Journal of Wine Economics, Volume 2, Number 1, Spring 2007, Pages 98–102

A Note on a Test for the Sum of Ranksums*

Richard E. Quandt[a]

I. Introduction

In wine tastings, in which several tasters (judges) taste several wines, it is important to insure objectivity to the extent possible. This is usually accomplished by holding the tasting "blind," i.e., covering the bottles so that the tasters do not know which wine is in which bottle. At some agreed upon point in the proceedings, the tasters reveal what they think about the various bottles. Ideally, this revelation would take place by secret ballot, lest a taster's choices be influenced by what he or she hears another taster say. But in any event, there are two standard ways of rating the wines. The older method is to assign them "grades" on a scale of, say, up to 100 points (Parker) or up to 20 points as in the famous face-off between California wines and French Bordeaux wines in 1976 (see Ashenfelter et al., 2007). As Ashenfelter at al. point out, this has the distinct disadvantage that a judge with greater dispersion in his or her grades will have a greater influence on the average score that each wine achieves.

A preferable method for rating the wines is to rank them, i.e., rank the most favored wine "1", the second most favored wine "2", etc. The winner, that is to say the wine that is liked "on the average" best by the group, is the one that achieves the lowest rank total. Of course, there will always be a wine that has a lower rank total than the other wines (with the proviso that there may be more than one wine tied for the lowest rank total) and we need to know whether this lowest rank total could have occurred by chance, even if there were no difference among the wines. Fortunately, a statistical significance test exists for the lowest rank total, originally due to Kramer (1956) and discussed in Quandt (2006).

But there are occasions when knowing which is the most favored wine and whether its performance is statistically significant is not all that we want to know. The typical case in point is when there are two types of wines being tasted, as in the California vs. French face-off, where we want to know whether the wines of type A are together and on the whole more favored than the wines of type B. The question we would want to ask then is whether the sum of the ranksums over the subset of wines comprising type A is significantly smaller

*I am indebted to Orley Ashenfelter and Burton G. Malkiel for comments. I alone am responsible for errors.
[a] Department of Economics, Princeton University, Princeton NY 08540, email: metrics@quandt.com.

than the sum of the ranksums over the wines comprising type B. The ranksums for the wines in any subgroup are not independent of each other and the distribution of the sum of the ranksums under the null hypothesis that the two groups are identical is not obvious. In the next section we produce critical values for the sum of ranksums by employing Monte Carlo experiments.

II. Monte Carlo Experiments

In each experiment, we fix the number of wines n and the number of judges m. In addition, we fix the number of wines of each of two types as n_1 and n_2, with $n = n_1 + n_2$. Since it is completely arbitrary which of the n wines are of one type and which of the other, we shall use the convention in the computations that the first n_1 wines are of type A and the remaining n_2 wines of type B. We then generate a random ranking of the wines for each of the m tasters and compute the sum of the ranksums over the first n_1 wines, denoted by R_1 and over the last n_2 wines, denoted in turn by R_2. Since the sum of the ranksums will generally be larger when more wines are in the group, we compute the statistic

$$R = \frac{R_1 / n_1}{R_2 / n_2}$$

This experiment is then replicated 100,000 times and from the cumulative sample distribution of R we determine the critical values corresponding to the 0.05 and 0.95 levels, i.e., the critical values corresponding to the lower and upper tails of the distribution.[1] If the observed value of R is less (greater) than the lower (upper) tail critical value, the corresponding group of wines would be said to be significantly good (bad). The critical lower tail and upper tail values for a variety of n, N_1, N_2 and m values are shown in Tables 1 and 2 respectively. It is noteworthy that the lower tail critical values uniformly increase as m, the number of judges, goes up and that the upper tail critical values uniformly decline in m. Finally, in Table 3 we display the mean values for each of the sampling distributions. Three features are of interest in this table: first that the mean values decrease (almost but not quite uniformly) in m and secondly that these values increase for each value of n (the number of wines) as the number of wines in each subgroup becomes more nearly equal and, finally, that all the means are (very slightly) larger than 1.0. As one would expect from this latter fact and Tables 1 and 2, the distribution of R-values has a slightly enlarged right tail.

