Contents

1. An unbiased population variance

An unbiased population variance s is obtained by

$$
\begin{aligned}
s^2 &= \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2 \\
&= \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i^2-2x_i\bar{x}+\bar{x}^2\right) \\
&= \frac{1}{N-1}\sum_{i=1}^{N}x_i^2-\frac{2}{N-1}\sum_{i=1}^{N}x_i\left(\frac{1}{N}\sum_{i=1}^{N}x_i\right)+\frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{1}{N^2}\sum_{i=1}^{N}x_i\right)^2 \\
&= \frac{1}{(N-1)}\left[\sum_{i=1}^{N}(x_i)^2-\frac{1}{N}\left(\sum_{i=1}^{N}x_i\right)^2\right] \qquad (1)
\end{aligned}
$$

2. Unbiased Estimators and Moments

In the earlier section, $\bar{x}$ and s^2 are unbiased estimators of the true population mean, μ, and variance, σ^2. That is, the expected values of x and $(x-\bar{x})^2$ are equal to μ and σ^2, respectively. It means that $E[\bar{x}]=\mu$ and $E[s^2]=\sigma^2$.

$$
\begin{aligned}
E[\bar{x}] &= E\left[\frac{1}{N}\sum_{i=1}^{N}x_i\right]=\frac{1}{N}\sum_{i=1}^{N}E[x_i] \\
&= \frac{1}{N}\sum_{i=1}^{N}\mu=\mu \qquad (2)
\end{aligned}
$$

Next,

$$
\begin{aligned}
E[s^2] &= E\left[\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2\right] \\
&= E\left[\frac{1}{N-1}\left\{\sum_{i=1}^{N}\left[(x_i-\mu)^2 - N(\bar{x}-\mu)^2\right]\right\}\right] \\
&= \frac{1}{N-1}\left\{\sum_{i=1}^{N}E\left[(x_i-\mu)^2\right] - NE\left[(\bar{x}-\mu)^2\right]\right\} \\
&= \frac{1}{N-1}\left\{\sum_{i=1}^{N}\sigma^2 - N\frac{\sigma^2}{N}\right\} = \sigma^2 \qquad (3)
\end{aligned}
$$

Here, we have used the relations $x_i-\bar{x} = (x_i-\mu)-(\bar{x}-\mu)$, $E[(x_i-\mu)^2] = V[x_i] = \sigma^2$, and $E[(x_i-\mu)^2] = V[x_i] = \sigma^2/N$.

3. Central limit theorem

Let $X_1, X_2, \ldots, X_i, \ldots$ be a sequence of independent random variables with $E[X_i] = \mu_i$ and $V[X_i] = \sigma_i^2$. Define a new random variable $X = X_1+X_2+\cdots+X_N$. Then, as N becomes large, the standard normalized variable,

$$
Z_N = \frac{\left(X - \sum_{i=1}^{N}\mu_i\right)}{\left(\sum_{i=1}^{N}\sigma^2\right)^{1/2}} \qquad (4)
$$

takes on a normal distribution regardless of the distribution of the original population variable from which the sample was drawn. A special form of the central limit theorem may be stated as: the distribution of mean values calculated from a suite of random samples $X_i(X_{i,1}, X_{i,2}, \ldots)$ taken from a discrete or continuous population having the same mean μ and variance σ^2 approaches the normal distribution with mean μ and variance σ^2/N as N goes to infinity. Consequently, the distribution of the arithmetic mean

$$
\bar{X} = \frac{1}{N}\sum_{i=1}^{N}X_i \qquad (5)
$$

is asymptotically normal with mean μ and variance σ^2/N when N is large. Ideally, we would like N to go to infinity but, for practical purposes, N greater than 30 will generally ensure that the population of X is normally distributed. When N is small, the shape of the sample distribution will depend mainly on the sampe of the parent population. However, as N becomes larger, the shape of the sampling distribution becomes increasingly more like that of a normal distribution no matter what the shape of the parent population. In many instances, the normal assumption for the sampling distribution for $\bar{X}$ is reasonably accurate for $N > 4$ and quite accurate for $N > 10$.

The central limit theorem has important implication for we often deal with average values in time or space.