Wilcoxon Signed Rank Test

The Mann-Whitney U Test and the Wilcoxon Signed-Rank Test are both non-parametric statistical tests used to compare two groups of data. They are similar in that they both do not require assumptions about the normality of the data, but they differ in their specific applications.

Mann-Whitney U Test:
- Used to compare two independent groups of data.
- Assumes that the two groups are drawn from the same population.
- The null hypothesis is that the variances (spread) of the two groups are equal.
- The test statistic is U, which is calculated based on the ranks of the data in the combined sample.

Wilcoxon Signed-Rank Test:
- Used to compare two paired or dependent groups of data.
- Assumes that the differences between the pairs are symmetrically distributed.
- The null hypothesis is that the median of the differences is zero.
- The test statistic is W, which is calculated based on the ranks of the absolute differences between the pairs.

Relationship between the two tests:
- The Mann-Whitney U Test is a generalization of the Wilcoxon Signed-Rank Test.
- When the two groups are paired, the Wilcoxon Signed-Rank Test can be used.
- If the pairing is not clear or appropriate, the Mann-Whitney U Test can be used instead.

For a Wilcoxon Signed-Rank Test, the hypotheses are generally formulated as follows:

- **Null Hypothesis (H_0)**: The median difference between the paired observations is zero. This means that there is no significant difference between the two related groups (e.g., before and after treatment).
- **Alternative Hypothesis (H_1)**: The median difference between the paired observations is not equal to zero. This indicates that there is a significant difference between the two related groups.

Calculating the Wilcoxon Signed-Rank Test Statistic

1. **Calculate the Differences:**
   - Subtract the post-treatment values from the pre-treatment values for each pair.
   - Record the absolute values of these differences.
2. **Rank the Differences:**
   - Rank the absolute differences from smallest to largest.
   - If there are ties, assign the average rank to each tied observation.
3. **Assign Signs to the Ranks:**
   - For each rank, assign a positive sign if the original difference was positive and a negative sign if it was negative.
4. **Sum the Ranks:**
   - Sum the ranks with positive signs.
   - Sum the ranks with negative signs.
   - The test statistic, W, is the smaller of these two sums.

Example:

Participant	Pre-app Stress Level	Post-app Stress Level	Difference	Absolute Difference	Rank	Signed Rank
1	8	5	3	3	3	3
2	7	6	1	1	1	1
3	9	4	5	5	5	5
4	6	7	-1	1	1	-1
5	10	8	2	2	2	2

- Sum of positive ranks: $T+ = 3 + 1 + 5 + 2 = 11$
- Sum of negative ranks: $T- = 1$
- $W = \min(11, 1) = 1$

Interpretation:

- The calculated W value is compared to a critical value obtained from a Wilcoxon signed-rank test table.
- If the calculated W is less than or equal to the critical value, it suggests a statistically significant difference between the pre-treatment and post-treatment values.

Use the following table to find the W critical value.

	alpha values						
n	0.001	0.005	0.01	0.025	0.05	0.10	0.20
5	--	--	--	--	--	0	2
6	--	--	--	--	0	2	3
7	--	--	--	0	2	3	5
8	--	--	0	2	3	5	8
9	--	0	1	3	5	8	10
10	--	1	3	5	8	10	14
11	0	3	5	8	10	13	17
12	1	5	7	10	13	17	21
13	2	7	9	13	17	21	26
14	4	9	12	17	21	25	31
15	6	12	15	20	25	30	36
16	8	15	19	25	29	35	42
17	11	19	23	29	34	41	48
18	14	23	27	34	40	47	55
19	18	27	32	39	46	53	62
20	21	32	37	45	52	60	69
21	25	37	42	51	58	67	77
22	30	42	48	57	65	75	86
23	35	48	54	64	73	83	94
24	40	54	61	72	81	91	104
25	45	60	68	79	89	100	113
26	51	67	75	87	98	110	124
27	57	74	83	96	107	119	134

