



WIKIPEDIA
The Free Encyclopedia

Bessel's correction

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,^[1] 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.

Formulation

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,^{[2][3]} 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 \bar{x} is the sample mean. While there are n independent observations in the sample, 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 size. 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.

Caveats

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.
^[4] 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, 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 observations 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

Most simply, to understand the bias that needs correcting, think of an extreme case. Suppose the population is (0,0,0,1,2,9), which has a population mean of 2 and a population variance of **31/3**. A sample of $n = 1$ is drawn, and it turns out to be $x_1 = 0$. The best estimate of the population mean is $\bar{x} = x_1/n = 0/1 = 0$. But what if we use the formula $(x_1 - \bar{x})^2/n = (0 - 0)/1 = 0$ to estimate the variance? The estimate of the variance would be zero – and the estimate would be zero for any population and any sample of $n = 1$. The problem is that in estimating the sample mean, the process has already made our estimate of the mean close to the value we sampled--identical, for $n = 1$. In the case of $n = 1$, the variance just cannot be estimated, because there is no variability in the sample.

But consider $n = 2$. Suppose the sample were (0, 2). Then $\bar{x} = 1$ and $[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2]/n = (1 + 1)/2 = 1$, but with Bessel's correction, $[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2]/(n - 1) = (1 + 1)/1 = 2$, which is an unbiased estimate (if all possible

samples of $n = 2$ are taken and this method is used, the average estimate will be 12.4, same as the sample variance with Bessel's correction.)

To see this in more detail, consider the following example. 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, 2053, 2055, 2050, 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:

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

But our estimate of the population average is the sample average, 2052. The actual average, 2050, is unknown. So the sample average, 2052, must be used:

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

The variance is now smaller, and it (almost) always is. The only exception occurs when the sample average and the population average are the same. To understand why, consider that variance *measures distance from a point*, and within a given sample, the average is precisely that point which minimises the distances. A variance calculation using *any* other average value must produce a larger result.

To see this algebraically, we use a simple identity:

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

With a representing the deviation of an individual sample from the sample mean, and b representing the deviation of the sample mean from the population mean. Note that we've simply decomposed the actual deviation of an individual sample from the (unknown) population mean into two components: the

deviation of the single sample from the sample mean, which we can compute, and the additional deviation of the sample mean from the population mean, which we can not. Now, we apply this identity to the squares of deviations from the population mean:

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

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

$$\begin{array}{l}
 \underbrace{(2051 - 2052)^2}_{\text{This is } a^2.} + \underbrace{2(2051 - 2052)(2052 - 2050)}_{\text{This is } 2ab.} + \underbrace{(2052 - 2050)^2}_{\text{This is } b^2.} \\
 (2053 - 2052)^2 + 2(2053 - 2052)(2052 - 2050) + (2052 - 2050)^2 \\
 (2055 - 2052)^2 + 2(2055 - 2052)(2052 - 2050) + (2052 - 2050)^2 \\
 (2050 - 2052)^2 + 2(2050 - 2052)(2052 - 2050) + (2052 - 2050)^2 \\
 (2051 - 2052)^2 + 2(2051 - 2052)(2052 - 2050) + (2052 - 2050)^2 \\
 \underbrace{\hspace{10em}}_{\text{The sum of the entries in this middle column must be 0.}}
 \end{array}$$

The sum of the entries in the middle column must be zero because the term a will be added across all 5 rows, which itself must equal zero. That is because a contains the 5 individual samples (left side within parentheses) which – when added – naturally have the same sum as adding 5 times the sample mean of those 5 numbers (2052). This means that a subtraction of these two sums must equal zero. The factor 2 and the term b in the middle column are equal for all rows, meaning that the relative difference across all rows in the middle column stays the same and can therefore be disregarded. The following statements explain the meaning of the remaining columns:

- The sum of the entries in the first column (a^2) is the sum of the squares of the distance from sample to sample mean;
- The sum of the entries in the last column (b^2) is the sum of squared distances between the measured sample mean and the correct population mean
- Every single row now consists of pairs of a^2 (biased, because the sample mean is used) and b^2 (correction of bias, because it takes the difference between the "real" population mean and the inaccurate sample mean into account). Therefore the sum of all entries of the first and last column now represents the correct variance, meaning that now the sum of squared distance between samples and population mean is used
- The sum of the a^2 -column and the b^2 -column must be bigger than the sum within entries of the a^2 -column, since all the entries within the b^2 -column 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 distance from samples to the *population* mean will always be bigger than the sum of squares of the distance to the *sample* mean, except when the sample mean happens to be the same as the population 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. The smaller the sample size, the larger is the difference between the sample variance and the population variance.

Terminology

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
- \bar{x} 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

The sample mean is given by

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

The biased sample variance is then written:

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

and the unbiased sample variance is written:

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

Proof

Suppose thus that X_1, \dots, X_n are independent and identically distributed random variables with expectation μ and variance σ^2 .

Knowing the values of the X_1, \dots, X_n at an outcome $\omega \in \Omega$ of the underlying sample space, we would like to get a good estimate for the variance σ^2 , which is unknown. To this end, we construct a mathematical formula containing the X_1, \dots, X_n such that the expectation of this formula is precisely σ^2 . This means that on average, this formula should produce the right answer.

The educated, but naive way of guessing such a formula would be

$$\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2,$$

where $x_k = X_k(\omega)$; this would be the variance if we had a discrete random variable on the discrete probability space $\{1, \dots, n\}$ that had value x_k at k . But let us calculate the expected value of this expression:

$$\begin{aligned} \mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2 \right] &= \mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n \left(x_k - \frac{1}{n} \sum_{j=1}^n x_j \right)^2 \right] \\ &= \mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n \left(\frac{1}{n} \sum_{j=1}^n (x_k - x_j) \right)^2 \right]; \end{aligned}$$

here we have (by independence, symmetric cancellation and equal distribution)

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{j=1}^n (x_k - x_j) \right)^2 \right] &= \mathbb{E} \left[\sum_{j=1}^n \sum_{l=1}^n (x_k - x_j)(x_k - x_l) \right] \\ &= n(n-1)\mathbb{E}[X_1^2] - n(n-1)\mathbb{E}[X_1]^2, \end{aligned}$$

and therefore

$$\mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2 \right] = \frac{n-1}{n} (\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2).$$

In contrast,

$$\operatorname{Var}(X_1) = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2.$$

Therefore, our initial guess was wrong by a factor of

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

and this is precisely Bessel's correction.

See also

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

Notes

- Radziwill, Nicole M (2017). *Statistics (the easier way) with R*. Lapis Lucera. ISBN 9780996916059. OCLC 1030532622 (https://www.worldcat.org/oclc/1030532622).
- W. J. Reichmann, W. J. (1961) *Use and abuse of statistics*, Methuen. Reprinted 1964–1970 by Pelican. Appendix 8.
- Upton, G.; Cook, I. (2008) *Oxford Dictionary of Statistics*, OUP. ISBN 978-0-19-954145-4 (entry for "Variance (data)")
- Rosenthal, Jeffrey S. (2015). "The Kids are Alright: Divide by n when estimating variance" (https://imstat.org/2015/11/17/the-kids-are-alright-divide-by-n-when-estimating-variance/). *Bulletin of the Institute of Mathematical Statistics*. December 2015: 9.

External links

- Weisstein, Eric W. "Bessel's Correction" (https://mathworld.wolfram.com/BesselsCorrection.html). *MathWorld*.
 - Animated experiment demonstrating the correction, at Khan Academy (https://www.khanacademy.org/cs/fishy-statistics-unbiased-estimate-of-population-variance/1183564841)
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Bessel%27s_correction&oldid=1211987825"

▪