A Table of Critical Values of a Rank Statistic Intended for Testing a Location-Scale Hypothesis

F. Rublík

Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia
Email: umerrubl@savba.sk

Abstract. *The exact critical points for selected sample sizes and significance levels are tabulated for the two-sample test statistic which is a combination of the Wilcoxon and the Mood test statistic. This statistic serves for testing the null hypothesis that two sampled populations have the same location and scale parameters.*

Keywords: Distribution-Free Rank Test, Combination of Rank Statistics, Small Sample Sizes

1. Introduction

Suppose that the random variables $X= \sigma_X \varepsilon + \mu_X$, $Y= \sigma_Y \varepsilon + \mu_Y$, where the real numbers μ_X, μ_Y denote the location parameters, $\sigma X >0$, $\sigma Y >0$ are scale parameters and the random variable ε has a distribution function continuous on the real line. Assume that $X_1 ,\dots, X_m$ is a random sample from the distribution of X, $Y_1 ,\dots, Y_n$ is a random sample from the distribution of Y and these random samples are independent. Let $(R_1, \dots ,R_N)$, $N=m+n$, denote the ranks of the pooled sample $X_1 ,\dots, X_m$, $Y_1 ,\dots, Y_n$ .The null hypothesis

$$H_0: \quad \mu_X = \mu_Y \ , \qquad \sigma_X = \sigma_Y \tag{1}$$

is sometimes called also the location-scale hypothesis. In practice the change in the location is often accompanied by the change in scale, and in such a case the statistics constructed for testing the location-scale null hypothesis (1) usually yield better results than the statistics constructed especially for the one type change of location or constructed especially for the one type change of scale. Readers interested in further discussion of the need of testing this hypothesis can found further arguments in Section 1 of [6] or in [3].

The null hypothesis (1) is against the alternative H_1 that at least one of the equalities (1) does not hold, tested usually by means of the Lepage test from [4]. This test is included also into the monograph [2]. The Lepage test statistic is given by the formula

$$T= T_K + T_B, \quad T_K = \frac{(S_W - E(S_W \mid H_0))^2}{Var(S_W \mid H_0)} \ , \quad T_B = \frac{(S_B - E(S_B \mid H_0))^2}{Var(S_B \mid H_0)} \ , \tag{2}$$

where $S_W = \sum_{i=1}^{m} R_i$ is the Wilcoxon rank test statistic and $S_B = \sum_{i=1}^{m} a_N(R_i)$ is the Ansari-Bradley rank tests statistic (i.e., the vector of scores $a_N=(1,2,\dots,k,k,\dots,2,1)$ if $N=2k$ and $a_N=(1,2,\dots,k,k+1,k,\dots,2,1)$ if $N=2k+1$). An analogous statistic has been formulated in the multisample setting in [7], another test statistics has been for this problem studied in [8], where also their non-centrality parameters for testing the null location-scale hypothesis are for some situations computed. The statistics in [8] are defined in the general multisample setting, but they can be used also in the two-sample setting, when the tables of the critical constants can be computed. However, for testing the location-scale hypothesis the only available tables of critical constants are those published in [5].

2. Basic Formulas

The topic of the paper is the computation of the critical values for the two-sample test statistic (3). In accordance with [8] label the combination of the Wilcoxon and the Mood statistic by T_{SQ}. Thus in the notation from the previous section

$$T_{SQ} = T_K + Q , \tag{3}$$

$$T_K = \frac{12}{mn(N+1)} \left(S_W - \frac{m(N+1)}{2} \right)^2 , \tag{4}$$

$$Q = \frac{180}{mn(N+1)(N^2-4)} \left(\tilde{S} - \frac{m(N^2-1)}{12} \right)^2 , \quad \tilde{S} = \sum_{i=1}^{m} \left(R_i - \frac{N+1}{2}\right)^2 . \tag{5}$$

Since according to the assumptions the distribution function of ε is continuous, the statistic T_{SQ} is distribution-free whenever the null hypothesis (1) holds. The null hypothesis (1) is rejected whenever $T_{SQ} \geq w_\alpha$. Values of $w_\alpha = w_\alpha\,(m,n)$ can be found from the table presented in the next section, for the sample sizes not included into the table instead of w_α use the (1-α)th quantile of the chi-square distribution with 2 degrees of freedom.