	alpha values						
n	0.001	0.005	0.01	0.025	0.05	0.10	0.20
28	64	82	91	105	116	130	145
29	71	90	100	114	126	140	157
30	78	98	109	124	137	151	169
31	86	107	118	134	147	163	181
32	94	116	128	144	159	175	194
33	102	126	138	155	170	187	207
34	111	136	148	167	182	200	221
35	120	146	159	178	195	213	235
36	130	157	171	191	208	227	250
37	140	168	182	203	221	241	265
38	150	180	194	216	235	256	281
39	161	192	207	230	249	271	297
40	172	204	220	244	264	286	313
41	183	217	233	258	279	302	330
42	195	230	247	273	294	319	348
43	207	244	261	288	310	336	365
44	220	258	276	303	327	353	384
45	233	272	291	319	343	371	402
46	246	287	307	336	361	389	422
47	260	302	322	353	378	407	441
48	274	318	339	370	396	426	462
49	289	334	355	388	415	446	482
50	304	350	373	406	434	466	503

p-value Approach: You may use statistical software to find the p-value corresponding to the computed W. If the p-value is less than your significance level (α), reject the null hypothesis.

Example One: Weight Loss Program

A veterinary nutritionist is interested in determining whether a new 6-week diet plan is effective in helping pets lose weight. The weights of 10 pets were recorded before and after the diet plan. The data collected are shown below.

Pet	Before	After
1	85	82
2	78	75
3	92	90
4	76	76
5	84	80
6	70	68
7	88	84
8	95	94
9	77	75
10	89	86

Null Hypothesis: The difference between the paired observations is zero.

Alternative Hypothesis: The difference between the paired observations is not zero.

Pet	Before	After	Difference
1	85	82	3
2	78	75	3
3	92	90	2
4	76	76	0
5	84	80	4
6	70	68	2
7	88	84	4
8	95	94	1
9	77	75	2
10	89	86	3

Step 1: Calculate the difference between before and after value. You will subtract the “after” from the “before” values. Before the next step, we are going to remove the row where the difference is zero.

Pet	Before	After	Difference	Count	Rank
8	95	94	1	1	1
3	92	90	2	2	3
6	70	68	2	3	3
9	77	75	2	4	3
1	85	82	3	5	6
2	78	75	3	6	6
10	89	86	3	7	6
5	84	80	4	8	8.5
7	88	84	4	9	8.5

The data has been sorted by "difference" for easier ranking.

Step 2: Assign ranks to the differences. In the case of ties, average the rank values.

Notice there is a tie for 5, 6, 7 rank, so the average rank assigned to the 3 differences will be 6. There is a tie at 8 and 9, so those 4 differences are assigned a rank of 8.5.

Pet	Before	After	Difference	Count	Rank	Pos Ranks	Neg Ranks
8	95	94	1	1	1	1	0
3	92	90	2	2	3	3	
6	70	68	2	3	3	3	
9	77	75	2	4	3	3	
1	85	82	3	5	6	6	
2	78	75	3	6	6	6	
10	89	86	3	7	6	6	
5	84	80	4	8	8.5	8.5	
7	88	84	4	9	8.5	8.5	
					Rank Sums	45	0

Step 3: Designate which ranks are positive and which are negative.

Step 4: Sum the positive ranks and the negative ranks. For this particular example, there are no negative ranks so the sum of the negative ranks will be zero. The sum of the positive ranks works out to be 45.

Step 5: Calculate the Wilcoxon Test Statistic W. It is written as W = min(T+, T-) where T+ is the sum of the positive ranks and T- is the sum of the negative ranks. For our example, W = min(45, 0) = 0. Essentially, the W-test value will always work out to be the smaller of the two signed rank sums.

Step 6: Compare your W-test value to the W-critical value. For alpha = .05 and N = 9, the W-critical for this example is 5. To determine this, you will use a W table like the one shown earlier in this document. Count the number of rows (differences) you have and cross reference to the alpha value established to find W-critical.

