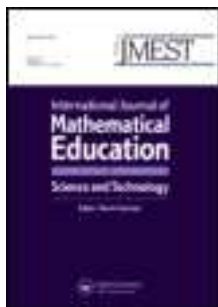


This article was downloaded by: [Mian Adnan]

On: 12 September 2013, At: 07:45

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



International Journal of Mathematical Education in Science and Technology

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/tmes20>

Shorter variation of standard deviation for small sample

M. Forhad Hossain^a, Mian Arif Shams Adnan^a & A.H. Joarder^b

^a Department of Statistics, Jahangirnagar University, Dhaka 1342, Bangladesh

^b Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

Published online: 11 Sep 2013.

To cite this article: M. Forhad Hossain, Mian Arif Shams Adnan & A.H. Joarder, International Journal of Mathematical Education in Science and Technology (2013): Shorter variation of standard deviation for small sample, International Journal of Mathematical Education in Science and Technology, DOI: 10.1080/0020739X.2013.822588

To link to this article: <http://dx.doi.org/10.1080/0020739X.2013.822588>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms &

CLASSROOM NOTE

Shorter variation of standard deviation for small sample

M. Forhad Hossain,^a Mian Arif Shams Adnan^{a*} and A.H. Joarder^b

^aDepartment of Statistics, Jahangirnagar University, Dhaka 1342, Bangladesh; ^bDepartment of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

(Received 10 March 2013)

The present study ensembles the variation of standard deviation in terms of range for small sample. Smaller ranges having minimum upper limit of the sample standard deviation and sample mean deviation have been found. A comparative study is conducted between their dispersions and evaluated by a couple of examples.

Keywords: mean deviation measured from median; range

1. Introduction

The measure of dispersion is one of the fundamental concepts for the readers of statistics. Unfortunately, the available inequality (as for example, Bhattacharya inequality) for explaining the variation of the measure of dispersion, like sample standard deviation, may be hard enough to understand especially for non-statisticians. Moreover, from the statistical point of view, some questions are unavoidable, which are as follows: (1) does the inequality provide the shorter variation of the sample standard deviation?, (2) is it possible to suggest a relatively more shorter range for the sample standard deviation? and (3) can the shorter range be easily grabbed by the users from various disciplines? A naïve range for the measure of dispersion may be an innovative alternative to the present statistics more simply to the multi-disciplinary readers. The determination of standard deviation for small sample is studied by authors like Joarder,[3–6] McLeod and Henderson,[7] Petocz,[8] Shiffler and Harsha,[9] and Thomson.[10] This study aims to suggest relatively smaller range of some dispersion measures.

2. Preliminaries

McLeod and Henderson,[7] Thompson,[10] and Joarder and Latiff [6] derived the lower bound for standard deviation such as

$$s \geq \frac{d}{\sqrt{2(n-1)}}, \quad (1)$$

where s and d stand for standard deviation and range of all the sample observations, respectively. For a sample of size 3, consisting of 0, a , ka (since variance does not depend

*Corresponding author. Email: julias284@yahoo.com

on the location), Corucher [1] has shown that

$$s^2 = \frac{d^2}{3} \left[1 - \frac{1}{k} + \frac{1}{k^2} \right]$$

known as Croucher conjecture based on the three samples and true for $k = 1$ such that $s \leq \frac{d}{\sqrt{3}} \approx 0.577d$.

For d_i to be the successive difference between the $(i + 1)$ th largest observation and the i th largest observation in a sample of size 4, Joarder and Latif [6] presented sample standard deviation for that sample as

$$s^2 = \frac{1}{4(4-1)} [d_1^2 + d_2^2 + d_3^2 + (d_1 + d_2)^2 + (d_2 + d_3)^2 + (d_1 + d_2 + d_3)^2].$$

