Simulating the Chi-Square Distribution©

Run the program **ProbChisq.sas**, which is found on my SAS Programs page. Look at the first part of the program. In the first data step, DATA one, I use a DO LOOP to sample 100,000 squared *z* scores, each drawn from a standard normal distribution. As you know, this should create a Chi-Square distribution on 1 *df*. A "do loop" simply tells SAS to do something a specified number of times. In this case, the number of times is 100,000. The specified action (SAS statements between the DO and the END) was randomly to draw a score from a standard normal distribution (using the NORMAL random number generator), square it (**2) and then OUTPUT it to the data set. PROC SGPLOT was then used to make a histogram of the resulting values with smooth curves overlayed. The kernel plot represents the actual scores with a smoothed line. The normal plot represents a normal distribution with the same mean and variance as the actual scores. PROC MEANS was then employed to get the mean, variance, and skewness of the sample of 100,000 squared *z* scores. Empirically producing a sampling distribution like this is referred to as doing a "Monte Carlo."

Look at the chart of the squared *z* scores. It has the expected distinctly skewed shape. Do keep in mind that the rightmost interval will include all of the scores in the upper tail, so you might get a little upturn there. From the output from proc means, note that the mean of these squared *z* scores is very close to the expected value (1, the *df*) and the variance also very close to its expected value (2, twice the *df*). Fisher's skewness measure should be high.

Look back at the program. In the second data step I obtained 100,000 samples, each with two squared *z* scores from a standard normal distribution. I added these two together. As you know, the resulting sum of pairs of squared *z* scores is the Chi-Square distribution on 2 *df*. The plot and Fisher's measure show that the skewness of the distribution has decreased, and the mean and variance have doubled, as expected.

In the final three data steps I simulated the Chi-Square distributions on 5. 10, and 50 *df*. Note that the mean and variance come out very close to their expected values and that the skewness of the distribution drops noticeably as the *df* increase. In fact, you should be able to discern that the shape of the distribution is approaching normal as the *df* increase.

Output from one run of the program can be found HERE.

Return to the Stats Lessons Page

Karl L. Wuensch, October, 2016.