TEACHING ABOUT SAMPLING IN A MARKETING RESEARCH CLASS

David Finn, Texas Christian University

ABSTRACT

In the spirit of sharing that is characteristic of ABSEL, this paper describes a classroom exercise that the author uses to convey the meaning of a sampling distribution. Although students enrolled in marketing research courses usually are required to have passed an introductory statistics course, the concept of sampling distributions (so important in understanding sampling) seems to have eluded them.

A description of the class exercise, instructions to students, and the source program is included.

INTRODUCTION

For years I have been struggling with the problem of how to get marketing research students to understand the reasoning behind sampling and sample size theory. Text-book treatments, although usually very thorough, tend to be beyond the comprehension of the average student. Many texts use a standard reference and a similar set of figures to describe how the distribution of the sample mean approaches a normal distribution with very small standard error as the sample size increases (e.g., see [1] and [3] who both cite [2]). This is shown to be true when the statistic being measured is the mean, and with larger sample sizes, when the statistic is the proportion of sample members in a particular category.

Feed-back tests reveal that students are able to memorize the rules and formulae associated with such coverage, but when they are asked to actually use or explain the phenomenon, it becomes clear that understanding is often totally absent. This is not surprising, given that fact that many of us did not understand the concept until graduate school! Yet it is not a very complex topic! I have come to believe that the problem is in the methods we use to “teach” the idea.

Texts that treat the topic with patience (e.g., [1]) develop the idea of a “derived population” of sample means by showing a listing of all possible sample means that can be computed from a stated population (necessarily small so that it can be listed in the book). Then those means are plotted to show the distribution of means. Generalizations about larger populations follow.

To a trained reader, these patient treatments look elegant. Only when we get into the classroom do we realize that our students are either totally confused or have decided to skip “all that explanation”. The elegance is lost on them!

As a member of ABSEL, I KNEW that if it were possible to let students generate observations of distributions of means by themselves, they would have a much better learning experience. The problem was that the marketing research course presumes that students already know and understand the topic (since introductory statistics is a prerequisite course) and is usually not designed to devote much time to such topics - - besides (at least at our school), marketing research emphasizes planning, using, and evaluating marketing research as a decision aid and de-emphasizes the statistical methods used in research. That means that any experiential learning of statistical concepts should be a very MINOR part of the course. All of the above led to the development and trial of the exercise that I am about to describe.

THE EXERCISE

One week before the concept of sampling is introduced, students are given an assignment to complete. They are told to sign on to the main frame (although any system will work) and to execute a program called SAMPLE (reproduced in Figure 1). I explain to them in class and via a conversational introduction to the program that the program samples from a very large population (millions) of beer drinkers whose ages range from 18 to 65 (the “beer drinker” description is purely arbitrary - - in keeping with an established reputation). The only information that the computer will extract (they are told) is the age of each population member. What they have to do is ask the computer to sample from that population and to tell them: 1) the ages of the sample elements; 2) the average of the sample; and 3) the percentage of the sample that is in each of 4 different age categories (18 to 24, 25 to 34, 35 to 49, and 50 to 65 years old).

The assignment also instructs them to draw several samples of size 10, of size 20, of size 50, of size 100, and of size 200. The actual number of samples they must draw depends on the class size. Generally, I try to get between 75 and 90 different samples among the entire class (e.g., in a class of 45 students, I would have each student ask for 2 samples of each sample size, for a total of 90 samples of each sample size in the class). Since the computer lists the samples and the associated statistics on the CRT, I supply the worksheet in Figure 2 as an aid in recording the results.

The program that I have used (and which is presented in Figure 1) is a simple simulation with an average age of 36.5 years, 45% between 18 and 24 years, 10% between 25 and 34 years, 10% between 35 and 49 years, and 35% between 50 and 65 years. These are the parameters that the class will try to discover.

