PRACTICAL ACTIVITIES

A Note on Illustrating the Central Limit Theorem

Thomas W. Woolley
Samford University, Birmingham, Alabama, USA.
e-mail: twwoolle@samford.edu

◆ INTRODUCTION ◆

A SHORT PAPER on "Illustrating the Central Limit Theorem" published several years back (Colwell and Gillet, 1994) provided a very nice example of how the generation of pseudorandom numbers can be used to help students empirically understand the Central Limit Theorem (CLT). As noted by Colwell and Gillet the CLT "states that if a random sample of size n is taken from any 'sensible' distribution with mean μ and variance σ^2 then the distribution of means of samples of size n has an approximate Normal distribution with mean μ and variance σ^2/n. The larger the value of n, the better the approximation. If the original distribution is itself Normal, the distribution of means has an exact Normal distribution."

It is true that students taking applied statistics courses (particularly nonmathematical service courses) are either explicitly or implicitly expected to "accept on faith" the truth of the CLT. However, apparently unlike Colwell and Gillet, I have found that *many of these students find empirical demonstrations using pseudorandom numbers unrealistic and intellectually intangible.* The purpose of this note is to provide an alternative source of data, useful in demonstrating the CLT, that seems to resonate with students.

◆ CLASSROOM EXERCISE ◆

In the States, telephone numbers (minus the regional Area Code) consist of a three-digit exchange followed by four digits; for example, 555-1234. It can be argued that, for the typical residential number, the last four digits are a random (or at least, haphazard) sample of a uniform distribution consisting of the digits 0 to 9, with an expected value of 4.5. Students are given the following assignment to complete at home:

1) Randomly (using a table of random numbers) select a page and a telephone number from the residential pages of the local telephone book;
2) Repeat this selection process until a total of 20 telephone numbers have been selected;
3) Record the last four digits of each telephone number. Bring back to class 20 samples of four digits and the corresponding sample means, each sample having been drawn from a uniform distribution with $E(X) = 4.5$.

Often we plot the means in class; after class I enter the data from all students into a Minitab worksheet.

◆ EXAMPLE ◆

Analysis of the individual digits shows an approximately uniform distribution, and students can estimate μ and σ^2. Graphical analysis shows the students that the means do, in fact, exhibit a unimodal, approximately Normal, distribution with mean about 4.5 and variance of roughly s^2/n. After a discussion of the impact of increasing n, students return to the telephone book to get samples of size 8 (i.e., two telephone numbers per sample) and continue their analyses.

Having gathered their own data and empirically tested their faith in the CLT, the students will often express an interest in illustrating the CLT with other distributions. At this point the Colwell and Gillet approach is introduced. Below is an example of one representative data set of means (440 means based on samples of $n = 4$) from a recent class and the subsequent Minitab analysis conducted by students:

Descriptive Statistics

Variable	N	Mean	Median	Tr Mean	St Dev	SE Mean
C5	440	4.5915	4.5000	4.5827	1.4516	0.0692

Variable	Min	Max	Q1	Q3
C5	1.2500	8.2500	3.5000	5.5000

Character Dotplot

Distribution of Telephone Number Means ($n = 4$) (Each dot represents 2 points)

```
                       :
                       : .    . :.
                 .     : :  : : ::
                 :  : .: : :: : ::
             .   :  : :: : :: : ::
            .: . :  : :: : :: : :: : :
          : :: : :: : :: : :: : :: : :. .
   .   .: : :: : :: : :: : :: : :: : :: : .  :
   : . :: : :: : :: : :: : :: : :: : :: : :: : :: .
-----+---------+---------+---------+---------+---------+-
   1.5       3.0       4.5       6.0       7.5       9.0
```

Telephone Number Means

◆ HISTORICAL FOOTNOTE ◆

I have been conducting this exercise in introductory statistics classes for 17 years, and students have always enjoyed and appreciated it. However, there is one unusual phenomenon that has occurred over this period. In 1981, when my first data sets were gathered by students, the mean of the sampling distribution was 3.7. This mean crept upward until 1990, when it achieved the expected value of 4.5, where it has remained for the past seven years. Clearly, the parent distribution was not uniform and μ was not 4.5 prior to about 1990. The most plausible explanation I have heard for this is the transition from rotary to touch-tone telephones. In the days of mechanical switching stations, small digits in telephone numbers allowed calls to be routed more efficiently. This led to the use of telephone numbers with small numbers whenever possible, and the parent distribution of digits represented in residential telephone numbers was non-uniform with expected value less than 4.5. In the digital era, there is no explicit attempt at assigning residential telephone numbers with small digits. Hence, we are now sampling from a more uniform distribution with an expected value of 4.5. The gradual transition from 3.7 to 4.5, it is hypothesised, reflected the slow change in the nature of the parent distribution. An interesting exercise would be to have students sample from "old" telephone books to see if they can detect this earlier bias.

Reference

Colwell, D.J. and Gillet, J.R. (1994). Illustrating the Central Limit Theorem. *Teaching Statistics,* **16**(2), 38.

Woolley, Thomas (Autumn, 1998). A Note on Illustrating the Central Limit Theorem. *Teaching Statistics, 20* (3), 89–90.
Reproduced with permission from *Teaching Statistics.* Copyright © 1998 by the Teaching Statistics Trust. All rights reserved.