
Estimating Average Variation About the Population Mean Using Geometric Measure of Variation

Troon John Benedict¹, Karanjah Anthony², Alilah Anekeya David³

¹Department of Mathematics and Physical Science, Maasai Mara University, Narok, Kenya

²Department of Mathematic, Multimedia University, Nairobi, Kenya

³Department of Mathematics and Statistics, Masinde Muliro University of Science and Technology, Kakamega, Kenya

Email address:

troonbenedict@gmail.com (T. J. Benedict), karanjah@mmarau.ac.ke (K. Anthony), aliladavid2010@gmail.com (A. A. David)

To cite this article:

Troon John Benedict, Karanjah Anthony, Alilah Anekeya David. Estimating Average Variation About the Population Mean Using Geometric Measure of Variation. *International Journal of Statistical Distributions and Applications*. Vol. 6, No. 2, 2020, pp. 23-35.

doi: 10.11648/j.ijstd.20200602.11

Received: September 17, 2019; **Accepted:** October 16, 2019; **Published:** August 25, 2020

Abstract: Measures of dispersion are important statistical tool used to illustrate the distribution of datasets. These measures have allowed researchers to define the distribution of various datasets especially the measures of dispersion from the mean. Researchers and mathematicians have been able to develop measures of dispersion from the mean such as mean deviation, variance and standard deviation. However, these measures have been determined not to be perfect, for example, variance give average of squared deviation which differ in unit of measurement as the initial dataset, mean deviation gives bigger average deviation than the actual average deviation because it violates the algebraic laws governing absolute numbers, while standard deviation is affected by outliers and skewed datasets. As a result, there was a need to develop a more efficient measure of variation from the mean that would overcome these weaknesses. The aim of the paper was to estimate the average variation about the population mean using geometric measure of variation. The study was able to use the geometric measure of variation to estimate the average variation about the population mean for un-weighted datasets, weighted datasets, probability mass and probability density functions with finite intervals, however, the function faces serious integration problems when estimating the average deviation for probability density functions as a result of complexity in the integrations by parts involved and also integration on infinite intervals. Despite the challenge on probability density functions, the study was able to establish that the geometric measure of variation was able to overcome the challenges faced by the existing measures of variation about the population mean.

Keywords: Standard Deviation, Geometric Measure of Variation, Deviation About the Mean, Average, Mean, Absolute Deviation, Estimation

1. Introduction

Geometric measure of variation about the mean, is a new measure of variation about the mean, that intends to solve the weaknesses of the current measures of variation about the mean, by using geometric averaging, to estimate the average deviation about the mean. Geometric averaging is a suitable measure of variation because based on past research it is not affected by outliers and skewed datasets. The technique also average product of numbers and not sum, a factor which makes it not to violate the algebraic laws [4, 19, 21].

Currently, there are three known measures of variation about the mean; Mean deviation which is the average

absolute deviation about the mean, variance which is the average of squared deviations about the mean and standard deviation which is the square-root of average squared deviation about the mean [17]. Past studies have established that the existing measures of variation about the mean are not 100% efficient, this is due to various issue that arises during their use in estimating the average variation about the mean [2, 3, 11-13, 16, 18, 20]. For example, mean deviation has been determined by past studies has to violate the algebraic number theory. Based on the algebraic number theory, given that the mean deviation is an average of absolute deviations, an absolute number on the field P such that $|\bullet|: P \rightarrow \Re^{>0}$ must satisfy

the following conditions [5];

1. $|a| = 0$ iff $a = 0 \forall a \in p$
2. $|ab| = |a||b| \forall a, b \in p$
3. $|a+b| < |a|+|b| \forall a, b \in p$

The mean deviation about the mean is given by the function [17];

$$MAD = \frac{\sum_{i=1}^n |v_i - \bar{v}|}{n} \quad (1)$$

where $|v_i - \bar{v}|$ is the absolute deviation from the mean

This measure of deviation from the mean as an absolute number, violates the algebraic laws as illustrated by the third condition (3), hence the average deviation about the mean estimated by the measure are not accurate because;

$$\frac{\sum_{i=1}^n |v_i - \bar{v}|}{n} > \frac{\sum_{i=1}^n (v_i - \bar{v})}{n}$$

Therefore, the measure always gives bigger estimates than the actual deviation about the mean. However, the measure argues on the basis of the theory behind measurement of average deviation about the mean, which assumes that for measuring of deviation from the mean, the metric (the distance from the mean) is more important than the sign of the deviation. Mean deviation has also been determined by past studies not to allow further algebraic application because of the absolution, as a result the measure is not considered as efficient [6].

A second measure of variation about the mean is variance which is given by the function [1, 17];

$$\text{var} = \frac{\sum_{i=1}^n (v_i - \bar{v})^2}{n} \quad (2)$$

Where $(v_i - \bar{v})^2$ is the squared deviation from the mean

This measure of variation from the mean allows for further algebraic manipulations which is an improvement from the mean deviation, it also do not violate the algebraic number theory, however, the average of deviation about the mean given by the measure are squared, hence, are not of the same unit as the initial datasets (squared). This makes the results given by the formula to be inappropriate [1, 17].

The last measure of variation about the mean is standard deviation, which is an improvement on variance by giving results which are of the same units as initial datasets. The measure is usually estimated by [1, 3, 13, 16];

$$SD = \sqrt{\frac{\sum_{i=1}^n (v_i - \bar{v})^2}{n}} \quad (3)$$

Over the years, standard deviation has been the most widely used measure of variation about the mean, because

it is a capable of further algebraic manipulation and it also solves the problem of variance by giving estimates which are of the same unit as the original datasets (square-root). However, past studies have determined that standard deviation is affected by outliers and skewed datasets, factors which makes this measure not to be efficient especially when dealing with datasets which have outliers and those that are skewed [2-3, 11-13, 16, 18, 20].

Given the shortcomings of the three existing measures of variation about the mean. The paper aimed at a estimate the average variation about the population mean using the geometric measure of variation, which is would overcome the weakness of the current measures of variation about the mean by not violating the algebraic laws, giving estimates which are of the same unit as the initial datasets, not affected by outliers and skewed datasets, and allows further algebraic manipulations to be carried on it. This is because, geometric measure of variation would use geometric averaging, an averaging technique that average products hence not violating the algebraic laws behind absolute number. And also, an averaging technique that has been determined not to be affected by outliers and skewed datasets.