On the assigned day, when the students come to class with their worksheets, the learning begins. To start, I ask one student to look at his/her first sample of size 10 and to tell me and the class what the average age of the sample is. When s/he reports it (perhaps something like “25.76 years old”), I say something like “Right!” or “Very Good! How many got that?” Needless to say, most of the students look at me like I am some kind of idiot and point out that they are all different. Which leads to THEIR teaching ME about sampling error.

When everyone agrees that every sample SHOULD be different, I ask “how different should they be?” Since they don’t know, I suggest that we plot all their samples on the board to see if we can make any sense of them. “How many average ages in your sample of size

FIGURE 1

THE SOURCE PROGRAM FOR THE EXERCISE

Note: the lower case comments are directed to the reader and are not part of the program.

```
100 RANDOMIZE
340 PRINT " "
       :                                        (In this section you should insert whatever cover story
       :                                        and instructions you want.  For example, you might explain
       :                                        how many samples to take, what to do when they are done, or
       :                                        anything else that helps with the assignment.)
350 INPUT "WHAT SAMPLE SIZE DO YOU WANT?" ;Z
360 IF Z<= 500 THEN 390                         (this is to assure that the
370 PRINT "TOO BIG"                             students don't get carried away)
380 GOTO 350
390 IF Z>= 0 THEN 420
400 PRINT "TOO SMALL"                           (we want positive numbers)
410 GO TO 350
420 IF Z = INT(Z) THEN 450                      (and whole numbers only)
430 PRINT "WHOLE NUMBERS ONLY"
440 GOTO 350
450 PRINT " "
460 LET X=A=B=C=D=0                             (initializes at each repetition)
510 FOR J=1 TO Z
520 LET R=INT(RND(T)*100)+1                     (choose random num 1 to 100)
540 IF R>=1 AND R<=45 THEN 850                  (45% of the nums)
550 IF R>=46 AND R<=55 THEN 880                 (10% of the nums)
560 IF R>=56 AND R<=65 THEN 910                 (another 10%)
570 GOTO 940                                    (the remaining 35% will be left)
580 PRINT AGE;                                  (this lists the sample on the screen)
600 LET X=X+AGE                                 (to keep a running sum of ages for all "Z" elements)
610 NEXT J
620 M=INT(X/Z*100)/100                          (computes the avg age of the sample to 2 decimals)
630 PRINT " "
640 PRINT ' '
650 PRINT 'THE AVERAGE (MEAN) AGE OF THIS SAMPLE'
660 PRINT  'IS ';M;' YEARS.  THE SAMPLE SIZE WAS ';Z;'.'
670 PRINT ' '
680 LET AA=INT((A/Z)*10000+.5)/100              (these compute percentages to 2 decimals)
690 LET BB=INT((B/Z)*10000+.5)/100                    |
700 LET CC=INT((C/Z)*10000+.5)/100                    |
710 LET DD=INT((D/Z)*10000+.5)/100                    |
720 PRINT ' ';AA;'% ARE 18 TO 24 YEARS OLD (';A;' PEOPLE).'
730 PRINT ' ';BB;'% ARE 25 TO 34 YEARS OLD (';B;' PEOPLE).'
740 PRINT ' ';CC;'% ARE 35 TO 49 YEARS OLD (';C;' PEOPLE).'
750 PRINT ' ';DD;'% ARE 50 TO 65 YEARS OLD (';D;' PEOPLE).'
755 PRINT ' '
760 PRINT 'WRITE DOWN THE SAMPLE SIZE, MEAN, AND % OF PEOPLE IN EACH CATEGORY'
765 PRINT 'ON YOUR WORKSHEET."
770 PRINT " "
780 INPUT "WOULD YOU LIKE TO TRY AGAIN? (TYPE Y OR N)";A$
790 IF A$="Y" THEN 350
800 IF A$="N" THEN 970
810 GOTO 780
850 LET AGE=INT(RND(Q)*7+18)                    (create num 18 to 24)
860 LET A=A+1                                   (keep a running sum of # in that category)
870 GOTO 580
880 LET AGE=INT(RND(Q)*10+25)                   (create num 25 to 34)
890 LET B=B+1                                   (running sum of how many)
900 GOTO 580
910 LET AGE=INT(RND(Q)*15+35)                   (create num 35 to 49)
920 LET C=C+1                                   (running sum of how many)
930 GOTO 580
940 LET AGE=INT(RND(Q)*16+50)                   (create num 50 to 65)
950 LET D=D+1                                   (running sum of how many)
960 GOTO 580
970 END
```

10 fall into the range 18 to 20 years old?" Count a show of hands and plot it on the board. "Between 21 and 23 years old?" And so on. Figure 3a is a reproduction of a plot of 90 student samples plotted from samples of size 10.

