

## Wilcoxin's Test -- Analysis of 2-Within-Group Data with a Quantitative Response Variable

**Application:** This statistic has two applications that can appear very different, but are really just two variations of the same statistical question. In one application the same quantitative variable is measured at two different times from the same sample (or from two samples that have been matched one or more important variables). In the other application, two comparable quantitative variables are measured from the same sample (usually at the same time). In both applications, Wilcoxin's test is used to compare the distributions of the two quantitative variables.

Thus, it is applied in the same data situation as a t-test or ANOVA for dependent samples except that it is used when the data are either from a too-small sample, are importantly non-normally distributed, or the measurement scale of the dependent variable is ordinal (not interval or ratio). There are two versions of the Wilcoxin test, one for small samples (i.e.,  $N < 50$ , but some suggest  $N < 10$ ), and one for large samples. It is important to remember the null hypothesis, and to differentiate it from the null for the dependent t-test.

There are two specific versions of the  $H_0$ : depending upon whether one characterizes the two conditions as representing a single population under two different circumstances (e.g., comparing treated vs. not treated or comparing two different treatments -- some consider this a representation of two different populations) or as representing comparable variables measured from a single population (as in the example below). Here are versions of the  $H_0$ : statement for each of these characterizations.

**$H_0$ :** The two populations represented by the different conditions of the have the same distribution of scores.

**To reject  $H_0$ :** is to say that the populations represented by the two conditions differ in some way, center, spread and/or shape. When the forms of the distributions from the two samples are similar (as is often the case -- check the size and symmetry of the IQR), then rejecting  $H_0$ : is interpreted to mean that one population tends to have larger scores (or a larger median) than the other.

**$H_0$ :** The population represented by the sample has the same distribution of scores on the two different measures represented by the conditions.

**To reject  $H_0$ :** is to say that the distributions of the variables are different in some way, center, spread and/or shape. When the forms of the distributions are similar (as is often the case -- check the size and symmetry of the IQR), then rejecting  $H_0$ : is interpreted to mean that one variable tends to have larger scores (or a larger median) than the other.

**The data:** In this analysis "grouping" variable is the type of animal (reptiles or fishes), and the response variable is the quality rating. From our database, we use two variables **reptgood** (reptile quality rated on a 1-10 scale) and **fishgood** (fish quality rated on a 1-10 scale). These scores are shown for the 12 stores below (reptqual, fishqual).

2,6   8,5   9,3   7,3   4,7   7,9   4,9   4,8   5,6   9,9   7,7   2,8

**Research Hypothesis:** The researcher hypothesized that a store's fish would be of higher quality than its reptiles, because of the greater difficulty obtaining and maintaining healthy reptiles.

**$H_0$ : for this analysis:** The quality ratings of reptiles and fish displayed by pet stores have the same distributions.

**Step 1** Rearrange the data so that scores from each subject are in the appropriate columns, one for each condition.

reptgood	fishgood
2	6
8	5
9	3
7	3
4	7
7	9
4	9
4	8
5	6
9	9
7	7
2	8

**Step 2** For each pair of scores, compute the difference between the scores and then find the absolute value of this difference.

reptgood	fishgood	difference	absolute difference
2	6	-4	4
8	5	3	3
9	3	6	6
7	3	4	4
4	7	-3	3
7	9	-2	2
4	9	-5	5
4	8	-4	4
5	6	-1	1
9	9	0	0
7	7	0	0
2	8	-6	6

**Step 3** Rank order the absolute differences, with the smallest absolute difference getting a value of 1. Cases with the same absolute difference each receive the average rank they would have received.

**By the way:** If an absolute difference has a value of 0, then the data from that subject is dropped from the analysis and the sample size is adjusted accordingly. Why? This statistical test is based upon comparisons of the number and direction of the inequalities in the data set (e.g., how many times reptgood > fishgood and how many times reptgood < fishgood), to determine if one variable has a different distribution than the other. Subjects for which reptgood = fishgood don't contribute to making this decision. What if there are several subjects with tied scores? The more subjects with tied scores (which are dropped), then it more difficult is to reject H0: - which should make sense, since tied scores suggest that the distributions are equivalent.

reptgood	fishgood	difference	absolute difference	rank of absolute difference
2	6	-4	4	6
8	5	3	3	3.5
9	3	6	6	9.5
7	3	4	4	6
4	7	-3	3	3.5
7	9	-2	2	2
4	9	-5	5	8
4	8	-4	4	6
5	6	-1	1	1
9	9	0	0	--
7	7	0	0	--
2	8	-6	6	9.5

**Step 4** Compute the signed rank absolute difference, as the rank of absolute difference (from 5<sup>th</sup> column) with the same sign as the difference (from 3<sup>rd</sup> column).

reptgood	fishgood	difference	absolute difference	rank of absolute difference	signed rank absolute difference
2	6	-4	4	6	-6
8	5	3	3	3.5	3.5
9	3	6	6	9.5	9.5
7	3	4	4	6	6
4	7	-3	3	3.5	-3.5
7	9	-2	2	2	-2
4	9	-5	5	8	-8
4	8	-4	4	6	-6
5	6	-1	1	1	-1
9	9	0	0	--	--
7	7	0	0	--	--
2	8	-6	6	9.5	-9.5

