SIMPLIFYING CONSULTING PROBLEMS FOR USE IN INTRODUCTORY STATISTICS LECTURES ®

Neil T. Diamond and Ewa M. Sztendur
Victoria University of Technology
Australia

Using real life examples in Introductory Statistics courses is now accepted as very desirable and even necessary. There are many resources available on the Internet and elsewhere that makes this particularly easy. In this paper we discuss some advantages of using homegrown examples, obtained mainly from consulting work. However, to be able to use these examples in undergraduate courses they must be framed in such a way that the students can understand them. We also give a number of examples where we have done this.

INTRODUCTION

There has been a substantial improvement in the quality of statistical education over the last ten years. One of the main reasons for this improvement has been the emphasis on exposing students to real data, as advocated by Cobb (1991). The use of actual examples motivates students and prepares them for using statistical techniques in the real world. In this paper, various sources of such data will be briefly discussed and the advantages of using the Instructor's own consulting problems will be outlined. Some examples of how we have done this in our own teaching will be presented.

SOURCES OF REAL LIFE EXAMPLES

The students themselves can generate some of the best real examples. When students produce their own data (e.g. Cobb (1991), Magel (1996)) they gain experience in both the design and analysis aspects of statistics, usually have a lot of fun, interact more with their classmates and learn more in depth than in a traditional lecture.

However, generating data is time-consuming so it should not be done all the time. In any case, there are many excellent and interesting real datasets available, which can be used by the Instructor. Singer and Willet (1992) give a guide to sources of such data. There are also a number of collections of data sets available in book form such as Andrews and Herzberg (1995), Hand et. al. (1994) and Chatterjee et.al. (1995).

There are also many online archives available on the Internet. Besides the data-set archive on StatLib (http://lib.stat.cmu.edu/), a very convenient source is DASL (Data and Story Library) (http://lib.stat.cmu.edu/DASL/), a collection of data sets and stories appropriate for illustrating the use of some statistical methods. The library serves as a resource for teachers of statistics providing them with real-life and interesting examples. Witmer (1996) presents an interesting review of DASL and gives examples of its use.

Another source of interesting datasets and problems can come from the Instructor's own consulting. Our own experience is that these problems are very valuable and add much to the learning experience of the student. Before giving some examples of how we have used some consulting problems in our own lectures, we outline the benefits and some of the drawbacks in using them.

ADVANTAGES AND DISADVANTAGES OF CONSULTING PROBLEMS

Using examples generated from consulting problems has many benefits. Since the instructor has been involved the background is easy to get across to the students. Even in a team teaching situation when one instructor has been connected with analysis of a particular dataset, the other instructors are able to get a much greater appreciation of the context in which the data was generated than they would by obtaining the dataset from an archive. In addition, most consulting problems in which we are involved are local, near where the students study and live, and hence the students can readily relate to the data and are keen to see how it should be analyzed, and what the implications are.

Using consulting problems is consistent with the calls of many statisticians to emphasize Statistical Thinking rather than techniques, (see e.g. Wild and Pfannkuch (1998)). In particular, the difference between Mathematics and Statistics can be emphasized. This is particularly important for our students who either major in Computer Science with a minor in Mathematical Sciences or do a combined degree with about equal amounts of Mathematical Sciences and Computer Science. We want to emphasize to our students how statistical problem solving is based on data and how the context of the problem is very important in solving the problem.

However, many consulting problems are quite complex and messy. If a problem with all this complexity is presented, particularly in very elementary courses, there is a risk that the students will be confused or worse, convinced that statistics is too difficult for them. What must be achieved is a balance between the excitement and pleasure of tackling real problems and extracting only the most important aspects of the problem to make it within the students' reach. Our approach is to take real consulting problems and, if necessary, to modify them so that the students are able to appreciate the most important aspects of the problem.

HOW TO SIMPLIFY CONSULTING PROBLEMS

We have set out below some advice, based on our own experience, for those considering introducing consulting problems in their lectures.

1. Use problems you enjoyed working on. You are more likely to get the students interest if you choose problems that interested you. If you show your obvious enjoyment this will have a very positive impact on the students.
2. Give plenty of background. This is very important. As Cobb (1999) has said "In Mathematics, context obscures Structure……In Statistics, context provides meaning." Explain the context using pictures, photographs or videos.
3. Ask yourself what was the real problem and how was it solved. Clarify all the steps you and your collaborators took in solving the problem. Remember that steps you now find obvious may be difficult for the students.
4. Remove messy aspects of the problem that are not within the student's reach. The amount of cleaning obviously depends on whether the problem is introduced in a lecture or used as an assignment where some of the messier aspects of the problem can be retained
5. Summarize the data using appropriate graphs, such as Box-plots or histograms.
6. Let the students have a go at solving the problem before you explain what you did. It's not much use explaining to the students what you did unless they have first had a go themselves. By doing this the students become actively involved.
7. Make the students work in groups. As Garfield (1995) reports "…students appear to learn better if they work cooperatively in small groups to solve problems and argue convincingly among conflicting ideas and methods." One of the benefits of working on consulting problems is the interaction fostered within the class. A series of written questions can provide a focus to their discussion. Encourage the students to use pictures as far as possible to develop and explain their solution.
8. Reinforce, revise and link. Using consulting problems is a great opportunity to show how the various aspects of the course fit together. The links that are made foster a much deeper understanding of the subject. This is a good time to discuss some of the messier aspects of the problem solving process such as what should be measured and how the sampling should be done. These discussions will prove very useful when the students collect their own data.

SOME EXAMPLES OF CONSULTING PROBLEMS

In this section we give a number of examples with discussion of modifications to the original problem we have sometimes used so that a simpler statistical technique is appropriate.

