International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 9(Special Issue), 2009, pp 209-215

Comparing Locations Using Medians: Exact and Asymptotics

Mezbahur Rahman and Larry M. Pearson
Minnesota State University
Mankato, MN 56001, USA
E-mail: mezbahur.rahman@mnsu.edu & larry.pearson@mnsu.edu

[Received June 9, 2008; Accepted May 30, 2009]

Abstract

Two medians are used in comparing location parameters in two independent populations in nonparametric testing. Three different exact procedures and two different approximation procedures are discussed. Performances are shown using simulation.

Keywords and Phrases: Binomial Distribution, Hypergeometric Distribution, Normal Approximation.

AMS Classification: 62F03.

1 Introduction

Let us consider $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ be two independent random ramples from two independent populations with equal shape parameters. We are interested in comparing their locations. One of the simplest and most widely used nonparametric procedures for testing the null hypothesis that two independent samples have been drawn from populations with equal medians is the *median test* attributed to Mood (1950) and Westenberg (1948). The first sample is from a population with unknown median M_1 and the second sample is from a population with unknown median M_2. The variables of interest are continuous and the measurement scales employed are at least ordinal. The null hypothesis is $H_0 : M_1 = M_2$ and the possible alternative hypotheses are $H_1 : M_1 \neq M_2$, $H_1 : M_1 < M_2$, and $H_1 : M_1 > M_2$. If the two populations have the same median, then for each population the probability that an observed value will exceed the combined population median are the same. The hypothesis tests combine the observations from the two samples and compute the median. Then, the observations are classified depending on whether they are above or equal to the combined median or below the combined median. The outcomes are displayed in the following table:

Table 1: Data Summary

	Sample 1	Sample 2	Total
Above	A	B	$A+B$
Equal or below	C	D	$C+D$
Total	$A+C=n_1$	$B+D=n_2$	$N=n_1+n_2$

If H_0 is true, we expect about one-half of the observations in each sample to fall above the combined sample median and one-half to fall below. Mood (1950) has shown that under the null hypothesis, the sampling distribution of A and B can be expressed as the *hypergeometric distribution*:

$$P(A,B)=\frac{\binom{n_1}{A}\binom{n_2}{B}}{\binom{N}{A+B}}. \tag{1.1}$$

For the two sided alternative hypothesis, the P-value can be computed using equation (1.1) as:

$$P_{HYP}=2\sum_{x=0}^{M}P(x,A+B-x) \tag{1.2}$$

where M is the minimum of A and B relative to n_1 and n_2. In cases of one sided tests, (1.2) can be computed for the respective critical regions and without multiplying by 2.

Since the samples are independent, it can also be argued that the sampling distribution is binomial, that is, for any p_1, probability of the observation being in the 'Above' group in Sample 1, and p_2 is the corresponding probability for Sample 2,

$$P(A,B)=\binom{n_1}{A}p_1^A(1-p_1)^{n_1-A}\binom{n_2}{B}p_2^B(1-p_2)^{n_2-B}. \tag{1.3}$$

For a two sided hypothesis, a higher absolute difference between A and B will indicate the rejection of the null hypothesis. The sampling distribution of $D=A-B$ under the null hypothesis is as follows: For $D=0$ or $A=B$,

$$P(A=B)=\binom{n_1}{A}\binom{n_2}{A}p^{2A}(1-p)^{n_1+n_2-2A} \tag{1.4}$$

where $p = 2A/N$. For $A \neq B$, and without loss of generality, we can consider $n_1 > n_2$ to obtain

$$P(D = d) = \binom{n_1}{d} p^d (1-p)^{n_1+n_2-d} \sum_{y=0}^{n_1-d} \binom{n_2}{y} \tag{1.5}$$

for $d = 1, 2, \ldots, n_1$, where p is the probability of the obsevation being 'Above' in either group. Then, the P-values can be computed for a two sided hypothesis as

$$P_{BNA} = 2P(D \geq R) \tag{1.6}$$

where $R = |A - B|$ as in Table 1 and the probabilities are computed using (1.4) and (1.5). BNA stands for binomial and $p = 2A/N$. Since the test is for the equality of two medians, under the null hypothesis $1/2$ can be substituted for p. Equations (1.4) and (1.5) can be written as

$$P(A = B) = \binom{n_1}{A} \binom{n_2}{A} \left(\frac{1}{2}\right)^N \tag{1.7}$$

For $A \neq B$, and without loss of generality, we can consider $n_1 > n_2$ which yields

$$P(D = d) = \binom{n_1}{d} \left(\frac{1}{2}\right)^N \sum_{y=0}^{n_1-d} \binom{n_2}{y} \tag{1.8}$$

for $d = 1, 2, \ldots, n_1$. Here we can compute P-values as in (1.6) using (1.7) and (1.8) and refer to them as P_{BNE}. One sided P-values can also be computed by adjusting (1.6) accordingly.

