Lab Handout — Lab 4.3: The Central Limit Theorem

The instructions below are keyed to the lab instructions found on pp. 223-225 of the text. Please use those instructions as well in preparing your report.

Experimental Procedure

First a word about the macros you will use in the simulations. When running the macro, don't worry if graphs pop up on the screen and disappear. They will reappear on a one-page template containing all four graphs that you called for. **CAUTION:** If you wish to print the template you must do it **BEFORE** moving on to the next macro. Submitting a new macro will overwrite the previous template and you'll have to run the first macro again.

The Central Limit Theorem for Rolls of a Die

1. The macro MAKEDATA macro will generate random data from the discrete uniform distribution having an equal probability of producing any of the integers 1,2,3,4,5, or 6, just like a fair die. The data will be put in the data set ROLLS. The macro will simulate the trial of rolling a fair die 50 times. It replicates this trial 250 times, producing a total of 250 times 50 or 12,500 simulated die rolls.

   MAKEDATA simulates the 50 rolls of the fair die, putting the result of the i^{th} roll in variable Ci. Thus each row in C1- C50 represents one replication of the trial. There are 250 such rows corresponding to the 250 replications of the trial. The macro also computes the means of the first 2, 10, 30 and 50 rolls from each trial, calling them MEAN2, MEAN10, MEAN30 and MEAN50, respectively.

   Run the macro now. A window will pop up informing you when the data set has been created. Click on the window as directed and hit return. The window will go away. Note that if the window fails to appear, something has gone wrong and you should ask for help.

   Use SAS or SAS/INSIGHT to look at ROLLS, which is a data set containing a portion of the data. Specifically, ROLLS has in each row the first 5 of the 50 original observations (die rolls) under variable names C1-C5 (Including all 50 would have been cumbersome). The mean of C1 and C2 is in MEAN2, the mean of C1-C10 is in MEAN10, the mean of C1-C30 is in MEAN30, and the mean of C1-C50 is in MEAN50.

2. Use SAS/INSIGHT to make a frequency and a density histogram of C1. Recall that in SAS/INSIGHT, *Analyze:Histogram/Bar Chart (Y)* will produce a frequency histogram, and *Analyze:Distribution (Y)* will produce a density histogram.

   To obtain density histograms of C1, MEAN2, MEAN10, and MEAN50 all plotted on the same scale, simply use the macro HISTREP. When you call HISTREP an input window will appear. Click on the green cursor and enter 'rolls' (without the quotes) as the data set name. Hit return and enter 'u', then successively the names C1, MEAN2 MEAN10 and MEAN50. Density histograms of these variables will appear in the SAS GRAPH window. You should print these now.

3. Make normal quantile plots of C1, MEAN2, MEAN10, and MEAN50. To do this, call the macro 'NORMREP' and proceed as you did for HISTREP. Print these graphs now.

4. In SAS/INSIGHT, *Analyze:Distribution (Y)* will produce the means and standard deviations of C1, MEAN2, MEAN10, and MEAN50. **IMPORTANT:** After printing or writing down these means and standard deviations, close the data set ROLLS in SAS/INSIGHT.

5. The macro SMEAN will compute the standardized means of C1, MEAN2, MEAN10, and MEAN50. These standardized means will be found in the data set ROLLS under the variable names SC1, SMEAN2, SMEAN10, and SMEAN50. Use SAS/INSIGHT to check that the means of each of these variables are nearly 0 and the standard deviations are nearly 1.

Use the macro HISTREP to generate density histograms and NORMREP to generate normal plots of the standardized means (don't forget to enter an 's' to to denote the fact that the data are standardized).

An Example Where the Central Limit Theorem Fails

The macro MAKECAU will generate 250 data sets each of 50 observations from a Cauchy distribution model. The data will be placed in CAU. C1 again denotes the first column of data, and MEAN2, MEAN10 and MEAN50 have the same meaning here as they did in ROLLS. Now do steps 2. and 3. on these data; don't forget to enter a 'c' to denote the fact that the data are Cauchy.