2. Method

2.1. Un-Weighted Dataset Geometric Measure of Dispersion

Consider a vectors V of data points v_i such that all the points are not weighted or coefficient by any weights. A geometric measure of deviation from the mean G for un-weighted datasets will be given by the function;

$$G = \begin{cases} \sqrt[n]{\prod_{i=1}^p |v_i - \bar{v}|} & \forall v_i \neq \bar{v} \\ 0 & \forall v_i = \bar{v} \end{cases} \quad (4)$$

where n is the total number of data points in the dataset V , \bar{v} is the mean for the dataset

p is the total number of data points which are not equal to the mean

$|v_i - \bar{v}|$ is the absolute deviation from the mean.

If all the data points $v_i = \bar{v}$ then the deviation from the mean will be zero, hence the mean deviation from the mean for such data points is hence zero, therefore, for data set with all points similar to each other the geometric deviation from the mean will be zero. For data set with at least p points not equal to the mean ($v_i \neq \bar{v}$), the geometric mean deviation from the mean will be given by the formula;

$$G = \sqrt[n]{\prod_{i=1}^p |v_i - \bar{v}|} \quad (5)$$

To make the formula applicable for large data sets, a simplification of the formula can be carried out using the using the logarithmic transformation as below;

$$G = \left(\prod_{i=1}^p |v_i - \bar{v}| \right)^{\frac{1}{n}}$$

Now let

$$v_i - \bar{v} = d_i \tag{6}$$

Then the function (5) can be written as;

$$G = \left(\prod_{i=1}^p |d_i| \right)^{\frac{1}{n}} \tag{7}$$

Now introducing natural logarithm on both sides we obtain the following function

$$\ln(G) = \ln \left(\prod_{i=1}^p |d_i| \right)^{\frac{1}{n}} \tag{8}$$

Equation (8) can be simplified as follows based on the logarithm rules;

$$\ln(G) = \frac{1}{n} \ln \left(\prod_{i=1}^p |d_i| \right)$$

Transforming back to get the original function by introducing exponential on both sides we obtain the following equation

$$G = \begin{cases} \exp\left(\frac{1}{n} \ln \left(\prod_{i=1}^p |d_i| \right)\right) & \forall d_i \neq 0 \\ 0 & \forall d_i = 0 \end{cases} \tag{9}$$

The function (9) will assist in reducing the problem that result from geometric rooting. The absolute value in the formula based on the algebraic theory helps in preventing us from having complex results when finding roots for negative numbers. The geometric mean is appropriate because it satisfies the algebraic formula [2];

$$\prod_{i=1}^n |d_i| = \left| \prod_{i=1}^n d_i \right| \tag{10}$$

Which is an improvement based on accuracy from the arithmetic averaging of absolute values [2];

$$\frac{\sum_{i=1}^n |d_i|}{n} > \frac{\left| \sum_{i=1}^n d_i \right|}{n}$$

Hence the geometric averaging gives more accurate results than arithmetic averaging. As a simplification and improvement, equation (9) can be simplified further by replacing the product with sum because of the logarithmic law which states that log of product is same as sum of

individual logs hence;

$$\ln(G) = \frac{1}{n} \sum_{i=1}^p \ln |d_i| \tag{11}$$

Based on the transformation in (11) we can now rewrite (9) as follows;

$$G = \begin{cases} \exp\left(\frac{1}{n} \sum_{i=1}^p \ln |d_i|\right) & \forall d_i \neq 0 \\ 0 & \forall d_i = 0 \end{cases} \tag{12}$$

2.2. Weighted Dataset Geometric Measure of Dispersion

Consider a data vector V such that $V = v_1, v_2, v_3, \dots, v_n$ and a weighted vector ζ of the same dimension as V , such that $\zeta = [\zeta_1, \zeta_2, \zeta_3, \dots, \zeta_n]$. The joint distribution of the two vectors results into a weighted vector of datasets $V\zeta = [\zeta_1 v_1, \zeta_2 v_2, \zeta_3 v_3, \dots, \zeta_n v_n]$. A geometric measure of deviation from the mean G for weighted datasets will be given by the function;

$$G = \begin{cases} \sqrt[\sum_{i=1}^n \zeta_i]{\prod_{i=1}^p |v_i - \bar{v}|^{\zeta_i}} & \forall v_i \neq \bar{v} \\ 0 & \forall v_i = \bar{v} \end{cases} \tag{13}$$

where ζ_i are the weights for every data point v_i .

\bar{v} is the mean for the dataset

P is the total number of data points which are not equal to the mean

$|v_i - \bar{v}|$ is the absolute deviation from the mean.

If all the data points $v_i = \bar{v}$ then the deviation from the mean will be zero, hence the mean deviation from the mean for such data points is hence zero, therefore, for data set with all points similar to each other the geometric deviation from the mean will be zero. For data set with at least p points not equal to the mean ($v_i \neq \bar{v}$), the geometric mean deviation from the mean will be given by the formula;

$$G = \sqrt[\sum_{i=1}^n \zeta_i]{\prod_{i=1}^p |v_i - \bar{v}|^{\zeta_i}} \tag{14}$$

To make the formula applicable for large data sets, a simplification of the formula can be carried out using the using the logarithmic transformation as below;

$$G = \left(\prod_{i=1}^p |v_i - \bar{v}|^{\zeta_i} \right)^{\frac{1}{\sum_{i=1}^n \zeta_i}}$$

Using (6) the function (14) can be written as;

$$G = \left(\prod_{i=1}^p |d_i|^{\zeta_i} \right)^{\frac{1}{\sum_{i=1}^n \zeta_i}} \quad (15)$$

Now introducing natural logarithm on both sides we obtain the following function

$$\ln(G) = \ln \left(\prod_{i=1}^p |d_i|^{\zeta_i} \right)^{\frac{1}{\sum_{i=1}^n \zeta_i}} \quad (16)$$

Equation (16) can be simplified as follows based on the logarithm rules;

$$\ln(G) = \frac{1}{\sum_{i=1}^n \zeta_i} \ln \left(\prod_{i=1}^p |d_i|^{\zeta_i} \right)$$

