Vol.5.Issue.1.2017 (Jan.-Mar)

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

http://www.bomsr.com
Email:editorbomsr@gmail.com

RESEARCH ARTICLE

ISSN:2348-0580

SAMPLE SIZE DETERMINATION OF 'ANOM' – TYPE GRAPHICAL METHOD FOR TESTING THE EQUALITY OF SEVERAL VARIANCES

M. PRAN KUMAR &G. V. S. R. ANJANEYULU
Department of Statistics, Acharya Nagarjuna University
GUNTUR, A.P., India
harini.pran@gmail.com,&profgvsranu@gmail.com

M. PRAN KUMAR

ABSTRACT

An expression to determine the sample size for the ANOM – type Graphical Method developed by Pran Kumar and Rao (1998) using log – transformation of Bartlett and Kendall (1946) for detecting the significant difference among k normal population variances by at least a specified amount D for fixed level of significance α and fixed power P in the case of equal sample sizes is derived. Tables of sample sizes for fixed power P = 0.8, 0.9, 0.95, 0.99, D = 1, 3, α = 0.05, 0.01 and for k = 3 (1) 20, 30, and 60 are presented.

Key words: Analysis of means, log-transformation, normality, power, sample size, variances.

1. INTRODUCTION

The problem of testing the equality of several normal population variances arises in many areas such as life sciences, physical sciences, engineering, medicine and agriculture. Several tests (like, Bartlett's test, Cochran's test, Hartley's test) are available in the literature to test the equality of several variances. These tests are non - graphical procedures and demonstrate only statistical significance of variances.

Pran Kumar and Rao (1998) developed a graphical procedure, namely ANOM – type graphical method for testing the equality of several normal population variances similar to Analysis of Means (ANOM) introduced by Ott (1967). Pran Kumar and Rao(1998) used log – transformation of Bartlett and Kendall (1946) in the derivation of decision lines. The advantage of graphical procedures over non – graphical procedures is that they demonstrate the presence of statistical significance as well as the magnitude or seriousness of statistical significance. Rao (2005) made a review of papers in the area of Analysis of Means. Section 2 presents graphical procedure developed by Pran Kumar and Rao (1998) for equal sample sizes.

Sample size determination is a vital aid in the problems relating to tests of hypotheses. By which one would know the number of units to be investigated from a population to test the hypothesis. This would allow one to control the cost, time and power of the test, that is the chance of correctly detecting the statistical significance when it actually exists.

Nelson (1983) generated tables of sample sizes for the analysis of means necessary of detecting differences among *k* treatment means differ by at least a specified amount for a fixed level of significance and fixed power. Motivated by Nelson's (1983) tables of sample sizes, in this paper an expression for sample size for the ANOM-type Graphical method developed by Prankumar and Rao (1998) is derived to detect the significant difference among *k* normal population variances by at least a specified amount *D* for fixed level of significance α and fixed power *P* in the case of equal sample sizes by using the expression (given in Chow et al, 2008, p.71) of sizes of samples drawn from normal populations and presented in Section 3. Tables of sample sizes for fixed power *P* = 0.8, 0.9, 0.95, 0.99, *D* = 1, 3 , α = 0.05, 0.01 and for *k* = 3 (1) 20, 30, and 60 are presented in Section 4.

2. ANOM-TYPE GRAPHICAL METHOD

Let $\{X_{ij}\}$, $i = 1, 2, ..., k;\ j = 1, 2, ..., n$ be *k* independent random sample drawn from *k* normal populations $N(\mu_i, \sigma_i^2)$. We wish to test the null hypothesis

$$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2 = \sigma^2 (unknown)$$

against the alternative hypothesis that at least one equality does not hold. The ANOM-type graphical method developed by Pran Kumar and Rao (1998) to test H_0 is given in the following steps and it is applicable when *n*> 10.

1. Calculate S_i^2, the sample variances for all (*i* =1, 2, …, *k*)
2. Calculate t_i' for all samples and then compute $\overline{t}'$

   where $t_i' = ln\, S_i^2 = \log_e S_i^2$ and $\overline{t}' = \frac{1}{k}\sum_{i=1}^{k} t_i'$
3. Calculate $SE(t_i' - \overline{t}') = \sqrt{\frac{2}{n-1}(1-\frac{1}{k})}$
4. The Lower Decision Line (LDL) and the Upper Decision Line (UDL) for the comparison of each of t_i' are given by

   LDL = $\overline{t}' - Z_{(1-\alpha/2k)} SE(t_i' - \overline{t}')$

   UDL = $\overline{t}' + Z_{(1-\alpha/2k)} SE(t_i' - \overline{t}')$

   Where $Z_{(1-\alpha/2k)}$ is the standard normal variable value for cumulative probability $(1-\alpha/2k)$.
5. Plot t_i' against the respective decision lines if any one the point plotted lies outside the respective decision lines, H_0 is rejected and conclude that the population variances are not equal.