We use three examples to illustrate the procedure. In the 1976 face-off between American and French wines there were two tastings: four French red wines were matched

[1] If no ties are allowed, as was the case in the present sampling experiments, the ranksums are integers and there is a finite number of possible outcomes for R. Where necessary, the critical values were obtained from the cumulative sample distributions by interpolation between adjacent values.

Table 1
Lower Tail Critical Values

$m \rightarrow$	*6*	*7*	*8*	*9*	*10*	*11*	*12*
$n = 6, n_1 = 2, n_2 = 4$	0.7040	0.7264	0.7427	0.7533	0.7673	0.7791	0.7874
$n = 6, n_1 = 3, n_2 = 3$	0.7314	0.7502	0.7643	0.7758	0.7879	0.7981	0.8044
$n = 7, n_1 = 2, n_2 = 5$	0.7083	0.7293	0.7448	0.7591	0.7717	0.7825	0.7907
$n = 7, n_1 = 3, n_2 = 4$	0.7419	0.7603	0.7755	0.7868	0.7965	0.8052	0.8132
$n = 8, n_1 = 2, n_2 = 6$	0.7106	0.7313	0.7486	0.7626	0.7741	0.7849	0.7941
$n = 8, n_1 = 3, n_2 = 5$	0.7503	0.7651	0.7800	0.7934	0.8040	0.8115	0.8202
$n = 8, n_1 = 4, n_2 = 4$	0.7618	0.7800	0.7938	0.8033	0.8142	0.8212	0.8272
$n = 9, n_1 = 2, n_2 = 7$	0.7149	0.7343	0.7501	0.7645	0.7774	0.7868	0.7960
$n = 9, n_1 = 3, n_2 = 6$	0.7533	0.7714	0.7862	0.7966	0.8066	0.8159	0.8230
$n = 9, n_1 = 4, n_2 = 5$	0.7717	0.7876	0.7988	0.8114	0.8206	0.8272	0.8345
$n = 10, n_1 = 2, n_2 = 8$	0.7148	0.7363	0.7538	0.7662	0.7775	0.7885	0.7971
$n = 10, n_1 = 3, n_2 = 7$	0.7559	0.7750	0.7883	0.7995	0.8087	0.8176	0.8252
$n = 10, n_1 = 4, n_2 = 6$	0.7765	0.7917	0.8060	0.8138	0.8251	0.8325	0.8398
$n = 10, n_1 = 5, n_2 = 5$	0.7864	0.8011	0.8113	0.8226	0.8303	0.8387	0.8432
$n = 11, n_1 = 2, n_2 = 9$	0.7148	0.7377	0.7521	0.7670	0.7796	0.7892	0.7995
$n = 11, n_1 = 3, n_2 = 8$	0.7587	0.7772	0.7897	0.8021	0.8117	0.8206	0.8278
$n = 11, n_1 = 4, n_2 = 7$	0.7810	0.7956	0.8082	0.8193	0.8273	0.8357	0.8423
$n = 11, n_1 = 5, n_2 = 6$	0.7921	0.8056	0.8183	0.8274	0.8353	0.8436	0.8496
$n = 12, n_1 = 2, n_2 = 10$	0.7168	0.7376	0.7549	0.7692	0.7808	0.7889	0.7978
$n = 12, n_1 = 3, n_2 = 9$	0.7597	0.7774	0.7927	0.8032	0.8129	0.8216	0.8296
$n = 12, n_1 = 4, n_2 = 8$	0.7846	0.8001	0.8106	0.8225	0.8300	0.8386	0.8440
$n = 12, n_1 = 5, n_2 = 7$	0.7960	0.8106	0.8218	0.8309	0.8404	0.8461	0.8547
$n = 12, n_1 = 6, n_2 = 6$	0.8024	0.8162	0.8273	0.8362	0.8435	0.8515	0.8572

against six California reds in one tasting and four French Chardonnays were matched against six American ones in the second tasting. For the four French and six California red wines $R_1 = 206$ for the French and $R_2 = 399$ for the American wines, yielding $R = 0.7744$. From Table 1, for 11 tasters, the critical value is 0.8325, and the result is significant; hence on the whole the French reds beat the American wines, even though the single best wine was American. In the Chardonnay tasting, $R_1 = 233$ for the French and $R_2 = 372$ for the Americans, yielding an R-value of 0.9395, which is not significant. The result is clearly due to the fact that while four of the five best wines were American, the two worst wines were also American, one of those by an overwhelming margin.