With an aim of finding the possible lower range, let us consider a sample of size 4 such as 0, a , ka , $2ka$ where $d_1 = a$, $d_2 = ka - a$, $d_3 = ka$, $d = 2ka$. Therefore, s can be written (in Equation (A1)) as

$$s = \frac{d}{4\sqrt{3}} \sqrt{11 - \frac{6}{k} + \frac{3}{k^2}} \text{ such that } s \leq 0.4787d \text{ for } k \geq 1.$$

Similarly, for a sample of size 5, let 0, a , ka , $2ka$ and $3ka$ be the observations for which (from Equation (A2))

$$s = \frac{d}{3\sqrt{10}} \sqrt{17 - \frac{6}{k} + \frac{2}{k^2}} \text{ iff } s \leq 0.4346d \text{ for } k \geq 1.$$

Now if we consider the data (60, 65, 70, 80) then by the change of the location parameter 60 (since variance does not depend on the change of location) we can rewrite the data as 0, 5, 10, 20 that is, 0, a , ka , $2ka$ where $k = 2$, $a = 5$ and s can be written (from Equation (A3)) as

$$s \leq 0.426956d \text{ for } k = 1.$$

Again, for the data set (60, 65, 70, 80, 90) after the shift of the location the data set for estimating sample variance can be thought 0, 5, 10, 20, 30 or 0, a , ka , $2ka$, $3ka$ where $k = 2$, $a = 5$ and s can be similarly written (from Equation (A4)) as

$$s \leq 0.4014d \text{ for } k = 1.$$

3. Main results

Since the measure of central value using mean is not always usable, especially when the data set contains some extreme values or is asymmetric, the mean does not represent the right central value of the distribution. On the other hand, median always represents the central value of a data set or a distribution. This is why some authors, nowadays, emphasize on median rather mean to measure the location. In such a case, the dispersions of the observations are suggested to be measured by mean deviation from median (md)

in lieu of standard deviation,[11] as md is less than standard deviation or any other type of measure of dispersion. Moreover, in these days of computer, the calculation of md is not a difficult task. Therefore, we are using md for measuring scatterness and will present some inequalities based on the range of the data set to compare standard deviation to md.

Let the sample size (n) is 3 and x_1, x_2 and x_3 are ordered data. Now if m refers to the median, then md will (from Equation (A5)) be

$$\text{md} = \frac{1}{n} \sum_{i=1}^n |x_i - m| = \frac{1}{3} [|x_1 - m| + |x_2 - m| + |x_3 - m|] = \frac{1}{3} d \forall \text{md} \leq 0.33d,$$

whereas from Croucher [1] conjecture it was obtained that

$$s \leq \frac{d}{\sqrt{3}} = \frac{3\text{md}}{\sqrt{3}} \forall s \leq \sqrt{3}\text{md}.$$

Now for any data set $x_1, x_2, \dots, x_n$ if n is odd, then (from Equation (A6))

$$\text{md} \leq \frac{d}{2} \left(1 - \frac{1}{n}\right).$$

Combining the aforementioned two successive inequalities, another inequality can be written as follows:

$$\frac{d}{3} = \text{md} \leq \frac{d}{2} \left(1 - \frac{1}{n}\right). \quad (2)$$

That is, md cannot exceed 50% of the range of the data set and is not less than one-third ($\frac{1}{3} = 33.33\%$) of the range of the data for any type (or any odd size) of the data set (or distribution).

4. Comparison

Now for evaluating the upper limit (in terms of range of the data) of the md, a typically distributed data set will be considered for checking whether the upper range reduce substantially from 50% of d . As for an example, if we have a sample of size n ($= 9$), $\text{md} \leq 0.3125d \left(1 - \frac{1}{n}\right)$ if the data are symmetric (from Equation (A7)); $\text{md} \leq 0.280d$ if uniform; $\text{md} \leq 0.203d \left(1 - \frac{1}{n}\right) = 0.107d$ if uniformly symmetric consistent; $\text{md} \leq 0.107d \left(1 - \frac{1}{n}\right) = 0.105d$ if uniformly symmetric more consistent; $\text{md} \leq 0.11d \left(1 - \frac{1}{n}\right) = 0.098d$ if normal; $\text{md} \leq 0.2d \left(1 - \frac{1}{n}\right) = 0.178d$ if positively skewed; and $\text{md} \leq 0.1d \left(1 - \frac{1}{n}\right) = 0.089d$ if extremely negatively skewed, which (the coefficients of the range ' d ' in the upper range of md) are much more smaller than the results we obtain for standard deviation ($s \leq 0.577d$) from Croucher conjecture (2004). Similar types of results may also be possible to obtain from any type of data set of size of even number.