P-values also can be approximated using

$$Z_{BNA} = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{1.9}$$

where $p_1 = A/n_1$, $p_2 = B/n_2$, and $p = (A+B)/(n_1+n_2)$ and denoted as P_{ZNA}. P-values also can be approximated using

$$Z_{BNE} = \frac{p_1 - p_2}{\sqrt{\frac{1}{2}\left(1 - \frac{1}{2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{1.10}$$

and denoted as P_{ZNE}. In the following section we compare the five different tests mentioned above using simulation.

2 Simulation

Independent random samples are generated from binomial variates. We consider $p = 0.5$ as the value of p when H_0 is true and $(p_1 = 0.45, p_2 = 0.55)$, $(p_1 = 0.3, p_2 = 0.7)$, and

$(p_1 = 0.1, p_2 = 0.9)$ are considered for power computations. For each case we take samples of sizes 5, 10, 20, 30, and 40. Ten thousand samples are selected for each case and each sample size and number of rejections are recorded using the 5% level of significance. In the table below (Table 2), we present the proportions of rejections along with mean and standard deviations of the P-values.

Table 2: Simulation Results

	N	P_{HYP}	P_{BNE}	P_{BNA}	P_{ZNE}	P_{ZNA}
$p_1 = 0.5, p_2 = 0.5$						
Rejection Rate	5	0.0225	0.0225	0.0225	0.0225	0.0606
Mean P-value		0.9131	0.6441	0.6266	0.5146	0.4966
St. Dev. P-value		0.4346	0.3023	0.3104	0.3238	0.3318
Rejection Rate	10	0.0129	0.0423	0.0428	0.0423	0.0428
Mean P-value		0.7815	0.6090	0.6015	0.5079	0.4995
St. Dev. P-value		0.3944	0.3085	0.3119	0.3084	0.3112
Rejection Rate	20	0.0226	0.0423	0.0424	0.0423	0.0467
Mean P-value		0.6939	0.5812	0.5775	0.5044	0.5003
St. Dev. P-value		0.3619	0.3069	0.3085	0.2983	0.2995
Rejection Rate	30	0.0299	0.0299	0.0403	0.0543	0.0543
Mean P-value		0.6584	0.5695	0.5671	0.5050	0.5023
St. Dev. P-value		0.3494	0.3070	0.3080	0.2966	0.2974
Rejection Rate	40	0.0308	0.0308	0.0316	0.0549	0.0549
Mean P-value		0.6409	0.5651	0.5633	0.5081	0.5061
St. Dev. P-value		0.3394	0.3041	0.3049	0.2940	0.2945
$p_1 = 0.45, p_2 = 0.55$						
Rejection Rate	5	0.0280	0.0280	0.0280	0.0280	0.0767
Mean P-value		0.8879	0.6263	0.6084	0.4968	0.4786
St. Dev. P-value		0.4410	0.3079	0.3155	0.3240	0.3314
Rejection Rate	10	0.0225	0.0629	0.0639	0.0629	0.0639
Mean P-value		0.7387	0.5751	0.5675	0.4762	0.4680
St. Dev. P-value		0.4041	0.3172	0.3202	0.3113	0.3136
Rejection Rate	20	0.0485	0.0803	0.0805	0.0803	0.0855
Mean P-value		0.6121	0.5114	0.5077	0.4401	0.4362
St. Dev. P-value		0.3792	0.3229	0.3240	0.3067	0.3075
Rejection Rate	30	0.0688	0.0686	0.0872	0.1164	0.1164
Mean P-value		0.5480	0.4722	0.4697	0.4149	0.4122
St. Dev. P-value		0.3669	0.3227	0.3234	0.3054	0.3058
Rejection Rate	40	0.1125	0.1125	0.1137	0.1592	0.1592
Mean P-value		0.4945	0.4340	0.4321	0.3861	0.3841
St. Dev. P-value		0.3588	0.3213	0.3217	0.3033	0.3035

