TEACHING SAMPLING METHODS USING ECOLOGICAL EXAMPLES

Bryan F. J. Manly, University of Otago, New Zealand
Lyman L. McDonald, WEST Inc., USA

Ecological examples provide a rich source of material for class exercises as part of a course on sampling methods. These examples can be used for teaching the classical theory for the estimation of the parameters of a finite population, in which case they help students to learn the practical aspects of the design of studies, and the appropriate use of the standard equations for estimation with simple random sampling and more complicated sampling strategies. In addition, they can be used to introduce the ideas of encounter sampling where observations cannot be identified except as a result of carrying out the sampling. We give examples of the types of ecological examples that can be used and argue that it is only through carrying out the design of sampling procedures, and collecting and analysing data themselves that students can properly understand statistical principles.

THE OBJECTIVES OF A COURSE ON SAMPLING

The type of course that we are discussing in this paper is one that follows on from the typical first year university course on statistical methods in general. Traditionally, such courses have tended to have emphasized the analysis of data rather than design aspects that are concerned with deciding how to do the sampling in the first place. We suspect that in the past many students who have just completed courses on sampling methods would feel very uncomfortable if asked to propose a sampling strategy for collecting data to answer some specific questions, particularly if, as is often the case, the first problem is to translate the questions into statistical terms.

Our courses on sampling are intended to have a different outcome. We stress the importance of all aspects of sampling procedures, from the interpretation of the original questions through to the production of the final report. If necessary, we prefer to limit the range of sampling designs studied rather than to compromise the ability of the students to use simple designs well. Indeed, because we believe that students can only understand and use sampling procedures properly through experience with practical projects that are fairly time-consuming we accept that the topics covered must be somewhat constrained in comparison with what could be done without such projects. It should be noted, however, that if students work in groups of 3-5 people for these projects then the amount of work per person can usually be kept within reasonable bounds.

Using ecological examples is clearly appropriate in a course on sampling that is mainly intended for biologists. However, even a course with a more general audience can benefit from the inclusion of some examples of this type. The reason for this is that many ecological examples are inherently interesting in their own right, and some special sampling methods such as mark-recapture are most easily described in this context. We emphasize ecological examples in this paper whilst, of course, accepting that examples from other areas of application will sometimes be more suitable.

CLASSICAL SAMPLING PROCEDURES

By classical sampling procedures we mean those that relate to the estimation of means, totals and proportions for a population that consists of a finite number of sample units, as described by Cochran (1977), Scheaffer et al. (1990, Chapters 1-10) and Manly (1992, Chapter 2), for example. A sampling course will usually cover these methods in some detail, starting with simple random sampling and including some or all of the techniques of stratified sampling, cluster sampling, multi-stage sampling, and ratio and regression estimation.

There are abundant opportunities for interesting student projects here, quite early in the course. For example, students can be asked to estimate the total number of blades of grass, and the mean length of these blades in an area on the campus as soon as the estimation of means and totals from a simple random sample has been covered in class. This seems on the face of it to be a simple exercise. However it is useful as a means of raising a number of important practical issues for the first time.

Some results that we see from carrying out this exercise are:

- The choice of the size of the sample unit (a plot of grass) is found to be important. With large units the counting of blades of grass becomes very tedious. On the other hand, sampling a large number of very small units is not convenient either.
- It is necessary to define what is meant by a ‘blade of grass’, and the ‘length’ of such a blade, demonstrating that sampling results usually depend on the definitions used for measuring variables.
- Sampling plots of grass at random over a large area requires marking out the area and dividing it up into plots. This is also necessary for determining the total size of the study area if it has an irregular shape. This setting up takes some time to do properly, demonstrating that the practical aspects of defining the

population and randomly sampling it are not necessarily easy.

- With several people doing sampling it is important to ensure that they are all using the same methods for counting and measuring. If a project group does not realize this in advance then they are likely to discover that they end up with inconsistent results from different members.
- Students get practice at taking a pilot sample and using this to decide on the sample size for the main study.

In addition, it may be visually apparent that the density or length of grass varies in different parts of the study area. This leads immediately to the idea of stratified sampling even if students have not met this concept yet in the class.

It is very instructive to have several groups carrying out the grass sampling independently at the same time. Typically they will adopt different approaches. For instance, if stratified sampling has been covered in the class then some groups will use this but others will not. Groups are also likely to end up with very different estimates, demonstrating clearly that non-sampling errors can be far more important than sampling errors. This is shown to a rather extreme extent by the results in Table 1 which were obtained by three groups independently sampling the grass over an area at the University of Otago with a river running through it. The results for the three groups were spectacularly different. With regard to the mean length, group 1 seems to have a realistic mean and standard error, group 2 clearly got its units of measurement wrong, and group 3 had an unrealistic standard error. With regard to the number of blades of grass, all that can be said is that the groups do not appear to have been working in the same place!

