Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.3, No.7, 2013
www.iiste.org

On Hypothesis Testing Under Unequal Group Variances: The Use of the Harmonic Variance

[1]Abidoye, A. O, [2]Jolayemi, E. T ,[3] Sanni, O. O. M and [4]Oyejola, B. A
Department of Statistics, University of Ilorin, Ilorin, Kwara State, Nigeria.
[1] abidoye@unilorin.edu.ng, [2]tejujola@unilorin.edu.ng ,[3] omsanni@unilorin.edu.ng, [4] boyejola@unilorin.edu.ng

Abstract
The assumptions of equality of group variances required for ANOVA often fail for real life data. However, because the major aim is to test equality of group means, a single summary value for these group variances is a necessity. Previous works in literature have zeroed in on the choice of either the Harmonic or Geometric mean as a proper mean especially when extreme observation(s) are present which renders a simple mean inappropriate.
In this work, the asymptotic sample distribution of harmonic mean of group variances is established to be a chi-square distribution though the degrees of freedom need not be an integer.
Keywords: Harmonic mean of group variances, three - parameter Beta distribution, chi – square distribution.

1. INTRODUCTION

Data, in real life situations, often contain outlier(s). Such outliers may be very important to the extent that it may often be misleading to exclude such from the data set while analyzing the data. Using the arithmetic mean in this context may inflate or deflate that summary as the case may be. Harmonic mean can be useful in such a situation or when the group means are largely unequal, since the harmonic mean is not usually unduly influenced by extreme values which characterize the simple (pooled) mean. This is more pronounced when the mean desired is that of the variances. When the grouped variances are unequal especially in hypothesis testing, the sample pooled variance is not suitable in the test statistic. The resulting F – test becomes highly conservative. As it is common in elementary statistics, the harmonic and /or the geometric means are more appropriate.
In this work, the harmonic mean of the variances is considered in hypothesis testing when equality of variances cannot be assumed. In the works such as Adegboye (1981), Adegboye and Gupta (1986) and Yahya and Jolayemi (2003) where ordered alternative of group means were considered, they settled on a single summary values of the variances. The harmonic mean of variances is preferred; because central limit theorem makes harmonic mean more appealing than the geometric mean see Abidoye (2012), Adegboye(2009) and Gupta (2004). Analytic distribution of the sample harmonic mean of variances has been shown to be generalized beta distribution when underlying distribution of normality has been assumed for the observations. Jolayemi et. al (2013).
The sampling distribution of the harmonic mean of group variances is confirmed to have the generalized beta in section two, while in section three, an approximation of generalized beta by chi – square distribution is demonstrated while section four contains a discussion and presents the conclusion.

2. METHODOLGY

Consider the beta distribution whose location and shape parameters are α and β respectively. By special choices of these parameters, the distribution may be made symmetric, skewed either way or even uniform. Such a beta random variable may be extended if the said random variable is multiplied by a constant $\lambda > 1$. In this regard, α, β and λ may be chosen such that $\alpha < \beta$ and
$\lambda > 1$ such that the distribution is skewed to the right.
The probability density function of $Y = \lambda X$ define in Abidoye et al (2007) is given as

$$f(Y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\lambda^{\alpha+\beta-1}} Y^{(\alpha-1)}(\lambda - Y)^{\beta-1}, \quad \beta > 0, \quad \alpha > 0, \quad 0 < Y < \lambda \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where X has the beta distribution. Note that if $\lambda = 1$, then Y has the regular beta distribution

$$i.e \qquad Y \sim Gb(\alpha, \beta, \lambda)$$

$$E(Y^r) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\lambda^{\alpha+\beta-1}} \int_0^{\lambda} Y^{(r+\alpha-1)}(\lambda - Y)^{\beta-1} \partial y$$

$$= \frac{\lambda \cdot B\left[(r+\alpha),\beta\right]}{B(\alpha,\beta)}$$

when r = 1

$$= \lambda\left[\frac{\Gamma(\alpha+1)\Gamma(\beta)/\Gamma(\alpha+\beta+1)}{\Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)}\right]$$

so that the mean E(Y) is given as

$$E(Y) = \frac{\alpha}{(\alpha+\beta)}\lambda \text{.......... ....}(2)$$

and variance

$$Var(Y) = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}\lambda^2 \text{.......... ....}(3)$$

by method of moments,

$$\hat{\lambda} = Y_{(n)}, \qquad \hat{\alpha} = \frac{v}{u}\bar{y}, \qquad \hat{\beta} = \frac{v}{u}(Y_{(n)} - \bar{y}) \qquad u = s_H^2 + \bar{y}(Y_{(n)} - \bar{y}) \qquad and$$

$$v = Y_{(n)}(s_H^2)\text{...........}(4)$$

where $Y_{(j)}$ is the j^{th} ordered value of Y.
If we let

$$Y = S_H^2 = \left(\frac{1}{g}\sum\frac{1}{S_i^2}\right)^{-1} \quad ,0 < S_H^2 < \lambda\text{...}(5)$$