3. SAMPLE SIZE DETERMINATION

Sample size is determined in such a way that the test statistic should give correct conclusion of either accepting H_0 when H_0 is true with a fixed confidence level $(1-\alpha)$ or rejecting H_0 when H_0 is false with a fixed power $P = (1-\beta)$. Using the expression (given in Chow et al., 2008, p.71) to determine the sample size where the samples are taken from normal populations, sample size expression is derived for the method developed by Pran Kumar and Rao (1998) to detect the significant difference between any one of t_i' from the grand mean $\overline{t}'$ by at least a specified difference *D* and all remaining

$t_l' = \bar{t}' - \dfrac{D}{k-1}$, $l = 1, 2, \ldots, k-1$ and $l \neq i$,

we consider

$$Z_i = Z_{(1-\alpha/2k)} + Z_p$$

where

$$\mathrm{Z_i} = \frac{t_i' - \overline{t}'}{SE\left(t_i' - \overline{t}'\right)} \tag{3.1}$$

and $Z_{(1-\alpha/2k)}$ and Z_P are the values of standard normal variate values for the cumulative probabilities of $(1-\alpha/2k)$ and power *P* respectively.

$$\frac{t_i' - \overline{t}'}{SE\left(t_i' - \overline{t}'\right)} = Z_{(1-\alpha/2k)} + Z_p$$

For one of t_i' we take $t_i' - \overline{t}' = D$

$$\frac{D}{\sqrt{\dfrac{2}{n-1}(1-\dfrac{1}{k})}} = Z_{(1-\alpha/2k)} + Z_p \tag{3.2}$$

$$D\sqrt{\frac{k(n-1)}{2(k-1)}} = Z_{(1-\alpha/2k)} + Z_P$$

$$\sqrt{\frac{k(n-1)}{2(k-1)}} = \frac{Z_{(1-\alpha/2k)} + Z_p}{D}$$

$$\frac{k(n-1)}{2(k-1)} = \left[\frac{Z_{(1-\alpha/2k)} + Z_p}{D}\right]^2$$

$$n-1 = 2\left(\frac{k-1}{k}\right)\left[\frac{Z_{(1-\alpha/2k)} + Z_p}{D}\right]^2$$

$$n = 1 + 2\left(\frac{k-1}{k}\right)\left[\frac{Z_{(1-\alpha/2k)} + Z_p}{D}\right]^2 \tag{3.3}$$

Hence, the expression (3.3) is the sample size for the method developed by Pran Kumar and Rao (1998) to detect the significant difference between any one of t_i' and $\bar{t}'$ by at least a specified difference *D* in the case of equal sample size for a fixed power *P*.

4. TABLES OF SAMPLE SIZES

Using the expression (3.3) the sample sizes computed for fixed power P, α, and D are presented in Tables 4. 1 through 4. 4.

Table 4. 1 $\alpha = 0.05$ $D = 3$

k \ P	0.80	0.90	0.95	0.99
3	**3**	**3**	**3**	**4**
4	**3**	**3**	**4**	**5**
5	**3**	**4**	**4**	**5**
6	**3**	**4**	**4**	**6**
7	**3**	**4**	**5**	**6**
8	**3**	**4**	**5**	**6**
9	**4**	**4**	**5**	**6**
10	**4**	**4**	**5**	**6**
11	**4**	**4**	**5**	**6**
12	**4**	**5**	**5**	**6**
13	**4**	**5**	**5**	**7**
14	**4**	**5**	**5**	**7**
15	**4**	**5**	**5**	**7**
16	**4**	**5**	**5**	**7**
17	**4**	**5**	**5**	**7**
18	**4**	**5**	**6**	**7**
19	**4**	**5**	**6**	**7**
20	**4**	**5**	**6**	**7**
30	**4**	**5**	**6**	**7**
60	**5**	**6**	**6**	**8**

