DETERMINING THE UNBIASED ESTIMATOR OF THE POPULATION GEOMETRIC MEASURES OF VARIATION ABOUT THE MEAN

Troon John Benedict[1], Obere Almadi[2], Karanjah Anthony[3], Alilah Anekeya David[4]

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

2Department of Economics, Maasai Mara University, Narok, Kenya

3Department of Mathematics, Multimedia University, Nairobi, Kenya

4Department of Mathematics and Statistics, Masinde Muliro University of Science and Technology, Kakamega, Kenya

Email address:

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

Abstract

Geometric Measures of Variation about the mean is a measure that uses the geometric averaging technique to average the deviations from the mean. From previous studies, it has been determined that the measure is more precise in estimating the average variation about the mean than the existing measures of variation about the mean. Given that the technique is a newly introduced technique of estimating the average variation about the mean, the actual sample estimator for the measure is still unknown, as a result, the study aimed at determining the unbiased estimator for the population geometric measure. The study used a mathematical estimation technique to determine the unbiased estimator among the existing possible estimators as it assumed a simple random sampling without replacement technique. The study determined that the unbiased estimator of the population estimator was the sample estimator which did not allow one degree of freedom.

Key Words: Estimator, Parameter, Unbiasedness, Sampling

Introduction

Geometric Measures of Variation about the mean is a measure that uses the geometric averaging technique to average the deviations from the mean [15, 16]. Unlike other measures of variation about the mean such as mean deviation, variance, and standard deviation [1,2,5]. The measure has been determined to be more efficient in estimating the average variation about the mean than the existing measures because of various reasons such as; unlike mean deviation, the measure average absolute products and not sums [9,11,13], which makes the estimates more precise based on the algebraic number theory on absolute numbers which shows that;

Given $a \neq 0, b \neq 0$ and $a, b \in \Re$ by the definition of an absolute number which is a transformation such that $|\bullet| : \Re \to +\Re$ (7).

1. $|ab| = |a||b|$
2. $|a+b| \leq |a| + |b|$

This shows that the averaging of deviation products by the geometric measures are more precise as illustrated by condition (1), however, the estimates given by mean deviation are not precise because there is a slight difference between the actual values and the estimates as illustrated by the triangular inequality in condition (2) [8,12].

The geometric measure gives estimates of average variation about the mean of same units of measurement as the original dataset, unlike, variance which gives squared units of measurement for the estimates of average variation about the mean [1,12]). Lastly, the measure uses geometric averaging which is nonresponsive to outliers and skewed datasets, as a result, the measure is not affected by outliers and skewed datasets, unlike standard deviation. All these shows that geometric measure is a more superior measure of variation about the mean [1,2, 12].

In most research work, due to various factors such as cost, time constrain, impossibility in accessing all respondents in the study area among others, most researchers always opt to carry out sample surveys as opposed to complete enumeration of all respondents in the population, as a result, most researchers dependents on the sample estimates to make inferential conclusions regarding the population. Therefore, a researcher is always required to pick an estimator that would precisely estimate the population parameter during the estimation process [3,8].

As newly formulated measures of variation about the mean, the unbiased sample estimator of the population parameter still unknown. As a result, the aim of this study was to determine the unbiased estimator of the population geometric measure of variation, which will assist in the precise estimation of the measure for various samples [15, 16].

Methods

Possible estimators

In determining the unbiased estimator of the population geometric measure of variation about the mean, the study compared two possible estimators of the population geometric measures mathematically [8].

