Estimating the Population Mean (μ) and Variance (σ^2)[1]

George H Olson, Ph. D.
Leadership and Educational Studies
Appalachian State University

Unbiased estimates of population parameters

A statistic, *w*, computed on a sample, is an unbiased estimate of a population parameter, θ, if its expected value [$\mathcal{E}(w)$] is the parameter, θ.

Unbiased estimate of the population mean

It is easy to show, for example, that the sample mean, $\bar{X}$, provides an unbiased estimate of the population mean, μ.

$$\begin{aligned}\mathcal{E}\left(\bar{X}\right) &= \mathcal{E}\frac{\left(X_1 + X_2 + X_3 + \cdots + X_N\right)}{N} \\ &= \frac{\mathcal{E}\left(X_1\right) + \mathcal{E}\left(X_2\right) + \mathcal{E}\left(X_3\right) + \cdots + \mathcal{E}\left(X_N\right)}{N}.\end{aligned} \tag{1}$$

However, any $\mathcal{E}(X_i)$ is, by definition, μ, for all observations taken from the same population. Therefore,

$$\mathcal{E}\left(\bar{X}\right) = \frac{N\mathcal{E}(X)}{N} = \mu\cdot \tag{2}$$

Unbiased estimate of the population variance

In contrast, it can be shown that the sample variance, s^2, is a *biased* estimate of the population variance, σ^2, i.e., that $\mathcal{E}(S^2) \neq \sigma^2$:

$$\begin{aligned}\mathcal{E}\left(s^2\right) &= \mathcal{E}\left(\frac{\sum_i^N X_i^2}{N} - \bar{X}^2\right) \\ &= \mathcal{E}\left(\frac{\sum_i^N X_i^2}{N}\right) - \mathcal{E}\left(\bar{X}^2\right).\end{aligned} \tag{3}$$

Consider, first, the first term to the right of the equal sign in (3), above.

[1] The material presented here is derived from Hays (1973, 2nd Ed, pp. 272-274

$$\mathcal{E}\left(\frac{\sum_i^N X_i^2}{N}\right)=\frac{\sum_i^N \mathcal{E}\left(X_i^2\right)}{N}. \tag{4}$$

By definition, the population variance is,

$$\sigma^2=\mathcal{E}\left(X^2\right)-\mu^2; \tag{5}$$

so that, for any observation, *i*,

$$\mathcal{E}\left(X_i^2\right)=\sigma^2+\mu^2. \tag{6}$$

Substituting (6) into (4) yields,

$$\begin{aligned}\mathcal{E}\left(\frac{\sum_i^N X_i^2}{N}\right)&=\frac{\sum_i^N \mathcal{E}\left(X_i^2\right)}{N}\\&=\frac{\sum_i^N\left(\sigma^2+\mu^2\right)}{N}\\&=\frac{N\left(\sigma^2+\mu^2\right)}{N}\\&=\sigma^2+\mu^2.\end{aligned} \tag{7}$$

Now, consider the second term to the right of the equal sign in (3), $\mathcal{E}\left(\bar{X}^2\right)$. The variance of the sampling distribution of means is:

$$\begin{aligned}\sigma_{\bar{X}}^2&=\mathcal{E}\left(\bar{X}^2\right)-[\mathcal{E}\left(\bar{X}\right)]^2\\&=\mathcal{E}\left(\bar{X}^2\right)-\mu^2\end{aligned}, \tag{8}$$

from which,

$$\mathcal{E}\left(\bar{X}^2\right)=\sigma_{\bar{X}}^2+\mu^2. \tag{9}$$

Substituting the expressions, (7) and (9) into (3) yields:

$$\mathcal{E}\left(s^2\right) = \mathcal{E}\left(\frac{\sum_i^N X_i^2}{N}\right) - \mathcal{E}\left(\bar{X}^2\right)$$
$$= \left(\sigma^2 + \mu^2\right) - \left(\sigma_{\bar{X}}^2 + \mu^2\right) \qquad (10)$$
$$= \sigma^2 - \sigma_{\bar{X}}^2.$$

In words, the expected value of the sample variance is the *difference* between the population variance, σ^2, and the variance of the distribution of sample means, $\sigma_{\bar{X}}^2$. Since the variance of the distribution of sample means typically is not zero, the sample variance *under-estimates* the population variance. In other words, the sample variance is a biased estimator of the population variance.

It has already been demonstrated, in (2), that the sample mean, $\bar{X}$, is an unbiased estimate of the population mean, μ. Now we need an unbiased estimate ($\hat{s}^2$) {note the tilde to imply *estimate*} of the population variance σ^2. In (10), it was shown that

$$\mathcal{E}\left(s^2\right) = \sigma^2 - \sigma_{\bar{X}}^2,$$

which, after substituting (11), yields

$$\mathcal{E}(s^2) = \sigma^2 - \frac{\sigma^2}{N}$$
$$= \frac{N}{N}\sigma^2 - \frac{\sigma^2}{N} \qquad (11)$$
$$= \left(\frac{N-1}{N}\right)\sigma^2.$$

Equation (13) shows that the average of the sample variances [$\mathcal{E}(s^2)$] is too small by a

factor of $\frac{N}{N-1}$.

Hence, an *unbiased* estimate of σ^2 is given by

$$\hat{s}^2 = \frac{N}{(N-1)}s^2.$$

It is easy to show that $\hat{s}^2$ is an unbiased estimate of σ^2:

$$
\begin{aligned}
\mathcal{E}\left(\hat{s}^2\right) &= \left(\frac{N}{N-1}\right)\mathcal{E}(s^{2)} \\
&= \left(\frac{N}{N-1}\right)\left(\frac{N-1}{N}\right)\sigma^2 \\
&= \sigma^2.
\end{aligned}
$$

Unbiased estimate of the standard error of the mean, $\sigma_{\bar{X}}$ **.**

The unbiased *estimate* of $\sigma_{\bar{X}}$ is given by $\hat{s}_{\bar{X}}$, where

$$
\begin{aligned}
\hat{s}_{\bar{X}} &= \sqrt{\frac{\hat{s}^2}{N}} \\
&= \sqrt{\left(\frac{1}{N}\right)\frac{N}{(N-1)}\hat{s}^2} \\
&= \sqrt{\frac{s^2}{N}} \\
&= \frac{\hat{s}}{\sqrt{N}}.
\end{aligned}
\tag{12}
$$

In words, the unbiased estimate of the *standard error of the mean* is the unbiased estimate of the population standard deviation divided by the square root of the sample size.