Conditions

When do we say "variances of the two samples are equal and not equal"? How do we know whether the variances of two samples are equal or not?

Solution

In probability theory and statistics, the **variance** is a measure of how far a set of numbers is spread out. It is one of several descriptors of a probability distribution, describing how far the numbers lie from the mean (expected value). In particular, the variance is one of the moments of a distribution. In that context, it forms part of a systematic approach to distinguishing between probability distributions. While other such approaches have been developed, those based on moments are advantageous in terms of mathematical and computational simplicity.

The **variance** of a random variable or distribution is the expectation, or mean, of the squared deviation of that variable from its expected value or mean. Thus the variance is a measure of the amount of variation of the values of that variable, taking account of all possible values and their probabilities or weightings (not just the extremes which give the range).

Population variance and sample variance

In general, the ***population variance*** of a *finite* population of size *N* is given by

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2 = \frac{1}{N}\sum_{i=1}^{N}x_i^2-\mu^2$$

where

$$\mu = \frac{1}{N}\sum_{i=1}^{N}x_i$$

is the population mean, and

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

In many practical situations, the true variance of a population is not known *a priori* and must be computed somehow. When dealing with extremely large populations, it is not possible to count every object in the population.

A common task is to estimate the variance of a population from a sample.] We take a sample with replacement of *n* values $y_1, \ldots, y_n$ from the population, where $n < N$, and estimate the variance on the basis of this sample. There are several good estimators. Two of them are well known

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 = \left(\frac{1}{n}\sum_{i=1}^{n} y_i^2\right) - \overline{y}^2,$$ and

$$\begin{aligned} s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 \\ &= \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i^2 - \overline{y}^2\right) \\ &= \frac{1}{n-1}\sum_{i=1}^{n} y_i^2 - \frac{n}{n-1}\overline{y}^2 \end{aligned}$$

The first estimator, also known as the second central moment, is called the ***biased sample variance***. The second estimator is called the ***unbiased sample variance***. Either estimator may be simply referred to as the ***sample variance*** when the version can be determined by context. Here, $\overline{y}$ denotes the sample mean:

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

Conclusions: The 2 sample variances are equal, when s^2 are equal for 2 samples with their n – number and y_i