A recent tasting (October 13, 2006) of the *Liquid Assets winetasting group* matched 1966 Bordeaux wines against the same Chateaux in 1970, with each group containing exactly the same four Chateaux (Ch. La Mission Haut Brion, Ch. Cheval Blanc,

Table 2
Upper Tail Critical Values

m	*6*	*7*	*8*	*9*	*10*	*11*	*12*
$n = 6, n_1 = 2, n_2 = 4$	1.3249	1.3003	1.2791	1.2625	1.2494	1.2370	1.2262
$n = 6, n_1 = 3, n_2 = 3$	1.3258	1.2967	1.2786	1.2601	1.2461	1.2323	1.2227
$n = 7, n_1 = 2, n_2 = 5$	1.3153	1.2904	1.2734	1.2571	1.2415	1.2316	1.2208
$n = 7, n_1 = 3, n_2 = 4$	1.2988	1.2750	1.2556	1.2398	1.2279	1.2152	1.2065
$n = 8, n_1 = 2, n_2 = 6$	1.3102	1.2849	1.2650	1.2496	1.2377	1.2277	1.2161
$n = 8, n_1 = 3, n_2 = 5$	1.2825	1.2616	1.2422	1.2307	1.2158	1.2081	1.1982
$n = 8, n_1 = 4, n_2 = 4$	1.2865	1.2615	1.2431	1.2277	1.2166	1.2035	1.1961
$n = 9, n_1 = 2, n_2 = 7$	1.3056	1.2814	1.2626	1.2467	1.2331	1.2230	1.2129
$n = 9, n_1 = 3, n_2 = 6$	1.2781	1.2549	1.2358	1.2222	1.2100	1.1994	1.1901
$n = 9, n_1 = 4, n_2 = 5$	1.2680	1.2459	1.2310	1.2143	1.2026	1.1938	1.1841
$n = 10, n_1 = 2, n_2 = 8$	1.3002	1.2763	1.2587	1.2427	1.2295	1.2217	1.2098
$n = 10, n_1 = 3, n_2 = 7$	1.2664	1.2467	1.2310	1.2161	1.2053	1.1949	1.1873
$n = 10, n_1 = 4, n_2 = 6$	1.2567	1.2356	1.2197	1.2066	1.2940	1.1860	1.1773
$n = 10, n_1 = 5, n_2 = 5$	1.2578	1.2343	1.2203	1.2058	1.1952	1.1853	1.1769
$n = 11, n_1 = 2, n_2 = 9$	1.2991	1.2736	1.2560	1.2404	1.2278	1.2173	1.2090
$n = 11, n_1 = 3, n_2 = 8$	1.2615	1.2413	1.2271	1.2133	1.2012	1.1909	1.1828
$n = 11, n_1 = 4, n_2 = 7$	1.2481	1.2298	1.2141	1.2009	1.1897	1.1789	1.1727
$n = 11, n_1 = 5, n_2 = 6$	1.2427	1.2253	1.2103	1.1964	1.1863	1.1746	1.1690
$n = 12, n_1 = 2, n_2 = 10$	1.2952	1.2711	1.2535	1.2388	1.2264	1.2161	1.2065
$n = 12, n_1 = 3, n_2 = 9$	1.2578	1.2401	1.2234	1.2092	1.1996	1.1889	1.1808
$n = 12, n_1 = 4, n_2 = 8$	1.2403	1.2240	1.2069	1.1947	1.1854	1.1761	1.1669
$n = 12, n_1 = 5, n_2 = 7$	1.2327	1.2168	1.2002	1.1880	1.1787	1.1694	1.1634
$n = 12, n_1 = 6, n_2 = 6$	1.2358	1.2171	1.2017	1.1886	1.1797	1.1693	1.1628

Ch. Latour, Ch. Palmer; see also http://www.liquidasset.com/report103.html). There were eight judges in this tasting, and computing the value of R with the 1966 wines in the numerator yields 0.8 while the critical value is 0.7938; hence the result just misses being significant. A somewhat earlier tasting (November 7, 2005), with nine judges, of five Opus One vintages against four Bordeaux first growths plus Cheval Blanc, with Opus One in the numerator, yielded an R-value of 0.6336, which is an extremely significant result in favor of the American wines.