Transforming back to get the original function by introducing exponential on both sides we obtain the following equation

$$G = \begin{cases} \exp \left(\frac{1}{\sum_{i=1}^n \zeta_i} \ln \left(\prod_{i=1}^p |d_i|^{\zeta_i} \right) \right) & \forall d_i \neq 0 \\ 0 & \forall d_i = 0 \end{cases} \quad (17)$$

The function (17) will assist in reducing the problem that result from geometric rooting. As a simplification and improvement, equation (17) can also be simplified further by replacing the product with sum based on logarithmic laws;

$$\ln(G) = \frac{1}{\sum_{i=1}^n \zeta_i} \sum_{i=1}^p \zeta_i \ln |d_i| \quad (18)$$

Based on the transformation in 18 we can now rewrite 17 as follows;

$$G = \begin{cases} \exp \left(\frac{1}{\sum_{i=1}^n \zeta_i} \sum_{i=1}^p \zeta_i \ln |d_i| \right) & \forall d_i \neq 0 \\ 0 & \forall d_i = 0 \end{cases} \quad (19)$$

2.3. Application on Probability Mass Functions

Based on equation (18) it can be determined that

$$G = \begin{cases} \exp \left(\frac{1}{10} \sum_{i=1}^p \ln |d_i| \right) & \forall d_i \neq 0 \\ 0 & \forall d_i = 0 \end{cases}$$

$\ln(G) = E(\ln |d_i|)$ and that $\ln(d_i)$ is distributed with the same weights as v_i . Therefore, extending this relationship on probability mass functions. Assume that the variable v_i is discrete with probability mass function $\zeta(v_i)$ for all $i = 1, 2, 3, \dots, n$ and 0 otherwise. Assume that $\ln |d_i|$ which is equal to $\ln |v_i - \bar{v}|$ where \bar{v} is the mean of the random variable v , is distributed in the same way as v_i with a probability mass function $\zeta(v_i)$. The geometric deviation for probability mass functions can be given as;

$$G = \begin{cases} \exp \left(\sum_{i=1}^n \zeta(v_i) \cdot \ln |d_i| \right) & \forall d_i \neq 0 \\ 0 & \forall d_i = 0 \end{cases} \quad (20)$$

2.4. Application on Probability Density Functions

Extending the relationship in equation (18) on continuous random variables. Assume that the variable v is continuous on the interval $a \leq v \leq b$ with probability density function $\zeta(v)$. Assume that $\ln |d|$ which is equal to $\ln |v - \bar{v}|$ where \bar{v} is the mean of the random variable v , is distributed in the same way as v_i with a probability density function $\zeta(v)$. The geometric deviation for probability density functions can be given as;

$$G = \begin{cases} \exp \left(\int_a^b \zeta(v) \cdot \ln |d| \cdot dd \right) & \forall d \neq 0 \\ 0 & \forall d = 0 \end{cases} \quad (21)$$

3. Results

3.1. Application on Un-weighted Datasets

Several simulations were conducted on both discrete and continuous data distributions of small populations of size 10 from Bernoulli, Binomial, Geometric, Normal, chi-square and F-distributed were. The results were as shown below;

3.1.1. Bernoulli Population

Consider a population of size 10 which is Bernoulli distributed with a probability of success 0.7, the geometric measure of variation from the mean for the population estimated as illustrated in table 1;

The geometric measure of variation from the mean for un-weighted dataset is given by the function;

Table 1. Estimation for un-weighted Bernoulli population

Population Number	Bernoulli Population	$d_i = x_i - 0.8$	$ d_i $	$\ln(d_i)$
1	1	0.2	0.2	-1.609437912
2	1	0.2	0.2	-1.609437912
3	1	0.2	0.2	-1.609437912
4	1	0.2	0.2	-1.609437912
5	0	-0.8	0.8	-0.223143551
6	1	0.2	0.2	-1.609437912
7	1	0.2	0.2	-1.609437912
8	1	0.2	0.2	-1.609437912
9	0	-0.8	0.8	-0.223143551
10	1	0.2	0.2	-1.609437912
Mean	0.8			-1.33217904

Geometric measure = $\exp(-1.33217904) = 0.263901582 \approx 0.2639$

3.1.2. Binomial Population

Consider a population of size 10 which is Binomial distributed with 30 trials and a probability of success 0.64, the geometric measure of variation from the mean was estimated as shown in table 2.

Table 2. Estimation for un-weighted Binomial population.

Population Number	Binomial Population	$d_i = x_i - 18.2$	$ d_i $	$\ln(d_i)$
1	18	-0.2	0.2	-1.609437912
2	17	-1.2	1.2	0.182321557
3	20	1.8	1.8	0.587786665
4	18	-0.2	0.2	-1.609437912
5	15	-3.2	3.2	1.16315081
6	16	-2.2	2.2	0.78845736
7	23	4.8	4.8	1.568615918
8	17	-1.2	1.2	0.182321557
9	20	1.8	1.8	0.587786665
10	18	-0.2	0.2	-1.609437912
Mean	18.2			0.023212679

Geometric measure = $\exp(0.023212679) = 1.02348419 \approx 1.0235$

3.1.3. Geometric Population

Consider a population of size 10 which is Geometric distributed with a probability of success 0.5, the average deviation from the mean can be estimated using geometric measure of deviation from the mean for the population as illustrated in table 3;

Table 3. Estimation for un-weighted Geometric population.

Population Number	Geometric Population	$d_i = x_i - 1.9$	$ d_i $	$\ln(d_i)$
1	1	-0.9	0.9	-0.105360516
2	1	-0.9	0.9	-0.105360516
3	2	0.1	0.1	-2.302585093
4	3	1.1	1.1	0.09531018
5	1	-0.9	0.9	-0.105360516
6	1	-0.9	0.9	-0.105360516
7	1	-0.9	0.9	-0.105360516
8	3	1.1	1.1	0.09531018
9	1	-0.9	0.9	-0.105360516
10	5	3.1	3.1	1.131402111
Mean	1.9			-0.161272572

Geometric measure = $\exp(-0.161272572) = 0.851060065 \approx 0.8511$

3.1.4. Chi-square Population

Population of size 10 from Chi-square distribution with 1 degree of freedom was simulated as shown in table 4. The calculation of geometric measure for the population was estimated as illustrated in table 4.