**Step 5** Compute the sum of the positive signed rank absolute differences (T+) and the sum of the negative signed rank absolute differences (T-).

$$T+ = 3.5 + 9.5 + 6 = 19 \quad T- = 6 + 3.5 + 2 + 8 + 6 + 1 + 9.5 = 36$$

**Step 6** The summary statistic, called **W**, is the **SMALLER** of T+ and T-

$$W = 19$$

**Step 7** Determine **N** as the number of subjects who contributed signed rank absolute difference values.

$$N = 10$$

**For small samples (N < 50 -- though some suggest N < 10):**

**Step 8** Determine the critical values of W using the table.

With N = 10, the critical value of W (p = .05) is 8

**Step 9** Compare the obtained W and the critical W values to determine whether to retain or reject the null hypothesis. **Be sure to notice that the decision rule for this statistics is "backwards" from most of the others -- smaller values of W are more likely to lead to rejecting H0: !!**

-- if the obtained W value (from Step 6) is **larger than** the critical value of W, then retain H0:

-- if the obtained W value is **less than or equal to** the critical value of W, then reject H0:

For the example data, we would decide to retain the null hypothesis, because the obtained value of W (19) is greater than the larger critical W value (10).

**For large samples (one or both groups with N > 20):**

**Step 10** With samples this large, the value of W has approaches a normal distribution, and so the null hypothesis can be tested by a Z-test.

$$Z = \frac{W - (N*(N + 1)/4)}{\sqrt{\frac{N*(N+1)*((2*N)+ 1)}{24}}} = \frac{19 - (10 * (10 + 1) / 4)}{\sqrt{\frac{10 * (10 + 1) * ((2*10) + 1)}{24}}}$$

$$= \frac{19 - 27.5}{\sqrt{2310 / 24}} = \frac{-8.5}{9.81} = -.87$$

**Step 11** Compare the obtained Z value and the critical Z value to determine whether to retain or reject the null hypothesis.

-- if the absolute value of the obtained Z is less than 1.96, then retain H0:

-- if the absolute value of the obtained Z is greater than 1.696, then reject H0:

For the example data, we would decide to retain the null hypothesis, because the absolute value of the obtained Z (.87) is less than the critical value of 1.96.

**For both small and large samples:**

**Step 12** IF you reject the null hypothesis, determine whether the data support or do not support the research hypothesis.

-- IF you reject the null hypothesis AND the condition that was hypothesized to have the larger scores does, then the research hypothesis is supported

-- IF you retain the null hypothesis OR you reject the null hypothesis BUT the condition that was hypothesized to have the larger scores actually has the smaller scores, then the research hypothesis is not supported.

**By the way:** Usually the researcher hypothesizes that there is a difference between the conditions. Sometimes, however, the research hypothesis is that there is NO difference between the conditions. If so, the research hypothesis and H0: are the same! When this is the case, retaining H0: provides support for the research hypothesis, whereas rejecting H0: provides evidence that research hypothesis is incorrect.

For the example data, we would decide that the research hypothesis is not supported, because the null hypothesis was retained.

**Step 12** Reporting the results

As for the other statistical tests, the report includes the "wordy" part and the statistical values upon which the you made your statistical decision. Be sure to tell which condition tended to have the larger scores, and whether or not this agrees with the research hypothesis. You should compute sample medians and IQR values to help describe the data,

Contrary to the research hypothesis, there was no difference between the distribution of quality ratings given to fish (Mdn = 6.00, IQR = 4.00 - 7.75) and those given to reptiles (Mdn =7.00, IQR = 5.25 - 8.75) in these stores, based on Wilcoxin test, W = 19, p > .05.

## W critical values for Wilcoxon's Test

N	.10	.05	.02	.01
5	0			
6	2			
7	3			
8	5	3	1	0
9	8	5	3	1
10	10	8	5	3
11	13	10	7	5
12	17	13	9	7
13	21	17	12	9
14	25	21	15	12
15	30	25	19	15
16	35	29	23	19
17	41	34	27	23
18	47	40	32	27
19	53	46	37	32
20	60	52	43	37
21	67	58	49	42
22	75	65	55	48
23	83	73	62	54
24	91	81	69	61
25	100	89	76	68
26	110	98	84	75
27	119	107	92	83
28	130	116	101	91
29	140	126	110	100
30	151	137	120	109
31	163	147	130	118
32	175	159	140	128
33	187	170	151	138
34	200	182	162	148
35	213	195	173	159
36	227	208	185	171
37	241	221	198	182
38	256	236	211	194
39	271	249	224	207
40	286	264	238	220
41	302	279	252	233
42	319	294	266	247
43	336	310	281	261
44	353	327	296	276
45	371	343	312	291
46	389	361	328	307
47	407	378	345	322
48	426	396	362	339
49	446	415	379	355
∞	466	434	397	373