Bessel's correction

From Wikipedia, the free encyclopedia

In statistics, **Bessel's correction** is the use of $n - 1$ instead of n in the formula for the sample variance and sample standard deviation, where n is the number of observations in a sample. This method corrects the bias in the estimation of the population variance. It also partially corrects the bias in the estimation of the population standard deviation. However, the correction often increases the mean squared error in these estimations. This technique is named after Friedrich Bessel.

In estimating the population variance from a sample when the population mean is unknown, the uncorrected sample variance is the *mean* of the squares of deviations of sample values from the sample mean (i.e. using a multiplicative factor $1/n$). In this case, the sample variance is a biased estimator of the population variance.

Multiplying the uncorrected sample variance by the factor

$$\frac{n}{n-1}$$

gives an *unbiased* estimator of the population variance. In some literature,[1][2] the above factor is called **Bessel's correction**.

One can understand Bessel's correction as the degrees of freedom in the residuals vector (residuals, not errors, because the population mean is unknown):

$$(x_1 - \bar{x}, \dots, x_n - \bar{x}),$$

where is the sample mean. While there are n independent samples, there are only $n - 1$ independent residuals, as they sum to 0. For a more intuitive explanation of the need for Bessel's correction, see § Source of bias.

Generally Bessel's correction is an approach to reduce the bias due to finite sample count. Such finite-sample bias correction is also needed for other estimates like skew and kurtosis, but in these the inaccuracies are often significantly larger. To fully remove such bias it is necessary to do a more complex multi-parameter estimation. For instance a correct correction for the standard deviation depends on the kurtosis (normalized central 4th moment), but this again has a finite sample bias and it depends on the standard deviation, i.e. both estimations have to be merged.

Contents

Caveats [edit]

Further information: Unbiased estimation of standard deviation and Mean squared error § Variance

There are three caveats to consider regarding Bessel's correction:

1. It does not yield an unbiased estimator of standard *deviation*.

2. The corrected estimator often has a higher mean squared error (MSE) than the uncorrected estimator. Furthermore, there is no population distribution for which it has the minimum MSE because a different scale factor can always be chosen to minimize MSE.
3. It is only necessary when the population mean is unknown (and estimated as the sample mean). In practice this generally happens.

Firstly, while the sample variance (using Bessel's correction) is an unbiased estimator of the population variance, its square root, the sample standard deviation, is a *biased* estimate of the population standard deviation; because the square root is a concave function, the bias is downward, by Jensen's inequality. There is no general formula for an unbiased estimator of the population standard deviation, though there are correction factors for particular distributions, such as the normal; see unbiased estimation of standard deviation for details. An approximation for the exact correction factor for the normal distribution is given by using $n - 1.5$ in the formula: the bias decays quadratically (rather than linearly, as in the uncorrected form and Bessel's corrected form).

Secondly, the unbiased estimator does not minimize mean squared error (MSE), and generally has worse MSE than the uncorrected estimator (this varies with excess kurtosis). MSE can be minimized by using a different factor. The optimal value depends on excess kurtosis, as discussed in mean squared error: variance; for the normal distribution this is optimized by dividing by $n + 1$ (instead of $n - 1$ or n).

Thirdly, Bessel's correction is only necessary when the population mean is unknown, and one is estimating *both* population mean *and* population variance from a given sample set, using the sample mean to estimate the population mean. In that case there are n degrees of freedom in a sample of n points, and simultaneous estimation of mean and variance means one degree of freedom goes to the sample mean and the remaining $n - 1$ degrees of freedom (the *residuals*) go to the sample variance. However, if the population mean is known, then the deviations of the samples from the population mean have n degrees of freedom (because the mean is not being estimated – the deviations are not residuals but *errors*) and Bessel's correction is not applicable.