In the computation of the tables of this paper the following lemma is useful.

Lemma 1. *Let J(m,n) denote the set of all m-tuples* $(i_1, \ldots ,i_m)$ *consisting of integers such that* $1 \leq i_1 < \ldots < i_m \leq m+n$. *Suppose that*

$$D(m,n,k_1,k_2) = \left\{ (i_1, \ldots ,i_m) \in J(m,n) ; \ \sum_{j=1}^{m} i_j = k_1 , \ \sum_{j=1}^{m} i_j^2 = k_2 \right\},$$

and $B(m,n,k_1,k_2)$ *denotes the number of elements of* $D(m,n,k_1,k_2)$.

(I) Let s >1. If at least one of the inequalities $k_1 \leq (r+s)$, $k_2 \leq (r+s)^2$ *holds, then* $B(s,r,k_1,k_2)= =B(s,r-1,k_1,k_2)$.

(II) If $k_1 > (s+r)$, $k_2 > (s+r)^2$, *then*

$$B(s,r,k_1,k_2) = B(s,r-1,k_1,k_2) + B(s-1, r, k_1 - (r+s) , k_2 - (r+s)^2) .$$

If $R=(R_1, \ldots ,R_N)$ is a random vector which is uniformly distributed over the set of all permutations of the set $\{1, \ldots ,N\}$, then according to Theorem 1 on p. 167 of [1] for any set $A \subset \{1, \ldots ,N\}$ consisting of *m distinct* integers

$$P(\{R_1, \ldots ,R_m\}=A) = 1 \Big/ \binom{n}{m} .$$

Combining this equality with Lemma 1 one can construct a program for computation of critical values of the statistic (3).

3. Tables of Critical Values

In this section we present the table 1 of the exact critical values of the statistic (3). First we describe the output of the table.

Since the set $V=V(m,n)$ of possible values of the statistic T_{SQ} is finite for every sample sizes m, n, in general one cannot find exact critical values for arbitrary prescribed probability α of the type I error. In the following table the number on the intersection of the column for $\underline{w}$ with the row for significance level α denote the quantity $\underline{w} = min\{ t \in V;\ P(T_{SQ} \geq t) \geq \alpha \}$, the entry corresponding to α and $\overline{w}$ is the quantity $\overline{w} = max\{t \in V;\ P(T_{SQ} \geq t) > \alpha \}$. Further, for given α, denote $\underline{p} = P(T_{SQ} \geq \underline{w})$, $\overline{p} = P(T_{SQ} \geq \overline{w})$ the corresponding probabilities of the type I error. Thus $\underline{p}$ is the largest available significance level not exceeding α and $\overline{p}$ is the smallest available significance level greater than α The value of critical constant yielding the significance level closer to the nominal level α is printed in boldface letter. If the difference in computed values exceeds the number of decimal places used to describe the result of computation, then the boldface symbol is used for the value corresponding to $\underline{p}$. For the space reasons only several combinations of sample sizes are given in the following table, a more detailed table will be presented in another paper.

Table 1 : Critical values of the test statistic T_{SQ} from (3).

