Why Stratify?

Ted Hodgson and John Borkowski
Montana State University, Bozeman, USA.
e-mail: Hodgson@mathfs.math.montana.edu

◆INTRODUCTION ◆

VALID statistical inference requires data that are representative of the population as a whole. Selecting a representative sample, however, is often easier said than done. For instance, classroom activities developed by Rouncefield and Holmes (1989) and Borkowski (personal communication) clearly illustrate the fact that, when left to their own devices, people tend to select "random" samples that are in fact distinctly non-random. It can be concluded, therefore, that there is a need to address sampling issues in the statistics classroom.

To assist with obtaining representative data, statisticians have developed an array of techniques, the most common being simple random sampling, stratified random sampling, cluster sampling and systematic sampling. Unfortunately, the distinctions between these techniques are often lost on students. Their performance in examinations indicates that they can collect samples of a designated type. Additionally, given information about a population, students can often identify an appropriate sampling technique. However, we contend that much of this behaviour is mimicry. Fundamental conceptual questions, such as "Why were the different techniques developed and what advantages does each offer?" for instance, are often met with blank stares. In other words, we have found that although students readily learn the mechanics of sampling procedures, they often lack any conceptual understanding of the procedures themselves.

In this paper, we offer an instructional approach to one of those procedures: stratified random sampling. Specifically, we have found that the following exercise serves as a useful introduction to stratified random sampling, helps students to develop a conceptual understanding of the procedure, and facilitates discussion regarding the characteristics and uses of the many sampling procedures. Moreover, the exercise accomplishes these objectives in an intuitive way, so that it can be used effectively in even the most elementary statistics course.

◆FURTHER BACKGROUND◆ AND DIRECTIONS

Stratified random sampling is an important statistical tool and is most appropriate when (1) the population can be divided into distinct strata, and (2) there is some reason to believe that the strata differ with regard to the characteristic of interest. For example, Shiver and Borders (1996) describe an "experiment" in which forest managers sought to estimate the potential harvest of conifer trees in a forested area. Since the forest was naturally divided into two strata - a 72-acre stand of Douglas fir and a 50-acre mixed stand of conifer and hardwood trees - and there was reason to believe that the potential harvests in each stratum differed, the researchers collected harvest samples from each and combined the results to estimate the overall potential harvest. Likewise, stratification plays an important role in the surveys of the United States Bureau of Labor Statistics (Scheaffer *et al.* 1990). The data used to calculate the consumer price index, for instance, are regional in nature and it is necessary to account for numerous regional differences, including geography, population size, percentage population increase, industrial descriptors, and levels of urbanisation and ethnicity.

To introduce stratified random sampling, we provide students with an unknown population (one contained in a paper bag) and instruct them to use sampling techniques to determine the population mean. In our course, which is based on Moore and McCabe's *(1993) Introduction to the Practice of Statistics,* students first consider simple random sampling and its characteristics. Subsequently, our presentation of stratified random sampling builds upon this knowledge. In particular, students collect both simple and stratified random samples of size 4 and compare the resulting estimates of the population mean and empirical sampling distributions.

The population, which is listed in table 1, consists of 40 cards (20 red and 20 black) each carrying a number with an overall mean of 18. In reality, one could conduct the activity with only ten cards - one card for each number - rather than 40 cards, as we use. However, if the former approach is used, then students can rightfully ask why one doesn't simply select the entire population as the sample. Moreover, since samples are collected without replacement, a smaller population is less likely to yield extreme cases (all red cards, for instance) than a larger one. A population of 40 cards promises to yield sufficiently diverse samples and encourages the use of sampling techniques (rather than a simple census), yet is of a size that is easily manageable in the classroom.

X	*Frequency*	*Card Colour*
6	4	Red
7	4	Red
8	4	Red
9	4	Red
10	4	Red
26	4	Black
27	4	Black
28	4	Black
29	4	Black
30	4	Black

To begin the activity, students draw simple random samples of size 4 and calculate the sample means. For example, the samples collected in a recent class of 34 students and the corresponding sample means are depicted in table 2 (the table is ordered by the sample mean). In the class itself, students' data are quickly compiled into tables and graphs, and these serve as the basis of whole-class discussion. Although we compiled students' data manually, this task lends itself to readily available technologies, such as a spreadsheet, statistics package or graphing calculator. With regard to our class, for instance, an initial inspection of the data revealed few trends. The sample means ranged from 12.25 to 24 and did not appear (at least from the table) to be centered about a particular value. Additional information regarding the underlying population can often be garnered from the sampling distribution. The students in our class,however, found that the distribution was uninformative (see figure 1). Overall, the consensus of the class was that the simple random samples provided little definitive information about the population mean.

Subsequently, we inform students that the deck actually consists of two subpopulations: cards that are red and cards that are black. Although they are not told that the subpopulations differ with regard to the characteristic of interest (which, in this case, is the numerical value of the card), students are informed that they *might* differ and that proper sampling procedures seek to account for potential differences.