Consider a population vector of observations $V = [v_1, v_2, v_3, ..., v_N]$. We can define the mean of the population vector $\bar{V}$ as

$$\bar{V} = \frac{\sum_{i=1}^{N} v_i}{N} \quad (1)$$

Further, define the ith population deviation D_i as

$$D_i = v_i - \bar{V} \quad (2)$$

Therefore, the population deviation vector will be given by $D = [D_1, D_2, D_3, ..., D_N]$. Now assume that all the observations in the population $v_i \neq \bar{V}$, therefore, all $D_i \neq 0$. Hence, by definition, the geometric measure of variation about the mean G for the population will be given by [15, 16]

$$G = \begin{cases} \sqrt[N]{\prod_{i=1}^{N} |D_i|} & \forall D_i \neq 0 \\ 0 & \forall D_i = 0 \end{cases} \quad (3)$$

Using natural logarithms, Equation (3) can be simplified as

$$G = \begin{cases} \exp\left(\frac{\sum_{i=1}^{N} \ln|D_i|}{N} \right) & \forall D_i \neq 0 \\ 0 & \forall D_i = 0 \end{cases} \quad (4)$$

Consider a random sample vector of size n from the above population vector and a dummy weight k_i such that [3]

$$k_i = \begin{cases} 1 \\ 0 \end{cases} \quad (5)$$

where $k_i = 1$ when the coefficient observation in the population v_i is selected into the sample and $k_i = 0$ when the coefficient observation in the population v_i is not selected into the sample. Based on this, the sample mean $\bar{v}$ is defined by

$$\bar{v} = \frac{\sum_{i=1}^{N} k_i v_i}{n} = \frac{1}{n}\left(\sum_{i=1}^{N} k_i v_i \right) \quad (6)$$

Based on the above sample, we can estimate the population geometric measure of variation using two methods. First, we can average the sample average variation about the mean using geometric average as follows [15, 16].

Define the ith deviation from the sample mean $\bar{v}$ as

$$d_i = v_i - \bar{v} \quad (7)$$

The sample geometric measure of variation about the mean g_r will be given by

$$g_r = \begin{cases} \sqrt[n]{\prod_{i=1}^{N} |d_i|^{k_i}} & \forall d_i \neq 0 \\ 0 & \forall d_i = 0 \end{cases} \quad (8)$$

The above estimate can be simplified using logarithm as

$$g_r = \begin{cases} \exp\left(\dfrac{\sum_{i=1}^{N} k_i \ln|d_i|}{n}\right) & \forall d_i \neq 0 \\ 0 & \forall d_i = 0 \end{cases} \tag{9}$$

Secondly, we can estimate the population geometric measure of variation about the mean by allowing a 1 degree of freedom in the sample estimation process, this will result into a new measure of variation about the mean g_q which is given by [15, 16]

$$g_q = \begin{cases} \sqrt[n-1]{\prod_{i=1}^{N} |d_i|^{k_i}} & \forall d_i \neq 0 \\ 0 & \forall d_i = 0 \end{cases} \tag{10}$$

Equation [10] can be simplified using logarithm to obtain

$$g_q = \begin{cases} \exp\left(\dfrac{\sum_{i=1}^{N} k_i \ln|d_i|}{n-1}\right) & \forall d_i \neq 0 \\ 0 & \forall d_i = 0 \end{cases} \tag{11}$$

The above two methods are all possible estimators of the population geometric measure. As a result, the study aims at proving which of the two estimators is an unbiased estimator of the population geometric measure.

Assuming a homogenous population distribution, the study assumes a simple random sampling technique without replacement during the estimation process of the population parameter using either of the two methods [14, 17].

Results

Theorem

Given the population geometric measure G, an estimator g of the parameter is said to be unbiased by definition if [4].

$$E(g) = G$$

Based on this theorem, if any of the two estimators g_r or g_q is an unbiased estimator of the population parameter G then by definition of an unbiased estimator

$$E(g_r) = G$$

or

$E\left(g_q\right)=G$

Proof

We can determine which one of these two estimators is the true unbiased estimator of the parameter G as follows;

