STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

SM 87 br

Wilcoxon's two-sample test.

A R Bloemena.

1959

APPENDIX

WILCOXON'S TWO-SAMPLE TEST

by A. R. Bloemena **

If one has a group of samples from one population and a group of samples from another population, one is often faced with the question whether both populations are the same or not. For this situation several statistical tests are available, one of these being the well-known *Student's test* (cf. Dixon and Massey, 1951, chapter 9). One of the assumptions underlying Student's test is that the quantities, of which observations are available, have a normal distribution. In many cases, however, it is not known whether or not this assumption is satisfied. In these cases it is advisable to use a statistical test, not based on the assumption of normal distributions. In the problem concerned one can use, e. g., Wilcoxon's two-sample test. The assumptions underlying this test are:

a. all observations are taken at random and are independent;
b. the observations in group I are taken from the same population;
c. the observation in group II are taken from the same population.

As an example we take the following situation. A type of rock has been found in two localities; at each locality one has taken 6 samples * at random. The sodium content (in percentages) of these samples is:

locality I; 6.3; 3.9; 3.5; 10.0; 2.5; 3.4.
locality II; 5.6; 5.2; 6.0; 3.3; 1.1; 3.0.

From the observations a test-statistic is calculated. In this case the test-statistic W is found as follows. Each observation of locality I is compared with each observation of locality II. Now W is equal to twice the number of pairs of observations for which the observation of locality I is larger than the observation of locality II plus (once) the number of pairs of observations for which the observation of locality I equals the observation of locality II. The calculation of W for the example is shown in scheme I. The groups of observations are ranked according to increasing size. The third column contains for each observation at locality I the contribution to W, i. e., twice the number of observations at locality II, which are smaller (plus once the number of observations at locality II which are equal to it). Adding the contribution gives W.

Now consider the case in which both populations are in fact different and more specifically that the rock at locality I has a higher Na content. In this case one might expect a high value of W. Thus a high value of W

* In statistics the word "sample" is used in another sense, i. e., in the sense of a group of observations, taken at random from a population. In geology one would rather call this "a group of samples".

** Report 1959-37 (1) of the Statistics Department of the Mathematical Centre, Amsterdam, Holland.

may be taken as an indication of the fact that the sodium percentage at locality I is higher than at locality II. In the same way a low value of W may be taken as an indication of a lower Na content at locality I, compared with locality II.

Now statistical theory gives a mean to determine critical values of W, the so-called *upper and lower α % critical values.* If a calculated value of W lies between these critical values one has to conclude that (as in the example) there is no evidence that both sodium percentages are different.

SCHEME I

The calculation of the test-statistic W for the example quoted

observations at locality I	observations at locality II	contribution to W
	1.1	
2.5		2
	3.0	
	3.3	
3.4		6
3.5		6
3.9		6
	5.2	
	5.6	
	6.0	
6.3		12
10.0		12
m = 6	n = 6	W = 44

This does not necessarily mean that they are equal, but it only means that in the presence of large random fluctuations, a significant difference could not be established. If a calculated value of W is not *inside* the critical limits, one may reject the hypothesis that at both localities the sodium content is the same. The α % critical limits have been chosen in such a way that if in fact both populations do not differ, the probability of finding a value of W not inside the critical limits is $\leqslant \alpha$ %, so that the probability of an incorrect conclusion, i. e., the conclusion that the sodium content is different, while in fact it is not different, is smaller than or equal to α %. In many fields one chooses as a rule $\alpha = 5$ %, in rough work one might use 10 %, but where incorrect conclusions can do serious harm, α should be taken to be 1 % or smaller. In the tables critical values have been given for group sizes $\leqslant$ 12 and $\alpha = 10$ % (upper table) and $\alpha = 5$ % (lower table). For other group sizes and the use of other values for α we refer to Constance van Eeden and Rümke (1958) and to Doraline Wabeke and Constance van Eeden (1955).

For the example W = 44 has been found. It appears from the tables (n = 6, m = 6) that no significant difference between the sodium content of the rock at both localities can be established.