Now for another inquiry of odd number sample size, let us have a sample of size $n = 5$ for which $\text{md} \leq \frac{3}{8}d \left(1 - \frac{1}{n}\right) = 0.300d$ if the data set is uniform; $\text{md} \leq \frac{9}{32}d \left(1 - \frac{1}{n}\right) = 0.230d$ if uniformly symmetric consistent; $\text{md} \leq \frac{17}{64}d \left(1 - \frac{1}{n}\right) = 0.210d$ if uniformly symmetric more consistent; $\text{md} \leq \frac{5}{16}d \left(1 - \frac{1}{n}\right) = 0.250d$ if normal

of symmetric; $md \leq \frac{9}{32}d \left(1 - \frac{1}{n}\right) = 0.225d$ if positively skewed; $md \leq \frac{33}{128}d \left(1 - \frac{1}{n}\right) = 0.206d$ if extremely negatively skewed. Still for any sample of small size, the md has a lower upper limit than that for standard deviation.

Finally, the limits of the sample md and sample standard deviation (s) in terms of the sample range (d) can be notified, respectively, as follows:

$$\frac{d}{n} \leq md \leq \frac{d}{2}, \frac{d}{\sqrt{2n}} \leq sd \leq d.$$

From the above relation (in terms of the range), this is evident that the interval of sample standard deviation is not smaller than that of md of the data set.

5. Conclusion

Dispersion is a very basic statistic for any data set. Maximum statistical tools depend on dispersion measure as well as its variation. Variation of the dispersion measure might not be instantly accessible. In this circumstance, the range-based variation of the dispersion measure can be very simple expression for assessing the measure of dispersion and its fluctuation.

References

- [1] Croucher JS. An upper bound on the value of the standard deviation. *Teach Stat.* 2004;26(2): 54–55.
- [2] Hossain F, Adnan MAS. A new approach to determine the asymmetry of a distribution. *J Appl Statist Sci.* 2007;15(1):127–134.
- [3] Joarder AH. On some representations of sample variance. *Int J Math Educ Sci Technol.* 2002;33(5):772–784.
- [4] Joarder AH. Sample variance and first-order differences of observations. *Math Sci.* 2003;28:129–133.
- [5] Joarder AH, Laradji A. Some inequalities in descriptive statistics. Saudi Arabia: Department of Mathematical Sciences, King Fahd University of Petroleum and Minerals; 2004. (Technical Report No. 321).
- [6] Joarder AH, Latif RM. Standard deviation for small samples. *Teach Stat.* 2006;28(2):40–43.
- [7] Mcleod AJ, Henderson GR. Bounds for the sample standard deviation. *Teach Stat.* 1984;6(3): 72–76.
- [8] Petocz P. An upper bound on standard deviation as a function of range. *Teach Stat.* 2005;27(2):42–44.
- [9] Shiffler RE, Harsha PD. Upper and lower bounds for the sample standard deviation. *Teach Stat.* 1980;2(3):84–86.
- [10] Thomson GW. Bounds for the ratio of range to standard deviation. *Biometrika* 1955;42:268–269.
- [11] Hossain F, Adnan MAS. A new approach to determine the asymmetry of a distribution. *J. Appl. Statist. Sci.* 2007;15:127–134.

