STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

SM 87 ar

Wilcoxon's two sample test.

C. van Eeden en Chr. L. Ruemke.

1958

Statistica Neerlandica 12 (1958), 275—280

Receptuur

Wilcoxon's two sample test

De toets van Wilcoxon voor twee steekproeven

Wilcoxon's two sample test may be used if one wants to investigate whether two samples are drawn from the same population. We first give an example of the problem of two samples.

Two different diets A and B are to be compared concerning their influence on the weight of rats; 11 rats are given diet A and 9 rats get diet B. At a certain moment the weight of each rat is determined. The results are (in gramms)

A: 155, 154, 149, 163, 146, 150, 154, 161, 148, 145, 149,
B: 156, 163, 153, 163, 165, 156, 151, 157, 167.

On account of these observations one wants to investigate whether, in general, a difference exists between the weight of rats with diet A and the weight of rats with diet B, i.e. one wants to investigate whether the observed weights of the rats can be regarded as two samples drawn from the same population.

In general the problem may be formulated as follows.

Let $x_1, \ldots, x_m$ be m independent observations of a random variable $\underline{x}$[1]), and let $y_1, \ldots, y_n$ be n independent observations of a random variable $\underline{y}$, where $\underline{x}$ and $\underline{y}$ are distributed independently. Then the hypothesis H_0 to be tested states that the probability distributions of $\underline{x}$ and $\underline{y}$ are identical.

The test to be described is a distributionfree test, i.e. for the applicability of this test no assumptions are necessary on the form of the distributions of $\underline{x}$ and $\underline{y}$. Thus e.g. $\underline{x}$ and $\underline{y}$ need not be normally distributed.

The test statistic of the Wilcoxon two sample test will be denoted by W and is calculated from the observations as follows. Each observation of $\underline{x}$ is compared with each observation of $\underline{y}$ and W is equal to twice the number of pairs of observations, for which the observation of $\underline{x}$ is larger than the observation of $\underline{y}$ plus (once) the number of pairs of observations for which the observation of $\underline{x}$ equals the observation of $\underline{y}$[2]).

It will be clear that $\underline{W}$ assumes a small value if the observations of $\underline{x}$ are

[1]) Random variables are distinguished from numbers (e.g. from the values they take in an experiment) by underlining their symbols.

[2]) Usually the statistic $U = \frac{1}{2}W$ is used for this test (cf [1]). In order to avoid fractions the statistic W is introduced in [2].

predominantly smaller than the observations of $\underline{y}$ and a large value in the reverse situation.

The calculation of W for the example mentioned above is shown in scheme 1. [1]) In this scheme the pooled samples of $\underline{x}$ and $\underline{y}$ are ranked according to increasing size (cf. column 1 and 2). Equal observations are placed on the same line.

Column 3 contains for each observation of $\underline{x}$ the contribution to W, i.e. twice the number of observations of $\underline{y}$ which are smaller than this observation of $\underline{x}$ plus the number of observations of $\underline{y}$ which are equal to it. Addition of the numbers in column 3 gives $W = 34$.

Column 4 contains the sizes of the ties, thus for each line the number of observations in the pooled samples on that line. Addition of the numbers in column 4 gives $N = m + n = 20$.

SCHEME 1

Calculation of W

1	2	3	4	5
observations of		contribution to W	sizes of ties (t)	t^3
$\underline{x}$	$\underline{y}$			
145		0	1	1
146		0	1	1
148		0	1	1
149,149		0	2	8
150		0	1	1
	151		1	1
	153		1	1
154,154		4,4	2	8
155		4	1	1
	156,156		2	8
	157		1	1
161		10	1	1
163	163,163	12	3	27
	165		1	1
	167		1	1
$m = 11$	$n = 9$	$W = 34$	$N = 20$	$D = 62$

Column 5 contains the cubes of the sizes of the ties; their sum D is 62.

The distribution of $\underline{W}$ assuming H_0 to be true is known. This distribution is symmetric if no ties are present. $\underline{W}$ then assumes even values and varies between 0 and 2 mn. The mean μ and variance σ^2 are given by (cf. [2] and [4])

[1]) An (analogous) scheme for cases with large samples and (or) large ties may be found in [2].

$$(1) \qquad \mu = mn$$

$$(2) \qquad \sigma^2 = \frac{mn\,(N^3 - D)}{3N\,(N-1)} = \tfrac{1}{3}\,mn\,(N+1) - \frac{mn}{3N\,(N-1)}(D-N).$$

If one wants to test H_0 against the alternative hypothesis that $\underline{x}$ is systematically larger or smaller than $\underline{y}$ a twosided test is applied. The critical region of this twosided test consists of large *and* small values of W.