Table 4. Estimation for un-weighted Chi-square population.

Population Number	Chi-square Population	$d_i = x_i - 1.580066995$	$ d_i $	$\ln(d_i)$
1	3.217024949	1.636957954	1.636957954	0.492839614
2	2.995690441	1.415623446	1.415623446	0.347570032
3	3.662891652	2.082824657	2.082824657	0.733724981
4	1.481728301	-0.098338694	0.098338694	-2.319337698
5	2.378023987	0.797956993	0.797956993	-0.225700577
6	0.189166369	-1.390900625	1.390900625	0.329951469
7	0.093799897	-1.486267098	1.486267098	0.396267673
8	1.601235263	0.021168268	0.021168268	-3.855251998
9	0.180945899	-1.399121096	1.399121096	0.335844251
10	0.000163189	-1.579903805	1.579903805	0.457363962
Mean	1.580066995			-0.330672829

$$\text{Geometric measure} = \exp(-0.330672829) = 0.718440183 \approx 0.7184$$

3.1.5. Normal Distribution Population

A population of size 10 from a Normal distribution with a mean of 10 and a standard deviation of 2 was simulated as illustrated in table 5. The estimation of average deviation from the mean using geometric measure of variation about the mean for the population was estimated as illustrated in table 5.

Table 5. Estimation for un-weighted Normal population.

Population Number	Normal Population	$d_i = x_i - 9.560952744$	$ d_i $	$\ln(d_i)$
1	9.4732888	-0.087663944	0.087663944	-2.434244594
2	10.0179588	0.457006053	0.457006053	-0.783058643
3	9.833743354	0.272790611	0.272790611	-1.299050771
4	9.843723505	0.282770761	0.282770761	-1.263118741
5	7.86179902	-1.699153723	1.699153723	0.530130317
6	10.54505528	0.98410254	0.98410254	-0.01602518
7	8.855622899	-0.705329844	0.705329844	-0.349089721
8	9.473922934	-0.087029809	0.087029809	-2.441504582
9	13.65741087	4.096458131	4.096458131	1.41012273
10	6.047001969	-3.513950775	3.513950775	1.256740981
Mean	9.560952744			-0.53890982

$$\text{Geometric measure} = \exp(-0.53890982) = 0.583383899 \approx 0.5834$$

3.1.6. F-distributed Population

A population of size 10 from a F-distribution with 2 numerator and 5 denominator degrees of freedom was simulated as illustrated in table 6. The calculation of geometric measure of variation about the mean for the population was calculated as illustrated in table 6.

Table 6. Estimation for un-weighted Fisher's population.

Population Number	F-distribution Population	$d_i = x_i - 2.273034666$	$ d_i $	$\ln(d_i)$
1	7.650678068	5.377643402	5.377643402	1.682250249
2	2.329789864	0.056755198	0.056755198	-2.869008029
3	5.536239586	3.26320492	3.26320492	1.182709817
4	1.738601369	-0.534433297	0.534433297	-0.626548351
5	0.771963586	-1.50107108	1.50107108	0.406178907
6	0.555808899	-1.717225767	1.717225767	0.540710062
7	1.502359971	-0.770674695	0.770674695	-0.260488921
8	0.31742318	-1.955611486	1.955611486	0.670702925
9	0.975575768	-1.297458898	1.297458898	0.260407658
10	1.351906368	-0.921128298	0.921128298	-0.082155949
Mean	2.273034666			0.090475837

$$\text{Geometric measure} = \exp(0.090475837) = 1.094695056 \approx 1.0947$$

Based on the illustration shown by the calculations, the geometric measure can be used to estimate the average

variation from the mean for both un-weighted discrete and continuous datasets.

3.2. Application on Weighted Datasets

The study intended to determine if the geometric measure could be used in the estimation of average deviations from the mean for weighted datasets with specific concentration on frequency distributions, where the frequencies were used as the weights of the respective data points. Simulated data for three discrete distributions (Bernoulli, Binomial and Geometric distributions) and three continuous distributions (Normal, Chi-square and Fishers/F-distributions) were used. 100 observations were simulated for each distribution after which the

data were summarized into frequency distributions before geometric measure for the average deviation about the mean was estimated for each frequency distributions. The results of the simulation and the estimation was as illustrated below;

3.2.1. Bernoulli Population

Consider a population of size 100 which is Bernoulli distributed with a probability of success 0.7, the frequency distribution and the estimation of the geometric measure variation from the mean for the population was as illustrated below;

The geometric measure of variation from the mean (G) for weighted datasets is given by the function;

$$G = \begin{cases} \exp \left(\frac{1}{\sum_{i=1}^n \zeta_i} \sum_{i=1}^n \zeta_i \ln(|d_i|) \right) & \forall d_i \neq 0 \\ 0 & \forall d_i = 0 \end{cases}$$

Table 7. Estimation for weighted Bernoulli population.

Number (v_i)	Frequency (ζ_i)	$\zeta_i v_i$	$d_i = v_i - \bar{v}$	$ v_i - \bar{v} $	$\ln d_i = \ln v_i - \bar{v} $	$\zeta_i \ln d_i = \zeta_i v_i - \bar{v} $
0	34	0	-0.66	0.66	-0.415515444	-14.12752509
1	66	66	0.34	0.34	-1.078809661	-71.20143765
Total	100	66				-85.32896275

$$Mean(\bar{v}) = \frac{\sum_{i=1}^n \zeta_i v_i}{\sum_{i=1}^n \zeta_i} = \frac{66}{100} = 0.66$$

$$\ln(G) = \frac{\sum_{i=1}^n \zeta_i \ln|d_i|}{\sum_{i=1}^n \zeta_i} = \frac{-85.32896275}{100} = -0.8532896275$$

Geometric measure = $\exp(-0.8532896275) = 0.426011206 \approx 0.4260$

3.2.2. Binomial Population

Consider a population of size 100 which is Binomial distributed with 30 trials and a probability of success 0.64, the frequency distribution and the estimation of the geometric measure variation from the mean for the population was as illustrated in table 8.

