UNIVERSITY OF NORTH CAROLINA

Department of Statistics

Chapel Hill, N. C.

COMPARING THE MEDIANS OF TWO SYMMETRICAL POPULATIONS

by

RICHARD F. POTTHOFF

February, 1962

Contract No. AF 49(638)-213

This paper is intended primarily to provide a relatively non-theoretical discussion corresponding to the material covered in Institute of Statistics Mimeograph Series No. 315 (reference [1]). The paper contains a numerical example and attempts to serve the needs of the practitioner rather than of the theoretician.

This research was supported partially by the Mathematics Division of the Air Force Office of Scientific Research and partially by Educational Testing Service.

Institute of Statistics
Mimeo Series No. 316

COMPARING THE MEDIANS OF TWO SYMMETRICAL POPULATIONS[1]

by

Richard F. Potthoff

1. Introduction. Suppose we have two populations with density functions $g(x)$ and $h(y)$, and with medians Θ_1 and Θ_2, respectively. Suppose that both populations are symmetrical: that is, for every c, $g(\Theta_1 + c) = g(\Theta_1 - c)$ and $h(\Theta_2 + c) = h(\Theta_2 - c)$. (Note that, for each population, the mean will be the same as the median, assuming that the mean exists.) We will assume that we are given a sample of m observations $X_1, X_2, \ldots, X_m$ from the first population, and a sample of n observations $Y_1, Y_2, \ldots, Y_n$ from the second population. With no loss of generality we may assume that $m \leq n$.

In this paper we will describe a method of testing the hypothesis that the two medians (means) are equal (i.e., the hypothesis that $\Theta_1 = \Theta_2$), and we will describe how to obtain confidence bounds on $(\Theta_2 - \Theta_1)$. The test we use will be quite similar to the Wilcoxon test, and, in fact, requires the computation of the same basic statistic as the Wilcoxon test.

Our test works as follows. First we determine how many out of the mn pairs (X_i, Y_j) are such that $X_i < Y_j$, and we divide this number by mn. Call the resulting fraction $W_{m,n}$. $W_{m,n}$ will be approximately normally distributed if m and n are sufficiently large, and will have expected value $\frac{1}{2}$ if $\Theta_1 = \Theta_2$. If we are (say) making a two-tailed test at the $5^{\circ}/\text{o}$ level, our test will consist of rejecting the hypothesis $\Theta_1 = \Theta_2$ if

(1.1) $$2\sqrt{m}\ \left|W_{m,n} - \frac{1}{2}\right| > 1.96 ,$$

---

[1]This research was supported partially by the Air Force Office of Scientific Research and partially by the Educational Testing Service.

and accepting otherwise. (Remember that m is the lesser of m,n.)

The ordinary Wilcoxon test rejects the null hypothesis when

$$(1.2) \quad \sqrt{\frac{12\, mn}{m + n + 1}} \; \left| W_{m,n} - \frac{1}{2} \right| > 1.96 \qquad ,$$

if we use a two-tailed test at the 5% level. The reason why it is not correct to use this ordinary Wilcoxon test (1.2) for the problem being considered in this paper is that the test (1.2) can be used only for testing a more restrictive null hypothesis. The test (1.2) is properly used only for testing the null hypothesis that the two populations (represented by g and h) are exactly the same; if it is used (improperly) for testing the null hypothesis that two (arbitrary) symmetrical populations have equal medians, then the probability of rejection when the null hypothesis is true may be much larger than 5°/o.

If we use the test (1.1), on the other hand, we need only assume that both populations are symmetrical. The two populations need not be the same type of distribution : e.g., one population might be normal and the other rectangular. The two populations could be the same type of distribution but with different variances or scale parameters: e.g., they could be two Cauchy distributions with different scale parameters. In other words, we can use the test (1.1) to test the equality of the medians of the two symmetrical populations even though their variances are different and/or even though they are different types of distributions; the test (1.2), on the other hand, cannot be used in any of these circumstances.

The test (1.1) will be a conservative test; that is, the probability of rejecting the null hypothesis when it is true will not be exactly equal to 5°/o, but will generally be somewhat less than 5°/o. The test (1.2) would always reject the null hypothesis whenever the test (1.1) rejects, but the reverse statement is not true. In using the test (1.1), we of course have to pay a price for the liberty of working under a less restrictive model.

When the two populations represented by g and h are identical, the variance of $W_{m,n}$ will always be equal to $(m + n + 1)/12mn$. But when the two populations are arbitrary symmetrical populations (with equal medians), the variance of $W_{m,n}$ may be as high as $1/4m$. This accounts for the difference in the multiplicative factors which are used in the tests (1.2) and (1.1).