Then Y has the generalized beta, that is ,

$$S_H^2 \approx Gb(\alpha,\beta,\lambda)$$

where (S_H^2) is the harmonic mean of the sample variances, s_i^2, $i = 1,2,\ldots,g$ and g is the number of groups. Jolayemi et. al (2013) justifies this, we therefore appeal to empirical approach.
In this regard we assume that $X_{ij} \sim N(\mu_i, \sigma_i^2)$ i= 1, 2, 3, 4 , j= 1, 2, …, $n_i(= 20)$. Note that if the distribution S_H^2 true for n_i =20 then it is equally true for all n_i> 20 by large sample theory. For each i , S_i^2 is obtained and S_H^2 evaluated. This is simulated 1000 times and α, β and λ derived using equation (4). The observed counts in a developed histogram were compared with expected counts according to equation (1). Goodness – of – fit was assessed using the Pearson X^2 as shown in Table 1 for the populations described there. Some other groups examined are shown in Table2, where P_i is the probability obtained using trapezium rule or method.
For instance the following three groups are among groups considered

1. $N_1(1,4)$, $N_2(2,5)$, $N_3(4,6)$, $N_4(8,25)$
2. $N_1(6,4)$, $N_2(6,7)$, $N_3(6,12)$, $N_4(6,37)$
3. $N_1(1,4)$, $N_2(2,7)$, $N_3(4,12)$, $N_4(6,18)$, $N_5(8,50)$

Detail of results of simulation for group 1 is as presented in Table 1
λ= 6.0018, α = 9.43, β = 5.66

Interval	Frequency	Probability (P_i)	Expected (np_i)	Chi – square
2.7105 – 3.03963	7	0.0057545	5.75545	0.27
3.03964 – 3.36877	43	0.03684226	36.84226	1.03
3.36878 – 3.6979	132	0.141297	141.1297	0.59
3.6980 – 4.02703	197	0.21541712	215.41712	1.58
4.02704 – 4.35616	227	0.23470455	234.70455	0.25
4.35617 – 4.68529	191	0.183992	183.992	0.27
4.68530 – 5.01442	125	0.1122794	112.2794	1.44
5.01443 – 5.34355	46	0.04225669	42.25669	0.33
5.34356 – 5.67269	24	0.021828546	21.828546	0.22
5.67270 – 6.0018	8	0.005799889	5.799889	0.84
Total				6.82

Table 1: Frequency distribution of S_H^2 along with their expected values

	Estimates			**Goodness**		
Groups	**Λ**	β	α	**X^2**	$\chi^2_{0.05,6}$	**P- value**
$N_1(6,4)$, $N_2(6,7)$, $N_3(6,12)$, $N_4(6,37)$	**8.05**	**6.48**	**18.96**	**5.05**	**12.59**	**>>0.05**
$N_1(1,4)$, $N_2(2,7)$, $N_3(4,12)$, $N_4(6,18)$, $N_5(8,50)$	**9.25**	**13.98**	**11.63**	**1.62**	**12.59**	**>>0.05**

Table 2: Summary of some other groups are shown in the table

Conclusion:

It is therefore concluded that the distribution of Y is the generalized beta of equation (1) as asserted by Jolayemi et . al (2013).

From the above two tables it seen that the random variable Y is a generalized beta distribution.

