African Journal of Mathematics and Computer Science Research Vol. 4(1), pp. 27-31, January 2011
Available online at http://www.academicjournals.org/AJMCSR
ISSN 2006-9731 ©2011 Academic Journals

Full Length Research Paper

Modified extended median test

Oyeka I. C. A. and Uzuke C. A.*

Department of Statistics, Nnamdi Azikiwe University Awka, Nigeria.

Accepted 27 April, 2010

This paper proposes a modified extended median test that adjusts for the effects of any possible ties in the specific data. The proposed method is illustrated with data. A comparison with the extended median test suggests that the proposed method is at least as powerful as the extended median test. In this study, the proposed extended median test presents an example indicating that the proposed modified extended median test is more powerful than the ordinary extended median test.

Key words: Median test, tied observation, chi-square, P-value, χ^2 distribution.

INTRODUCTION

Let x_{ij} be the i^{th} observation in a random sample of size n_j independently drawn from population j for i = 1, 2… n_j and j = 1, 2,...,k. It is further assumed that the k populations are independent and continuous (Krauth, 2003; Cohen, 1983). The objective is to determine whether these samples could have been drawn from populations with a common median.

The procedure often adopted is to first pool the j sample into one combined sample and determine the median M of this combined sample. Then one is able to determine for each sample how many of its values are above the common median M and how many are below M, and the resulting data are then presented in a 2xk contingency table and the usual chi-square test is applied (Miller, 1996; Vargha et al., 1996).

Thus, let

$$U_{ij} = \begin{cases} 1, \; if\ xij > M, i = 1,2,\dots,nj\ ; j = 1,2,\dots,k \\ 0 \quad otherwise \end{cases} \tag{1}$$

where M is the common median of the pooled sample data.

Now,

$$\text{let } \theta j = P(Uij = 1) \tag{2}$$

Also

$$\text{Let } Uj = \sum_{i=1}^{nj} Uij \tag{3}$$

$$\text{Then } E(Uij) = \theta j, Var\ (Uij) = \theta j(1-\theta j) \tag{4}$$

$$\text{Also } E(Uj) = nj\theta j,\ Var\ (Uj) = nj\theta j(1-\theta j) \tag{5}$$

$$Pj = \hat{\theta}j = \frac{Uj}{nj}, \quad \text{and } \bar{P} = \hat{\theta} = \frac{t}{n} \tag{6}$$

Where $t = \sum_{j=1}^{k} Uj,\ n = \sum_{j=1}^{k} nj$

It is possible to derive a test statistic based on equation (1) to (6) by method of complete enumeration. However, a less tedious method is suggested and followed here in terms of contingency tables.

Now the observed frequencies are

$$O_{1j} = U_j\ ;\ O_{2j} = n_j - U_j \tag{7}$$

*Corresponding author. E-mail: chinwe_uzuke @yahoo.com.

And the corresponding expected frequencies are

$$e_{1j} = \frac{nj.t}{n}, \; e_{2j} = \frac{nj(n-t)}{n} \quad (8)$$

Hence the usual chi-square test statistic is

$$\chi 2 = \sum_{i=1}^{2}\sum_{j=1}^{k}\frac{(Oij-eij)^2}{eij}$$

$$= \sum_{j=1}^{k}\frac{\left(Uj-nj\frac{t}{n}\right)^2}{nj\frac{t}{n}} + \sum_{j=1}^{k}\frac{\left((nj-Uj)-nj\left(\frac{n-t}{n}\right)\right)^2}{nj\left(\frac{n-t}{n}\right)}$$

$$= \sum_{j=1}^{k} nj\frac{(Pj-\bar{P})^2}{\bar{P}\,(1-\bar{P})} \quad (9)$$

Which under the null hypothesis (Ho) has a chi-square distribution with k-1 degrees of freedom and can be used to test the null hypothesis of equal population medians. A necessary assumption in the above presentation is that the population from which samples are drawn are continuous. This makes it possible to assume at least theoretically that the probability of ties in the data is zero. However, in practice ties do occur. The presence of ties in the data especially if they are not few seriously affect the result of the analysis since they often reduce the power of the test (Westerberg, 1948; Gart, 1963). In the following we present a method that does not require that the populations be necessarily continuous. The proposed method by specification structurally adjusts for the possibilities of ties in the data.