Table 3
Mean Values of the Sample *R*s

$m \rightarrow$	*6*	*7*	*8*	*9*	*10*	*11*	*12*
$n = 6, n_1 = 2, n_2 = 4$	1.0124	1.0112	1.0093	1.0079	1.0070	1.0062	1.0059
$n = 6, n_1 = 3, n_2 = 3$	1.0161	1.0138	1.0111	1.0108	1.0090	1.0085	1.0083
$n = 7, n_1 = 2, n_2 = 5$	1.0099	1.0090	1.0070	1.0072	1.0048	1.0055	1.0048
$n = 7, n_1 = 3, n_2 = 4$	1.0122	1.0098	1.0090	1.0074	1.0072	1.0059	1.0061
$n = 8, n_1 = 2, n_2 = 6$	1.0077	1.0074	1.0066	1.0054	1.0050	1.0045	1.0043
$n = 8, n_1 = 3, n_2 = 5$	1.0097	1.0086	1.0072	1.0065	1.0062	1.0055	1.0055
$n = 8, n_1 = 4, n_2 = 4$	1.0118	1.0111	1.0098	1.0083	1.0075	1.0071	1.0063
$n = 9, n_1 = 2, n_2 = 7$	1.0081	1.0052	1.0048	1.0043	1.0037	1.0035	1.0033
$n = 9, n_1 = 3, n_2 = 6$	1.0090	1.0074	1.0062	1.0052	1.0048	1.0042	1.0035
$n = 9, n_1 = 4, n_2 = 5$	1.0100	1.0092	1.0073	1.0059	1.0067	1.0062	1.0052
$n = 10, n_1 = 2, n_2 = 8$	1.0063	1.0058	1.0043	1.0043	1.0045	1.0031	1.0028
$n = 10, n_1 = 3, n_2 = 7$	1.0071	1.0061	1.0061	1.0046	1.0042	1.0037	1.0037
$n = 10, n_1 = 4, n_2 = 6$	1.0083	1.0070	1.0061	1.0067	1.0049	1.0048	1.0038
$n = 10, n_1 = 5, n_2 = 5$	1.0103	1.0083	1.0072	1.0063	1.0061	1.0052	1.0056
$n = 11, n_1 = 2, n_2 = 9$	1.0063	1.0045	1.0043	1.0036	1.0032	1.0031	1.0035
$n = 11, n_1 = 3, n_2 = 8$	1.0062	1.0055	1.0045	1.0044	1.0043	1.0035	1.0030
$n = 11, n_1 = 4, n_2 = 7$	1.0077	1.0065	1.0056	1.0050	1.0049	1.0039	1.0038
$n = 11, n_1 = 5, n_2 = 6$	1.0083	1.0078	1.0060	1.0057	1.0051	1.0046	1.0041
$n = 12, n_1 = 2, n_2 = 10$	1.0045	1.0055	1.0038	1.0029	1.0027	1.0020	1.0027
$n = 12, n_1 = 3, n_2 = 9$	1.0048	1.0052	1.0034	1.0037	1.0033	1.0028	1.0033
$n = 12, n_1 = 4, n_2 = 8$	1.0069	1.0057	1.0048	1.0039	1.0040	1.0031	1.0029
$n = 12, n_1 = 5, n_2 = 7$	1.0068	1.0061	1.0053	1.0054	1.0048	1.0039	1.0042
$n = 12, n_1 = 6, n_2 = 6$	1.0087	1.0070	1.0063	1.0061	1.0057	1.0047	1.0043

References

Ashenfelter, O., Quandt, R.E., and Taber, G. (2007). Wine tasting epiphany: an analysis of the 1976 California vs. France tasting. In Allhoff, F. (ed.), *Wine and Philosophy*. Oxford: Blackwell Publishing. forthcoming.

Kramer, A. (1956). A quick rank test for significance in multiple comparisons. *Food Technology*, 10, 391–392.

Quandt, R.E. (2006). Measurement and inference in wine tasting. *Journal of Wine Economics*, 1, 7–30.