3. Approximation of generalized beta by chi – square distribution

Since $Y = S_H^2 \sim Gb(\alpha,\beta,\lambda)$

Then according to equation (1)

$$f(Y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\lambda^{\alpha+\beta-1}} Y^{(\alpha-1)}(\lambda - Y)^{\beta-1}; \qquad 0 < Y < \lambda,\ \beta > 0, \alpha > 0$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\lambda^{\alpha+\beta-1}} \lambda^{\beta-1} Y^{(\alpha-1)}\left(1-\frac{Y}{\lambda}\right)^{\beta-1}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\lambda^{\alpha}} Y^{(\alpha-1)}\left(1-\frac{Y}{\lambda}\right)^{\beta-1}$$

$$\doteq \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\lambda^{\alpha}} Y^{(\alpha-1)} e^{-\beta\frac{Y}{\lambda}}$$

since by a Taylor series expansion, $\left(1-\frac{Y}{\lambda}\right)^{\beta-1} \doteq e^{-\beta Y/\lambda}$, provided $\frac{Y}{\lambda} \to 0$

$$= kY^{(\alpha-1)} e^{-\beta\frac{Y}{\lambda}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6)$$

where

$$k = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\lambda^{\alpha}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7)$$

is the appropriate normalization of the chi – square distribution.

Thus equation (6) establishes that the sample harmonic mean of variances is known to be approximately distributed chi – square distribution

4. Discussion of Results

The results of simulated data described in sections 2 and 3 were plotted as shown in Fig 1 , Fig 2 and Fig 3. Clearly the chi-square distribution is a good approximation for the generalized beta distribution, though the chi – squared distribution has degrees of freedom equaling 2α where α need not be an integer.

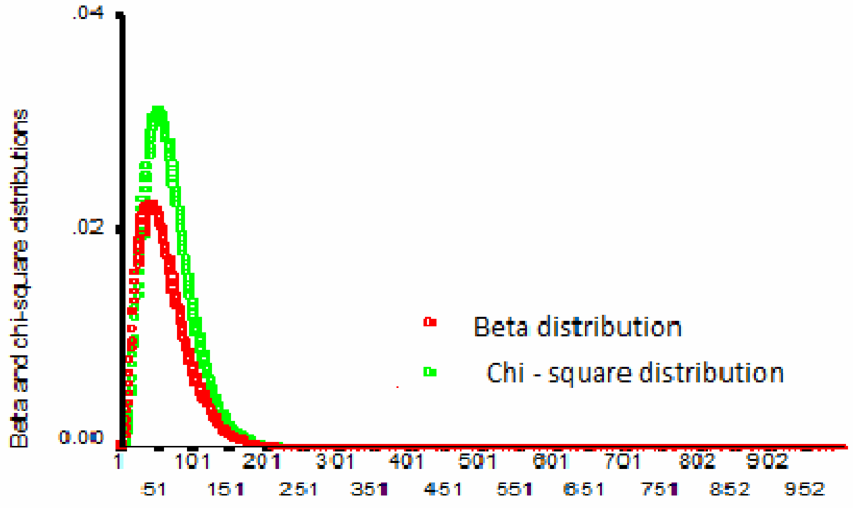

Fig. 1: Graph for Beta and approximation chi - square

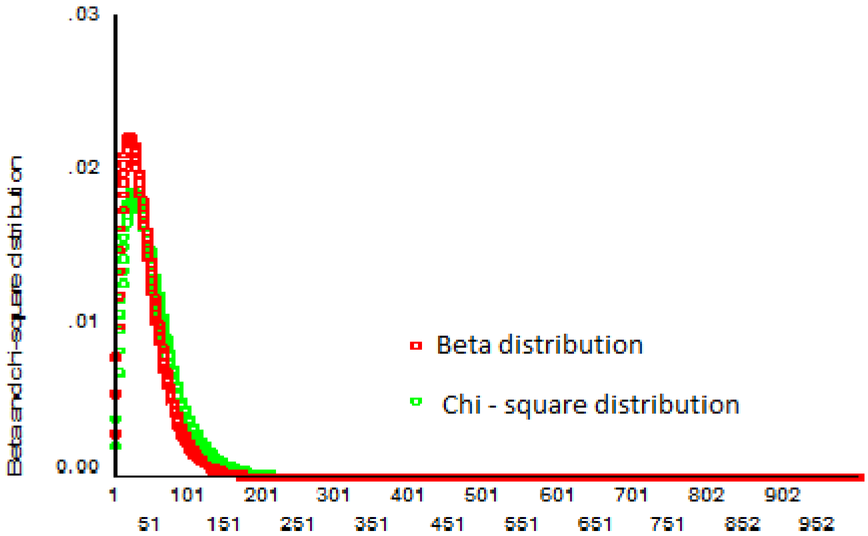

Fig. 2: Graph for Beta and approximation chi - square.