Source of bias [edit]

Suppose the mean of the whole population is 2050, but the statistician does not know that, and must estimate it based on this small sample chosen randomly from the population:

$$2051,\quad 2053,\quad 2055,\quad 2050,\quad 2051$$

One may compute the sample average:

$$\frac{1}{5}(2051 + 2053 + 2055 + 2050 + 2051) = 2052$$

This may serve as an observable estimate of the unobservable population average, which is 2050. Now we face the problem of estimating the population variance. That is the average of the squares of the deviations from 2050. If we knew that the population average is 2050, we could proceed as follows:

$$\begin{aligned}&\frac{1}{5}\left[(2051-2050)^2+(2053-2050)^2+(2055-2050)^2+(2050-2050)^2+(2051-2050)^2\right]\\=&\frac{36}{5}=7.2\end{aligned}$$

But our estimate of the population average is the sample average, 2052, not 2050. Therefore, we do what we can:

$$\begin{aligned}&\frac{1}{5}\left[(2051-2052)^2+(2053-2052)^2+(2055-2052)^2+(2050-2052)^2+(2051-2052)^2\right]\\=&\frac{16}{5}=3.2\end{aligned}$$

This is a substantially smaller estimate. Now a question arises: is the estimate of the population variance that arises in this way using the sample mean *always* smaller than what we would get if we used the population mean? The answer is *yes* except when the sample mean happens to be the same as the population mean.

We are seeking the sum of squares of distances from the population mean, but end up calculating the sum of squares of differences from the sample mean, which, as will be seen, is the number that minimizes that sum of squares of distances. So unless the sample happens to have the same mean as the population, this estimate will always underestimate the sum of squared differences from the population mean.

To see why this happens, we use a simple identity in algebra:

$$(a+b)^2 = a^2 + 2ab + b^2$$

With representing the deviation from an individual to the sample mean, and representing the deviation from the sample mean to the population mean. Note that we've simply decomposed the actual deviation from the (unknown) population mean into two components: the deviation to the sample mean, which we can compute, and the additional deviation to the population mean, which we can not. Now apply that identity to the squares of deviations from the population mean:

$$\begin{aligned}
[\underbrace{2053-2050}_{\begin{smallmatrix}\text{Deviation from} \\ \text{the population} \\ \text{mean}\end{smallmatrix}}]^2 &= [\overbrace{(\underbrace{2053-2052}_{\begin{smallmatrix}\text{Deviation from} \\ \text{the sample mean}\end{smallmatrix}})}^{\text{This is } a.} + \overbrace{(2052-2050)}^{\text{This is } b.}]^2 \\
&= \overbrace{(2053-2052)^2}^{\text{This is } a^2.} + \overbrace{2(2053-2052)(2052-2050)}^{\text{This is } 2ab.} + \overbrace{(2052-2050)^2}^{\text{This is } b^2.}
\end{aligned}$$

Now apply this to all five observations and observe certain patterns:

$$\begin{matrix}
\overbrace{(2051-2052)^2}^{\text{This is } a^2.} & + & \overbrace{2(2051-2052)(2052-2050)}^{\text{This is } 2ab.} & + & \overbrace{(2052-2050)^2}^{\text{This is } b^2.} \\
(2053-2052)^2 & + & 2(2053-2052)(2052-2050) & + & (2052-2050)^2 \\
(2055-2052)^2 & + & 2(2055-2052)(2052-2050) & + & (2052-2050)^2 \\
(2050-2052)^2 & + & 2(2050-2052)(2052-2050) & + & (2052-2050)^2 \\
(2051-2052)^2 & + & \underbrace{2(2051-2052)(2052-2050)}_{\begin{smallmatrix}\text{The sum of the entries in this} \\ \text{middle column must be 0.}\end{smallmatrix}} & + & (2052-2050)^2
\end{matrix}$$

The sum of the entries in the middle column must be zero because the sum of the deviations from the sample average must be zero. When the middle column has vanished, we then observe that

- The sum of the entries in the first column (a^2) is the sum of the squares of the deviations from the sample mean;
- The sum of *all* of the entries in the two columns (a^2 and b^2) is the sum of squares of the deviations from the population mean, because of the way we started with $[2053 - 2050]^2$, and did the same with the other four entries;
- The sum of *all* the entries must be bigger than the sum of the entries in the first column, since all the entries that have not vanished are positive (except when the population mean is the same as the sample mean, in which case all of the numbers in the last column will be 0).

Therefore:

- The sum of squares of the deviations from the *population* mean will be bigger than the sum of squares of the deviations from the *sample* mean (except when the population mean is the same as the sample mean, in which case the two are equal).