A	$\underline{w}$	$\underline{p}$	$\overline{w}$	$\overline{p}$	$\underline{w}$	$\underline{p}$	$\overline{w}$	$\overline{p}$	$\underline{w}$	$\underline{p}$	$\overline{w}$	$\overline{p}$
		m=3	n=6			m=3	n=7			m=3	n=8	
0.200	3.5520	.179	**3.2796**	.202	**3.1298**	.200	3.0519	.217	3.3803	.176	**3.2435**	.212
0.100	4.2077	.083	**3.6121**	.107	**4.2532**	.100	4.0454	.117	4.1880	.091	**4.0213**	.103
0.050	**6.3272**	.048	4.7342	.071	**5.5000**	.050	4.2980	.067	6.2435	.036	**4.9359**	.061
0.020			**6.4519**	.024	**7.3181**	.017	7.0000	.033	8.1367	.012	**7.5897**	.024
0.010			**6.4519**	.024			**7.3181**	.017			**8.1367**	.012
0.005			**6.4519**	.024			**7.3181**	.017			**8.1367**	.012
		m=6	n=10			m=6	n=ll			m=6	n=12	
0.200	**3.3170**	.200	3.3086	.201	**3.2812**	.200	3.2748	.202	**3.2998**	.200	3.2998	.200
0.100	**4.4683**	.100	4.4487	.101	**4.4572**	.100	4.4550	.100	**4.4561**	.100	4.4473	.100
0.050	5.4700	.049	**5.4683**	.050	**5.4694**	.050	5.4630	.051	**5.5087**	.050	5.5065	.050
0.020	**6.8313**	.020	6.8016	.020	**6.8474**	.020	6.8410	.020	**6.8837**	.020	6.8771	.020
0.010	**7.7563**	.010	7.6605	.010	**7.8872**	.010	7.8745	.010	**7.9539**	.010	7.9100	.010
0.005	**8.5019**	.0049	8.4733	.0054	8.9718	.0048	**8.8718**	.0050	**9.0235**	.005	8.9183	.0051
		m=7	n=7			m=7	n=8			m=7	n=9	
0.200	**3.3020**	.200	3.2857	.203	**3.3223**	.200	3.3189	.200	3.3445	.198	**3.3328**	.200
0.100	**4.4653**	.100	4.4500	.101	**4.4931**	.100	4.4921	.100	**4.5012**	.100	4.4756	.100
0.050	5.5602	.049	**5.5551**	.050	**5.5029**	.050	5.5029	.050	**5.5694**	.050	5.5480	.050
0.020	**6.7183**	.020	6.6163	.021	**6.7162**	.020	6.6980	.020	**6.7635**	.020	6.7581	.020
0.010	**7.2500**	.010	7.0499	.011	**7.4921**	.010	7.4686	.010	**7.5478**	.010	7.5424	.010
0.005	**8.0500**	.0047	7.8622	.0058	8.0855	.0048	**7.9873**	.0051	8.4740	.004	**8.4441**	.0051
		m=7	n=10			m=7	n=11			m=7	n=12	
0.200	**3.3223**	.199	3.2982	.201	**3.2836**	.200	3.2831	.200	**3.2905**	.200	3.2871	.201
0.100	**4.5132**	.100	4.4972	.100	**4.5140**	.100	4.5140	.100	**4.4804**	.100	4.4789	.101
0.050	**5.5739**	.050	5.5719	.050	**5.5428**	.050	5.5428	.050	**5.6012**	.050	5.6000	.050
0.020	**6.8912**	.020	6.8711	.020	**6.8373**	.020	6.8373	.020	**6.9289**	.020	6.9159	.020
0.010	**7.7453**	.010	7.7052	.010	**7.7821**	.010	7.7805	.010	**7.9493**	.010	7.9373	.010
0.005	**8.5894**	.0050	8.5714	.0051	**8.7792**	.0050	8.7648	.0050	**8.9306**	.005	8.9087	.0050

4. Some Simulation Results

The aim of the following simulation is to obtain a picture of the power of tests based on the statistics (2) and (3) for small sample sizes covered by the Table 1. To consider power of the concerned tests in the case of distributions with various tail behavior, the sampling from normal, logistic and Cauchy distribution is employed, in each case $\mu_X = 0$, $\sigma_X = 1$. The simulations are based in each case on 10000 trials. The critical constants w_α of the statistic T_{SQ} defined in (3) are taken from the previous table, the critical constants t_α of the statistic (2) are those computed in [5]. The power better of the two considered cases is emphasized by the boldface type.

Table 2: Simulation estimates of the power.