Sample Number					***Sample Mean***
1	26	7	10	6	12.25
2	10	26	9	6	12.75
3	29	6	7	10	13
4	6	8	9	29	13
5	6	9	8	30	13.25
6	9	8	7	29	13.25
7	7	7	30	9	13.25
8	9	9	10	26	13.5
9	9	8	8	30	13.75
10	9	10	8	29	14
11	10	9	29	9	14.25
12	6	27	6	26	16.25
13	7	7	26	27	16.75
14	28	8	6	26	17
15	7	6	29	26	17
16	6	29	26	8	17.25
17	9	6	26	29	17.5
18	26	9	8	28	17.75
19	7	10	26	29	18
20	27	6	30	9	18
21	6	29	28	10	18.25
22	8	29	26	10	18.25
23	6	8	30	30	18.5
24	26	9	30	10	18.75
25	8	11	28	30	19.25
26	26	27	6	27	21.5
27	30	26	27	6	22.25
28	8	26	29	28	22.75
29	10	26	26	30	23
30	29	6	30	27	23
31	28	9	30	26	23.25
32	27	26	30	10	23.25
33	30	10	29	26	23.75
34	29	30	7	30	24

Table 2. Students' simple random samples

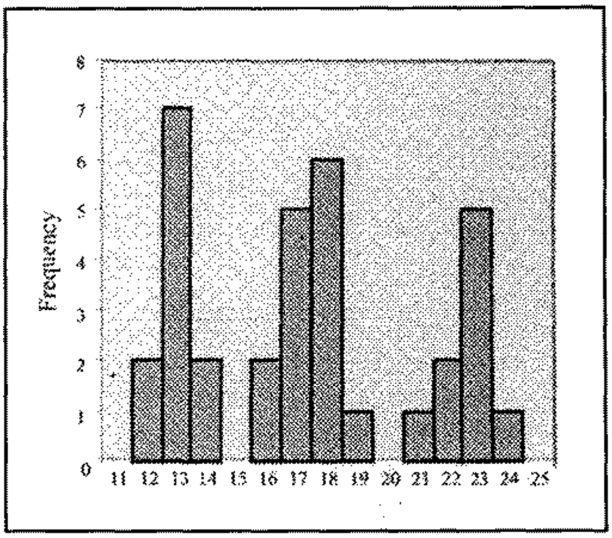

Fig 1. A frequency histogram of the means of students' simple random samples

Fig 1.
A frequency histogram of the means of students' simple random samples

One such procedure involves selecting red and black cards so that the percentage of each in the sample mirrors their percentages in the population. In other words, students are introduced to stratified random samples. Since the population consists of 20 red and 20 black cards, students draw samples of size two from each subpopulation (which we place in separate bags) and average the results.

Sample Number					***Sample Mean***
1	6	7	27	28	17
2	8	9	26	27	17.5
3	8	6	28	28	17.5
4	7	8	29	26	17.5
5	9	9	26	26	17.5
6	6	9	29	27	17.75
7	8	10	26	27	17.75
8	10	6	27	28	17.75
9	9	9	28	26	18
10	6	8	28	30	18
11	10	8	27	28	18.25
12	10	7	28	29	18.5
13	9	9	27	29	18.5
14	8	9	29	29	18.75
15	9	10	28	29	19

Table 3. Students' stratified samples

As it turns out, the subpopulations do differ with regard to the characteristic of interest. That is, the red cards correspond to "small" numerical values, whereas the value of all black cards is large. Thus, as data generated by the same class of students and contained in table 3 demonstrate, stratified random samples promise to yield more precise estimates of the population mean than simple random samples. In the class, fifteen stratified samples were drawn and, contrary to the averages obtained through simple random sampling, the resulting averages were consistently "close" to the population average of 18 (which is only revealed after students collect their stratified samples and calculate the second estimate of the population mean). Moreover, the distribution of stratified sample means was, at least in this case, less variable than that for simple random sampling (see figure 2). The educational impact of the simulation, therefore, is that students can *observe* the potential increases in precision that accompany stratified random sampling.

