## TEACHING SAMPLING CONCEPTS IN MARKETING RESEARCH

Gopala "GG" Ganesh, University of North Texas

This is a proposal for a special session at MEA 2017 that would describe, in detail, an easily implemented exercise to facilitate student understanding of sampling concepts. Students' grasp of the very basic ideas of sampling: estimating the population value of a characteristic through a sample of that population, the resulting sample distribution, the sampling distribution of which that single sample is a part and hence the associated confidence interval, can all be improved through the exercise. In a typical undergraduate marketing research class, sampling is typically taught after the measurement, scaling and questionnaire design topics and immediately before basic data analysis.

The approach involves intentionally setting up a small population. The author has implemented sizes of just 6 or 8 members. The population has just two characteristics: minutes reading the newspaper (summarized by the mean) and gender (summarized by the proportion). Given its small size taking a census to determine the two characteristics is easy and no sample-based inference is needed. Let us first consider a population of 6 members. Students are asked to go ahead and determine the population mean and standard deviation for minutes reading the newspaper and % female. They then draw the 15 unique or "all possible samples" of size 2 from the population. For each of these 15 samples, they determine the average minutes and % female and asked to verify that few of the 15 samples are "exactly on the money" in estimating the true population characteristics. They are then asked to calculate the "mean of the 15 sample means and proportions", and see for themselves that THESE, i.e., mean of the two sampling distributions, exactly match the two population characteristics. Next, they construct column charts of the two sampling distributions and visually verify they are "reasonably" normal, even for such a small population size (N=6) and sample size (n=2). They quickly realize that for much larger populations and typical sample sizes in market research, the sampling distribution becomes a smoother bell curve given the very large population size N, the literally astronomical number of unique samples of size  $n (=N^{c}n)$ possible from it, and consequently the equally numerous points available for plotting the column chart.

Next, students calculate the standard error of the mean = standard deviation of the population of 6 elements divided by square root of n, the sample size (= 2 in the example). They also calculate standard error of the proportion as  $\sqrt{(pq/n)}$  where p=proportion female in the population of 6, q=1-p, and n = 2, again. They go on to construct three ranges: population mean for minutes ± 1, 2, 3 standard errors of the mean and population proportion ± 1, 2, 3 standard errors of the proportion. Once they complete this, they are able to verify that approximately 68%, 95% and 99% of the 15 sample means and proportions DO fall within the three ranges, respectively, re-confirming that the sampling distribution is normal. This sets the setting the stage for the author to explain and the students to understand the central limit theorem easily.

Depending upon availability of time, the exercise is extended to N=6 and n=3. This time, there are 20 unique samples and with the additional data points, the column chart and the three ranges become even easier to understand. In the past, the author has assigned for extra credit, N=8, n=2 (28 unique samples) and N=8, n=3 (56 unique samples), both of which result in much smoother bell curves.

The author has been teaching marketing research as an online MBA class for the past 11 years and since there is no face to face interaction, has incorporated an Excel-and PowerPoint-based implementation of the above exercise as a part of the sampling module of the class.

Once the students get a good grasp of the central limit theorem, it is easy for them to understand that estimated mean or proportion p from a single sample is unlikely to equal the true mean  $\mu$  or proportion  $\pi$  of the population because, well, the odds are against it: most of the samples produce off estimates. And even if a researcher happened to have one of the rare samples that coincided with the population, they will

not know it because determining the population's  $\mu$  or  $\pi$  requires a census, not feasible in most marketing research instances and if a census is possible, why sample? Therefore, simply transposing and  $\mu$  (or  $\pi$  and p), the relationship can be re-written as:

 $\mu = \pm$  Error for estimating a population mean from a single sample mean, OR,

 $\pi = p \pm Error$ , for estimating a population proportion from a single sample proportion.

The Error, of course equals 1, 2, or 3 (the approximate "z value" for 68%, 95%, or 99% confidence. In the class, the author puts up the standard normal curve on the screen and students see that this is indeed the case) \* the standard error of the mean or proportion.

Since Error = z \* Standard Error, deriving and understanding the sample size formula for means and proportions becomes very easy:

Error =  $z * \sigma/\sqrt{n}$  for means, and, Error =  $z * \sqrt{pq} / \sqrt{n}$  for proportions.

Therefore,  $n = (z^2 * \sigma^2)/E^2$  for means, and,  $n = (z^2 * pq)/E^2$  for proportions.

During the MEA 2017 special session, the author will take the audience through the entire assignment, for N = 6 and 8 and n = 2 and 3.