Table 8. Estimation for weighted Binomial population.

Number (v_i)	Frequency (ζ_i)	$\zeta_i v_i$	$d_i = v_i - \bar{v}$	$ v_i - \bar{v} $	$\ln d_i = \ln v_i - \bar{v} $	$\zeta_i \ln d_i = \zeta_i v_i - \bar{v} $
11	2	22.0	-8.02	8.02	2.081938422	4.163876844
12	1	12.0	-7.02	7.02	1.948763218	1.948763218
13	1	13.0	-6.02	6.02	1.795087259	1.795087259
14	3	42.0	-5.02	5.02	1.613429934	4.840289801
15	4	60.0	-4.02	4.02	1.391281903	5.565127611
16	5	80.0	-3.02	3.02	1.105256831	5.526284157
17	13	221.0	-2.02	2.02	0.703097511	9.140267648
18	7	126.0	-1.02	1.02	0.019802627	0.138618391
19	21	399.0	-0.02	0.02	-3.912023005	-82.15248311
20	9	180.0	0.98	0.98	-0.020202707	-0.181824366
21	16	336.0	1.98	1.98	0.683096845	10.92954952
22	8	176.0	2.98	2.98	1.091923301	8.735386404
23	7	161.0	3.98	3.98	1.381281819	9.668972735

Number (v_i)	Frequency (ζ_i)	$\zeta_i v_i$	$d_i = v_i - \bar{v}$	$ v_i - \bar{v} $	$\ln d_i = \ln v_i - \bar{v} $	$\zeta_i \ln d_i = \zeta_i v_i - \bar{v} $
24	1	24.0	4.98	4.98	1.605429891	1.605429891
25	2	50.0	5.98	5.98	1.788420568	3.576841136
Total	100	1902.0				-14.69981287

$$Mean(\bar{v}) = \frac{\sum_{i=1}^n \zeta_i v_i}{\sum_{i=1}^n \zeta_i} = \frac{1902}{100} = 19.02$$

$$\ln(G) = \frac{\sum_{i=1}^n \zeta_i \ln|d_i|}{\sum_{i=1}^n \zeta_i} = \frac{-14.69981287}{100} = -0.1469981287$$

Geometric measure = $\exp(-0.1469981287) = 0.863295593 \approx 0.8633$

3.2.3. Geometric Population

Consider a population of size 100 which is Geometric distributed with a probability of success 0.5, the frequency distribution and the estimation of the geometric measure variation from the mean for the population was as illustrated in table 9.

Table 9. Estimation for weighted Geometric population.

Number (v_i)	Frequency (ζ_i)	$\zeta_i v_i$	$d_i = v_i - \bar{v}$	$ v_i - \bar{v} $	$\ln d_i = \ln v_i - \bar{v} $	$\zeta_i \ln d_i = \zeta_i v_i - \bar{v} $
1	52	52	-1.12	1.12	0.113328685	5.893091636
2	22	44	-0.12	0.12	-2.120263536	-46.6457978
3	14	42	0.88	0.88	-0.127833372	-1.789667201
4	4	16	1.88	1.88	0.631271777	2.525087107
5	2	10	2.88	2.88	1.057790294	2.115580588
6	2	12	3.88	3.88	1.355835154	2.711670307
7	1	7	4.88	4.88	1.58514522	1.58514522
8	1	8	5.88	5.88	1.771556762	1.771556762
9	1	9	6.88	6.88	1.928618652	1.928618652
12	1	12	9.88	9.88	2.290512512	2.290512512
Total	100	212.0				-27.61420221

$$Mean(\bar{v}) = \frac{\sum_{i=1}^n \zeta_i v_i}{\sum_{i=1}^n \zeta_i} = \frac{212}{100} = 2.12$$

$$\ln(G) = \frac{\sum_{i=1}^n \zeta_i \ln|d_i|}{\sum_{i=1}^n \zeta_i} = \frac{-27.61420221}{100} = -0.2761420221$$

Geometric measure = $\exp(-0.2761420221) = 0.75870517 \approx 0.7587$

3.2.4. Chi-square Population

Population of size 100 from Chi-square distribution with 1 degree of freedom was simulated, the frequency distribution and the estimation of the geometric measure variation from the mean for the population was as illustrated in table 10.

Table 10. Estimation for weighted Chi-square population.

Number (v)	Midpoint (v)	Frequency (ζ)	ζv	$ d = v - \bar{v} $	$\ln d = \ln v - \bar{v} $	$\zeta \ln d = \zeta v - \bar{v} $
0 – 2	1	82	82.0	0.5	-0.69314718	-56.83806881
2 – 4	3	12	36.0	1.5	0.40546511	4.865581297
4 – 6	5	5	25.0	3.5	1.25276297	6.263814842
6 – 8	7	1	7.0	5.5	1.70474809	1.704748092
Total		100	150.0			-44.00392457

$$Mean(\bar{v}) = \frac{\sum \zeta^v}{\sum \zeta} = \frac{150}{100} = 1.5$$

$$\ln(G) = \frac{\sum \zeta \ln|d|}{\sum \zeta} = \frac{-44.00392457}{100} = -0.4400392457$$

Geometric measure = $\exp(-0.4400392457) = 0.644011146 \approx 0.6440$

3.2.5. Normal Distribution Population

A population of size 100 from a Normal distribution with a mean of 10 and a standard deviation of 2 was simulated. The frequency distribution and the estimation of the geometric measure variation from the mean for the population was as illustrated in table 11.

Table 11. Estimation for weighted Normal Population.

