Introduction of Geometric Deviation as a New Measure of Dispersion For Proportional Data

Dr Syed Mohd Haider Zaidi[1], Dr Syed Zeashan Haider Zaidi [2], Dr Syed Qaim Akbar Rizvi[3],Arshita Srivastava[4]

1.Associate Professor, [2]Assisstant Professor, [3]Assisstant Professor, [4]Research scholar

[1]Department of Statistics,

[1]Shia P.G.College, Lucknow, India

Abstract

The paper introduces a new measure of dispersion, the **Geometric Deviation** for proportional data & comparisons between proportional relationship and linear relationship using geometric deviation are discussed. Two numerical examples are also illustrated.

Introduction :

Measures of dispersion shows how the various observations of a set of data are dispersed or scattered in relation to the average or central value. These measures attain large values if the observations are distantly scattered else if the data is close to average then they have small values.

Some measures of dispersion are the Range, the Quartile Deviation, the Mean Deviation and the Standard Deviation, in which the Standard Deviation is considered to the best. The standard deviation (SD) is a measure that is used to quantify the amount of variation or dispersion of a set of data values [1]. A low standard deviation indicates that the data points tend to be close to the average of the set, while a high standard deviation indicates that the data points are more dispersed over a wider range. The term standard deviation was first used [2] in writing by Karl Pearson in 1894, following his use of it in lectures.

But Standard Deviation always measures variation of the data about Arithmetic Mean [4]. Standard Deviation is not taken into consideration when there is some other type of average suited best to the given data. For example in case of averaging ratio or percentages we use geometric mean [3]. There are many cases in real life where the data observations show proportional changes or trend. Here Geometric Mean is the best for averaging these observations. Clearly in these cases Standard Deviation does not provide satisfactory results for dispersion. So there is need for a measure of dispersion, which is more suited for proportional changes.

The Geometric Deviation :

Now we introduce a new measure of dispersion. We called it the Geometric Deviation. It is defined as below :

Let x_i (i=1,2,.....,n) be the set of n observations, all are positive. Let A be an average of these observations.

Define r_i as follows

$$r_i = \begin{cases} x_i / A & \text{when} \quad x_i > A \\ A / x_i & \text{when} \quad x_i < A \end{cases} \qquad \text{---(1)}$$

$r_i = 1$ when $x_i = 1$

Now the geometric deviation about point *A* of these n observations is given by

$$Gd(A) = \left(\prod_{i=1}^{n} r_i \right)^{1/n} \qquad \text{---(2)}$$

And if the i^{th} observation x_i has frequency f_i then the geometric deviation is given by

$$Gd(A) = \left(\prod_{i=1}^{n} r_i^{f_i} \right)^{1/N}, \text{ where } N = \sum_{i=1}^{n} f_i$$

Since in proportion data, we generally use the geometric mean (*G*) as the average. So the geometric deviation about geometric mean will be the best measure of dispersion for proportion data (in case of growth, Index Numbers etc.) and it can be proved.

Properties of Geometric Deviation :

Now we discuss some important properties of our new measure of dispersion

(1) Qualities of geometric deviation

a. It is rigidly defined.

b. It is based on all the observations.

c. It is capable of mathematical treatment.

d. It is unit-less. Hence we can compare variability of two data containing different units.

e. It is easily understood.

It has some limitations also. We can't use geometric deviation for non positive observations. Though the calculation of geometric deviation is quite tedious but it is computer age and it is possible to deal with such tedious calculations.

(2) Relation between geometric deviation and mean deviation.

We have

$$\begin{aligned} \log Gd(A) &= \frac{1}{N} \sum_{i=1}^{n} f_i \log r_i \\ &= \frac{1}{N} \sum_{i=1}^{n} f_i \left| \log x_i - \log A \right| \\ &= \text{Mean deviation of } \log x_i^s \text{ about the point } \log A \end{aligned}$$

(3) Let A and G are the arithmetic mean and geometric mean of a given data set $x_i \ (i = 1, 2, ..., n)$ respectively. Then a fascinating result about geometric deviation is that

$Gd(A) < Gd(G) \Leftrightarrow$ There is more linear relationship among data values

The above result can be used as a criterion for testing whether the given data set is linear or proportional and which average (arithmetic mean or geometric mean) is best suited for it. We can prove this result into the following manner:

MATHEMATICAL PROOF:

Case I :

If the observations $x_i^{'s}\ \left(i=1,2,...,n\right)$ are proportionally related then they can be written as

$x_1=a,\ x_2=ar,\ \ldots,x_n=ar^{n-1};$ where $a>0$

without loss of generality assumption can be made that $r>1$.

Consider the case when n is odd. i.e. $n = 2m+1$ (m is a non negative integer). For even n, the result can be proved in a similar way.