Remarks

1. From the tables it is clear that if only two groups of three samples were available, the conclusion that there is a difference could never be reached by means of the described test-procedure.

TABLE

Two-sided critical values of the test-statistic W for group sizes 12 and $\alpha = 10\%$ and 5% *)

	m \ n	1	2	3	4	5	6	7	8	9	10	11	12
	1	•	•	•	•	•	•	•	•	•	•	•	•
	2	•	•	•	•	0—20	0—24	0—28	2—30	2—34	2—38	2—42	4—44
	3	•	•	0—18	0—24	2—28	4—32	4—38	6—42	8—46	8—52	10—56	10—62
	4	•	•	0—24	2—30	4—36	6—42	8—48	10—54	12—60	14—66	16—72	18—78
	5	•	0—20	2—28	4—36	8—42	10—50	12—58	16—64	18—72	22—78	24—86	26—94
	6	•	0—24	4—32	6—42	10—50	14—58	16—68	20—76	24—84	28—92	32—100	34—110
$\alpha = 10\%$	7	•	0—28	4—38	8—48	12—58	16—68	22—76	26—86	30—96	34—106	38—116	42—126
	8	•	2—30	6—42	10—54	16—64	20—76	26—86	30—98	36—108	40—120	46—130	52—140
	9	•	2—34	8—46	12—60	18—72	24—84	30—96	36—108	42—120	48—132	54—144	60—156
	10	•	2—38	8—52	14—66	22—78	28—92	34—106	40—120	48—132	54—146	62—158	68—172
	11	•	2—42	10—56	16—72	24—86	32—100	38—116	46—130	54—144	62—158	68—174	76—188
	12	•	4—44	10—62	18—78	26—94	34—110	42—126	52—140	60—156	68—172	76—188	84—204
	1	•	•	•	•	•	•	•	•	•	•	•	•
	2	•	•	•	•	•	•	•	0—32	0—36	0—40	0—44	2—46
	3	•	•	•	•	0—30	2—34	2—40	4—44	4—50	6—54	6—60	8—64
	4	•	•	•	0—32	2—38	4—44	6—50	8—56	8—64	10—70	12—76	14—82
	5	•	•	0—30	2—38	4—46	6—54	10—60	12—68	14—76	16—84	18—92	22—98
	6	•	•	2—34	4—44	6—54	10—62	12—72	16—80	20—88	22—98	26—106	28—116
$\alpha = 5\%$	7	•	•	2—40	6—50	10—60	12—72	16—82	20—92	24—102	28—112	32—122	36—132
	8	•	0—32	4—44	8—56	12—68	16—80	20—92	26—102	30—114	34—126	38—138	44—148
	9	•	0—36	4—50	8—64	14—76	20—88	24—102	30—114	34—128	40—140	46—152	52—164
	10	•	0—40	6—54	10—70	16—84	22—98	28—112	34—126	40—140	46—154	52—168	58—182
	11	•	0—44	6—60	12—76	18—92	26—106	32—122	38—138	46—152	52—168	60—182	66—198
	12	•	2—46	8—64	14—82	22—98	28—116	36—132	44—148	52—164	58—182	66—198	74—214

*) • means that for the respective values of m and n, the hypothesis that both populations are equal cannot be rejected (using this test with probability level α).

2. The tables are valid only if no two observations are equal. In the case of equal observations ("ties") a correction has to be applied. If the number of tied observations is small compared to the total number, this correction is negligible.

3. The testing-procedure has been described as a two-sided one. In some cases one-sided tests can be used. We refer to the above mentioned publications for further details.

References:

DIXON, W. J. and MASSEY, F. J., 1951. "Introduction to statistical analysis". New York, Toronto, London, McGraw-Hill, 370 p., 26 tables.

EEDEN, CONSTANCE VAN and RÜMKE, CHR. L., 1958. "Wilcoxon's two-sample test". Statistica Neerlandica, **12**, p. 275—280.

WABEKE, DORALINE and EEDEN, CONSTANCE VAN, 1955. "Handleiding voor de toets van Wilcoxon". Report S 176 (M 65), Statistics Dept., Mathematical Centre, Amsterdam, 37 p., 11 tables.