THE PROPOSED METHOD

As above, let xij be the i^{th} observation in a random sample of size nj independently drawn from the jth population for $i = 1, 2, \ldots, n_j, j = 1, 2, \ldots, k$. The populations may consist of data based on ordinal, interval or ratio scale of measurement.

Let as seen above M be taken as the common median of the pooled sample data.

Now let

$$Uij = \begin{cases} 1 \;\; if \; x_{ij} > M \\ 0 \;\; if \; x_{ij} = M \\ -1 \;\; if \; x_{ij} < M \end{cases} \quad (10)$$

Let $\theta j^+ = P(Uij = 1)$

$$\theta j^0 = P(Uij = 0) \quad (11)$$

$$\theta j^- = P(Uij = -1)$$

Where $\theta j^+ + \theta j^0 + \theta j^- = 1$ (12)

Let $Uj = \sum_{i=1}^{nj} Uij$ (13)

Now

$$E(Uij) = \theta j^+ - \theta j^- \; and \; Var(Uij) = \theta j^+ + \theta j^- - (\theta j^+ - \theta j^-)^2 \quad (14)$$

The test statistic for the equality of population medians based on equations (10) to (14) can be obtained by complete enumeration. We will, however, base our test statistic on the chi-square test for goodness of fit using a 3xk contingency table.

Let fj^+, fj^0, fj^-, be the number of 1's 0's and -1's obtained in the frequency distribution of the nj values of Uij.

Let

$$t^+ = \sum_{j=1}^{k} fj^+$$
$$t^- = \sum_{j=1}^{k} fj^- \quad (15)$$
$$t^0 = \sum_{j=1}^{k} fj^0 = n - t^+ - t^-$$

Where $n = \sum_{j=1}^{k} n_j$

Let

$$P_j^+ = \hat{\theta}_j^+ = \frac{f_j^+}{n_j}$$
$$P_j^- = \hat{\theta}_j^- = \frac{f_j^-}{n_j} \quad (16)$$
$$P_j^0 = \hat{\theta}_j^0 = \frac{f_j^0}{n_j} = 1 - P_j^+ - P_j^-$$

Also let $P^+ = \hat{\theta}^+ = \frac{t^+}{n}$

$$P^- = \hat{\theta}^- = \frac{t^-}{n} \quad (17)$$
$$P^0 = \hat{\theta}^0 = \frac{t^0}{n} = 1 - P^+ - P^-$$

Therefore the observed frequencies are:

$$O_{1j} = f_j^+,$$
$$O_{2j} = f_j^-,$$
$$O_{3j} = f_j^0 = n_j - f_j^+ - f_j^- \quad (18)$$

And the expected frequencies are:

$$e_{1j} = n_j \frac{t^+}{n},$$
$$e_{2j} = n_j \frac{t^-}{n},$$
$$e_{3j} = n_j \frac{t^0}{n} = n_j \frac{(n-t^+ - t^-)}{n} \quad (19)$$

Table 1. Frequency distribution of students' grade by departments and its relationship with the median grade (Use of extended median test).

Relationship with the median grade D	Students' departments						
	Physics	PAE	Zoology	HPE	MCB	BCH	Total
>Median	35	72	16	41	83	126	373
≤Median	34	62	26	46	194	162	524
Total n_j	69	134	42	87	277	288	897

Hence the χ^2 test statistic is

$$\chi^2 = \sum_{i=1}^{3}\sum_{j=1}^{k}\frac{(o_{ij}-e_{ij})^2}{e_{ij}}$$

$$= \sum_{j=1}^{k}\frac{\left(f_j^+ - n_j\frac{t^+}{n}\right)^2}{n_j\frac{t^+}{n}} + \sum_{j=1}^{k}\frac{\left(f_j^- - n_j\frac{t^-}{n}\right)^2}{n_j\frac{t^-}{n}} + \sum_{j=1}^{k}\frac{\left((n_j - f_j^+ - f_j^-) - n_j\frac{(n-t^+-t^-)}{n}\right)^2}{n_j\frac{(n-t^+-t^-)}{n}} \qquad (20)$$