Here the geometric mean $G=ar^m$

So

$$\left[Gd\left(G\right)\right]^n=\frac{ar^{2m}\cdot ar^{2m-1}\cdot\ldots\cdot ar^{m+1}}{a\cdot ar\cdot ar^2\cdot\ldots\cdot ar^{m-1}}=r^{m(m+1)} \qquad \text{---(3)}$$

Now let the arithmetic mean of the above series is A. Let in the series of $x_i^{'s}$ the two consecutive values $ar^{2m-(d-1)}$ and ar^{2m-d} then

$$\begin{aligned}\left[Gd\left(A\right)\right]^n&=\frac{ar^{2m}\cdot ar^{2m-1}\cdot\ldots\cdot ar^{2m-(d-1)}}{a\cdot ar\cdot ar^2\cdot\ldots\cdot ar^{2m-d}}\left(A\right)^{2m+1-2d}\\&>\frac{ar^{2m}\cdot ar^{2m-1}\cdot\ldots\cdot ar^{2m-(d-1)}}{a\cdot ar\cdot ar^2\cdot\ldots\cdot ar^{2m-d}}\left(ar^{2m-d}\right)^{2m+1-2d}\\&=\frac{r^{2m}\cdot r^{2m-1}\cdot\ldots\cdot r^{2m-(d-1)}}{r\cdot r^2\cdot\ldots\cdot r^{2m-d}}\left(r^{2m-d}\right)^{2m+1-2d}\\&=\frac{r^d\cdot r^d\cdot\ldots d\text{ times}}{r\cdot r^2\cdot\ldots\cdot r^{2m+1-2d-1}}\left(r^{2m-d}\right)^{2m+1-2d}\\&=\frac{r^d\cdot r^d\cdot\ldots d\text{ times}}{r\cdot r^2\cdot\ldots\cdot r^{2m+1-2d-1}}\left(r^{2m-d}\right)^{2m+1-2d}\\&=\frac{r^{d^2}}{r^{\frac{1}{2}(2m-d)(2m-d+1)}}\left(r^{2m-d}\right)^{2m+1-2d}\\&=r^{d^2}\cdot r^{(2m-2d+1)(2m-d-m+d)}\end{aligned}$$

$$\Rightarrow\left[Gd\left(A\right)\right]^n>r^{2m^2+d^2+m-2md} \qquad \text{---(4)}$$

Now consider the ratio

$$\begin{aligned}\frac{r^{2m^2+d^2+m-2md}}{r^{m(m+1)}}&=r^{m^2+d^2-2md}\\&=r^{(m-d)^2}\\&>1\end{aligned}$$

So,

$$r^{2m^2+d^2+m-2md}>r^{m(m+1)}$$

Hence,

$$\left[Gd\left(A\right)\right]^n > \left[Gd\left(G\right)\right]^n$$

So it comes out for the proportional data,

$$Gd\left(A\right) > Gd\left(G\right)$$

Case II :

If the observations $x_i^{'s}$ $\left(i=1,2,...,n\right)$ are linearly related then they can be written as

$$x_1 = a,\ x_2 = a+d,\ \ldots,\ x_n = a+\left(n-1\right)d;\ \text{where } a>0$$

Without loss of generality assumption can be made that $d>0$.

Consider the case when n is odd. *i.e.* $n=2m+1$(m is non negative integer). For even n, the result can be proved in a similar way.

Here the arithmetic mean $A=a+md$

So,

$$\left[Gd\left(A\right)\right]^n = \frac{\left[a+2md\right]\left[a+\left(2m-1\right)d\right]\ldots\left[a+\left(m+1\right)d\right]}{a\left[a+d\right]\left[a+2d\right]\ldots\left[a+\left(m-1\right)d\right]} \qquad \text{---(5)}$$

Now let in the series of $x_i^{'s}$ the two consecutive values $a+\left(r-1\right)d$ and $a+rd$ are such that

$$a+\left(r-1\right)d < G < a+rd\ ,$$

Clearly $a+\left(r-1\right)d \le a+md$

Then

$$\left[Gd\left(G\right)\right]^n = \frac{\left[a+2md\right]\left[a+\left(2m-1\right)d\right]\ldots\left[a+\left(r+1\right)d\right]}{a\left[a+d\right]\left[a+2d\right]\ldots\left[a+\left(r-1\right)d\right]} \cdot \frac{\left[a+rd\right]}{G^{2(m-r)+1}}$$

$G < a+rd$ implies $\frac{1}{G^2} > \frac{1}{\left(a+rd\right)^2}$

Thus

$$\left[Gd\left(G\right)\right]^n > \frac{\left[a+2md\right]\left[a+\left(2m-1\right)d\right]\ldots\left[a+\left(r+1\right)d\right]}{a\left[a+d\right]\left[a+2d\right]\ldots\left[a+\left(r-1\right)d\right]\left[a+rd\right]} \cdot \frac{1}{G^{2(m-r-1)+1}}$$

