing experiments before they are done, discuss how the experiment might be improved, and suggest ways of presenting the data that were collected. This would be an interesting example to discuss in a class format. The set-up is easy to understand, and the example could also be used to consider how to tell a client tactfully that there is little that can be done with the data.

An experiment where there was no easy and obvious way to analyse the data also involved insects. Here an insect was dropped on to the centre of a revolving disk which was divided into four equal-sized zones, with different treatments in each zone. The times that the insect stayed in each zone were recorded. The experiment was repeated several times with the same configuration of zones, and also with different configurations. As the total time an insect stayed on the disk varied with the insect, the researcher had converted the times to percentages. The research project was concerned with which lethal treatments attracted the insects, and this experiment was one of several done, others involving random allocation. Randomization tests (Manly, 1997) might well be a suitable approach. This too would be a good problem to use in teaching. Although a consultant can get a lot of detail about a situation in a first consultation, sometimes after cross-questioning the client, she might realise later, as in this case, that even more information is needed. This particular problem remains unfinished as the client, a research student who was writing up her thesis at the time, did not respond to emailed requests asking for more details. Nor was she in touch with her supervisor, who was alerted to the situation.

An interesting problem from the statistical point of view was concerned with which cassava plants farmers in a West African country would choose as best from a large number of plants grown under controlled conditions. This was part of a larger project. Four hundred and thirty-nine seeds out of 600 that had been planted had germinated, and about 13 farmers were each asked to say, before the plants were cut, which 10 plants they thought were best, though some named fewer than 10. The heights and yields were known for all the (numbered) plants in the finite population. The researchers hypothesised that the farmers would tend to choose tall plants of high yield, and that there would be less variability in each of heights and yields in the choices than in the whole group of plants. There is little relation between heights and yields. Some farmers had discussed which plants they would choose with one another so their choices were not independent. The choices of the different farmers had been combined into a single file listing 121 choices, with no record as to which farmer had chosen which plant, but only 42 different plants were listed, that is, there were many repetitions in these 121 choices. Thus the sample was in some ways a with replacement sample from a finite population. Ignoring the repetitions in the 121 choices gives insufficient weight to the popular plants. Obviously there is a super-population underlying the population of plants. In cases like this one has to turn a blind eye to some of the problems with the design and data. The researchers' original question had been in connection with a query from another scientist as to why they had done a two-sample F-test as well as a two-sample two-sided *t*-test (one-sided would have been better here), where their two samples were the 439 plants and the 42 plants chosen from these 439. Answering the question as asked would have been easy, but answering this without probing into the situation would have left many other misunderstandings as to whether the procedure itself was correct. This would be an excellent example to illustrate the risk of jumping to a hasty conclusion as to a client's needs, and could lead into a discussion of how to deal with the data.

STATISTICAL INVESTIGATIONS REVISITED

Two of the aims of a unit called *Statistical Investigations* (described more fully in Jolliffe, 2002) were "To give students experience of a range of communication skills in a statistical consultancy context" and "To enable the student to understand and apply a range of

problem solving techniques in practical contexts." The unit was intended to equip students for work as a practical statistician and formed part of an MSc degree in Applied Statistics designed to convert graduates in other disciplines into statisticians. In the year when the author taught the unit she hoped to give the three students taking the unit an overview of statistics, and to help them see statistics as a whole rather than as sets of techniques packaged in short and largely independent modules. About 10% of the 37 hours available for the unit was spent entirely on consultancy, but other sessions, such as those on investigation of web resources and on case studies were also relevant, as were many of the assessment tasks the students were asked to do. About a third of the course consisted of lectures on topics which had not been covered elsewhere in the degree course, including general issues in data analysis, an overview of the design of experiments, and a brief introduction to simulation.

The students took part in three activities which specifically involved consulting skills. One was a group activity writing a response to a scientist who had asked for comments on some notes she had written on sample surveys. Another was observing another lecturer giving statistical help to undergraduate psychology students who were writing up their projects. The third activity involved the students taking turns to play the role of client or consultant on mock consultancy problems based on real questions asked on an email broadcast system, watched by the author and the third student (see Jolliffe, 2002). Had the group of students been larger this activity would have been difficult to organize, but worked well with the small group. We had hoped to arrange for them to act as consultants to the psychology students. This would have given them a chance to practise the skills needed, and where wrong advice would have mattered in a way that it did not in the class exercise, but timetabling constraints and the fact that the psychology department was on a different site meant we could not organise this.

The MSc degree has not run since the author taught the unit Statistical Investigations, so there has been no opportunity for her to use her experience in teaching it, or her later experience in the practice of consultancy, to make any changes to it. The following discussion considers what changes she would make if she were to teach it or a similar unit again.