It may occur that one wants to test H_0 against the alternative hypothesis that $\underline{x}$ is systematically smaller than $\underline{y}$. In this case a lower onesided critical region is used consisting of small values of W. If one wants to test H_0 against the alternative hypothesis that $\underline{x}$ is systematically larger than $\underline{y}$ the upper onesided critical region is used consisting of large values of W.

Table 1 and 2 contain the lower critical values of W for the twosided test with $\alpha = 0{,}1$ respectively $\alpha = 0{,}05$ for $m + n \leqq 40$ and $m \leqq n$. The upper critical values are found by substracting the tabulated value from $2\,mn$. The sample sizes m and n are interchangeable. The tables may also be used for the onesided test with $\alpha = 0{,}05$ respectively $\alpha = 0{,}025$. These tables are taken from [2], where also tables are given of the exact tailprobabilities for $m \leqq n \leqq 10$ and a table of the critical values with $\alpha = 0{,}02$ (twosided) for $m + n \leqq 40$, $m \leqq n$ and $n \geqq 11$. Strictly speaking these tables only hold for cases without ties, but they give a reasonable approximation for cases with small ties.

If m and n are large and if moreover the difference between m and n and the differences between the sizes of the ties are not too large, $\underline{W}$ is approximately normally distributed with mean and variance according to (1) and (2). This fact may be used if $m + n > 40$ and also if $m + n \leqq 40$ and the sizes of the ties are too large to use the tables 1 and 2. In this case one calculates

$$u = \frac{|\,W - \mu\,| - 1}{\sigma}\ ^{1)}$$

for the twosided test,

$$u = \frac{W - \mu + 1}{\sigma}$$

for the lower onesided test, and

$$u = \frac{W - \mu - 1}{\sigma}$$

for the upper onesided test.

1) The term $\pm\,1$ in the numerator is the correction for continuity.

TABLE 1

Critical values for Wilcoxon's two sample test for $m + n \leqq 40$, $m \leqq n$ and $\alpha = 0{,}1$ (twosided) [1])

n \ m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	—	—	0																	
4	—	—	0	2																
5	—	0	2	4	8															
6	—	0	4	6	10	14														
7	—	0	4	8	12	16	22													
8	—	2	6	10	16	20	26	30												
9	—	2	8	12	18	24	30	36	42											
10	—	2	8	14	22	28	34	40	48	54										
11	—	2	10	16	24	32	38	46	54	62	68									
12	—	4	10	18	26	34	42	52	60	68	76	84								
13	—	4	12	20	30	38	48	56	66	74	84	94	102							
14	—	6	14	22	32	42	52	62	72	82	92	102	112	122						
15	—	6	14	24	36	46	56	66	78	88	100	110	122	132	144					
16	—	6	16	28	38	50	60	72	84	96	108	120	130	142	154	166				
17	—	6	18	30	40	52	66	78	90	102	114	128	140	152	166	178	192			
18	—	8	18	32	44	56	70	82	96	110	122	136	150	164	176	190	204	218		
19	0	8	20	34	46	60	74	88	102	116	130	144	160	174	188	202	216	232	246	
20	0	8	22	36	50	64	78	94	108	124	138	154	168	184	200	214	230	246	260	276
21	0	10	22	38	52	68	82	98	114	130	146	162	178	194	210	226	242	258	276	
22	0	10	24	40	56	72	88	104	120	136	154	170	188	204	222	238	256	272		
23	0	10	26	42	58	74	92	108	126	144	162	180	196	214	232	250	268			
24	0	12	26	44	60	78	96	114	132	150	170	188	206	224	244	262				
25	0	12	28	46	64	82	100	120	138	158	176	196	216	236	254					
26	0	12	30	48	66	86	106	124	144	164	184	204	226	246						
27	0	14	30	50	70	90	110	130	150	172	192	214	234							
28	0	14	32	52	72	92	114	136	156	178	200	222								
29	0	14	34	54	76	96	118	140	162	186	208									
30	0	14	34	56	78	100	122	146	170	192										
31	0	16	36	58	80	104	128	152	176											
32	0	16	38	60	84	108	132	156												
33	0	16	38	62	86	112	136													
34	0	18	40	64	90	114														
35	0	18	42	66	92															
36	0	18	42	68																
37	0	20	44																	
38	0	20																		
39	2																			

[1]) „—" means that, for these values of m and n, no value of W esists with a twosided tail probability $\leqq \alpha$.