The discussion in this paper is relatively non-technical and is intended primarily to provide potential users of the test (1.1) with a guide which indicates when to use the test and how to perform the computations. Hence many of the technical details are omitted. For a more technical discussion, see [1].

A practical situation in which the problem treated by this paper could arise might be as follows. Suppose that we have available $(m + n)$ classes, and suppose that we select at random m out of these $(m + n)$ classes to receive curriculum number 1; the remaining n classes receive curriculum number 2. We assume that, for each class, some standard measure of achievement after completion of the course is available. For the i-th class under curriculum number 1, this measure would be X_i, and for the j-th class under curriculum number 2, this measure would be what we call Y_j. (In this example we are assuming that there is no concomitant variate available representing a measure of the ability of each class before starting the course, or at least we are assuming that no such variate is to be utilized in the analysis. If such a concomitant variate is to be utilized, we have a regression problem, the treatment of which is discussed in a separate paper.) Now if we are willing to assume that the X_i's and the Y_j's are both normal and have the same variance, then we can use the ordinary t-test for comparing the effects of the two curriculums. If the observations are non-normal but we are willing to assume that the two populations have the same shape (including the same variance), then we can use the ordinary Wilcoxon test (1.2). But if the X's and the Y's have different variances, and/or if they come from different types of populations, then neither the t-test nor the test (1.2) can be used, and we can employ the test (1.1) to compare the effects of the two curriculums so long as we are willing to assume that both populations are symmetrical.

2. Numerical example. We consider now a numerical example which was constructed (artificially) by drawing $m = 16$ random observations (the X_i's) from a certain rectangular distribution and $n = 20$ random observations (the Y_j's) from a certain Cauchy distribution. Obviously, such a strange combination of distributions is not likely to arise in practice, but we chose to use this combination for the example in order to emphasize the wide variety of conditions under which the test (1.1) may be used.

The X_i's (in order of increasing magnitude) are 2.09, 2.16, 2.18, 2.51, 2.82, 2.83, 2.90, 3.35, 3.68, 3.69, 3.86, 3.95, 4.08, 4.11, 4.24, and 4.61; the Y_j's are -6.82, -5.02, -4.10, -.79, .40, .94, .97, 1.08, 1.26, 1.55, 1.60, 1.63, 1.98, 2.35, 2.54, 3.11, 4.25, 5.45, 6.25, and 9.98. The experimenter is assumed to know nothing about either the median or the scale parameter of either distribution, and in fact he may not even know that the first population is rectangular and the second Cauchy.

To start, let us form a 20 x 16 chart with $mn = 320$ entries (see Table I.). Each row in the chart pertains to a Y_j, and each column to an X_i. The 320 entries represent the 320 values of $(Y_j - X_i)$.

We count up and find that, out of these 320 entries, 77 are greater than 0, while the remaining 243 are less than 0. Hence

$$(2.1) \qquad W_{m,n} = \frac{77}{320} = .241 \quad .$$

Thus

$$2\sqrt{m}\,(W_{m,n} - \tfrac{1}{2}) = 2\sqrt{16}\,(.241 - .500) = 2 \times 4 \times (-.259) = -2.07 \quad ,$$

and since -2.07 is greater in absolute value than 1.96, we reject the null hypothesis $\Theta_1 = \Theta_2$ if we are using a two-tailed test at the $5^{\circ}/\text{o}$ level.

Suppose that, instead of making a two-tailed test at the $5^{\circ}/\mathrm{o}$ level, we had wanted to make a one-tailed test of $\theta_1 = \theta_2$ against $\theta_2 < \theta_1$ at the $1^{\circ}/\mathrm{o}$ level. Then our decision rule would be to reject if

$$2 \sqrt{m} \, (W_{m,n} - \tfrac{1}{2}) \; < \; -2.33 \quad .$$

Since -2.07 is larger than -2.33, we would not quite be able to reject in the example above.

None of the 320 entries in Table I. are exactly zero; all are either positive or negative. If there had been some zero entries, we could have handled the situation by counting $\frac{1}{2}$ for each zero entry. For example, suppose there were 63 positive entries, 254 negative entries, and 3 zero entries. Then we could calculate

$$W_{m,n} = \frac{63 + 3(\frac{1}{2})}{320} = \frac{64.5}{320} = .202 \quad .$$

3. <u>Alternative calculating technique</u>. There is another possible way of calculating $W_{m,n}$ which is often used. Let the combined sample of $(m + n)$ observations be ranked, so that the highest observation receives rank 1, the second highest rank 2, and so forth, with the lowest observation receiving rank $(m + n)$. Let $R_1, R_2, \ldots, R_m$ represent the ranks associated with $X_1, X_2, \ldots, X_m$ respectively. Then

$$(3.1) \qquad W_{m,n} = \frac{1}{mn} \left[\sum_{i=1}^{m} R_i - \frac{1}{2} m (m + 1) \right] \quad .$$