Example 1 gives an unusual application of regression analysis. We like it as a motivation of regression as it is very clear how the company must use a regression analysis to determine how to produce the delay detonators. We have sometimes used a video to illustrate how a blast is staggered prior to discussing the problem in depth, and pictures of detonators and detonator relay connectors have been used to add interest. One complication is that with the actual data a weighted regression analysis is required, since the variation in times increases as the length of the delay element increases. For a first year class we fudge the data so that the variation is constant, while for a second year class we give the real data with a promise that we will be able to appropriately analyze the data using weighted least squares later on in the course.

Example 1: Cutting Length for Delay Detonators

XYZ manufacture delay detonators. They are used to stagger the timing of blast hole detonation, thus increasing the rock moving capacity of heavy explosives. Delay detonators are manufactured in various nominal delay times (such as 17 ms, 25 ms, 35 ms, …) and individual detonators need to adhere to these times as precisely as possible.

Delay element composition is imported. The composition is drawn into long lengths and ten elements each of three different lengths are cut, added to detonators, and timed. Based on a regression analysis of the results, a production length of elements is determined.

Problem in Example 2 has a number of levels of complexity to it. It has proved useful for introducing the standard deviation of the mean, as well as giving another application of inverse normal distribution problems. In order to do this, many of the complexities can be removed. There is no need to introduce the second clause of the weights and measures act since, in practice, the mean requirement is the more stringent. In addition, the fact that a density measurement is required makes the problem too complex as too does the details of how the machine is set up. It does not detract for the student to imagine that the filling machine has a dial giving the target weight required.

Example 2: Overfill Problem

The (old) Australian Weights and Measures act specifies that the mean of 12 cans of a liquid should be greater than the nominal amount printed on the can, and every can should be at least 95% of the nominal amount. A paint manufacturer needs to determine the target volume so that the chance of violating the weights and measures act is very small, say 0.001. Although the paint is sold by volume it is manufactured by weight. The density of the paint is determined using an necessarily imprecise measurement, the desired weight determined and then machine is adjusted until paint delivered in a can balances a check can with the specified weight in it.

One way of introducing this problem is to cut out shapes from a piece of cardboard of the normal density both for individual cans and for means of size 12. We mark on an overhead transparency a scale corresponding to the volume of paint in the cans. We then ask a student to set an appropriate target by placing the normal distribution shape for the means on the transparency so that there is a very small chance of the mean being less than the nominal amount. The normal distribution shape for the individuals can be used to show that with this target there are quite a significant percentage of cans that will be underfilled. Later in the lecture, the students can obtain a more exact answer for the appropriate target using inverse normal calculations.

Example 3: Compliance Problem

A water company samples a treated effluent once a week and measures the concentration of a pollutant. The company had a license with the Environment Authority that 50 percent of the samples in a year have to be below 20 mg/L while 90% of the samples have to be below 40 mg/L. The distribution of the effluent can be considered to be lognormal. Determine the probability that the water company complies with the license.

This is a very rich problem, involving conditional probability, binomial and normal distribution calculations. Introduced wisely, it is suitable at the end of the probability section of a course. However since the problem is quite complex, modifications must be made to make the problem suitable for an introductory statistics course. The simplest change would be to amend the license requirement so that 90% of the samples have to be below 40 mg/L, with no requirement that 50% of the samples need to less than 20 mg/L. This simplifies the problem so that it involves only normal and binomial calculations. However, with appropriate modification and introduction even the original problem can be made accessible to an introductory statistics course.

DISCUSSION

Real world examples are an essential part of statistical education. Such examples motivate students and make their learning meaningful since they can see the purpose and real use of the material that is presented. There are many advantages if these problems come from the teachers own consulting. In cases where the consulting problem is too complex for the students, appropriate simplifications can be made to make it easier for the students to understand and still get an appreciation of the use of statistics in solving real problems.

REFERENCES

Andrews and Herzberg (1985*), Data: A Collection of Problems from Many Fields for the Student and Research Worker*, New York: Springer.

Chatterjee, S., Handcock, M.S., and Simonoff, J.S. (1995*), A Casebook for a First Course in Statistics and Data Analysis*, New York: John Wiley.

Cobb, G. W. (1991), Teaching Statistics: More Data, Less Lecturing, *Amstat News*, 1, 4.

Cobb, G. W. (1999), Discussion of Special Section-Undergraduate Statistics Education: What Should Change? *The American Statistician*, 53, pp. 16-21.

Garfield, J. (1995), How Students Learn Statistics. *International Statistical Review*, 63, 1, pp. 25-34.

Hand, D. J., Daly, F., Lunn, A. D., McConway, K. J., and Ostrowski, E. (1994*), A Handbook of Small Data Sets*, New York: Routledge, Chapman Hall.

Magel, R. C. (1996), Increasing Students Participation in Large Introductory Statistics Classes, *The American Statistician*, 50, pp. 51-56.

Singer, J. D., and Willett, J. B. (1992). Annotated Bibliography of Sources of Real-World Datasets Useful for Teaching Applied Statistics, *Statistics for the 21st Century,* F. Gordon and S. Gordon, Washington, DC: MAA, pp. 298-313.

Wald, C., and Pfannkuch, M. (1998). What is Statistical Thinking? In Proceedings of the 5th International Conference on the Teaching of Statistics (ICOTS 5), pp. 333-339. Eds. Pereira-Mendoza, L., Kea, L.S., Kee, T.W., and Wong, W.-K. Vooburg:ISI.

Witmer, J. A. (1996), DASL – The Data and Story Library*, The American Statistician*, 51, p. 98.