	$m = 3, n = 7$					
	$\mu_Y = 1.5, \sigma_Y = 3$		$\mu_Y = 3.5, \sigma_Y = 4$		$\mu_Y = 7, \sigma_Y = 5$	
	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
$P(T_{SO} \geq w_\alpha \mid Normal)$	**0.057**	0.07	**0.202**	0.224	**0.533**	**0.547**
$P(T \geq t\alpha \mid Normal)$	0.032	**0.14**	0.136	**0.251**	0.435	0.539
$P(T_{SO} \geq w_\alpha \mid Logistic)$	**0.032**	0.04	**0.073**	0.086	**0.191**	0.208
$P(T \geq t\alpha \mid Logistic)$	0.017	**0.13**	0.047	**0.170**	0.137	**0.247**
$P(T_{SO} > w_a \mid Cauchy)$	**0.041**	0.06	**0.084**	0.104	**0.174**	0.197
$P(T > ta \mid Cauchy)$	0.025	**0.114**	0.060	**0.165**	0.138	**0.233**
	$m = 7, n = 8$					
	$\mu_Y = 1.5, \sigma_Y = 3$		$\mu_Y = 3.5, \sigma_Y = 4$		$\mu_2 = 7, \sigma_2 = 5$	
	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
$P(T_{SO} \geq w_\alpha \mid Normal)$	**0.448**	**0.640**	**0.745**	**0.886**	0.950	**0.987**
$P(T \geq t\alpha \mid Normal)$	0.415	0.61	0.730	0.868	**0.952**	0.984
$P(T_{SO} \geq w_\alpha \mid Logistic)$	**0.363**	**0.540**	**0.586**	**0.765**	**0.796**	**0.915**
$P(T \geq t\alpha \mid Logistic)$	0.326	0.51	0.547	0.732	0.774	0.900
$P(T_{SO} > w_a \mid Cauchy)$	**0.228**	0.32	**0.431**	0.566	**0.659**	**0.759**
$P(T > ta \mid Cauchy)$	0.219	**0.36**	0.419	**0.568**	0.643	0.757

5. Discussion and conclusions

As shown on p. 283 of [8], the bounds for the asymptotic efficiency of the test statistics (2) and (3) in the case of there considered sampled distributions do not depend on the number of sampled populations, i.e., they are the same in the two-sample case and in the multisample case. As concluded in [8], taking into account computed values of the asymptotic efficiencies, one sees that a combination of the multisample Kruskal-Wallis statistic (in the two-sample case the Wilcoxon test statistic) with the Mood test statistic appears to be a good choice when one considers symmetric distributions whose type of tail weight is unknown. However, these considerations are related to the asymptotic case, when both m and n tend to infinity. The simulation results, given in the previous section do not contradict the mentioned asymptotic results. After inspecting the Table 2 it can be said that for small sample sizes and $\alpha=0.05$ testing based on (3) is preferable to (2), but for $\alpha=0.1$ it is advisable to use the Lepage test. This suggests that the test based on (3) can be considered as a useful competitor to the Lepage test.

Acknowledgements

The research was supported by the grants VEGA 1/0077/09, APVV-SK-AT-0003-09, and APVV-RPEU-0008-06.

References

[1] Hájek, J, Šidák, Z. and Sen, P. K. *Theory of rank tests*. Academic Press, San Diego 1999.

[2] Hollander, M. and Wolfe, D. A. *Nonparametric statistical methods*. Wiley, New York 1999.

[3] Hutchison, T. P. Should we routinely test for simultaneous location and scale changes? *Ergonomics*, 45(2002), 248-251.

[4] Lepage, Y. A combination of Wilcoxon's and Ansari-Bradley's statistics. *Biometrika* 58(1971), 213-217.

[5] Lepage, Y. A table for a combined Wilcoxon Ansari-Bradley statistic. *Biometrika* 60(1973), 113-116.

[6] Podgor, M. J. and Gastwirth, J. L. On non-parametric and generalized tests for the two-sample problem with location and scale change alternatives. *Statistics in Medicine* 13(1994), 747-758.

[7] Rublík, F. The multisample version of the Lepage test. *Kybernetika* 41(2005), 713-733.

[8] Rublík, F. On the asymptotic efficiency of the multisample location-scale rank tests and their adjustment for ties. *Kybernetika* 43(2007), 279-306.