



planetmath.org

Math for the people, by the people.

trimmed mean

Canonical name	TrimmedMean
Date of creation	2013-03-22 14:42:02
Last modified on	2013-03-22 14:42:02
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	5
Author	CWoo (3771)
Entry type	Definition
Classification	msc 62F10
Classification	msc 62F35
Defines	winsorized mean
Defines	outlier
Defines	robust estimation

Let x_1, x_2, \dots, x_n be a set of real-valued data observations. Let $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ be the order statistics of the observations. The k th *trimmed mean* \bar{x}_k is defined as:

$$\bar{x}_k = \frac{x_{(k+1)} + x_{(k+2)} + \dots + x_{(n-k)}}{n - 2k} = \frac{1}{n - 2k} \sum_{i=k+1}^{n-k} x_{(i)}.$$

By ordering the original observations, and taking away the first k smallest observations and the first k largest observations, the trimmed mean takes the arithmetic average of the resulting data. The idea of a trimmed mean is to eliminate *outliers*, or extreme observations that do not seem to have any logical explanations in calculating the overall mean of a population.

For example, suppose 10 new lightbulbs are drawn from a population of 100 to find the average lifetime of a typical lightbulb, measured in number of hours. The measurements are 802, 854, 823, 428, 815, 840, 833, 809, 843, 821. The (arithmetic) mean of the measurement is

$$\frac{802 + 854 + 823 + 428 + 815 + 840 + 833 + 809 + 843 + 821}{10} = 786.8,$$

with sample standard deviation = 127.1, whereas the 1st trimmed mean gives:

$$\frac{802 + 823 + 815 + 840 + 833 + 809 + 843 + 821}{8} = 823.25,$$

with sample standard deviation = 14.6, greatly reducing the sample deviation.

The trimmed mean gives a much more robust estimation (an estimation not greatly affected by outliers) of the average than the arithmetic mean.

Another robust estimator of a mean is the *winsorized mean*. Like the trimmed mean, the winsorized mean eliminates the outliers at both ends of an ordered set of observations. Unlike the trimmed mean, the winsorized mean replaces the outliers with observed values, rather than discarding them. The formal definition of the k th winsorized mean w_k is:

$$w_k = \frac{(k+1)x_{(k+1)} + x_{(k+2)} + \dots + x_{(n-k-1)} + (k+1)x_{(n-k)}}{n} = \frac{kx_{(k+1)} + (n-2k)\bar{x}_k + kx_{(n-k)}}{n}.$$

From the definition, we see that the winsorized mean is the average of the observations where the first k smallest values are replaced by the $k+1$ th

smallest value, $x_{(k+1)}$, and the first k largest values are replaced by the $k+1$ th largest value, $x_{(n-k)}$.

From the above example, the 1st winsorized mean is

$$\frac{802 + 843 + 823 + 802 + 815 + 840 + 833 + 809 + 843 + 821}{10} = 823.1,$$

with sample standard deviation = 16.1, fairly close to the answer given by the trimmed mean.