Sample Space : Practical Experiments for Teaching Sampling

Peter Petocz - Sydney, Australia

1. Introduction

A traditional view of statistics is that it is *numerical detective work.* An important part of this detective work is the collection of the data which will form the basis of inferences and action. Selecting a sample is a key activity in statistics. If the data collected are defective in some way - for instance, because of bias - then the inferences and actions based on them can be seriously flawed.

I believe that every course in statistics, even one at an introductory level, should include some inferential statistics. This can be presented in the usual formal way via estimation and hypothesis testing. Alternatively, it can be presented in an informal way. For instance, general questions about a situation can be posed on the basis of a histogram of sample values. Whichever approach is used, it is then necessary to introduce some of the ideas of sampling as a method of checking on the plausibility of the inferences.

However, in many courses it is not necessary or desirable to go too deeply into the mathematical aspects of sampling. Students need an idea of the importance of the topic, a knowledge of various methods of selecting a sample, and a comparison of the efficiency of alternative methods.

An experimental approach can be very effective in this situation. This paper presents a series of practical experiments that can be used to demonstrate the ideas of sampling at all levels, from introductory to advanced. The material presented consists of a brief introduction to sampling, the "Sample Space" star map, and two sampling experiments based on the map. It is presented in a form that can be used directly for teaching classes or for individual study. Further experiments can be found in Petocz (1990). Three of these were described in the original draft of this paper, as presented at the ICOTS Conference, and are available from the author on request.

2. The experiments

Any experiment in sampling from a finite population relies on a complete listing of the elements of the population and some random method for selecting samples. Samples are selected and are used to estimate a population parameter such as the mean or the total. Estimates obtained using repetitions of various methods can be compared, and conclusions drawn about the efficiency of the sampling methods.

There are several pedagogical advantages in using an experiment for teaching sampling. First, it provides an opportunity for an interesting and practical alternative to traditional teaching or lecturing. Second, it facilitates the development of intuitive ideas about sampling - ideas which are otherwise difficult to acquire. Third, it allows the examination of sampling plans whose mathematical details would be beyond the capabilities of the students: alternatively, for mathematically advanced students, it enables the mathematical details to be illustrated in a practical way. Fourth, it leads naturally into an investigation of key statistical concepts such as variability and sampling error. Finally, it can provide motivation for further study of the mathematics needed in sampling theory.

"Sample Space" has all the features for a successful teaching experiment. The complete population - the numbers of stars in each of the 100 squares - can be seen at one glance. It would seem to be an easy matter to enumerate the whole population, and hence obtain population characteristics such as the mean or the total. However, a complete enumeration is not as easy as it seems. Although the population cannot be easily enumerated, the pictorial nature of the experiment makes it possible to make visual checks on the results of the various sampling methods.

The experiments can be used in a variety of ways, and at a variety of levels.

For a group of complete beginners with no statistical background, for instance, a high school class, they can be used to illustrate the statistical approach to a problem - an approach that usually results in reasonable but not exact answers.

For a group of students taking a first statistics course at tertiary or pre-tertiary level, they can be used to introduce key statistical concepts such as variability and sampling error in a non-technical way; to introduce the basic terminology and techniques of sampling; and as a starting point to develop statistical intuition.

For a more advanced group, for instance, tertiary students studying sampling, they can be used to introduce, evaluate and compare various methods of sampling; to illustrate in a practical way the theory of sampling from a finite population; to motivate the further study of the mathematics needed in sampling theory.

The experiments always seem to generate a lot of interest, whether used with groups that have no statistical background at all or with more statistically advanced groups. Students like the strong visual aspects of the problem - all the information is there in the star map, yet it is not as easy as it seems at first sight to find the total number of stars by direct counting.

The strengths and weaknesses of the statistical approach are very evident in the experiments. Reasonable answers are obtained without too much effort, and although in this situation we have the option of counting all the stars, it is apparent that in other situations this might not be possible.

The nature of variability and its connection with sampling theory comes across well. Classes are often involved in discussions about why all their results are different, and they appreciate that variability is one of the key ideas for comparing sampling methods.

3. Procedure

We will examine various methods for selecting a sample. Each method will be illustrated by a sampling experiment using the "Sample Space" star map. This map, shown on a following page, has been prepared from a photograph of part of the night sky. The negative was used, so the stars appear as black dots and the background sky is white. We will be trying to find the total number of stars pictured. This will be done by selecting some of the squares, counting the number of stars in each, and then using this information to guess the total number of stars.