In our example, the combined sample looks like this when ranked from high to low:

			4.61		4.24	4.11	4.08	3.95	3.86	3.69	3.68
9.98	6.25	5.45		4.25							

3.35		2.90	2.83	2.82		2.51		2.18	2.16	2.09	
	3.11				2.54		2.35				1.98

1.63	1.60	1.55	1.26	1.08	.97	.94	.40	-.79	-4.10	-5.02	-6.82

Hence the X_i's have ranks 4, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 19, 21, 22, and 23. The sum of these 16 ranks is 213, so that (3.1) becomes

$$W_{m,n} = \frac{1}{320}\left[213 - \frac{1}{2}(16)(16+1)\right] = \frac{213 - 136}{320} = \frac{77}{320} ,$$

which is the same as (2.1).

This second way of calculating $W_{m,n}$ (3.1) will generally be easier than the way described in Section 2. However, the method of Section 2 can be used without calculating any of the 320 entries in Table I.; it is necessary only to determine how many of these entries will be positive.

Although one reason why we calculated the 320 entries in Table I. was simply that we wanted to demonstrate what was going on, the main reason was that these 320 entries are convenient to have if we want to obtain confidence bounds on $(\Theta_2 - \Theta_1)$. Thus the technique associated with formula (3.1) is good to use if we are interested only in computing $W_{m,n}$ and testing the null hypothesis; but if we want confidence bounds, we have to resort to something else.

4. Confidence bounds. Suppose, in our example, that we want to obtain a two-sided 95°/o confidence interval for $(\Theta_2 - \Theta_1)$, the difference in the medians of the two populations. To do this, we find that value of $\delta = \Theta_2 - \Theta_1$ which, when subtracted from every entry in Table I., will cause $W_{m,n}$ to be on the threshhold of significance. Now $W_{m,n}$ will be on the threshhold of being significantly large if 238 of the 320 entries are positive and 1 entry is zero, since

$$\left(1.960 \times \frac{1}{2\sqrt{16}} + \frac{1}{2}\right) \times 320 = (.245 + .500) \times 320 = 238.4 ;$$

and it will be on the threshhold of being significantly small if 81 of the 320 entries are positive and 1 entry is zero, since

$$\left(\frac{1}{2} - 1.960 \times \frac{1}{2\sqrt{16}}\right) \times 320 = (.500 - .245) \times 320 = 81.6 \quad .$$

Upon examining Table I., we see that 238 entries are larger than -3.03 and 81 entries are smaller than -3.03 (with 1 entry being equal to -3.03). Thus, if -3.03 were subtracted from every element of Table I. (i.e., if +3.03 were added to every element of the table), then the resulting new $W_{m,n}$ would be on the threshhold of being significantly large. Hence -3.03 is the lower end of our confidence interval. To obtain the upper end of the confidence interval, we find that entry in the table which is such that 81 of the entries are larger than it and 238 are smaller; the answer turns out to be -.24. Hence, with confidence coefficient $\geq 95^{\circ}/o$, we can state that

$$-3.03 \leq (\Theta_2 - \Theta_1) \leq -.24 \quad .$$

Suppose we desire a one-sided upper $99^{\circ}/o$ confidence bound for $(\Theta_2 - \Theta_1)$. First we find that, at the $99^{\circ}/o$ level with a one-tailed test, $W_{m,n}$ would be on the threshhold of being significantly small if 66 of the 320 entries in Table I. were positive and 1 entry were zero, since

$$\left(\frac{1}{2} - 2.326 \times \frac{1}{2\sqrt{16}}\right) \times 320 = 66.96 \quad .$$

Now note that 66 of the elements of Table I. exceed .30 and 253 elements are less than .30. Hence we have as our bound

$$(\Theta_2 - \Theta_1) \leq .30 \quad ,$$

with confidence coefficient $\geq 99^{\circ}/o$.

It is not claimed that the clerical technique described here for obtaining confidence bounds is necessarily the easiest technique; it might be possible to devise a short-cut to arrive at the answer. The discussion above attempted not only to describe the mechanics of obtaining confidence bounds, but also to indicate to some extent the reasons behind the mechanics.