Table 2: Simulation Results Continued						
	N	P_{HYP}	P_{BNE}	P_{BNA}	P_{ZNE}	P_{ZNA}
$p_1 = 0.3$, $p_2 = 0.7$						
Rejection Rate	5	0.1578	0.1578	0.1578	0.1578	0.2619
Mean P-value		0.5773	0.4011	0.3845	0.2973	0.2822
St. Dev. P-value		0.4594	0.3315	0.3322	0.3032	0.3034
Rejection Rate	10	0.2381	0.4098	0.4117	0.4098	0.4117
Mean P-value		0.3057	0.2316	0.2251	0.1772	0.1714
St. Dev. P-value		0.3492	0.2766	0.2756	0.2396	0.2380
Rejection Rate	20	0.6047	0.7078	0.7081	0.7078	0.7198
Mean P-value		0.0999	0.0794	0.0775	0.0626	0.0610
St. Dev. P-value		0.1909	0.1597	0.1586	0.1373	0.1362
Rejection Rate	30	0.8403	0.8401	0.8655	0.8965	0.8966
Mean P-value		0.0362	0.0292	0.0286	0.0233	0.0227
St. Dev. P-value		0.0993	0.0846	0.0839	0.0720	0.0714
Rejection Rate	40	0.9382	0.9382	0.9388	0.9632	0.9632
Mean P-value		0.0144	0.0117	0.0114	0.0095	0.0092
St. Dev. P-value		0.0551	0.0471	0.0467	0.0409	0.0405
$p_1 = 0.1$, $p_2 = 0.9$						
Rejection Rate	5	0.7351	0.7351	0.7351	0.7351	0.8205
Mean P-value		0.1022	0.0611	0.0554	0.0362	0.0323
St. Dev. P-value		0.1682	0.1183	0.1128	0.0828	0.0788
Rejection Rate	10	0.9577	0.9879	0.9880	0.9879	0.9880
Mean P-value		0.0081	0.0051	0.0047	0.0033	0.0030
St. Dev. P-value		0.0334	0.0245	0.0236	0.0173	0.0166
Rejection Rate	20	0.9999	0.9999	0.9999	0.9999	0.9999
Mean P-value		0.0001	0.0000	0.0000	0.0000	0.0000
St. Dev. P-value		0.0014	0.0010	0.0009	0.0007	0.0007
Rejection Rate	30	1.0000	1.0000	1.0000	1.0000	1.0000
Mean P-value		0.0000	0.0000	0.0000	0.0000	0.0000
St. Dev. P-value		0.0000	0.0000	0.0000	0.0000	0.0000
Rejection Rate	40	1.0000	1.0000	1.0000	1.0000	1.0000
Mean P-value		0.0000	0.0000	0.0000	0.0000	0.0000
St. Dev. P-value		0.0000	0.0000	0.0000	0.0000	0.0000

In Table 2 we notice that assuming H_0 is true, that is, for $p_1 = p_2 = 0.5$, rejection rates are closer to 0.05 in normal approximations irrespective of sample sizes. When H_0 is not true, the rejection rates are similar for all tests except they are higher for Z_{BNA} at $n = 5$. It is to be noted that in all cases, the rejection rates are smaller for P_{HYP} at $n = 10$ and $n = 20$.

3 Conclusion

Most nonparametric texts introduce P_{HYP} as the test statistic and then mention approximation methods for larger sample sizes. But here we have seen that P_{BNE} and P_{BNA} perform better than P_{HYP}. After analyzing overall performances, the Z_{BNA} test is preferred. That is, even for smaller sample sizes Z_{BNA} is preferred instead of exact procedures and the more tempting procedure Z_{BNE}.

4 Application

Newmark et al. (1973) have reprted the results of an attempt to assess the predictive validity of Klopfer's Prognostic Rating Scale (PRS) with subjects who received behavior modification psychotherapy. Following psychotherapy, the subjects were separated into two groups: improved and unimproved. Table 3 shows the PRS score for each subject before therapy. We wish to see whether we can conclude on the basis of this data that the two represented populations are different with respect to their medians.

Table 3: PRS scores for improved and unimproved subjects

Improved subjects		Improved subjects		Unimproved subjects	
1	11.9	11	6.9	1	6.6
2	11.7	12	6.8	2	5.8
3	9.5	13	6.3	3	5.4
4	9.4	14	5.0	4	5.1
5	8.7	15	4.2	5	5.0
6	8.2	16	4.1	6	4.3
7	7.7	17	2.2	7	3.9
8	7.4			8	3.3
9	7.4			9	2.4
10	7.1			10	1.7

Table 4: Data Summary

	Improved subjects	Unimproved subjects	Total
Above	12	1	13
Equal or below	5	9	14
Total	17	10	27

The P-values are computed for $H_0 : M_1 = M_2$ versus $H_1 : M_1 \neq M_2$, where M_1 and M_2 are corresponding medians for the scores for improved subjects and unimproved subjects. The results are $P_{ZNE} = 0.0024$, $P_{ZNA} = 0.0023$, $P_{BNE} = 0.0059$, $P_{BNA} = 0.0068$, and $P_{HYP} = 0.0118$, using the procedures mentioned in Section 1.

References

[1] Mood, Alexander M., *Introduction to the Theory of Statistics*, New York: McGraw-Hill, 1950.

[2] Newmark, Charles S., William Hetzel, Lilly Walker, Steven Holstein, and Martin Finklestein (1973). Predictive Validity of the Research Prognostic Rating Scale with Behavior Modification Techniques, *Journal of Clinical Psychology*, **29**, 246-248.

[3] Westenberg, J. (1948). Significance Test for Median and Interquartile Range in Samples from Continuous Populations of Any Form, *Akad. Wetensch. Afdeeling Voor de Wis.*, **51**.