In statistical terms, we have a population whose values consist of the number of stars in each of the 100 squares in the map. The counts we make in the selected squares form the sample. In this particular experiment the population is finite and not too large. It is possible to find the total number of stars directly by counting up the stars in each of the 100 squares. However, this is not as easy as it seems: it is quite simple to make mistakes in the counting, especially in the squares containing many stars!

For each experiment in this section, we will sample 10 squares from the 100 available, and use these to estimate the total number of stars. If you are working in a class group, you will each carry out the experiments, and the combined results will give you an idea of how good the method of sampling is. If you are working individually you may repeat the experiments to obtain "class" results. The main aim of the experiments is not to find out how many stars there are, but to look at the methods of selecting a sample and the influence of these methods on the estimates that we obtain.

There are a number of ways of selecting a sample:

(i) *Non-random sampling*: the sample is not chosen in any random way. Often this is because there are only a limited number of values available. For example, in studying a rare illness, measurements can only be made on the people actually suffering from the illness. In other cases, the sample may be obtained by using volunteers; the experiment may be tedious or unpleasant, for example. Non-random sampling can cause serious problems with bias, but it may represent the only practical way of getting a sample.

(ii) *Simple random sampling*: here the sample is chosen in such a way that every member of the population has an equal chance of being included in the sample. Sample members are chosen independently of each other.

(iii) *Probability sampling*: each member of the population has a known (but not necessarily equal) chance of being included in the sample. The sample is chosen by some random method consistent with these probabilities.

(iv) *Stratified random sampling*: the sample is chosen randomly, but it is made to reflect the population in some specified characteristic. For instance, in a mining community consisting of 80% males and 20% females, respondents for a questionnaire may be chosen to reflect the unusual sex ratio. Thus, 80% of the sample will be obtained from the men, and 20% of the sample will be obtained from the women.

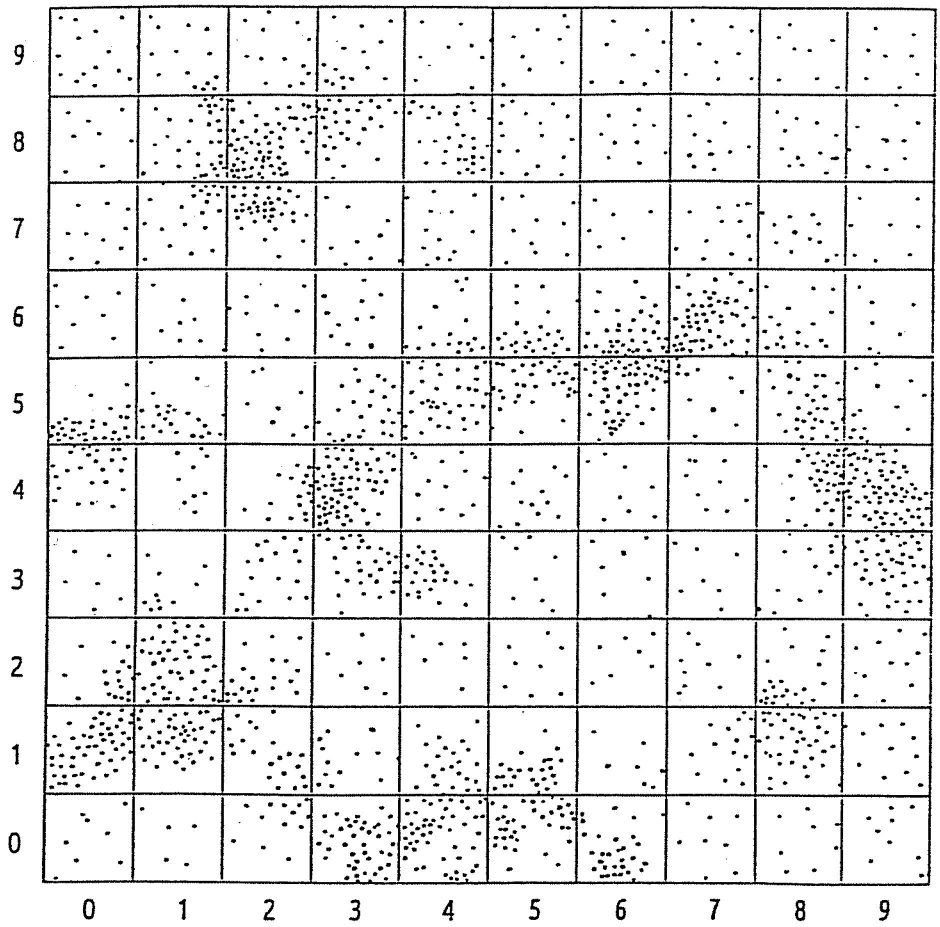

FIGURE 1
Sample space star map

(v) *Systematic sampling*: the first member of the sample is chosen randomly, and thereafter other members are chosen at regular intervals. This is often used in industrial quality control, for example. A particular item is randomly selected for inspection, and (say) every hundredth item is inspected after that.