Following the same format, we then plot the average ages from samples of size 20 and notice that it is almost a normal curve! Then 50, 100, and 200 (although the students see the pattern by the time we have plotted samples of size 50 and it is efficient to skip to 200). Figure 3 reproduces actual student data for several sample sizes. Notice that the result is very similar to the text book figures that this exercise is meant to reproduce!

The concepts that I try to get them to notice are:

1. with small sample sizes, our samples are all very different from each other.
2. as sample size increases, all our samples get more and more similar.
3. even with sample sizes of 20, the distribution we have been drawing on the board looks like a normal curve (then discuss what they know about the standard deviation of a normal distribution, etc. - - this reminds them of confidence intervals).

Next, we replicate the exercise with percentages. First is the percentage of people in the age category 18 to 24 years old (recall that the "truth" in the simulation is 45%. This was a deliberate choice, so that this distribution could be compared with a 10% category.) - - then with the other age categories. Actual student data is reproduced in Figures 4 and 5. What I try to get them to notice here is:

1. when the "truth" is around the middle of the scale (e.g., 457.), the range of sample percentages is much wider than when the "truth" is closer to one end of the scale (e.g., 10% for the 25 to 34 year old category).
2. when the sample sizes are 20 to 50, the sample distributions do not look like normal curves (as they did when we plotted sample means). We need much larger sample sizes with percentages than with means before we can rely on any knowledge we have about the normal curve (which reminds them of the Central Limit Theorem).

By now the students usually have a pretty good "feel" for why we can have so much confidence in confidence intervals. They also SEE why sample sizes don't have to be super big and that a lot of knowledge about the population can be retrieved from samples as small as 200. They understand that if the sample size is big enough, every sample they can draw will have almost the same mean as any other sample. And the experience enhances their understanding of the statistical principles that they have memorized in other courses.

The exercise takes approximately 40 minutes of class time; considerably less when fewer sample sizes are plotted. One word of caution on that subject: the instructor must move through the exercise rapidly. Students will get bored quickly if too much time is spent counting raised hands, etc.

STUDENT RESPONSE

The only measure of success of this simulation exercise is from student comments during the semester. A few students have approached me and commented that they never understood what a sampling distribution was before this class and that they feel like they know it now. Obviously, a true test of its worth would be an experiment using one set of students exposed to it and one set taught in a more traditional method. However, the ease with which we get through the material, and the absence of blank faces during the exercise precludes me from changing things in the direction of the old ways. Like many ABSEL members, I believe that each of us must "tune in" to what is best for our idiosyncratic style. The ABSEL conference offers us all an opportunity to share and to choose that which fits our own methods. I hope that this exercise is of value to many.

REFERENCES

[1] Churchill, Gilbert A., Jr., Marketing Research (3rd edition) (New York: The Dryden Press, 1983).

[2] Kurnow, Ernest, Gerald J. Glasser, and Frederick R. Ottman, Statistics for Business Decisions (Homewood, IL: Richard D. Irwin 1959), pp. 182-183

[3] Zikmund, William G., Exploring Marketing Research (2nd edition) (New York: The