Again

$$\frac{1}{G^2} > \frac{1}{\left[a+\left(r+1\right)d\right]^2}$$

So

$$\left[Gd\left(G\right)\right]^n > \frac{\left[a+2md\right]\left[a+\left(2m-1\right)d\right]...\left[a+\left(r+2\right)d\right]}{a\left[a+d\right]\left[a+2d\right]...\left[a+\left(r-1\right)d\right]\left[a+rd\right]\left[a+\left(r+1\right)d\right]} \cdot \frac{1}{G^{2(m-r-2)+1}}$$

Continuing this process, finally we get

$$\left[Gd\left(G\right)\right]^n > \frac{\left[a+2md\right]\left[a+\left(2m-1\right)d\right]...\left[a+\left(m+1\right)d\right]}{a\left[a+d\right]\left[a+2d\right]...\left[a+\left(m-1\right)d\right]} \cdot \frac{a+md}{G}$$

$$> \frac{\left[a+2md\right]...\left[a+\left(m+1\right)d\right]}{a\left[a+d\right]\left[a+2d\right]...\left[a+\left(m-1\right)d\right]} = \left[Gd\left(A\right)\right]^n$$

So for linear data

$$Gd\left(G\right) > Gd\left(A\right)$$

Numerical Illustration

Application on real data (some examples) :

(1). During the last decade, the annual exports of India [5] are given in the following table

Year	Exports (million US$)
1999-2000	36,822.49
2000-2001	44,560.29
2001-2002	43,826.73
2002-2003	52,719.43
2003-2004	63,842.55
2004-2005	83,535.94
2005-2006	1,03,090.54
2006-2007	1,26,262.68

The total exports during eight years of last decade were 554,660.65 million US$ and the annual average imports taken to be arithmetic mean was comes out 69332.58 million US$, while the geometric mean of exports comes out to be 63464.53 million US$. (Calculations can be done by using MS-Excel formulae *=AVERAGE()* and *=GEOMEAN()*.

Now the geometric deviation method can be applied on this data to check whether the data shows a proportional trend or linear trend.

Arithmetic mean A = 69332.58

Geometric mean G = 63464.53

Now table for calculating geometric deviation is shown below :

x	about A	about G
	r_i	R_i
36,822.49	1.8829	1.7235
44,560.29	1.5559	1.4242
43,826.73	1.5820	1.4481
52,719.43	1.3151	1.2038
63,842.55	1.0860	1.0060
83,535.94	1.2049	1.3163
1,03,090.54	1.4869	1.6244
1,26,262.68	1.8211	1.9895

Therefore

$$\left[Gd\left(A\right)\right]^8 = \prod_{i=1}^{8} R_i = 21.60,$$

$$\left[Gd\left(G\right)\right]^8 = \prod_{i=1}^{8} r_i = 18.31$$

Since $Gd\left(A\right) > Gd\left(G\right)$, So it is concluded that the data is showing a proportional trend.

(2). The crude birth rate (CBR) record of India [6] since 1950 is shown below :

Year	CBR
1950-1955	43.3
1955–1960	42.1
1960–1965	40.4
1965–1970	39.2
1970–1975	37.5
1975–1980	36.3
1980–1985	34.5
1985–1990	32.5
1990–1995	30.0
1995–2000	27.2
2000–2005	25.3
2005–2010	22.9
2010–2015	20.4

Calculations :

Here arithmetic mean $A = 33.2$

Geometric mean $G = 32.3$

Now table for calculating geometric deviation is shown below :

x	about A	about G
	r_i	R_i
43.3	1.3042	1.3385
42.1	1.2681	1.3014
40.4	1.2169	1.2488
39.2	1.1807	1.2117
37.5	1.1295	1.1592
36.3	1.0934	1.1221
34.5	1.0392	1.0665
32.5	1.0215	1.0046
30.0	1.1067	1.0783
27.2	1.2206	1.1893
25.3	1.3123	1.2787
22.9	1.4498	1.4127
20.4	1.6275	1.5858

Therefore

$$\left[Gd\left(A\right)\right]^8 = \prod_{i=1}^{8} R_i = 13.03,$$

$$\left[Gd\left(G\right)\right]^8 = \prod_{i=1}^{8} r_i = 13.50$$

Here Gd(A)<Gd(G) implies that data shows a linear trend.

Conclusion :

So we can use geometric deviation as a measure of dispersion in case if data shows a proportion trend. We can also check the data trend with the help of geometric deviation by calculating it about arithmetic mean and geometric mean and then comparing.

References:

1. Bland, J.M.; Altman, D.G. (1996). "Statistics notes: measurement error". BMJ. 312 (7047)

2. Dodge, Yadolah (2003). The Oxford Dictionary of Statistical Terms. Oxford University Press.

3. H.Mulholland, Jones C.R. (1968). Fundamentals of Statistics, Springer Science+Business Media New York

4. Joe Kennedy Adams (1955). Basic Statistical Concepts, The Maple Press Company, York, PA.

5. MANORAMA yearbook 2008

6. United Nations, Department of Economic and Social Affairs website, Population Division > World Population Prospects: The 2015 Revision