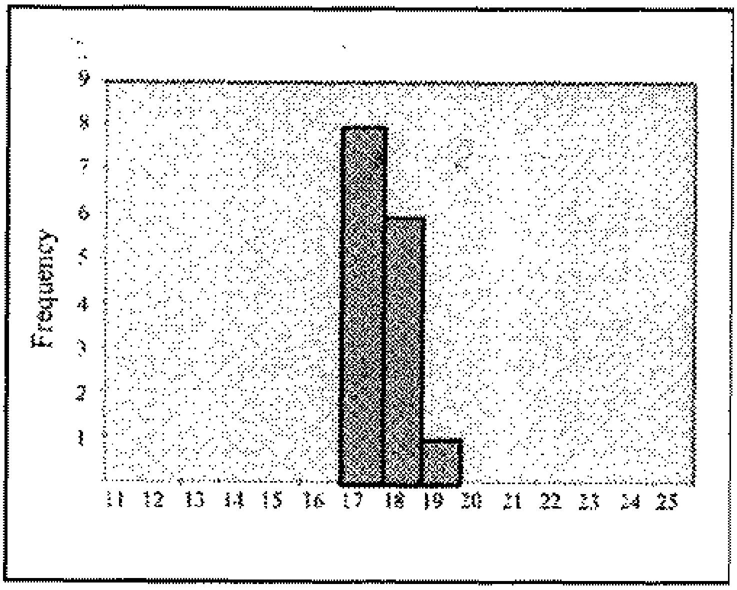

Fig 2. **A frequency histogram of the means of students' stratified samples**

◆CONCLUSIONS ◆ AND EXTENSIONS

In conclusion, we have found that simple simulation exercises, such as the one described above, help students to develop a greater appreciation for, and understanding of, commonly-used sampling techniques. In particular, these exercises encourage students to attend to two critical features of sampling procedures: the precision of the sample statistic and the variability of the resulting sampling distribution. Moreover, we have found the approach (simulation) to be an extraordinarily rich source of problems and investigations.

As an example, the preceding exercise focuses on a population in which there is an equal number of red and black cards. In most real world situations, however, the numbers of individuals in each of the identifiable subpopulations differ. Thus, a natural extension of the initial exercise is to vary the percentages of red and black cards in the population. Likewise, another extension is to ask what happens when the population can be stratified, but the subpopulations actually resemble one another with regard to the characteristic of interest. A simple simulation reveals that stratified random sampling and simple random sampling yield similar estimates in

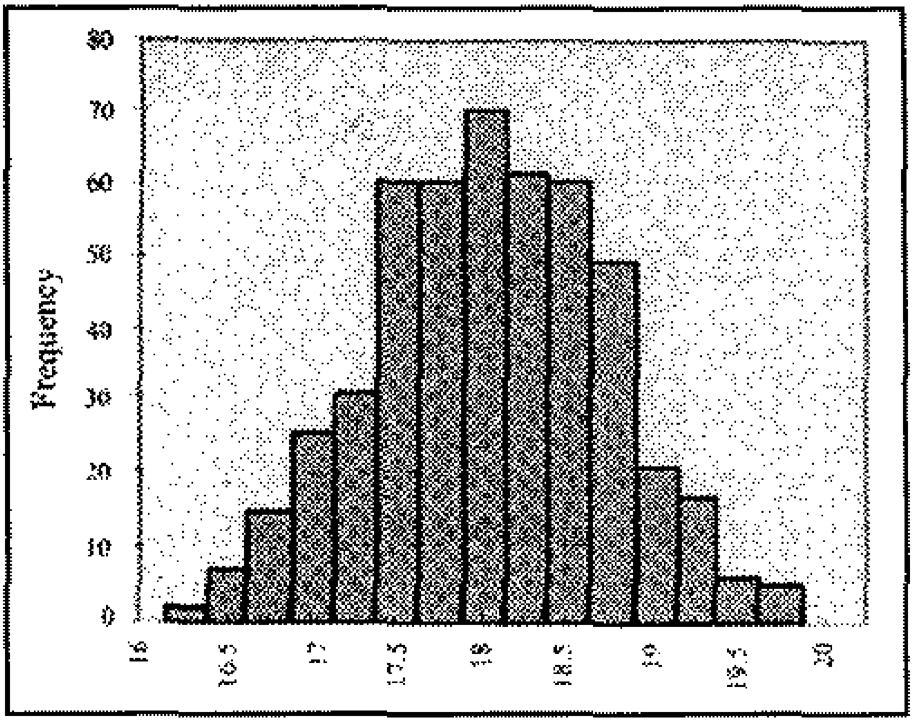

Fig 3. A frequency histogram of the means of 500 simulated stratified samples

this situation.

Lastly, the results of by-hand simulations can be compared with those of computer or calculator simulation. For example, figure 3 displays the results of a calculator simulation of 500 stratified samples. Note that the larger number of samples facilitates a finer grain of analysis and is suggestive of the true sampling distribution for this situation. In general, large-scale simulations can reveal the "true" differences between sampling techniques. Additionally, large-scale simulations can lead to discussions of other important statistical concepts, such as the effect of sample size on the results and the asymptotic behaviour of empirical sampling distributions.

References

Moore, D.S. and McCabe, G.P. *(1993).Introduction to the Practice of Statistics (Second Edition).* New York: W.H. Freeman.

Rouncefield, M. and Holmes, P. (1989). *Practical Statistics.* London: Macmillan.

Scheaffer, R.L., Mendenhall, W. and Ott, L. (1990). *Elementary Survey Sampling.* Boston: PWSKent.

Shiver, B.D. and Borders, B.E. (1996). *Sampling techniques for Forest Resource Inventory.* New York: John Wiley & Sons.