7. The Distribution of the Sample Mean When the Population is Not Normal

In Section 5.3 we established that if $Y_i \sim N\left(\beta, \sigma^2\right)$, then $b \sim N\left(\beta, \sigma^2/T\right)$. Based on our assumption of normality we can also determine the sampling distribution of $\hat{\sigma}^2$. These sampling distributions are the basis for tests of economic hypotheses and other forms of statistical inference which we will explore in *Statistical Inference II* and also throughout *UE/2.* A question we need to ask is "If the population is not normal, then what is the sampling distribution of the sample mean?" The *Central Limit Theorem* provides an answer to this question.

Central Limit Theorem: If $Y_1, \ldots, Y_T$ are independent and identically distributed random variables with mean β and variance σ^2, and

$$b = \bar{Y} = \frac{1}{T}\left(Y_1 + Y_2 + \cdots + Y_T\right) = \sum Y_i / T \tag{7.1}$$

then

$$Z = \frac{\bar{Y} - \beta}{\sigma / \sqrt{T}} \tag{7.2}$$

has a probability distribution that converges to the standard normal $N(0,1)$ as $T \to \infty$.

This Theorem says that the sample average of T independent random variables from *any* probability distribution (as long as they have a mean and variance) will have an approximate standard normal distribution after standardizing, by subtracting its mean and dividing by its standard deviation, if the sample is sufficiently large.

To illustrate this, we carry out a simulation experiment. Let the continuous random variable Y have probability density function

$$f(y) = \begin{cases} 2y & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases} \tag{7.3}$$

The mean of Y is $E(Y) = 2/3$ and its variance is $\operatorname{var}(Y) = 1/18$. The Central Limit Theorem says that if $Y_1, \ldots, Y_T$ are independent and identically distributed with density $f(y)$ then

$$Z = \frac{\bar{Y} - 2/3}{\sqrt{\dfrac{1/18}{T}}} \tag{7.4}$$

approaches standard normal as T approaches infinity.

We use a random number generator to create random values from the density in (7.3). Plotting 5000 values gives the histogram in Figure 6