Class	Midpoint (v)	Frequency (ζ)	ζ ^v	d = v - \bar{v}	d = v - \bar{v}	ln d = ln v - \bar{v}	ζ ln d = ζ ln v - \bar{v}
4 - 6	5	1	5.0	-5.3	5.3	1.6677	1.667706821
6 - 8	7	10	70.0	-3.3	3.3	1.1939	11.93922468
8 - 10	9	38	342.0	-1.3	1.3	0.2624	9.96984205
10 - 12	11	26	286.0	0.7	0.7	-0.3567	-9.273548542
12 - 14	13	24	312.0	2.7	2.7	0.9933	23.83804255
14 - 16	15	1	15.0	4.7	4.7	1.5476	1.547562509
Total		100	1030.0				39.68883007

$$Mean(\bar{v}) = \frac{\sum \zeta^v}{\sum \zeta} = \frac{1030}{100} = 10.3$$

$$\ln(G) = \frac{\sum \zeta \ln|d|}{\sum \zeta} = \frac{39.68883007}{100} = 0.3968883007$$

Geometric measure = $\exp(0.3968883007) = 1.487189803 \approx 1.4872$

3.2.6. F-distributed Population

A population of size 100 from a F-distribution with 2 numerator and 5 denominator degrees of freedom was simulated as illustrated in table 13. The calculation of geometric measure of variation about the mean for the population was calculated as illustrated in table 12.

Table 12. Estimation for weighted Fisher's population.

Class	Midpoint (v)	Frequency (ζ)	ζ ^v	d = v - \bar{v}	ln d = ln v - \bar{v}	ζ ln d = ζ ln v - \bar{v}
0 - 2	1	69	69.0	1.055	0.05354077	3.694312918
2 - 4	3	18	54.0	0.945	-0.05657035	-1.018266327
4 - 6	5	7	35.0	2.945	1.08010882	7.56076172
6 - 8	7	5	35.0	4.945	1.59837697	7.991884825
8 - 10	9	0	0.0	6.945	1.93802198	0
10 - 15	12.5	1	12.5	10.445	2.34612340	2.346123395
Total		100	205.5			20.57481653

$$Mean(\bar{v}) = \frac{\sum \zeta^v}{\sum \zeta} = \frac{205.5}{100} = 2.055$$

$$\ln(G) = \frac{\sum \zeta \ln|d|}{\sum \zeta} = \frac{20.57481653}{100} = 0.2057481653$$

Geometric measure = $\exp(0.2057481653) = 1.2284438 \approx 1.2284$

Based on the illustration shown by the calculations, the geometric measure can be used to estimate the average variation from the mean for both weighted discrete and continuous datasets.

3.3. Application on Probability Mass Functions

The study considered various probability mass functions to illustrate how the function can be used in estimating the average variation from the mean for probability mass functions. The results were as illustrated below;

$$G = \begin{cases} \exp\left(\sum_{i=1}^n \zeta(v_i) \cdot \ln|d_i|\right) & \forall d_i \neq 0 \\ 0 & \forall d_i = 0 \end{cases}$$

The calculation of the estimate was as follows;

Table 13. Estimation for coin tossing gambling game.

Gain (v_i)	Probability $\zeta(v_i)$	$v_i \zeta(v_i)$	d_i	$ d_i $	$\ln d_i $	$\zeta(v_i) \ln d_i $
-1	0.5	-0.5	-1.5	1.5	0.405465108	0.202732554
2	0.5	1	1.5	1.5	0.405465108	0.202732554
Total	1	0.5				0.405465108

$$G = \exp(0.405465108) = 1.5$$

3.3.2. Dice

When an individual roll a die, he can get one of the numbers ranging from 1 to 6, estimate the variation of obtaining a number when a die is rolled.

Table 14. Estimation for rolling a dice.

Number (v_i)	Probability $\zeta(v_i)$	$v_i \zeta(v_i)$	d_i	$ d_i $	$\ln d_i $	$\zeta(v_i) \ln d_i $
1	0.166666667	0.166666667	-2.5	2.5	0.916290732	0.152715122
2	0.166666667	0.333333333	-1.5	1.5	0.405465108	0.067577518
3	0.166666667	0.5	-0.5	0.5	-0.693147181	-0.11552453
4	0.166666667	0.666666667	0.5	0.5	-0.693147181	-0.11552453
5	0.166666667	0.833333333	1.5	1.5	0.405465108	0.067577518
6	0.166666667	1	2.5	2.5	0.916290732	0.152715122
Total		3.5				0.20953622

$$G = \exp(0.20953622) = 1.233106037 \approx 1.2331$$

3.3.3. Bernoulli Distribution

Consider a Bernoulli distribution with probability mass function;

$$\zeta(V) = \begin{cases} \zeta^{v_i} (1-\zeta)^{1-v_i} & V = 0,1 \\ 0 & \text{Otherwise} \end{cases} \quad (22)$$

Where ζ is the probability of success, the natural logarithm of the geometric measure of variation about the mean can be estimated for the function by;

$$\ln(G) = \sum_{i=0}^1 \zeta^{v_i} (1-\zeta)^{1-v_i} \ln|v_i - \zeta| \quad (23)$$

Hence,

3.3.1. Coin Tossing Gambling Game

In a coin tossing gambling game, a player loses 1 shilling whenever a fair coin tosses head and gains 2 shillings whenever the fair coin tosses tail. Estimate the average variation of the gain using geometric measure of variation from the mean.

The geometric variation for the probability mass function is estimated using the functions;

$$G = \begin{cases} \exp\left(\sum_{i=0}^1 \zeta^{v_i} (1-\zeta)^{1-v_i} \ln|v_i - \zeta|\right) & \forall v_i \neq \zeta \\ 0 & \forall v_i = \zeta \end{cases} \quad (24)$$

3.3.4. Binomial Distribution

Consider a binomial distribution with probability mass function

$$\zeta(V) = \begin{cases} \binom{t}{v_i} \zeta^{v_i} (1-\zeta)^{t-v_i} & V = 0,1,2,\dots,t \\ 0 & \text{Otherwise} \end{cases} \quad (25)$$

Where ζ is the probability of success, the natural logarithm of the geometric measure of variation about the mean can be estimated for the function by;

$$\ln(G) = \sum_{i=0}^t \binom{t}{v_i} \zeta^{v_i} (1-\zeta)^{t-v_i} \ln|v_i - t\zeta| \quad (26)$$

Hence,

$$G = \begin{cases} \exp\left(\sum_{i=0}^t \binom{t}{v_i} \zeta^{v_i} (1-\zeta)^{1-v_i} \ln|v_i - t\zeta|\right) & \forall v_i \neq t\zeta \\ 0 & \forall v_i = t\zeta \end{cases} \quad (27)$$