That is why the sum of squares of the deviations from the *sample* mean is too small to give an unbiased estimate of the population variance when the average of those squares is found.

Terminology [edit]

This correction is so common that the term "sample variance" and "sample standard deviation" are frequently used to mean the corrected estimators (unbiased sample variation, less biased sample standard deviation), using $n - 1$. However caution is needed: some calculators and software packages may provide for both or only the more unusual formulation. This article uses the following symbols and definitions:

μ is the population mean

is the sample mean

σ^2 is the population variance

s_n^2 is the biased sample variance (i.e. without Bessel's correction)

s^2 is the unbiased sample variance (i.e. with Bessel's correction)

The standard deviations will then be the square roots of the respective variances. Since the square root introduces bias, the terminology "uncorrected" and "corrected" is preferred for the standard deviation estimators:

s_n is the uncorrected sample standard deviation (i.e. without Bessel's correction)

s is the corrected sample standard deviation (i.e. with Bessel's correction), which is less biased, but still biased

Formula [edit]

The sample mean is given by

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

The biased sample variance is then written:

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{\sum_{i=1}^{n}(x_i^2)}{n} - \frac{(\sum_{i=1}^{n} x_i)^2}{n^2}$$

and the unbiased sample variance is written:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{\sum_{i=1}^{n}(x_i^2)}{n-1} - \frac{(\sum_{i=1}^{n} x_i)^2}{(n-1)n} = \left(\frac{n}{n-1}\right) s_n^2.$$

Proof of correctness – Alternate 1 [edit]

Click [show] to expand

As a background fact, we use the identity $E[x^2] = \mu^2 + \sigma^2$ which follows from the definition of the standard deviation and linearity of expectation.

A very helpful observation is that for any distribution, the variance equals half the expected value of $(x_1 - x_2)^2$ when x_1, x_2 are independent samples of that distribution. To prove this observation we will use that $E[x_1 x_2] = E[x_1]E[x_2]$ (which follows from the fact that they are independent) as well as linearity of expectation:

$$E[(x_1 - x_2)^2] = E[x_1^2] - E[2x_1x_2] + E[x_2^2] = (\sigma^2 + \mu^2) - 2\mu^2 + (\sigma^2 + \mu^2) = 2\sigma^2$$

Now that the observation is proven, it suffices to show that the expected squared difference of two samples from the sample population $x_1, \ldots, x_n$ equals $(n-1)/n$ times the expected squared difference of two samples from the original distribution. To see this, note that when we pick x_u and x_v via *u*, *v* being integers selected independently and uniformly from 1 to *n*, a fraction $n/n^2 = 1/n$ of the time we will have *u* = *v* and therefore the sampled squared difference is zero independent of the original distribution. The remaining $1 - 1/n$ of the time, the value of $E[(x_u - x_v)^2]$ is the expected squared difference between two unrelated samples from the original distribution. Therefore, dividing the sample expected squared difference by $(1 - 1/n)$, or equivalently multiplying by $1/(1 - 1/n) = n/(n-1),$ gives an unbiased estimate of the original expected squared difference.

Proof of correctness – Alternate 2 [edit]

Click [show] to expand

Recycling an identity for variance,

$$\begin{aligned}\sum_{i=1}^{n}(x_i - \overline{x})^2 &= \sum_{i=1}^{n}\left(x_i^2 - 2x_i\overline{x} + \overline{x}^2\right) \\ &= \sum_{i=1}^{n} x_i^2 - 2\overline{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\overline{x}^2 \\ &= \sum_{i=1}^{n} x_i^2 - 2n\overline{x}^2 + n\overline{x}^2 \\ &= \sum_{i=1}^{n} x_i^2 - n\overline{x}^2\end{aligned}$$

so

$$\begin{aligned}\operatorname{E}\left(\sum_{i=1}^{n}[(x_i - \mu) - (\overline{x} - \mu)]^2\right) &= \operatorname{E}\left(\sum_{i=1}^{n}(x_i - \mu)^2 - n(\overline{x} - \mu)^2\right) \\ &= \sum_{i=1}^{n}\operatorname{E}((x_i - \mu)^2) - n\operatorname{E}((\overline{x} - \mu)^2) \\ &= \sum_{i=1}^{n}\operatorname{Var}(x_i) - n\operatorname{Var}(\overline{x})\end{aligned}$$

and by definition,

$$\begin{aligned}\operatorname{E}(s^2) &= \operatorname{E}\left(\sum_{i=1}^n \frac{(x_i-\overline{x})^2}{n-1}\right)\\ &= \frac{1}{n-1}\operatorname{E}\left(\sum_{i=1}^n [(x_i-\mu)-(\overline{x}-\mu)]^2\right)\\ &= \frac{1}{n-1}\left[\sum_{i=1}^n \operatorname{Var}(x_i) - n\operatorname{Var}(\overline{x})\right]\end{aligned}$$