Students find comparisons like this for the results from their first sampling exercise to be a good illustration of how badly things can go wrong with something that at first sight did not appear to be at all difficult. The hope is that they will learn to be more careful in the future.

Other ecological sampling projects are very easy to devise. For example, one that has been used at the University of Wyoming involved estimating the density, mean number of stems, mean height, and the frequency of occurrence of mountain mahogany in an abandoned limestone quarry using stratified sampling. Similar exercises can be used with any plant and have been applied on University of Wyoming grounds for estimation of 'population' parameters of a certain species of shrub used in landscaping . If it is desired projects can also easily involve more complicated statistical methods. An example is to estimate the biomass of a plant in an area using 'eye estimates' for all plots in the area and precise assessments of biomass on a random sample of the plots.

Table 1 Estimates and their estimated standard errors as obtained for three groups working on the grass project.

Group	Sampling Method	Sample Size	Estimated Mean Length (mm)	Standard Error	Estimated Total Blades of Grass (millions)	Standard Error
1	Quadrats (5m by 5m), subsampled	40 quadrats	28.3	2.2	961	87
2	Cluster sampling	10 clusters	254	28.0	13	1.4
3	Quadrats (5m by 5m), subsampled	40 quadrats	14.3	0.1	338	33

ENCOUNTER SAMPLING

One of the characteristics of many plant populations is that the individuals plants are rare and hard to find. For such populations classical sampling methods are often not suitable because there is no satisfactory way to take a simple random sample from the population. Likewise, animals are mobile and it is again often not possible to take a random sample from a population.

To handle situations where conventional random sampling of biological populations is not possible, encounter sampling methods have been developed, where essentially these involve sampling the individuals that are encountered by an observer moving through the study area (Scheaffer et al., 1990; Manly, 1992, Chapter 3; Thompson, 1992, Chapters 16-19). Often the primary purpose of such methods has been to estimate the total number of individuals in the population, but an estimate of the mean per individual or the population total can also be determined for a variable that is measured on the individuals.

A student project was conducted at the University of Wyoming where the objective was to estimate the total number of vehicles using the parking lots on campus during the hours 6:00 pm to 7:00 pm of a five day period. The analogy of the vehicles with animals using a certain biological resource is clear. There are various approaches to design of a study, but the one promoted by the instructor was to estimate the ratio of total "vehicle hours" to the "mean time per vehicle" during the five hour period.

Other student projects conducted by the authors include: (1) a poll of a subset of the student body (or faculty) for opinions on a controversial environmental topic, (2) borrowing equipment from the botany or forestry department and estimating the mean height of trees on the campus, and (3) arranging for an entomologist to make available an aquarium of easy to handle insects or a fisheries biologist to make available an aquarium of easy to handle fish, and estimating the number of insects or fish in the aquarium by capture-recapture methods. It has been our experience that interesting student projects can be devised to illustrate the design and analysis issues connected with all topics in a statistics course.

DESIGN-BASED AND MODEL-BASED INFERENCE

One thing that we like to emphasize in a sampling course is the difference between design-based and model-based inference (e.g., Dorfman, 1993, Ghosh and Sinha, 1990). This difference has not in the past received much emphasis. However, we believe that potentially it is of considerable importance, particularly in controversial situations, and is not something that should be glossed over in a sampling course.

The difference is, of course, that with design based inference we consider the variation in estimates that is inherent in the random sampling process, whereas with model-based inference the interest is in the variation in estimates that can be expected because of the random variation in the error terms in the assumed model. An advantage of the design-based approach is that the validity of inferences is guaranteed providing that the sampling process has been properly carried out. On the other hand, model-based inference has a wider range of applications but does require the assumption that the model is correct.

A student project that will illustrate these issues was conducted at the University of Wyoming. Teams of students were instructed to estimate the total number of stems in a dense planting of shrubs (on the edge of the soccer field) by counting all the stems on a relatively small set of randomly selected shrubs and estimating by 'eye' the number of stems on a much larger sample. In an example like this interesting discussions can revolve around the issue of: (1) where does design-based analysis stop and model based analysis start, (2) which model is best for expanding the results in this double sampling exercise to the population, and (3) how to evaluate the variance of the estimated total.

REFERENCES

Cochran, W. G. (1977). *Sampling Techniques*, 3rd Edit. Wiley, New York.

Dorfman, A. H. (1993). A comparison of design-based and model-based estimators of the finite population distribution function. *Australian Journal of Statistics 3,* 29-41.

Ghosh, M. and Sinha, B. K. (1990). On the consistency between model- and design-based estimators in survey sampling. *Communications in Statistics: Theory and Methods,19*, 689-702.

Manly, B. F. J. (1992). *The Design and Analysis of Research Studies*. Cambridge University Press, Cambridge.

Scheaffer, R. L., Mendenhall, W. and Ott, L. (1990). *Elementary Survey Sampling*, 4th Edit. PWS-Kent, Boston.

Thompson, S. K. (1992). *Sampling*. Wiley, New York.