FIGURE 2

EXAMPLE STUDENT WORKSHEET

Marketing Research
Prof. XXXXXXXXX

Worksheet for
"Sample" assignment

SAMPLE SIZE=10

Sample #	Avg. age	% 18-24	% 25-34	% 35-49	% 50-up
1					
2					

SAMPLE SIZE=20

Sample #	Avg. age	% 18-24	% 25-34	% 35-49	% 50-up
1					
2					

SAMPLE SIZE=50

Sample #	Avg. age	% 18-24	% 25-34	% 35-49	% 50-up
1					
2					

SAMPLE SIZE=100

Sample #	Avg. age	% 18-24	% 25-34	% 35-49	% 50-up
1					
2					

SAMPLE SIZE=200

Sample #	Avg. age	% 18-24	% 25-34	% 35-49	% 50-up
1					
2					

FIGURE 3
DISTRIBUTION OF STUDENTS' SAMPLE MEANS

Vertical axis is # of samples (of 90)
Horizontal axis is age ranges
(Mean of population is 36.5 yrs.)

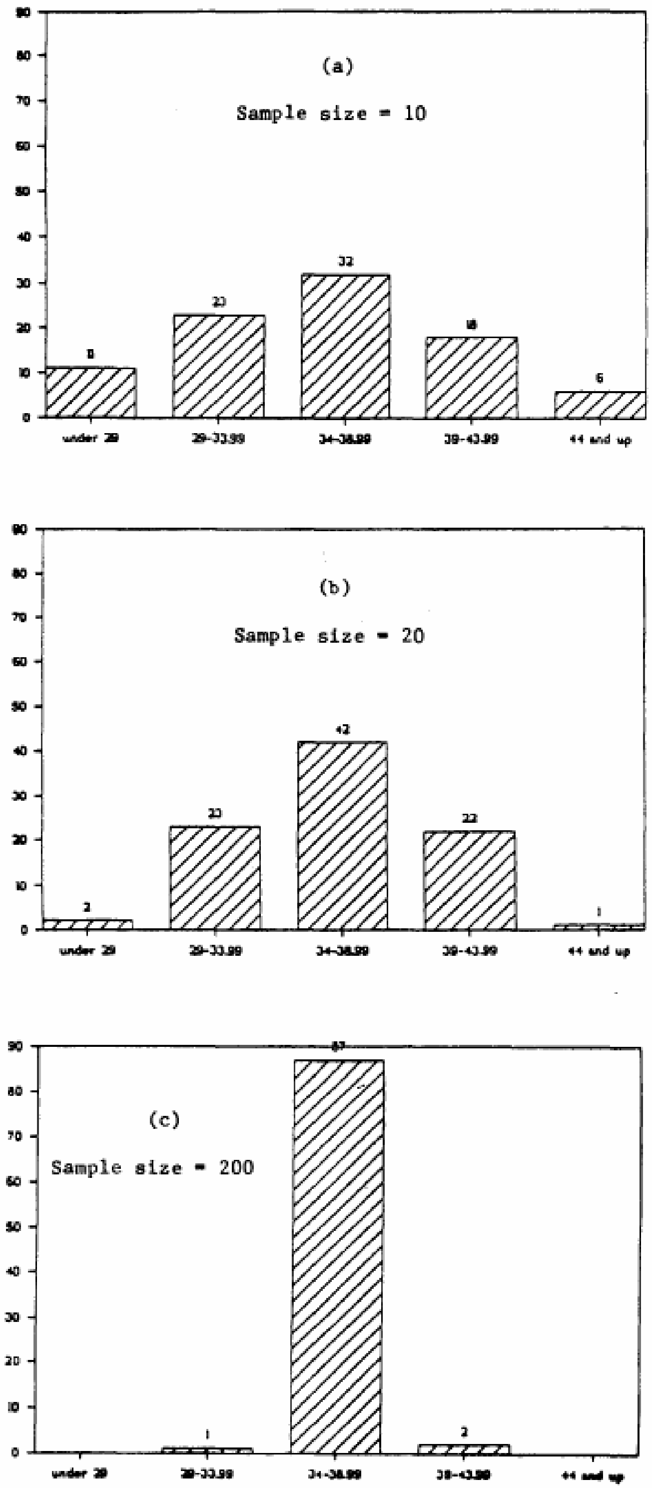

FIGURE 4
DISTRIBUTION OF STUDENTS' SAMPLE PROPORTIONS
(% 18-24 years old in sample)

Vertical axis is # of samples (of 90)
Horizontal axis is the % in the sample that is 18-24 years old
(Population proportion is 45%)

*For table 4(c) the bars report the # of samples whose proportion falls within 2.5% of the horizontal label.