One final note: we never mentioned what the actual parameters were of the populations from which the X_i's and the Y_j's were drawn. In a practical situation, of course, this generally would never be known. But it should not cause any harm if we mention that, for the numerical illustration treated in this paper, the X_i's came from the rectangular population

$$g(x) = \frac{1}{3} \qquad 2 \leq x \leq 5$$
$$= 0 \qquad \text{otherwise} \qquad ,$$

while the Y_j's came from the Cauchy population

$$h(y) = \frac{1}{\pi[1 + (y - 1)^2]} \qquad .$$

Thus the true value of $(\theta_2 - \theta_1)$ was $1 - 3.5$, or -2.5 .

REFERENCE

[1] Richard F. Potthoff, "Use of the Wilcoxon Statistic for a Generalized Behrens-Fisher Problem," Institute of Statistics Mimeograph Series No. 315, University of North Carolina.

TABLE I.

Y_j \ X_i	2.09	2.16	2.18	2.51	2.82	2.83	2.90	3.35	3.68	3.69	3.86	3.95	4.08	4.11	4.24	4.61
-6.82	-8.91	-8.98	-9.00	-9.33	-9.64	-9.65	-9.72	-10.17	-10.50	-10.51	-10.68	-10.77	-10.90	-10.93	-11.06	-11.43
-5.02	-7.11	-7.18	-7.20	-7.53	-7.84	-7.85	-7.92	-8.37	-8.70	-8.71	-8.88	-8.97	-9.10	-9.13	-9.26	-9.63
-4.10	-6.19	-6.26	-6.28	-6.61	-6.92	-6.93	-7.00	-7.45	-7.78	-7.79	-7.96	-8.05	-8.18	-8.21	-8.34	-8.71
-.79	-2.88	-2.95	-2.97	-3.30	-3.61	-3.62	-3.69	-4.14	-4.47	-4.48	-4.65	-4.74	-4.87	-4.90	-5.03	-5.40
.40	-1.69	-1.76	-1.78	-2.11	-2.42	-2.43	-2.50	-2.95	-3.28	-3.29	-3.46	-3.55	-3.68	-3.71	-3.84	-4.21
.94	-1.15	-1.22	-1.24	-1.57	-1.88	-1.89	-1.96	-2.41	-2.74	-2.75	-2.92	-3.01	-3.14	-3.17	-3.30	-3.67
.97	-1.12	-1.19	-1.21	-1.54	-1.85	-1.86	-1.93	-2.38	-2.71	-2.72	-2.89	-2.98	-3.11	-3.14	-3.27	-3.64
1.08	-1.01	-1.08	-1.10	-1.43	-1.74	-1.75	-1.82	-2.27	-2.60	-2.61	-2.78	-2.87	-3.00	-3.03	-3.16	-3.53
1.26	-.83	-.90	-.92	-1.25	-1.56	-1.57	-1.64	-2.09	-2.42	-2.43	-2.60	-2.69	-2.82	-2.85	-2.98	-3.35
1.55	-.54	-.61	-.63	-.96	-1.27	-1.28	-1.35	-1.80	-2.13	-2.14	-2.31	-2.40	-2.53	-2.56	-2.69	-3.06
1.60	-.49	-.56	-.58	-.91	-1.22	-1.23	-1.30	-1.75	-2.08	-2.09	-2.26	-2.35	-2.48	-2.51	-2.64	-3.01
1.63	-.46	-.53	-.55	-.88	-1.19	-1.20	-1.27	-1.72	-2.05	-2.06	-2.23	-2.32	-2.45	-2.48	-2.61	-2.98
1.98	-.11	-.18	-.20	-.53	-.84	-.85	-.92	-1.37	-1.70	-1.71	-1.88	-1.97	-2.10	-2.13	-2.26	-2.63
2.35	.26	.19	.17	-.16	-.47	-.48	-.55	-1.00	-1.33	-1.34	-1.51	-1.60	-1.73	-1.76	-1.89	-2.26
2.54	.45	.38	.36	.03	-.28	-.29	-.36	-.81	-1.14	-1.15	-1.32	-1.41	-1.54	-1.57	-1.70	-2.07
3.11	1.02	.95	.93	.60	.29	.28	.21	-.24	-.57	-.58	-.75	-.84	-.97	-1.00	-1.13	-1.50
4.25	2.16	2.09	2.07	1.74	1.43	1.42	1.35	.90	.57	.56	.39	.30	.17	.14	.01	-.36
5.45	3.36	3.29	3.27	2.94	2.63	2.62	2.55	2.10	1.77	1.76	1.59	1.50	1.37	1.34	1.21	.84
6.25	4.16	4.09	4.07	3.74	3.43	3.42	3.35	2.90	2.57	2.56	2.39	2.30	2.17	2.14	2.01	1.64
9.98	7.89	7.82	7.80	7.47	7.16	7.15	7.08	6.63	6.30	6.29	6.12	6.03	5.90	5.87	5.74	5.37