E C O N O M I C S   B U L L E T I N

Median as a weighted arithmetic mean of all sample observations

SK Mishra
Dept. of Economics, NEHU, Shillong (India)

Abstract

It is generally held that median does not use all sample observations. However, median may be expressed as a weighted arithmetic mean of all sample observations. Some Monte Carlo studies have been conducted to show that the method works perfectly well.

Citation: Mishra, SK, (2004) "Median as a weighted arithmetic mean of all sample observations." *Economics Bulletin,* Vol. 3, No. 17 pp. 1−3
Submitted: May 29, 2004. **Accepted:** June 1, 2004.
URL: http://www.economicsbulletin.com/2004/volume3/EB−04C10009A.pdf

Median as a Weighted Arithmetic Mean of All Sample Observations

SK Mishra
Dept. of Economics
NEHU, Shillong

1. Introduction: Innumerably many textbooks in Statistics explicitly mention that one of the weaknesses (or properties) of ***median*** (a well known measure of central tendency) is that it is not computed by incorporating all sample observations. That is so because if the sample $x = (x_1,\ x_2, ..., x_n)$, where the variate values are ordered such that $x_1 \le x_2 \le ... \le x_n$ then $median(x) = (x_k + x_{n-k})/2;\ k = \text{int}((n+1)/2).$ Here int(.) is the integer value of (.). For example int(10 ≤ (n+1)/2 < 11) = 10. This formula, although queer and expressed in a little roundabout way, applies uniformly when n is odd or even. Evidently, $median(x)$ is not obtained by incorporating all the values of x, and so the alleged weakness of the median as a measure of central tendency.

2. The Median Minimizes the Absolute Norm of Deviations: It is a commonplace knowledge in Statistics that the statistic $\bar{x}$ (the arithmetic mean of x) minimizes the (squared) Euclidean norm of deviations of the variate values from itself or explicitly stated, it minimizes $S = \sum_{i=1}^{n} \left| x_i - c \right|^2$ since S attains its minimum when $c = \bar{x}$. To obtain this result, one may minimize $\sqrt{S}$ (the Euclidean norm per se) also. On the other hand the *median* minimizes the Absolute norm of deviations of the variate from itself, expressed as $M = \sum_{i=1}^{n} \left| x_i - c \right|$ which yields $c = median(x)$. In a general framework, we obtain arithmetic mean or median by minimizing the general Minkowski norm $\left[\sum_{i=1}^{n} \left| x_i - c \right|^p \right]^{1/p}$ for p=2 or p=1 respectively. This view of the arithmetic mean and the median gives them the meaning of being the measures of central tendency.

3. Indeterminacy of Median when the Number of Values in the Sample is Even: When in the sample $x = (x_1,\ x_2, ..., x_n)$, the number of observations, n, is odd, the value of $median(x) = (x_k + x_{n-k})/2;\ k = \text{int}((n+1)/2)$ is determinate; $x_k = x_{n-k}$ minimizes the absolute norm, M. However, when n is an even number, x_k and x_{n-k} are (very often) different. As a matter of fact, any number z for which the relationship ($x_k \le z \le x_{n-k}$) holds, minimizes the absolute norm of deviations. Thus, the median is indeterminate. It has been customary, therefore, that in absence of any other relevant information, one uses the ***principle of insufficient reason*** and obtains $median(x) = (x_k + x_{n-k})/2.$ However, it remains a truth that any number z for which the relationship ($x_k \le z \le x_{n-k}$) holds, is the value of the median as much as $z = (x_k + x_{n-k})/2.$

4. Median as a Weighted Arithmetic Mean of Sample Observations: If $x = (x_1,\ x_2,...,x_n)$ are ordered such that $x_1 \le x_2 \le ... \le x_n$, it is possible to express median as a weighted arithmetic mean $\left(\sum_{j=1}^{n} x_j w_j \Big/ \sum_{j=1}^{n} w_j\right)$ where $w_j = w_{n-j} = 0.5;$ for $j = \text{int}((n+1)/2)$ else $w_j = 0$ for $j \ne \text{int}((n+1)/2)$. However, this is trivial.