$$= \sum_{j=1}^{k} n_j\frac{(P_j^+ - P^+)^2}{P^+} + \sum_{j=1}^{k} n_j\frac{(P_j^- - P^-)^2}{P^-} + \sum_{j=1}^{k} n_j\frac{\left((P_j^+ - P^+) - (P_j^- - P^-)\right)^2}{1-P^+-P^-}$$

$$= \sum_{j=1}^{k} n_j\frac{(P_j^+ - P^+)^2}{P^+} + \sum_{j=1}^{k} n_j\frac{(P_j^- - P^-)^2}{P^-} + \sum_{j=1}^{k} n_j\frac{(P_j^+ - P^+)^2}{1-P^+-P^-} + \sum_{j=1}^{k} n_j\frac{(P_j^- - P^-)^2}{1-P^+-P^-}$$

$$- 2\sum_{j=1}^{k} n_j\frac{\left((P_j^+ - P^+) - (P_j^- - P^-)\right)}{1-P^+-P^-}$$

$$\frac{1}{(P^+)(P^-)(1-P^+-P^-)}\left[P^-(1-P^-)\sum_{j=1}^{k} n_j(P_j^+ - P^+)^2 + P^+(1-P^+)\sum_{j=1}^{k} n_j(P_j^- - P^-)^2 - 2P^+P^-\sum_{j=1}^{k} n_j\,(P_j^+ - P^+)(P_j^- - P^-)\right] \qquad (21)$$

Which, under the null hypothesis has a χ^2 distribution with 2(k-1) degrees of freedom and can be used to test the null hypothesis of equal population medians provided as in all chi-square test; the cell frequencies are each at least equal to 5. If this requirement cannot be satisfied, it is recommended that categories and cell frequencies be pooled as appropriate.

Note that with the modified extended median test, the probability of an individual having exactly the median score can be obtained unlike M, the extended median test where this probability cannot be determined.

ILLUSTRATIONS

Below are the grades of students from six (6) departments who took an introductory course in statistics. Interest is to ascertain whether these students perform equally well in the course; the grades were pooled together to obtain the common median, which was found to be a grade of D. For each department, every student grade was compared with the median grade and was scored 1, 0, or -1 depending on whether the grade was greater than the median grade, equal, or less than the median grade D of the pooled sample data for the six (6) departments. If we are to use the extended median test, the scores zero may be collapsed with "less than" category. The result is shown in Table 1.

The data in Table 1 was converted into proportions for the use of equation (9) in extended median test and presented as in Table 2

Applying equation (9) we have,

$$\chi^2 = \sum_{j=1}^{k} nj\frac{(Pj - \bar{P})^2}{\bar{P}\,(1-\bar{P})}$$

$$= 69\frac{(0.507-0.416)^2}{0.2429} + 134\frac{(0.537-0.417)^2}{0.2429} + 42\frac{(0.0.381-0.416)^2}{0.2429} + 87\frac{(0.471-0.416)^2}{0.2429} + 277\frac{(0.0.300-0.416)^2}{0.2429} + 288\frac{(0.0.438-0.416)^2}{0.2429}$$

= 2.3524 + 8.0771 + 0.2118 + 1.0835 + 15.345 + 0.5379

= 27.64

p-value = 0.000043.

Now to apply the modified extended median test, the data may be presented as in Table 3 which has now three categories namely: less than, equal to and greater than the median M.

The data in Table 3 were converted into proportions for

Table 2. Proportions of student's grade by department and its relationship with the median grade.

Relationship with the median grade D	Students' departments						
	Physics	PAE	Zoology	HPE	MCB	BCH	Total
>Median	0.507	0.537	0.381	0.471	0.300	0.438	$\bar{P}$ = 0.416
≤Median	0.493	0.463	0.619	0.529	0.700	0.562	1- $\bar{P}$ = 0.584
Total n_j	69	134	42	87	277	288	897

Table 3. Frequency distribution of students' grade by departments and its relationship with the median grade (Use of modified extended median test).

Relationship with the median grade D		Students' departments						
		Physics	PAE	Zoology	HPE	MCB	BCH	Total
>Median	+	35	72	16	41	83	126	373
= Median	0	15	8	4	14	32	43	116
< Median	-	19	54	22	32	162	119	408
Total n_j		69	134	42	87	277	288	897

Table 4. Proportions of student's grade by department and its relationship with the median grade.