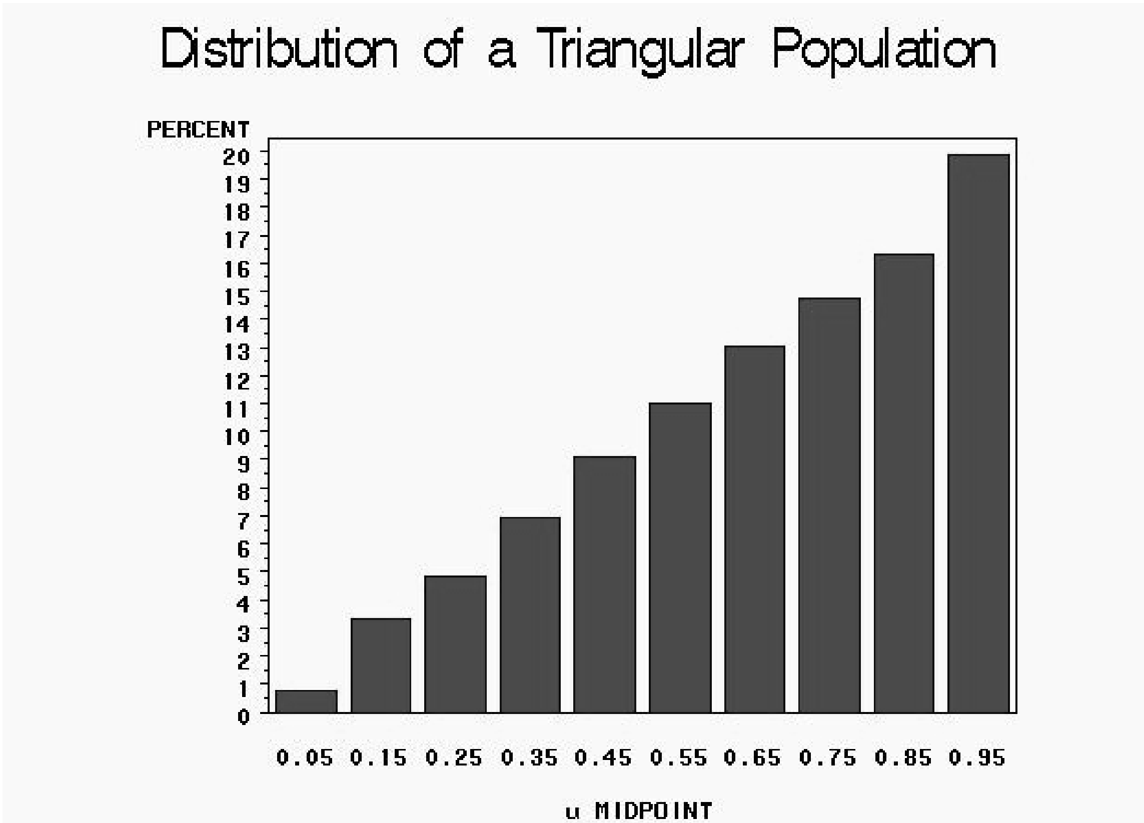

Figure 6 Random Values from a Triangular Distribution

We generate 5000 samples of sizes T = 3, 10 and 30, compute the sample means of each sample, and standardize them as shown in (7.4). The histograms are given in Figures 7-9.