Starting with g_r considering a random sample of size n, from a population $V=\left[V_1,V_2,V_3,\ldots,V_N\right]$. The sample vector v selected at random using simple random sampling without replacement from the population is given by $v=\left\{k_iV_i \middle| i=1,2,3,4,\ldots N\right\}$ where k_i is a dummy weight such that [3, 14, 17]

$$k_i=\begin{cases}1\\0\end{cases}$$

where $k_i=1$ when the coefficient observation in the population V_i is selected into the sample and $k_i=0$ when the coefficient observation in the population V_i is not selected into the sample. The sample mean is given by [3, 6]

$$\overline{v}=\frac{\sum_{i=1}^{N}k_iV_i}{n}$$

The i[th] deviation from the mean is defined by

$$d_i=V_i-\overline{v}$$

The geometric measure of variation about the mean g_r is therefore given by

$$g_r=\begin{cases}\sqrt[n]{\prod_{i=1}^{N}\left|d_i\right|^{k_i}} & \forall d_i\neq 0\\ 0 & \forall d_i=0\end{cases}$$

This can be simplified as

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

Given that the population geometric measure G is defined by

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

Then if g_r is an unbiased estimator of G then

$$E(g_r) = G = \begin{cases} \exp\left(\frac{\sum_{i=1}^{N} k_i \ln|D_i|}{n}\right) & \forall d_i \neq 0 \\ 0 & \forall d_i = 0 \end{cases} \quad (12)$$

Now given the unique nature of exponential such that

$$\exp(K) = \exp(R), iff K = R \quad (13)$$

Then if $\ln(g_r)$ is an unbiased estimator of $\ln(G)$, then g_k is an unbiased estimator of G. Now, if $\ln(g_r)$ is an unbiased estimator of $\ln(G)$ then

$$E(\ln(g_r)) = \ln(G) \quad (14)$$

By definition

$$\ln(g_r) = \frac{\sum_{i=1}^{N} k_i \ln|d|_i}{n} = \frac{1}{n}\left(\sum_{i=1}^{N} k_i \ln|d_i|\right) \quad (15)$$

Then

$$E(\ln(g_r)) = E\left(\frac{1}{n}\sum_{i=1}^{N} k_i \ln|d_i|\right) = \frac{1}{n}\sum_{i=1}^{N} E(k_i \ln|d_i|) \quad (16)$$

It can be shown that

$$E(\ln|d_i|) = \ln|D_i| \quad (17)$$

This is because

$$E(\ln|d_i|) = \ln(E|d_i|) = \ln(E|V_i - \bar{v}|) = \ln(|E(V_i) - E(\bar{v})|)$$

Because the sample mean $\bar{v}$ is an unbiased estimator of the population mean, therefore

$= \ln|V_i - \bar{V}| = \ln|D_i|$ therefore,

$$E\left(\ln\left(g_r\right)\right) = E\left(\frac{1}{n}\sum_{i=1}^{N} k_i \ln\left|d_i\right|\right) = \frac{1}{n}\sum_{i=1}^{N} E\left(k_i \ln\left|d_i\right|\right) = \frac{1}{n}\sum_{i=1}^{N} \ln\left|D_i\right| E\left(k_i\right) \tag{18}$$

Now, given that k_i is a binary dummy variable, it can be seen that k_i is Bernoulli distributed with probability mass function

$$P\left(k_i\right) = \begin{cases} p^{k_i}\left(1-p\right)^{1-k_i} & k_i = 0,1 \\ 0 & otherwise \end{cases} \tag{19}$$

Therefore, the expected value of k_i by definition is p. we can determine the actual value of p through obtaining the maximum likelihood estimator of p from the probability mass function as follows [4];

Let L be the likelihood function of the probability mass function of k_i which is given as

$$L = \prod_{i=1}^{N} p^{k_i}\left(1-p\right)^{1-k_i} = p^{\sum_{i=1}^{N} k_i}\left(1-p\right)^{N-\sum_{i=1}^{N} k_i} \tag{20}$$

Now we get the log-likelihood function, which is given by

$$\begin{aligned} \ln L &= \ln p^{\sum_{i=1}^{N} k_i}\left(1-p\right)^{N-\sum_{i=1}^{N} k_i} = \sum_{i=1}^{N} k_i \ln p + N - \sum_{i=1}^{N} k_i \ln\left(1-p\right) \\ \ln L &= \sum_{i=1}^{N} k_i \ln p + N\ln\left(1-p\right) - \sum_{i=1}^{N} k_i \ln\left(1-p\right) \end{aligned} \tag{21}$$