Relationship with the median grade D	Students' departments						
	Physics	PAE	Zoology	HPE	MCB	BCH	Total
>Median	0.507	0.537	0.381	0.471	0.300	0.438	0.416
= Median	0.218	0.06	0.085	0.161	0.115	0.149	0.128
< Median	0.275	0.403	0.534	0.368	0.585	0.413	0.456
Total n_j	69	134	42	87	277	288	897

the use of equation (21) in modified extended median test and presented as in Table 4

Applying equation (21) we have,

$$\chi^2 = \frac{1}{(P^+)(P^-)(1-P^+-P^-)}\left[P^-(1-P^-)\sum_{j=1}^{k} n_j\left(P_j^+ - P^+\right)^2 + P^+(1-P^+)\sum_{j=1}^{k} n_j\left(P_j^- - P^-\right)^2 - 2P^+P^-\sum_{j=1}^{k} n_j\left(P_j^+ - P^+\right)\left(P_j^- - P^-\right)\right]$$

$$\frac{1}{(0.416)(0.456)(0.128)}[(0.456)(0.544)[69(0.507-0.416)^2+134(0.537-0.416)^2+42(0.381-0.4$$
$$+87(0.471-0.416)^2+277(0.300-0.416)^2+288(0.438-0.416)^2]$$
$$+(0.416)(0.584)[69(0.275-0.456)^2+134(0.403-0.456)^2+42(0.524-0.456)^2$$
$$+87(0.368-0.456)^2+277(0.585-0.456)^2+288(0.413-0.456)^2]+2\times 0.416$$
$$\times 0.456[69(0.507-0.416)(0.275-0.456)+134(0.537-0.419)(0.403-0.456)$$
$$+42(0.381-0.416)(0.524-0.456)+87(0.471-0.416)(0.36-0.456)$$
$$+277(0.300-0.416)(0.585-0.456)+288(0.433-0.416)(0.413-0.456)]]$$

$$= \frac{1}{0.02428}[1.6186 + 2.1007 - 2.4530]$$

= 52.15

P-value = 0.000000.

Note that the probability a randomly selected individual earns exactly the median score P^o = 0.128 or 12.8%. Also the attained probability of rejecting a true null hypothesis that is type I error is 0.000043 for the extended median test and only 0.000000 for the proposed extended median test indicating that the proposed modified extended median test is at least in this example more powerful than the ordinary extended median test. This is further vindicated by the fact that even though the two chi-square values here both lead to the rejection of the null hypothesis, the chi-square value obtained using the modified extended median test (χ^2 = 52.15) is 1.89 times or nearly twice the chi-square value obtained using the ordinary extended median test (χ^2 =

26.74), an indication that the former yields a much more statistically significant result than the latter.

Conclusion

A modified extended median test is developed in such a way that it structurally adjusts for the possible presence of ties in the data. This method does not require the population to be continuous unlike the extended median test.

However, because the proposed method has been structurally adjusted for ties in the data, the method is not affected by ties unlike the extended median test which is not adjusted for ties. Hence as demonstrated by the illustrative example, the proposed method is as powerful as the extended median test. It also enables one to determine the probability that a randomly selected subject earns exactly the median score of the population unlike the case with the ordinary extended median test.

REFERENCES

Cohen J, (1983). The Cost of Dichotomisation, Appl. Psychol. Measure., 7: 249-253.

Gart JJ (1963). A Median Test with Sequential Application. Biometrika, 50: 55-62.

Krauth J (2003). Median Dichotomisation in Configural Frequency Analysis: Is it Allowed? Appl. Psychol. Measure., 7: 249-253.

Miller J (1996). Statistics for Advanced Level (2nd Edition) Cambridge University Press.

Vargha A, Rudas T, Delany HD, Maxwel SE (1996). Dichotomisation, Partial Correlation, and Conditional Independence. J. Educ. Behav. Stat., 21: 264–282.

Westerberg J (1948). Significance Test for Median and Interquartile Range in Samples from Continuous Populations of Any Form. Proceedings of the Konmklijke Naderladse Akademie Van Wetenschappen, Amsterdam, Series A: Math. Sci., pp. 252-261.