(vi) *Two-stage sampling*: the population is divided into primary and secondary units; a sample of primary units is selected, and from each of these a sample of secondary units is selected. For example, if a sample of the apples in an orchard is required, a sample of trees might be selected first, followed by a sample of apples from each tree.

In each of the random sampling methods, the actual selections are made using a table of *random numbers*, or some physical equivalent like rolling dice or tossing coins.

Random number tables consist of digits randomly chosen from {0,1,2,3,4,5,6,7,8,9}: they should not show any systematic patterns in the digits. Often these tables are generated with a computer using specific formulas that produce numbers which cannot be distinguished from random numbers: these are called pseudo-random numbers.

The rows and columns in the star map are numbered 0 to 9. Let us agree to refer to each square by a two-digit number consisting of the number on the horizontal axis followed by the number on the vertical axis. So, for instance, the bottom right hand square will be '90', the bottom left hand square will be '00', and the top left hand square will be '09'. Remember to give the horizontal number first, and then the vertical number.

To use the random number table to pick a square, shut your eyes and stab at the table with a pencil. The digit you have selected, and the one next to it, will give you the number of the square you have picked. If you wish to select 10 distinct squares, use the numbers in the table starting from the point you selected and proceeding across or down the page. You should not select a new starting point for each square. As an example, if the digits from your starting point are 5167848432, the first five numbers you pick will be '51', '67', '84', '84' and '32'. You would ignore the second '84', and so the first four squares you count will be '51', '67', '84' and '32'.

You are now ready to start the experiments.

4. Experiment 1 : Simple random sampling

Use the table of random numbers to select the two-digit code numbers of ten squares. Remember that the first digit represents the number along the horizontal axis, and the second digit represents the number along the vertical axis. For example, square '42' is 5 squares along and 3 squares up, and contains 6 stars.

Write down the code numbers of the squares that you select. Simple random sampling is done *without* replacement, so you can't use the same square more than once. Ignore any repeated squares when selecting your sample.

Count the number of stars in each square you selected and fill in your results in the following table. If you are working as part of a class, fill in the section showing class results. If you are working individually, you may repeat the process a few times to get "class results". Finally, answer the questions under the tables.

	Squares selected	*Number of stars*
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

Individual Results:
$n = 10$, $N = 100$
Sample Mean $\overline{X} =$
Sample StDev $S =$
Estimate of Total $N\overline{X} =$

Class Results:
Estimates of Total Number of Stars for each group in the class:

(A) Mean of these values =
(B) StDev of these values =

Questions:

1. Why would people's individual results be more accurate if 20 squares had been sampled rather than 10? How much more accurate would they be, and how could you measure this accuracy? Would the extra accuracy have been worth the extra work?
2. What would happen to the accuracy of the estimates of the total number of stars if all 100 squares were counted?
3. On the basis of these results, what is your final guess for the total number of stars in the picture?
4. What value would you use to measure the variability in the estimates of the total number of stars made using simple random sampling?

5. Experiment 2 : Stratified random sampling

We divide the 100 squares into two subgroups or strata. Stratum 1 is the "high density" stratum. It contains the *twenty* squares with most stars in them (squares 01, 11, 12, 18, 27, 28, 30, 34, 40, 45, 50, 55, 65, 66, 76, 81, 84, 85, 93, 94). Remember, the first digit refers to the horizontal axis, the second digit to the vertical axis. Stratum 2 is the "low density" stratum. It contains the other *eighty* squares.

You will now take a sample of 10 squares with "proportional allocation", that is, two squares from the high density stratum and 8 squares from the low density stratum. Carry on using the random number table from where you left off previously. Go down the list of random numbers using the first two from the high density stratum and the first eight from the low density stratum, and neglecting all others. Once again, you cannot use repeated numbers, but it doesn't matter if you use any of the numbers from the previous experiment.

Count the number of stars in each square you selected and fill in your results in the following table. If you are working as part of a class, fill in the section showing class results. If you are working individually, you may repeat the process a few times to get "class results". Finally, answer the questions under the tables.

