Estimation of mean and variance

If μ is the population mean, the sample mean $\hat{M}$ is

$$\hat{M} = \frac{X_1 + X_2 + \cdots + X_n}{n} \tag{1}$$

The $\hat{M}$ is a random variable which is an estimator of the population mean. The expectation value of $\hat{M}$ is

$$\begin{aligned} E[\hat{M}] &= \frac{1}{n}E[X_1 + X_2 + \cdots + X_n] \\ &= \frac{E[X_1] + E[X_2] + \cdots + E[X_n]}{n} \end{aligned} \tag{2}$$

Because $E[X_j]$ is the expectation value for the population, $E[X_j] = \mu$ and

$$E[\hat{M}] = \frac{n\mu}{n} = \mu \tag{3}$$

therefore the expectation value $E[\hat{M}]$ of the sample mean $\hat{M}$ is μ, the population mean.

Let us consider sample variance

$$\hat{S}^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n} = \frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})^2 \tag{4}$$

where $\bar{X}$

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \tag{5}$$

is the sample mean. The expectation value $E[\hat{S}^2]$ of sample variance $\hat{S}^2$ is

$$\begin{aligned} E[\hat{S}^2] &= \frac{1}{n}E\left[\sum_{j=1}^{n}(X_j - \bar{X})^2\right] = \frac{1}{n}E\left[\sum_{j=1}^{n}[(X_j - \mu) - (\bar{X} - \mu)]^2\right] \\ &= \frac{1}{n}E\left[\sum_{j=1}^{n}[(X_j - \mu)^2 - 2(X_j - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2]\right] \\ &= \frac{1}{n}E\left[\sum_{j=1}^{n}(X_j - \mu)^2 - 2(\bar{X} - \mu)\sum_{j=1}^{n}(X_j - \mu) + \sum_{j=1}^{n}(\bar{X} - \mu)^2\right] \end{aligned} \tag{6}$$

In equation (6) we have three terms. The first term

$$\frac{1}{n}E\left[\sum_{j=1}^{n}(X_j-\mu)^2\right]=\frac{n\sigma^2}{n}=\sigma^2 \tag{7}$$

where σ^2 denotes the variance of the population, and $(X_j-\mu)^2$ is the variance of the population. Because we have $n\bar{X}=\sum_{j=1}^{n}X_j$ In the second term we have the sum

$$\sum_{j=1}^{n}(X_j-\mu)=n\bar{X}-n\mu=n(\bar{X}-\mu) \tag{8}$$

and the second term equals to

$$-2E[(\bar{X}-\mu)^2] \tag{9}$$

The third term equals

$$E[(\bar{X}-\mu)^2] \tag{10}$$

so we have

$$E[\hat{S}^2]=\sigma^2-E[(\bar{X}-\mu)^2] \tag{11}$$

We need to compute $E[(\bar{X}-\mu)^2]$.

$$E[(\bar{X}-\mu)^2]=E\left[\left(\frac{X_1+X_2+\cdots+X_n}{n}-\mu\right)^2\right] \tag{12}$$

$$=\frac{1}{n^2}E\left[\Big((X_1-\mu)+(X_2-\mu)+\cdots+(X_n-\mu)\Big)^2\right]$$

$$=\frac{1}{n^2}\sum_{j=1}^{n}E[(X_j-\mu)^2]+\frac{1}{n^2}\sum_{\substack{i=1\\i\neq j}}\sum_{\substack{j=1\\i\neq j}}E[(X_i-\mu)(X_j-\mu)]$$

The last term in the above equation is equal to zero due to independence of the random variables X_i and X_j. Due to equation (7) we have the relation

$$E[(\bar{X}-\mu)^2]=\frac{\sigma^2}{n} \tag{13}$$

The expectation value $E[\hat{S}^2]$ of sample variance $\hat{S}^2$ is then

$$E[\hat{S}^2]=\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2 \tag{14}$$

based on equation (11) and (13). We have that

$$\frac{n}{n-1}E[\hat{S}^2] = \sigma^2 \tag{15}$$

so

$$\hat{S}_X^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1} \tag{16}$$

is the estimator of the variance of the population. For the sample we compute the variance as

$$s_x^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1} \tag{17}$$

Square root of s_x^2 is called the standard deviation of the sample.

Pawel J. Piskorz (paweljs@gmail.com)