In general there were too many lectures in the unit so the number would be cut substantially. Students could easily read up some topics, particularly on the presentation of data, for themselves, making them more responsible for their own learning. It was unfortunate that the design of experiments was not covered elsewhere in the degree. This was mainly because the degree was very intensive and it was not easy to fit in everything thought to be important, but also because the degree had grown out of a course given to trainee UK government statisticians for whom this topic was not greatly important. Giving an overview of the topic is better than nothing, but to be really useful in training statisticians to act as consultants needs to backed up by discussion of the practical issues involved in design, and by analysis of real data. This is one area where the presentation by the instructor would be improved by practice as a consultant. Even in designing an experiment there are often constraints causing departure from the standard designs, for example there are severe limitations on the amount and type of land available on farms in developing countries. Ensuring randomness when counting pests on a sample of leaves is little more than a statistician's dream. There might well be other problems with the data, such as missing or fabricated data. Text-books tend not to discuss these non-standard cases.

Simulation is important when dealing with unusual problems and this is something requiring more time than it was given. This would usefully lead into randomization tests (Manly, 1997), a powerful technique when the assumptions needed for standard tests do not hold, or where there is no standard test. Here too, if the techniques are illustrated on problems that the instructor has met as a consultant, and the students work on such problems, they are more likely to see the importance of this topic and to remember it. Knowing something about how to deal with qualitative data and observational studies is useful for any statistician who does consulting, but in a short course giving out a reading list and some warnings about potential problems could be sufficient.

About a quarter of the unit was spent on case studies and data analysis and this seems a reasonable proportion, but rather than using published and perhaps out of date studies done by others, referring to studies and data on which the instructor had worked recently as a consultant would be preferable. Not only would the instructor know more about the studies - published ones

sometimes lack details which one would like to know - but, as Diamond and Sztendur (2002) remark, choosing problems that interest you, and showing an obvious enjoyment at having worked on them, will have a positive impact on students. More effort would be made to give students an opportunity to act as consultants, and existing links with clients would be used to facilitate this. After suitable preparation in the classroom, students might first work with the instructor in providing help to clients, and later might be able to act as the main consultant under a minimum of supervision.

The unit could be adapted for other types of student, with the lecture content adjusted to suit the students' background and level. More emphasis could be placed on consulting if the students were acquiring their knowledge of statistics elsewhere or if they already had a good knowledge of statistical methods. If necessary, the problems the instructor had met as a consultant might be simplified if too complex for the group (see Diamond and Sztendur, 2002).

CONCLUSION

The instructor's main role in teaching consultancy skills to statisticians is to guide them as to what skills are appropriate and, perhaps, to observe them acting as consultants. Teaching of consultancy skills involves not only emphasis on the communication skills needed in interchanges between statisticians and clients, but also in making statisticians aware of the wide range of statistical problems that they might meet, telling them about specialised and less well-known statistical methods, and, most importantly, in showing them how to deal with those problems which are not standard text-book problems. It is in the last of these that an instructor who has practised as a consultant will add value in the way that an instructor who has no practical experience of statistics or consulting cannot.

REFERENCES

Belli, G. (2001). The teaching/learning process in university statistical consulting labs in the United States. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics*, 325-338. International Association for Statistical Education and International Statistical Institute.

Cabrera, J. and McDougall, A. (2002). *Statistical Consulting.* New York: Springer-Verlag.

Cordani, L. (2003). Consultancy can be of service to teaching, research, and extension activities. In *Proceedings, Bulletin of the International Statistical Institute 54th Session.* Published as a CD by the International Statistical Institute, www.stat.auckland.ac.nz/~iase/publications/.

Diamond, N. T. and Sztendur, E. M. (2002). Simplifying consulting problems for use in introductory statistics lectures. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics*, Cape Town. Voorburg, The Netherlands: International Statistical Institute.

Jersky, B. (2002). Statistical consulting with undergraduates - a community outreach approach. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics*, Cape Town. Voorburg, The Netherlands: International Statistical Institute.

Jolliffe, F. R. (2002.) Statistical investigations – Drawing it all together. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics*, Cape Town. Voorburg, The Netherlands: International Statistical Institute.

Jolliffe, F. R. (2003). Communication, collaboration, and consulting. In *Proceedings, Bulletin of the International Statistical Institute, 54th Session.* Published as a CD by the International Statistical Institute, www.stat.auckland.ac.nz/~iase/publications/.

Manly, B. F. J. (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology* (2nd edition). London: Taylor and Francis CRC Press.