3.3.5. Geometric Distribution

Consider a geometric distribution with probability mass function

$$\zeta(V) = \begin{cases} \zeta(1-\zeta)^{V-1} & V = 0, 1, 2, \dots \\ 0 & \text{Otherwise} \end{cases} \quad (28)$$

Where ζ is the probability of success, the natural logarithm of the geometric measure of variation about the mean can be estimated for the function by;

$$\ln(G) = \sum_{i=0}^{\infty} \zeta(1-\zeta)^{v_i} \ln\left|v_i - \frac{1-\zeta}{\zeta}\right| \quad (29)$$

$$\ln(G) = \sum_{i=0}^{\infty} \zeta(1-\zeta)^{v_i} \ln\left|\frac{\zeta v_i - 1 + \zeta}{\zeta}\right| \quad (30)$$

$$\ln(G) = \sum_{i=0}^{\infty} \zeta(1-\zeta)^{v_i} \ln\left|\frac{\zeta v_i + \zeta - 1}{\zeta}\right| \quad (31)$$

$$\ln(G) = \sum_{i=0}^{\infty} \zeta(1-\zeta)^{v_i} \ln\left|\frac{\zeta(v_i + 1) - 1}{\zeta}\right| \quad (32)$$

Hence,

$$G = \begin{cases} \exp\left(\sum_{i=0}^{\infty} \zeta(1-\zeta)^{v_i} \ln\left|\frac{\zeta(v_i + 1) - 1}{\zeta}\right|\right) & \forall v_i \neq \frac{1-\zeta}{\zeta} \\ 0 & \forall v_i = \frac{1-\zeta}{\zeta} \end{cases} \quad (33)$$

3.3.6. Poisson Distribution

Consider a Poisson distribution with probability mass function;

$$\zeta(V) = \begin{cases} \frac{e^{-\zeta} \zeta^{v_i}}{v_i!} & V = 0, 1, 2, 3, \dots \\ 0 & \text{Otherwise} \end{cases} \quad (34)$$

Where ζ is the average rate of event occurrence, the natural logarithm of the geometric measure of variation about the mean can be estimated for the function by;

$$\ln(G) = \sum_{i=0}^{\infty} \frac{e^{-\zeta} \zeta^{v_i}}{v_i!} \ln|v_i - \zeta| \quad (35)$$

Hence,

$$G = \begin{cases} \exp\left(\sum_{i=0}^{\infty} \frac{e^{-\zeta} \zeta^{v_i}}{v_i!} \ln|v_i - \zeta|\right) & \forall v_i \neq \zeta \\ 0 & \forall v_i = \zeta \end{cases} \quad (36)$$

The above illustrations show how the geometric measure about the mean can be used to estimate the average variations about the mean of probability mass functions.

3.4. Application on Probability Density Functions

The study considered various probability density functions to illustrate how the function can be used in estimating the average variation from the mean for probability density functions. The results were as illustrated below;

3.4.1. Simple Probability Density Function

Consider a random v variable which is distributed in the interval $[0, 1]$ with a probability density function;

$$\zeta(v) = \begin{cases} 1 & 0 \leq v \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

We can estimate the average deviation from the mean for v using the geometric measure as follows;

The geometric average deviation from the mean is given by the function;

$$G = \begin{cases} \exp\left(\int_a^b \zeta(v) \bullet \ln|v - \bar{v}| \bullet dv\right) & \forall v \neq \bar{v} \\ 0 & \forall v = \bar{v} \end{cases}$$

Therefore, for the example pdf, we can obtain the expected value of the distribution as follows;

$$E(V) = \int_0^1 v \bullet dv = \frac{1}{2} v^2 + c \Big|_0^1 = \left[\frac{1}{2} - 0\right] = \frac{1}{2}$$

The natural logarithm of the geometric measure of variation from the mean by definition is given by the formula;

$$\ln(G) = E(\ln|v - \bar{v}|) = \int_a^b \zeta(v) \bullet \ln|v - \bar{v}| \bullet dv$$

Hence, for the above probability density function;

$$\begin{aligned} \ln(G) &= E\left(\ln\left|v - \frac{1}{2}\right|\right) = \int_0^1 1 \bullet \ln|v - \bar{v}| \bullet dv \\ &= \int_0^1 \ln\left|v - \frac{1}{2}\right| \bullet dv = \int_0^1 \left[\left(v - \frac{1}{2}\right) \ln\left|v - \frac{1}{2}\right| - v + \frac{1}{2}\right] \bullet dv \\ &= \int_0^1 \left[\frac{1}{2} \ln\left(\frac{1}{2}\right) - 1 + \frac{1}{2}\right] - \int_0^1 \left[-\frac{1}{2} \ln\left(\frac{1}{2}\right) + \frac{1}{2}\right] \bullet dv \\ &= [-0.84657] - [0.84657] \\ &= -1.69315 \end{aligned}$$

Therefore, the geometric measure of variation from the mean will be given by;

$$G = \exp(-1.69315) = 0.18394$$

3.4.2. Uniform Distribution

Consider a random variable v which is distributed uniformly in the interval $[2 \ 10]$. By definition, the probability density function of v will be given by;

$$\zeta(v) = \begin{cases} \frac{1}{10-2} & 2 \leq v \leq 10 \\ 0 & \text{otherwise} \end{cases}$$

The expectation of the function is given by;

$$E(v) = \frac{10+2}{2} = 6$$

Therefore, the natural logarithm of geometric measure of variation from the mean for the distribution will be given by;

$$\begin{aligned} \ln(G) &= E(\ln|v-6|) = \int_2^{10} \frac{1}{10-2} \cdot \ln|v-6| \cdot dv \\ &= \frac{1}{8} \int_2^{10} \ln|v-6| \cdot dv = \frac{1}{8} [(v-6)\ln|v-6| - v + 6]_2^{10} \\ &= \frac{1}{8} [(4\ln(4) - 10 + 6) - (-4\ln(4) - 2 + 6)] \\ &= \frac{1}{8} [1.54518 + 1.54518] = 0.386294 \end{aligned}$$

Therefore, the geometric measure of variation from the mean will be given by;

$$G = \exp(0.386294) = 1.471518 \approx 1.4715$$

3.4.3. Exponential Probability Density Function

Consider an exponential distribution with probability density function;

$$\zeta(v) = \begin{cases} \zeta e^{-\zeta v} & v \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