Now, differentiating Equation (21) with respect to p in order to maximize the estimator we get that

$$\begin{aligned} \frac{\partial \ln L}{\partial p} &= \sum_{i=1}^{N} k_i\left(\frac{1}{p}\right) + N\left(\frac{-1}{1-p}\right) - \sum_{i=1}^{N} k_i\left(\frac{-1}{1-p}\right) \\ \frac{\partial \ln L}{\partial p} &= \frac{\sum_{i=1}^{N} k_i}{p} - \frac{N}{1-p} + \frac{\sum_{i=1}^{N} k_i}{1-p} = \frac{\sum_{i=1}^{N} k_i\left(1-p\right) - Np + \sum_{i=1}^{N} k_i\left(p\right)}{p-p^2} \\ \frac{\partial \ln L}{\partial p} &= \frac{\sum_{i=1}^{N} k_i - \sum_{i=1}^{N} k_i\left(p\right) - Np + \sum_{i=1}^{N} k_i\left(p\right)}{p-p^2} \\ \frac{\partial \ln L}{\partial p} &= \frac{\sum_{i=1}^{N} k_i - Np}{p-p^2} \end{aligned} \tag{22}$$

Equating the above Equation (22) to zero we have that

$$\frac{\partial \ln L}{\partial p} = \frac{\sum_{i=1}^{N} k_i - Np}{p - p^2} = 0$$

$$\therefore \sum_{i=1}^{N} k_i - Np = 0$$

$$\therefore Np = \sum_{i=1}^{N} k_i$$

$$p = \frac{\sum_{i=1}^{N} k_i}{N}$$

But there are n k_i terms which are equivalent to 1 with the rest being equivalent to 0, therefore, it can be shown that

$$\sum_{i=1}^{N} k_i = n \tag{23}$$

Hence, the maximum likelihood estimator of $p = \frac{n}{N}$. Therefore

$$E(k_i) = \frac{n}{N} \tag{24}$$

Hence, going back to Equation (18) we can now show that

$$E(\ln(g_r)) = \frac{1}{n}\sum_{i=1}^{N} \ln|D_i| E(k_i) = \frac{1}{n}\sum_{i=1}^{N} \ln|D_i| \frac{n}{N} = \frac{1}{n} \bullet \frac{n}{N}\sum_{i=1}^{N} \ln|D_i| = \frac{1}{N}\sum_{i=1}^{N} \ln|D_i| = \ln(G)$$

Now because

$$E(\ln(g_r)) = \ln(G)$$

Introducing exponential both sides

$$\exp(E(\ln(g_r))) = \exp(\ln(G))$$

$$E(\exp(\ln(g_r))) = G \tag{25}$$

$$E(g_r) = G$$

Hence, the estimator g_r if an unbiased estimator of the population parameter G.

Second, Consider the estimator g_q which is given by the function

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

Similar to g_r, if g_q is an unbiased estimator of G then

$$E(g_q) = G$$

$$E(g_q) = G = \begin{cases} \exp\left(\frac{\sum_{i=1}^{N} k_i \ln|D_i|}{n-1}\right) & \forall d_i \neq 0 \\ 0 & \forall d_i = 0 \end{cases} \tag{26}$$

Given the unique character of exponentiation illustrated in Equation (13), then if Equation (26) is true, then

$$E(\ln(g_q)) = \ln(G)$$

$$E(\ln(g_q)) = E\left(\frac{1}{n-1}\sum_{i=1}^{N} k_i \ln|d_i|\right) = \frac{1}{n}\sum_{i=1}^{N} E(k_i \ln|d_i|)$$