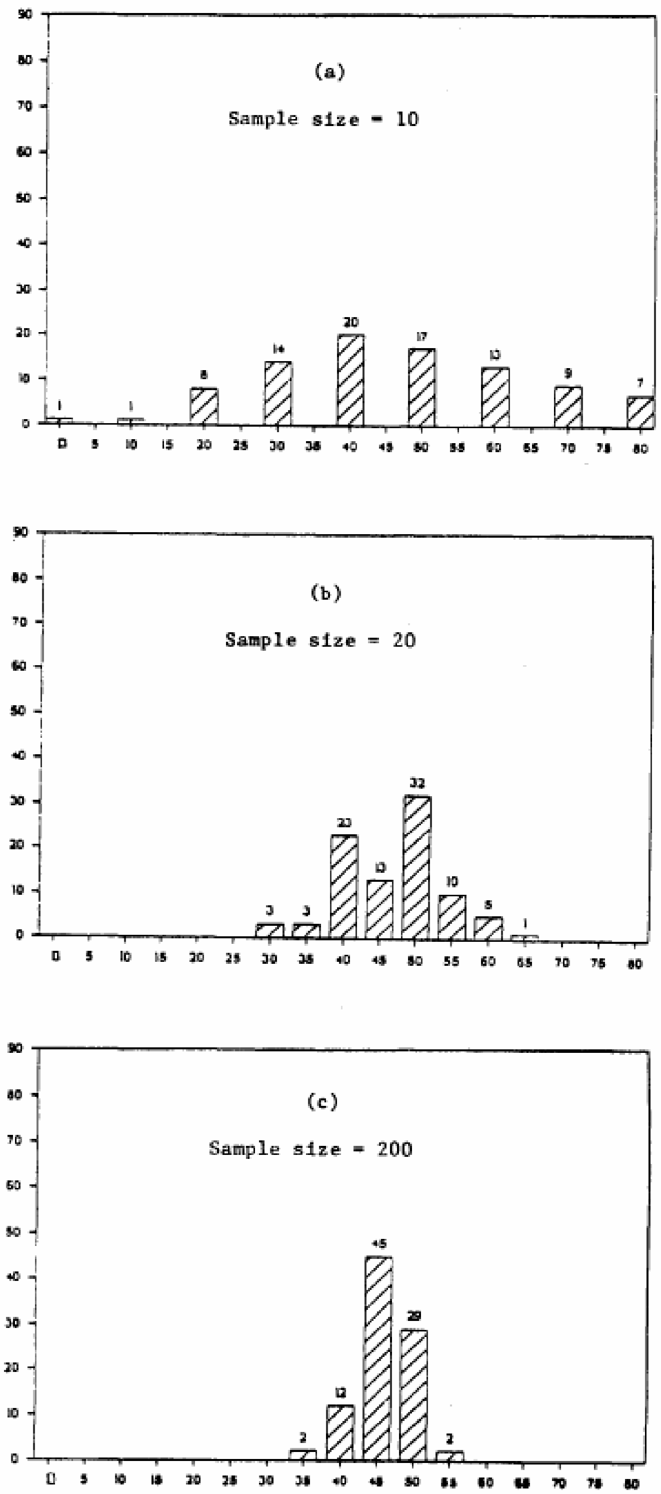

FIGURE 5

DISTRIBUTION OF STUDENTS' SAMPLE PROPORTIONS
(% 25-34 years old in sample)

Vertical axis is # of samples (of 90)
Horizontal axis is the % in the sample that is 25-34 years old
(Population proportion is 10%)

*For table 5(c) the bars report the # of samples whose proportion falls within 2.5% of the horizontal label.