Appendix

$$\begin{aligned} s^2 &= \frac{1}{4(4-1)} [d_1^2 + d_2^2 + d_3^2 + (d_1 + d_2)^2 + (d_2 + d_3)^2 + (d_1 + d_2 + d_3)^2] \\ &= \frac{1}{12} [a^2 + a^2(k-1)^2 + a^2k^2 + a^2k^2 + a^2(2k-1)^2 + (2ak)^2] \\ &= a^2 [3 - 6k + 11k^2] = \frac{d^2}{48} \left[11 - \frac{6}{k} + \frac{3}{k^2} \right]. \end{aligned} \quad (A1)$$

$$d_1 = a, d_2 = a(k-1), d_3 = ka, d_4 = ka, d = 3ka;$$

$$\begin{aligned} s^2 &= \frac{1}{5(5-1)} \left[\frac{d_1^2 + d_2^2 + d_3^2 + d_4^2 + (d_1 + d_2)^2 + (d_2 + d_3)^2 + (d_3 + d_4)^2}{+(d_1 + d_2 + d_3)^2 + (d_2 + d_3 + d_4)^2 + d^2} \right] \\ &= \frac{1}{5(5-1)} \left[\frac{a^2 + a^2(k-1)^2 + k^2a^2 + k^2a^2 + k^2a^2 + a^2(2k-1)^2 + (2ak)^2}{+(2ka)^2 + a^2(3k-1)^2 + (3ka)^2} \right] \\ &= \frac{a^2}{20} [34k^2 - 12k + 4] = \frac{1}{10} \cdot \frac{d^2}{9} \left[17 - \frac{6}{k} + \frac{2}{k^2} \right]. \end{aligned} \quad (A2)$$

Using Equation (A1), we get

$$s^2 = \frac{d^2}{48} \left[11 - \frac{6}{k} + \frac{3}{k^2} \right] = 0.182292d^2. \quad (A3)$$

Using Equation (A2), we get

$$s^2 = \frac{1}{10} \cdot \frac{d^2}{9} \left[17 - \frac{6}{k} + \frac{2}{k^2} \right] = 0.161111d^2. \quad (A4)$$

For the ordered data x_1, x_2, x_3 ; median (m) is x_2 , range (d) is $x_3 - x_1$ and $|x_1 - x_2| = d_1$, $|x_3 - x_2| = d_2$.

$$\text{Therefore, md} = \frac{1}{3} [|x_1 - x_2| + |x_2 - x_3| + |x_3 - x_1|] = d_1 + d_2 = \frac{d}{3}. \quad (A5)$$

For the data set since n is odd, the median (m) is the $\frac{n+1}{2}$ th observation and there are $(n-1)$ non-negative d_i 's where

$$\left| x_1 - x_{(\frac{n+1}{2})} \right| = d_1, \left| x_2 - x_{(\frac{n+1}{2})} \right| = d_2, \dots, \left| x_{(\frac{n+1}{2})} - x_{(\frac{n+1}{2})} \right| = 0, \dots, \left| x_n - x_{(\frac{n+1}{2})} \right| = d_n. \quad (A6)$$

$$\text{Since } \sum_{i=1}^n d_i < \frac{(n-1)d}{2}; \text{ therefore, md} = \frac{1}{n} \sum |x_i - m| = \frac{\sum_{i=1}^n d_i}{n} < \frac{(n-1)d}{n} < \frac{(n-1)d}{2}.$$

$$\begin{aligned} \text{md} &\leq \frac{1}{n} \sum |x_i - m| \leq \frac{1}{n} \left[\frac{n-1}{8} \cdot \frac{d}{2} + \frac{n-1}{8} \cdot \frac{3d}{8} + \frac{n-1}{8} \cdot \frac{d}{4} + \frac{n-1}{8} \cdot \frac{d}{8} \right] \times 2 \\ &\leq \frac{(n-1)d}{n} \left[\frac{1}{8} + \frac{3}{32} + \frac{1}{16} + \frac{1}{32} \right] \leq \frac{d(n-1)}{n} \left[\frac{10}{32} \right]. \end{aligned} \quad (A7)$$