TABLE 2

Critical values for Wilcoxon's two sample test for $m + n \leqq 40$, $m \leqq n$ and $\alpha = 0{,}05$ (twosided) [1])

n \ m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	—	—	—																	
4	—	—	—	0																
5	—	—	0	2	4															
6	—	—	2	4	6	10														
7	—	—	2	6	10	12	16													
8	—	0	4	8	12	16	20	26												
9	—	0	4	8	14	20	24	30	34											
10	—	0	6	10	16	22	28	34	40	46										
11	—	0	6	12	18	26	32	38	46	52	60									
12	—	2	8	14	22	28	36	44	52	58	66	74								
13	—	2	8	16	24	32	40	48	56	66	74	82	90							
14	—	2	10	18	26	34	44	52	62	72	80	90	100	110						
15	—	2	10	20	28	38	48	58	68	78	88	98	108	118	128					
16	—	2	12	22	30	42	52	62	74	84	94	106	118	128	138	150				
17	—	4	12	22	34	44	56	68	78	90	102	114	126	138	150	162	174			
18	—	4	14	24	36	48	60	72	84	96	110	122	134	146	160	172	186	198		
19	—	4	14	26	38	50	64	76	90	104	116	130	142	156	170	184	198	210	224	
20	—	4	16	28	40	54	68	82	96	110	124	138	152	166	180	194	210	224	238	254
21	—	6	16	30	44	58	72	86	100	116	130	146	160	176	190	206	222	236	252	
22	—	6	18	32	46	60	76	90	106	122	138	154	170	186	202	218	234	250		
23	—	6	18	34	48	64	80	96	112	128	144	162	178	194	212	228	246			
24	—	6	20	34	50	66	84	100	118	134	152	170	186	204	222	240				
25	—	6	20	36	54	70	88	106	124	140	158	178	196	214	232					
26	—	8	22	38	56	74	92	110	128	148	166	186	204	224						
27	—	8	22	40	58	76	96	114	134	154	174	194	214							
28	—	8	24	42	60	80	100	120	140	160	180	202								
29	—	8	26	44	64	84	104	124	144	166	188									
30	—	10	26	46	66	86	108	130	150	172										
31	—	10	28	48	68	90	112	134	156											
32	—	10	28	48	70	92	116	138												
33	—	10	30	50	74	96	120													
34	—	10	30	52	76	100														
35	—	12	32	54	78															
36	—	12	32	56																
37	—	12	34																	
38	—	12																		
39	0																			

[1]) cf. the footnote at table 1.

The tailprobabilities may then be found in a table of the normal distribution.

If H_0 is rejected one concludes that $\underline{x}$ is systematically larger than $\underline{y}$ if $W > \mu$ and that $\underline{x}$ is systematically smaller than $\underline{y}$ if $W < \mu$.

In our example we have $m = 11$, $n = 9$, $W = 34$, $D = 62$. In table 2 [1]) one finds for the lower critical value with $\alpha = 0{,}05$ (twosided) 46; thus W being smaller than 46, the twosided tailprobability is smaller than 0,05.

The approximation with the normal distribution gives (cf. the tables 1 and 2 in [4])

$$\mu = 11 \times 9 = 99,$$

and

$$\sigma^2 = 693 - \frac{8{,}684}{100}(62 - 20) = 689{,}35, \quad \sigma = 26{,}26.$$

Thus for the twosided case

$$u = \frac{|W - \mu| - 1}{\sigma} = 2{,}44.$$

which gives a twosided tailprobability of 0,015.

If e.g. $W = 250$ in a case with $m = 10$, $n = 15$ and small ties, one first calculates (W being larger than μ) the upper critical value. For $\alpha = 0{,}05$ (twosided) this critical value is (cf. table 2) 300—78 = 222. Thus $W = 250$ lies in the twosided critical region with $\alpha = 0{,}05$. More detailed data about this test may be found in [2].

References

[1] Mann, H. B. and D. R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat. **18** (1947), 50-60.

[2] Wabeke, Doraline en Constance van Eeden, Handleiding voor de toets van Wilcoxon, Report S 176 (M 65) of the Statistical Department of the Mathematical Centre, Amsterdam, 1955.

[3] Wilcoxon, F., Individual comparisons by ranking methods, Biometrics **1** (1945), 80—82.

[4] Zaalberg, J., Auxiliary tables for Wilcoxon's two sample test, Statistica Neerlandica **12** (1958), 265—273.

Constance van Eeden and Chr. L. Rümke

Statistical Dept. of the Mathematical Centre, Amsterdam
resp. Pharmacological and Neurobiochemical Lab. of the
„Vrije Universiteit", Amsterdam

1) The sample sizes m and n are interchangeable. Thus the critical values for $m = 11$, $n = 9$ are found at $m = 9$, $n = 11$.