But according to Equation (18)

$$\ln|V_i - \bar{V}| = \ln|D_i|$$

And according to Equation (24)

$$E(k_i) = \frac{n}{N}$$

Therefore,

$$E(\ln(g_q)) = E\left(\frac{1}{n-1}\sum_{i=1}^{N} k_i \ln|d_i|\right) = \frac{1}{n-1}\sum_{i=1}^{N} \ln|D_i| \frac{n}{N} = \frac{n}{Nn-N}\sum_{i=1}^{N} \ln|D_i| \neq \frac{1}{N}\sum_{i=1}^{N} \ln|D_i|$$

$$E(\ln(g_q)) \neq \ln(G)$$

$$\therefore \exp(E(\ln(g_q))) \neq \exp(\ln(G))$$

$$\therefore E(\exp(\ln(g_q))) \neq \exp(\ln(G))$$

$$\therefore E(g_q) \neq G$$

Therefore, the estimator g_q is a biased estimator of the population geometric measure G.

Conclusion

In conclusion, considering a simple random sample without replacement, the unbiased estimator of the population geometric measure is given by the estimator which does not allow one degree of freedom during the estimation process of the population parameter. The unbiased estimator is given by

$$g_r = \begin{cases} \exp\left(\dfrac{\sum_{i=1}^{N} k_i \ln|d_i|}{n} \right) & \forall d_i \neq 0 \\ 0 & \forall d_i = 0 \end{cases}$$

Recommendation

When estimating the average variation about the mean for a sample using the geometric measure, one should use the above estimator g_r because it is unbiased and more precise in estimating the average variation about the mean for a sample.

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]. Banning, R, Camstra, A. and Knottnerus, P. (2012). Sampling theory: Sampling design and estimation methods. Statistics Netherlands. Henri Faasdreef 312. 2492 JP The Hague.

[4]. Bensoussan, A., Bertrand, P. and Brouste, A. (2015). Future Perspectives in Risk Models and Finance. International Series in Operations Research & Management Science 211. Springer International Publishing Switzerland DOI 10.1007/978-3-319-07524-2_1

[5]. Bhardwaj, A., (2013). Comparative Study of Various Measures of Dispersion. Journal of Advances in Mathematics. Vol 1, No 1.

[6]. Camstra, A. and Nieuwenbroek, N.J. (2002), Syllabus for the sampling theory course. CBS-Academie, Voorburg

[7]. Clark, P. L., (2012). Number Theory: A Contemporary Introduction. Available at http://math.uga.edu/~pete/4400FULL.pdf

[8]. 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.

[9]. Hu, S., (2010). Simple Mean, Weighted Mean, or Geometric Mean?. Presented at the 2010 ISPA/SCEA Joint Annual Conference and Training Workshop

[10]. Manikandan, S., (2016). Measures of dispersion. Journal of Pharmacology and Pharmacotherapeutics. October-December 2011. Vol 2. Issue 4

[11]. Mindlin, D., (2011). On the Relationship between Arithmetic and Geometric Returns. Cdi Advisors Research. LLC.

[12]. 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.
[13]. Roenfeldt, K., (2018). Better than average: Calculating Geometric Means Using SAS. Henry. M. Foundation for the Advancement of Military Medicine.
[14]. Singh, A. and Masuku, M. (2014). Sampling Techniques & Determination of Sample Size in Applied Statistics Research: An Overview. International Journal of Economics, Commerce and Management. Vol. II, Issue 11.
[15]. Troon, J. B., Karanjah, A., and Alilah, A. D.(2019). Modeling Geometric Measure of Variation About the Population Mean. *American Journal of Theoretical and Applied Statistics*. Vol. 8, No. 5, pp. 179-184. doi: 10.11648/j.ajtas.20190805.13
[16]. Troon, J. B., Karanjah, A., and Alilah, A. D.(2019). Empirical Comparison of Relative Precision of Geometric Measure of Variation about the Mean and Standard Deviation. International Journal of Recent Engineering Research and Development (IJRERD). Volume 04 – Issue 10. pp. 01-12.
[17]. Yates, F. and Grundy, P.M. (1953), Selection without replacement from within strata with probability proportional to size. Journal of the Royal Statistical Society B, 15, 253-261.