Table 4.2 $\alpha = 0.01$ $D = 3$

k \ P	0.80	0.90	0.95	0.99
3	**3**	**4**	**4**	**5**
4	**3**	**4**	**5**	**6**
5	**4**	**4**	**5**	**6**
6	**4**	**5**	**5**	**7**
7	**4**	**5**	**5**	**7**
8	**4**	**5**	**6**	**7**
9	**4**	**5**	**6**	**7**
10	**4**	**5**	**6**	**7**
11	**4**	**5**	**6**	**7**
12	**5**	**5**	**6**	**8**
13	**5**	**5**	**6**	**8**
14	**5**	**5**	**6**	**8**
15	**5**	**6**	**6**	**8**
16	**5**	**6**	**6**	**8**
17	**5**	**6**	**6**	**8**
18	**5**	**6**	**6**	**8**
19	**5**	**6**	**7**	**8**
20	**5**	**6**	**7**	**8**
30	**5**	**6**	**7**	**9**
60	**6**	**7**	**7**	**9**

Table 4.3 $\alpha = 0.01$ $D = 1$

k \ P	0.8	0.9	0.95	0.99
3	20	25	29	38
4	23	29	34	44
5	26	32	37	48
6	27	34	39	51
7	29	35	41	53
8	30	37	43	55
9	31	38	44	56
10	32	39	45	58
11	32	39	46	59
12	33	40	47	60
13	34	41	47	61
14	34	41	48	62
15	35	42	49	62
16	35	42	49	63
17	35	43	50	64
18	36	43	50	64
19	36	44	51	65
20	36	44	51	65
30	39	47	54	69
60	43	51	59	74

Table 4.4 $\alpha = 0.05$ $D = 1$

K \ P	0.8	0.9	0.95	0.99
3	15	19	23	31
4	18	22	27	36
5	20	25	30	39
6	21	27	32	42
7	22	28	33	44
8	23	29	35	46
9	24	30	36	47
10	25	31	37	48
11	26	32	38	49
12	26	33	38	50
13	27	33	39	51
14	27	34	40	52
15	28	34	40	53
16	28	35	41	53
17	28	35	41	54
18	29	35	42	54
19	29	36	42	55
20	29	36	42	55
30	32	39	45	59
60	35	43	50	64

5. CONCLUDING REMARKS

In Tables 4.1, and 4.2, , the observations (in bold numerals) of sample sizes should be replaced by sample size 11, since the log – transformation of Bartlett & Kendall (1946) used in the graphical method developed by Pran Kumar and Rao(1998) follows normality only when n>10.

In general, higher sample sizes are required for lower significance level α, lower difference D, higher power P, and higher number of populations$k$. Lower sample sizes are enough to consider for higher significance level α, higher difference D, lower power P and lower number of populations k.

The lowest sample size n = 11, is enough to use for α = 0.05, 0.01, D = 3, P = 0.8, 0.9, 0.45, 0.99 and k = 3(1) 20, 30, 60. In general, a lowest sample size of 11 is considered for $\alpha \geq 0.01$, $D \geq 3$, $P \leq 0.99$ and $\forall\ 3 \leq k \leq 60$.

6. REFERENCES

[1]. Bartlett, M. S and Kendall, O. D. (1946). "The statistical analysis of variance –heterogeneity and the logarithmic transformation". Journal of Royal Statistical Society, 8, pp. 126 – 128.

[2]. Bratcher, T. L, Moran, M.A and Zimmer, W.J. (1970). "Tables of sample sizes in the analysis of variance". Journal of Quality Technology, 2, pp. 156 – 154.

[3]. Chow S, Shao, J and Wang H. (2008). "Sample size calculations in clinical research". 2nd ed. Chapman and Hall / CRC, Biostatistics series, pp. 71.

[4]. Nelson, P. R. (1983). "A comparison of sample sizes for the analysis of means and the analysis of variance". Journal of Quality Technology , 15, pp. 33 – 39.

[5]. Ott, E. R. (1967). "Analysis of Means – a graphical procedure". Industrial Quality Control, 24, pp. 101 – 109.

[6]. Pran Kumar, M and Rao, C. V. (1998). "Anom – type graphical method for testing the equality of several variances". Communications in statistics – Simulation and Computation, 27, pp. 459– 468.

[7]. Rao, C. V. (2005). "Analysis of means – a review". Journal of Quality Technology, 37, pp. 308 – 315.