By definition,

$$E(v) = \frac{1}{\zeta}$$

Therefore, the natural logarithm of geometric measure about the mean is given by;

$$\ln(G) = E\left(\ln\left|v - \frac{1}{\zeta}\right|\right) = \int_0^{\infty} \zeta e^{-\zeta v} \cdot \ln\left|v - \frac{1}{\zeta}\right| \cdot dv \tag{37}$$

Hence,

$$G = \begin{cases} \exp\left(\int_0^{\infty} \zeta e^{-\zeta v} \cdot \ln\left|v - \frac{1}{\zeta}\right| \cdot dv\right) & \forall v \neq \frac{1}{\zeta} \\ 0 & \forall v = \frac{1}{\zeta} \end{cases} \tag{38}$$

The above illustrations show that the geometric measure of

variation from the mean can be used to estimate the average deviation from the mean, however, the function faces a challenge in the estimation especially when the limits are infinite. Also, the challenge arises when conducting partial integration for complex probability density functions which do not have finite integrals for the integration by parts. For example, consider the case of standard normal probability density function;

3.4.4. Standard Normal

Consider a standard normal distribution for a random variable v illustrated as follows;

$$\zeta(v) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} & -\infty \leq v \leq \infty \\ 0 & \text{otherwise} \end{cases}$$

Therefore, the natural logarithm of geometric measure about the mean is given by;

$$\ln(G) = E(\ln|v|) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{v^2}{2}} \cdot \ln|v| \cdot dv \tag{39}$$

Hence,

$$G = \begin{cases} \exp\left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{v^2}{2}} \cdot \ln|v| \cdot dv\right) & \forall v \neq 0 \\ 0 & \forall v = 0 \end{cases} \tag{40}$$

This shows that it is almost impossible to integrate such functions. This is a short coming of the geometric measure of variation especially during its application on probability density functions

4. Conclusion

In conclusion, geometric measure of variation about the mean can be used to estimate the average deviation from the population mean for un-weighted datasets, weighted datasets, probability mass and probability density functions with finite intervals, however, the function faces serious integration problems when estimating the average deviation for probability density functions as a result of complexity in the integrations by parts involved and also integration on infinite intervals. The geometric measure of variation was also determined to give smaller estimates of the variation from the mean compared to standard deviation.

References

[1] Ahn, S., & Fessler, J. A., (2003). Standard Errors of Mean, Variance, and Standard Deviation Estimators. EECS Department. The University of Michigan. U.S.A.
 [2] Altman, D. G., & Bland, J. M. (2005). Standard deviations and standard errors. *BMJ* Volume 331.

- [3] Bhardwaj, A., (2013). Comparative Study of Various Measures of Dispersion. *Journal of Advances in Mathematics*. Vol 1, No 1.
- [4] Buckland, S. T., A. C. Studeny, A. E. Magurran, J. B. Illian, and S. E. Newson. (2011). The geometric mean of relative abundance indices: a biodiversity measure with a difference. *Ecosphere* 2 (9): 100. doi: 10.1890/ES11-00186.1.
- [5] Clark, P. L., (2012). *Number Theory: A Contemporary Introduction*. Available at <http://math.uga.edu/~pete/4400FULL.pdf>.
- [6] Deshpande, S., Gogtay, N. J., Thatte, U. M., (2016). Measures of Central Tendency and Dispersion. *Journal of the Association of Physicians of India*. Vol. 64. July 2016.
- [7] Grechuk, B., Molyboha, A., & Zabaranin M., (2011). Mean-Deviation Analysis in The Theory of Choice. *Risk Analysis*.
- [8] Hu, S., (2010). Simple Mean, Weighted Mean, or Geometric Mean?. Presented at the 2010 ISPA/SCEA Joint Annual Conference and Training Workshop.
- [9] Kum, S., & Lim, Y., (2012). A Geometric Mean of Parameterized Arithmetic and Harmonic Means of Convex Functions. *Hindawi Publishing Corporation*. Volume 2012, Article ID 836804.
- [10] Lawson, J. D., & Lim, Y., (2001). *The Geometric Mean, Matrices, Metrics, and More*. *The American Mathematical Monthly*.
- [11] Lee, D., In, J., & Lee, S., (2015). Standard deviation and standard error of the mean. *Korean journal of anesthesiology*. 68. 220-3. 10.4097/kjae.2015.68.3.220.
- [12] Leys, C., Klein, O., Bernard, P., & Licata, L., (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology* 49 (2013) 764–766.
- [13] Manikandan, S., (2016). Measures of dispersion. *Journal of Pharmacology and Pharmacotherapeutics*. October-December 2011. Vol 2. Issue 4.
- [14] McAlister, D., (1879). *The Law of Geometric Mean*. The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Proceedings of the Royal Society of London*.
- [15] Mindlin, D., (2011). *On the Relationship between Arithmetic and Geometric Returns*. Cdi Advisors Research. LLC.
- [16] Mohini, P. B., & Prajakt, J. B., (2012). What to use to express the variability of data: Standard deviation or standard error of mean?. *Perspectives in clinical research* July 2012.
- [17] Raymondo, J., (2015). *Measures of Variation from Statistical Analysis in the Behavioral Sciences*. Kendall Hunt Publishing.
- [18] Roberson, Q. M., Sturman, M. C., & Simons, T. L., (2007). Does the Measure of Dispersion Matter in Multilevel Research? A Comparison of the Relative Performance of Dispersion Indexes. *Cornell University School of Hotel Administration*. The Scholarly Commons.
- [19] Roenfeldt, K., (2018). *Better than average: Calculating Geometric Means Using SAS*. Henry. M. Foundation for the Advancement of Military Medicine.
- [20] Schuetter, J. (2007). Chapter 1. In J. Schuetter, *measures of dispersion* (pp. 45-54).
- [21] Thenwall, M. (2018). The precision of the arithmetic mean, geometric mean and percentiles for citation data: An experimental simulation modelling approach. *Statistical Cybermetrics Research Group, School of Mathematics and Computer Science, University of Wolverhampton, Wulfruna Street, Wolverhampton, UK*.