Now we present a non-trivial alternative algorithm to obtain $median(x)$. In order to use this algorithm it is not necessary that the values of x be arranged in an ascending (or descending) order, that is $x_1 \le x_2 \le ... \le x_n$ condition is relaxed. The steps in the algorithm are as follows:

(i) Set $w_i = 1\ \forall\ i = 1,2,...,n.$ Obviously, $\sum_{i=1}^{n} w_i = n.$

(ii) Find $v_1 = \dfrac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}$, the weighted arithmetic mean of $(x_1,\ x_2,...,x_n)$.

(iii) Find new $w_i = 1\Big/|d_i|$ if $d_i = |x_i - v_1| \ge \varepsilon$ ($\varepsilon \ge 0$ is a small number, say 0.000001), else $w_i = 0.000001$ or any such small number.

(iv) Find $v_2 = \dfrac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}$ using the weights obtained in (iii) above.

(v) If $|v_1 - v_2| \ge \tau$ (where τ is a very small number, say, 0.00001 or so, controlling the accuracy of result) then v_1 is replaced by v_2 (that is, v_2 is renamed as v_1) and go to step (iii); **else**

(vi) Median is v_2 and $w = (w_1, w_2,..., w_n)$ are the weights associated with $(x_1,\ x_2,...,x_n)$. Stop.

This algorithm yields non-trivial weights $w = (w_1, w_2,..., w_n)$. It yields the median identical to that obtained by the conventional formula if n is odd. If n is even, it gives a number z : ($x_k \le z \le x_{n-k}$), which is median as mentioned in section 2.

5. Some Monte Carlo Experiments: We have conducted some Monte Carlo experiments to study the performance of the alternative method (weighted arithmetic mean representation) vis-à-vis the conventional method of obtaining median. Three sample sizes (of n = 10, 21 and 50) have been considered. Samples have been drawn from five distributions (Normal, $Beta_1$, $Beta_2$, Gamma and Uniform). In each case 10,000 experiments have been carried out. A success of the alternative estimator is there if it obtains median identical to that obtained by the traditional method in case n is odd and obtains median = z : ($x_k \le z \le x_{n-k}$) in case n is even. The summary of results is presented in table 1.

Table 1. Performance of the Alternative method to obtain Median

Distribn.	Sample Size = n	Arithmetic Mean	Median (Traditional)	Median (Alternative)	Inclination to Mean	Success Rate (%)
Uniform	10	50.00490	50.04172	50.00207		100
	21	49.98979	50.02352	50.02352		100
	50	49.99520	50.05171	50.07819		100
Gamma	10	2.50630	1.35170	1.57886	yes	100
	21	2.50647	1.23411	1.23415		100
	50	2.50656	1.17245	1.22250	yes	100
$Beta_1$	10	251.66043	251.45144	251.47560		100
	21	251.63371	252.62551	252.62551		100
	50	251.64002	253.49261	252.82187		100
$Beta_2$	10	3343.83487	614.92200	741.89412	yes	100
	21	3346.41080	526.19264	526.19264		100
	50	3346.43339	500.91137	519.64713	yes	100
Normal	10	0.00062	-0.00149	0.02596		100
	21	0.04474	0.39990	0.39990		100
	50	0.07170	-0.11195	-0.14733		100

We find that when n is odd, irrespective of the distribution or the sample size both the methods yield identical results. When the distribution is skewed (i.e. there is a significant divergence between median and mean) and n is even, the alternative median is slightly pulled by the mean (its inclination is towards the mean). This appears justified because it is expected that the values lying between x_k and x_{n-k} (for $x = (x_1, x_2, ..., x_n)$: $x_1 \le x_2 \le ... \le x_n$; $k = \text{int}((n+1)/2)$) must be more densely distributed in the side of the mean. The traditional method, however, considers them uniformly distributed in want of information. The alternative method appears to exploit the information contained in the sample.

6. Conclusion: This study establishes that median may be expressed as a weighted arithmetic mean of all sample observations. If the traditional formula does not incorporate all sample values, it is the property of the specific method of computation and not of median per se, as often alleged to it. Additionally, the alternative method of computation is easily extended to other median type estimators - such as Least Absolute Deviation (LAD) estimator of the regression model $y = X\beta + u$ - as shown by Fair (1974) and Schlossmacher (1973).

References

Fair, RC (1974). "On the Robust Estimation of Econometric Models", *Annals of Economic and Social Measurement*, 3 (667-677).

Schlossmacher, EJ (1973). "An Iterative Technique for Absolute Deviations Curve Fitting", *Journal of the American Statistical Association*, 68 (857-859).