	Squares selected	*Number of stars*
Str 1		
1		
2		
Str 2		
3		
4		
5		
6		
7		
8		
9		
10		

Individual Results:
$n = 10 \quad (n_1 = 1, n_2 = 8)$
$N = 100 \quad (N_1 = 20, N_2 = 80)$
Sample Mean $\overline{X}$ =
Sample StDev S =
Estimate of Total $N\overline{X}$ =

Class Results:
Estimates of Total Number of Stars for each group in the class:

(A) Mean of these values =
(B) StDev of these values =

Questions:

5. What reasons could you suggest for using stratified random sampling rather than simple random sampling?
6. Would you expect the class mean estimate (A) of the total number of stars to be higher, lower, or about the same as the value obtained with simple random sampling?
7. Would you expect the variability of the class estimates (B) to be larger, smaller, or about the same as the value obtained with simple random sampling?
8. The *design effect (deff)* of a sampling scheme measures its performance relative to simple random sampling. If deff is less than one, the sampling scheme is better than simple random sampling: the smaller the value of deff, the better the scheme is. Calculate the design effect for this method of stratified random sampling using your values of the standard deviation S at (B) and the formula:

$$\text{deff} = \frac{S^2 \text{ under the sampling scheme}}{S^2 \text{ under simple random sampling}} .$$

6. Additional remarks

Population values: To carry out theoretical work, and to check the answers of people doing the experiments, it is necessary to have a description of the population. The histogram below shows the 100 values graphically.

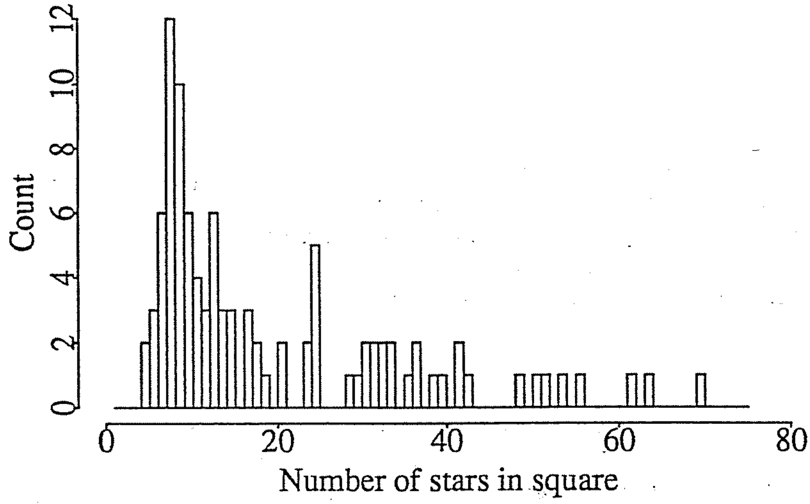

FIGURE 2
Histogram of the numbers of stars in each square of "sample space"

The summary statistics are:

$$N = 100 \quad \mu = 20.01 \quad \sigma^2 = (15.137)^2 \quad N\mu = 2001(\text{total})$$

With two strata (1 = high density, 2 = low density, and note that one of the squares containing 32 stars goes into each group):

$$N_1 = 20 \quad \mu_1 = 45.75 \quad \sigma_1^2 = (11.040)^2 \quad N_1\mu_1 = 915$$

$$N_2 = 80 \quad \mu_2 = 13.575 \quad \sigma_2^2 = (6.992)^2 \quad N_2\mu_2 = 1086$$

Answers to selected questions are as follows:

1. The accuracy of individual estimates of the total number of stars is measured by the *standard error*, and is given by the formula:

$$se(N\bar{X}) = N \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}.$$

This represents the standard deviation of the estimates of the total number of stars, and should be close to the value obtained at (B).
If n = 10, se = 456.40, and if n = 20, se =304.27. This represents a one-third reduction for double the amount of work.

2. If all 100 squares were counted then n = 100 and se = 0. Thus, the *sampling error* would be zero, although there may still be non-sampling errors from mistakes in counting.

8. The theoretical value is:

$$\text{deff} = \frac{242.37^2}{456.40^2} = 0.28.$$

A note on bibliography: Standard texts are Cochran (1967), or for readers with less mathematical background, Williams (1978). Many books on statistical methods contain chapters on sampling. For example, Snedecor and Cochran (1980) presents a one-chapter summary of sampling theory, and Freedman et al. (1978) contains chapters describing the ideas of sampling in non-technical language.

References

Cochran, W G (1967) *Sampling Techniques* (3rd ed). Wiley, New York.

Freedman, D, Pisani, R and Purves, R (1978) *Statistics*. Norton, New York.

Petocz, P (1990) *Introductory Statistics*. Nelson, Melbourne.

Snedecor, G W and Cochran, W G (1980) *Statistical Methods* (7th ed). Iowa State University Press.

Williams, B (1978) *A Sampler on Sampling*. Wiley, New York.