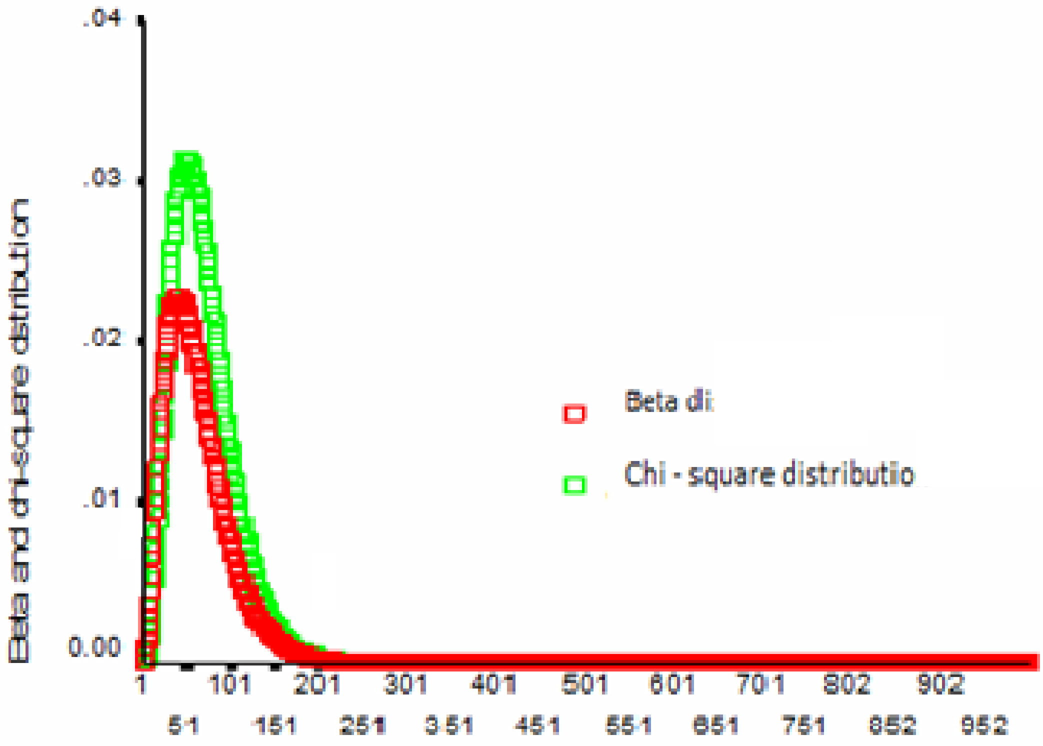

Fig 3: Graph for Beta and approximation chi - square

4.1 Conclusion

The immediate consequence of this result is that in the hypothesis test for equality of group means where equal group variance is violated, pair wise t – test or Bonferoni – type of simultaneous comparison can be achieved when the sample pooled variance is replaced by the harmonic mean of variances.

References:

(1) Abidoye, A. O (2012): Development of Hypothesis Testing Technique for Ordered Alternatives under heterogeneous variances. Unpublished Ph.D Thesis submitted to Dept. of Statistics, University of Ilorin, Ilorin.
(2) Abidoye, A. O, Jolayemi, E.T and Adegboye, O.S (2007): On the Distribution of a Scaled Beta Random Variable. JNSA, 19, 1 – 3.
(3) Adegboye, O. S (2009): Descriptive Statistics for Students, Teachers and Practitioners. First Edition. Olad Publishing.
(4) Adegboye, O.S (1981) : On Testing Against Restricted Alternative In Gaussian Models. Ph.D dissertation. BGUS
(5) Adegboye, O.S and Gupta, A.K (1986) : "On Testing Against Restricted Alternative About The Means of Gaussian Models With common Unknown Variance." Sankhya, The Indian Journal of Statistics, series B, 48, pp 331-341.
(6) Jolayemi, E.T, Oyejola, B. A, Sanni, O.O.M, Abidoye, A. O (2013): On the Distribution of Harmonic Mean of Group Sample Variances. To appear in International Journal of Statistics and Application.
(7) Gupta, S.C (2004): Fundamentals of Statistics. Sixth edition. Himalaya Publishing.
(8) Yahya, W.B and Jolayemi, E. T (2003): Testing Ordered means against a control. JNSA. 16, 40– 51.