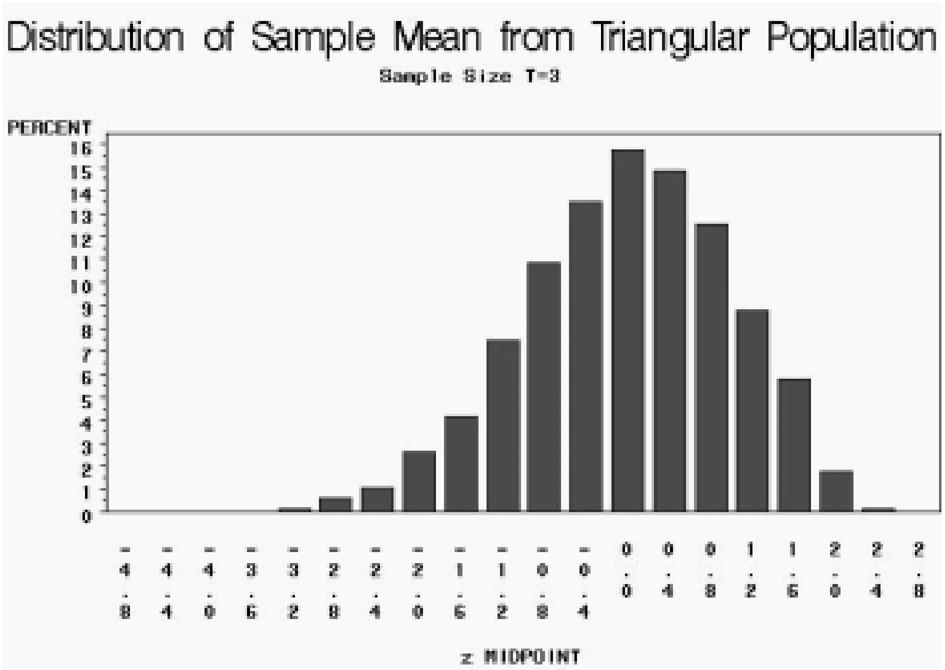

Figure 7 Histogram of Standardized Sample Means for $T = 3$

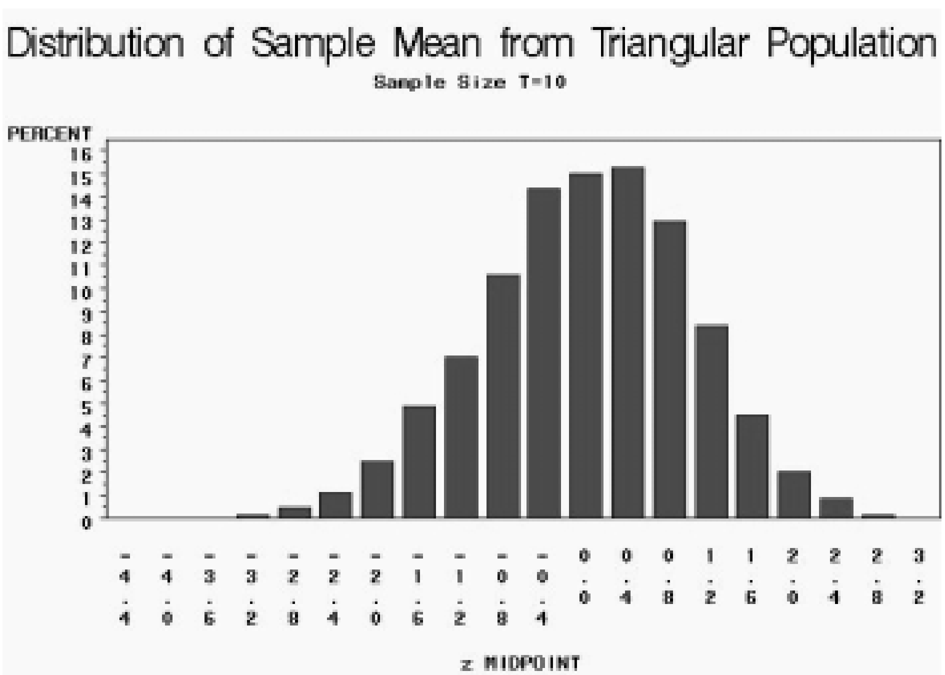

Figure 8 Histogram of Standardized Sample Means for $T = 10$

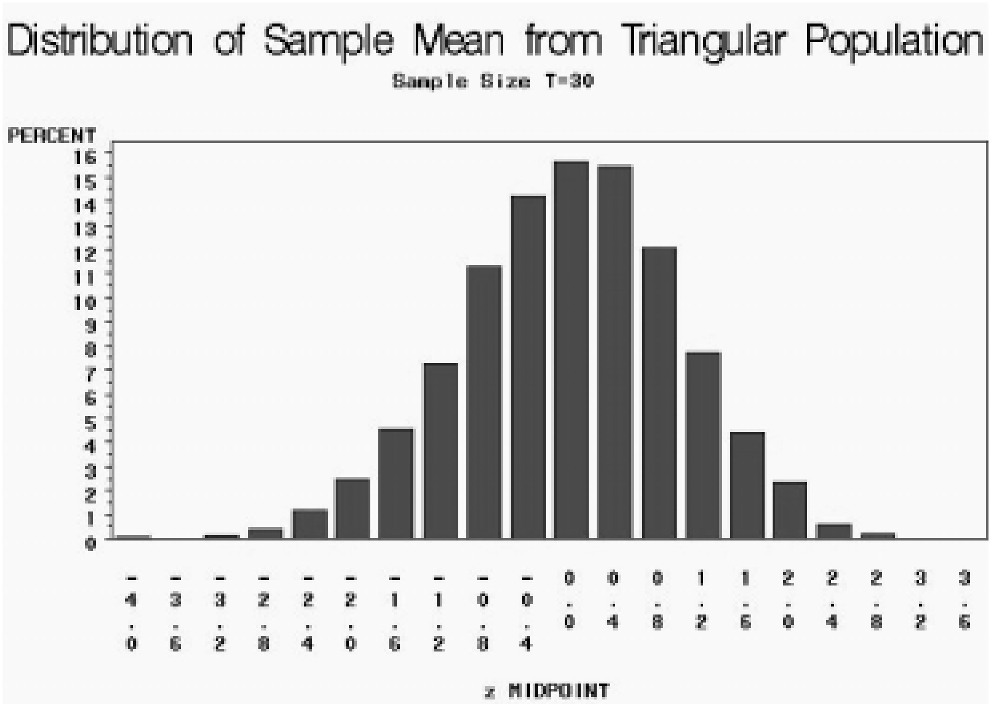

Figure 9 Histogram of Standardized Sample Means for $T = 30$

You see the amazing convergence of the sample mean to a distribution that is bell-shaped, centered at zero, symmetric, with almost all values between -3 and 3, just like a standard normal distribution. We use this remarkable fact throughout *UE/2* when we can not justify the assumption of normality.