Note that, since $x_1, x_2, \cdots, x_n$ are a random sample from a distribution with variance σ^2, it follows that for each $i = 1, 2, \ldots, n$:

$$\operatorname{Var}(x_i) = \sigma^2$$

and also

$$\operatorname{Var}(\overline{x}) = \sigma^2/n$$

This is a property of the variance of uncorrelated variables, arising from the Bienaymé formula. The required result is then obtained by substituting these two formulae:

$$\operatorname{E}(s^2) = \frac{1}{n-1}\left[\sum_{i=1}^n \sigma^2 - n\sigma^2/n\right] = \frac{1}{n-1}(n\sigma^2-\sigma^2) = \sigma^2.$$

Proof of correctness – Alternate 3 [edit]

Click [show] to expand

The expected discrepancy between the biased estimator and the true variance is

$$\begin{aligned}\operatorname{E}[\sigma^2 - s^2_{\text{biased}}] &= \operatorname{E}\left[\frac{1}{n}\sum_{i=1}^n (x_i-\mu)^2 - \frac{1}{n}\sum_{i=1}^n (x_i-\overline{x})^2\right]\\ &= \frac{1}{n}\operatorname{E}\left[\sum_{i=1}^n \left((x_i^2 - 2x_i\mu+\mu^2) - (x_i^2 - 2x_i\overline{x} + \overline{x}^2)\right)\right]\\ &= \operatorname{E}\left[\mu^2 - \overline{x}^2 + \frac{1}{n}\sum_{i=1}^n (2x_i(\overline{x}-\mu))\right]\\ &= \operatorname{E}\left[\mu^2 - \overline{x}^2 + 2(\overline{x}-\mu)\overline{x}\right]\\ &= \operatorname{E}\left[\mu^2 - 2\overline{x}\mu + \overline{x}^2\right]\\ &= \operatorname{E}\left[(\overline{x}-\mu)^2\right]\\ &= \operatorname{Var}(\overline{x})\\ &= \frac{\sigma^2}{n}\end{aligned}$$

So, the expected value of the biased estimator will be

$$\operatorname{E}[s^2_{\text{biased}}] = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2$$

So, an unbiased estimator should be given by

$$s^2_{\text{unbiased}} = \frac{n}{n-1}s^2_{\text{biased}}$$

Intuition [edit]

In the biased estimator, by using the sample mean instead of the true mean, you are underestimating each $x_i - \mu$ by $\overline{x} - \mu$. We know that the variance of a sum is the sum of the variances (for uncorrelated variables). So, to find the discrepancy between the biased estimator and the true variance, we just need to find the variance of $\overline{x} - \mu$.

This is just the variance of the sample mean, which is σ^2/n. So, we expect that the biased estimator underestimates σ^2 by σ^2/n, and so the biased estimator = $(1 - 1/n)$ × the unbiased estimator = $(n - 1)/n$ × the unbiased estimator.

See also [edit]

- Bias of an estimator
- Standard deviation
- Unbiased estimation of standard deviation
- Jensen's inequality

Notes [edit]

1. ^ W. J. Reichmann, W. J. (1961) *Use and abuse of statistics*, Methuen. Reprinted 1964–1970 by Pelican. Appendix 8.
2. ^ Upton, G.; Cook, I. (2008) *Oxford Dictionary of Statistics*, OUP. ISBN 978-0-19-954145-4 (entry for "Variance (data)")

External links [edit]

- Weisstein, Eric W. "Bessel's Correction". *MathWorld*.
- Animated experiment demonstrating the correction, at Khan Academy

Categories: Statistical deviation and dispersion | Estimation methods