Using Monte-Carlo Simulations and an Interactive White-Board for Teaching Sampling Distribution Characteristics

Presented at the United States Conference on Teaching Statistics (USCOTS)
Spotlight Session on Pedagogy
May 19th – May 21st, 2005 • Columbus, Ohio

By

Fabio F. Santos, Ed.D.
Assistant Professor of Statistics
Raymond Walters College, University of Cincinnati
9555 Plainfield Rd; Cincinnati, OH 45236
(513)745-5758; fabio.santos@uc.edu

This presentation shows a practical guide for using Monte-Carlo simulation studies and an interactive white-board for the teaching and learning of sampling variability, sampling distribution characteristics, and the central limit theorem. The presenter will show the use of both SPSS and Stata statistical software for performing Monte Carlo simulations. Below is an outline of the presentation. First, a definition of Monte Carlo simulations will be given; second, there will be a brief discussion about the role of simulations in statistics; and third, two examples of Monte-Carlo simulation studies will be demonstrated.

1. Definition of Monte Carlo simulations.

2. Importance of Monte Carlo simulations in statistics.

3. Example #1. Use of Monte Carlo simulations for investigating properties of location estimates.

   Here we compare the sample mean $\theta^{(1)}$, sample 20% trimmed mean $\theta^{(2)}$, and sample median $\theta^{(3)}$ as estimates of the mean μ of a specified distribution (normal or skewed). We follow the Monte Carlo simulation procedure indicated below to make the comparison.

   - Generate independent draws $X_1, \ldots, X_n$ from the distribution.
   - Compute $\theta^{(1)}, \theta^{(2)}$, and $\theta^{(3)}$.
   - Repeat the first two steps B times (i.e. $B = 1000$) to obtain
     $\theta_1^{(1)}, \ldots, \theta_B^{(1)}$; $\theta_1^{(2)}, \ldots, \theta_B^{(2)}$; $\theta_1^{(3)}, \ldots, \theta_B^{(3)}$
   - For $k = 1, 2, 3$, calculate

$$\bar{\theta}^{(k)} = \frac{\sum_{i=1}^{B} \theta_i^{(k)}}{B} \qquad S_{\hat{\theta}^{(k)}} = \sqrt{\frac{\sum_{i=1}^{B} (\theta_i^{(k)} - \bar{\theta}^{(k)})^2}{B-1}}$$

$$\hat{BIAS} = \bar{\theta}^{(k)} - \mu \qquad \hat{MSE} = \frac{\sum_{i=1}^{B} (\theta_i^{(k)} - \mu)^2}{S} \approx \hat{BIAS}^2$$

- Compute the relative efficiency of estimate j to estimate i for $i = 1 \wedge j = 2, i = 1 \wedge j = 3, i = 2 \wedge j = 3$.
  a) If $E(\theta^{(i)}) = E(\theta^{(j)}) = \mu \Rightarrow RE = \frac{\text{var}(\theta^{(i)})}{\text{var}(\theta^{(j)})}$
  b) If the estimates are not unbiased, compute the relative efficiency by using the formula $RE = \frac{MSE(\theta^{(i)})}{MSE(\theta^{(j)})}$
  If $RE < 1$, then estimate i is more efficient than estimate j.
- Investigate the sampling distributions of these estimates of μ through the use of sample histograms.

4. Example #2. Use of Monte Carlo simulations for investigating the sampling distribution of a sample mean.

Here we investigate the behavior of the sample mean $\bar{x}$ as described by its sampling distribution. First, we consider that the underlying population of x-values is normally distributed with parameters μ and σ. The Monte Carlo simulation procedure to do the investigation follows below.

- Generate independent draws $X_1, \ldots, X_n$ from the population under consideration.
- Compute its sample mean.
- Repeat the first two steps B times (i.e. $B = 1000$) to obtain $\bar{X}_1, ..., \bar{X}_B$.
- Calculate $\bar{X} = \frac{\sum_{i=1}^{B} \bar{X}_i}{B}$; compare this value to the population mean μ.
- Calculate $\sigma_{\bar{X}} = \sqrt{\frac{\sum_{i=1}^{B} (\bar{X}_i - \bar{X})^2}{B}}$; compare this value to the population standard deviation σ.
- Construct a sample histogram of the $\bar{X}$ - values (i.e. $\bar{X}_1, ..., \bar{X}_B$). Because B is reasonable large, the sample histogram should rather closely resemble the true sampling distribution of $\bar{X}$ (obtained from an infinite sequence of $\bar{X}$ - values).
- Repeat the experiment for several values of n to determine how the choice of sample size affects the sampling distribution.
- Repeat the experiment for a skewed distribution with parameters μ and σ.
- State the central limit theorem based on the above observations.