Here's an online version of the table: https://real-statistics.com/statistics-tables/wilcoxon-signed-ranks-table/

Since the W-test value of 0 is less than the W-critical value of 5, we will reject the null hypothesis in favor of the alternative hypothesis. This indicates that the diet plan may be effective in helping these pets lose weight.

Here is the R-code to conduct the hypothesis test:

```
x1 <- c(85,78,92,76,84,70,88,95,77,89) #Before
x2 <- c(82,75,90,76,80,68,84,94,75,86) #After
wilcox.test(x2, x1, alternative = "two.sided", mu = 0, paired = TRUE, exact = FALSE, correct = TRUE)
```

Here are the results:

```
Wilcoxon signed rank test with continuity correction
        data:  x2 and x1
        V = 0, p-value = 0.00861
alternative hypothesis: true location shift is not equal to 0
```

Interpretation:

Since p-value < alpha of .05, then we will reject the null hypothesis in favor of the alternative hypothesis. The V value is the sum of the differences of the negative ranks.

Note: The message ”alternative hypothesis: true location shift is not equal to 0” is not an interpretation of the result. It is just a statement by R regarding the alternative hypothesis.

Example Two: Reaction Times

An animal psychologist is studying the reaction times of a group of animals. She measures their reaction times to a stimulus in the morning and in the evening resulting the following data:

SUBJECT	MORNING	EVENING
C	25	24
H	31	30
E	36	36
I	52	52
F	27	29
D	40	43
G	44	47
A	22	26
B	39	45
J	48	59

Null Hypothesis: The difference between the paired observations is zero.

Alternative Hypothesis: The difference between the paired observations is not zero.

SUBJECT	MORNING	EVENING	DIFF
C	25	24	1
H	31	30	1
E	36	36	0
I	52	52	0
F	27	29	-2
D	40	43	-3
G	44	47	-3
A	22	26	-4
B	39	45	-6
J	48	59	-11

Step 1: Calculate the difference between before and after value. You will subtract the "after" from the "before" values.

Before the next step, we are going to remove the rows where the difference is zero.

SUBJECT	BEFORE	AFTER	DIFF	COUNT	RANK
C	25	24	1	1	1.5
H	31	30	1	2	1.5
F	27	29	-2	3	3
D	40	43	-3	4	4.5
G	44	47	-3	5	4.5
A	22	26	-4	6	6
B	39	45	-6	7	7
J	48	59	-11	8	8

The data has been sorted by "difference" for easier ranking.

Step 2: Assign ranks to the differences. In the case of ties, average the rank values.

Notice there is a tie for 1 & 2 rank, so the average rank assigned to the 1 differences will be 1.5. There is a tie at 4 & 5, so those -3 differences are assigned a rank of 4.5.

SUBJECT	BEFORE	AFTER	DIFF	COUNT	RANK	POS RANK	NEG RANK
C	25	24	1	1	1.5	1.5	
H	31	30	1	2	1.5	1.5	
F	27	29	-2	3	3		3
D	40	43	-3	4	4.5		4.5
G	44	47	-3	5	4.5		4.5
A	22	26	-4	6	6		6
B	39	45	-6	7	7		7
J	48	59	-11	8	8		8
					RANK SUMS	3	33

Step 3: Designate which ranks are positive and which are negative.

Step 4: Sum the positive ranks and the negative ranks. For this particular example, the sum of the positive ranks is 3 and the sum of the negative ranks is 33.

Step 5: Calculate the Wilcoxon Test Statistic W. For this example, the statistic works out to be W = min(33, 3) = 3. As mentioned in the earlier example, the W-test value will always work out to be the smaller of the two signed rank sums.

Step 6: Compare your W-test value to the W-critical value. For alpha = .05 and N = 8, the W-critical is 3. To determine this, you will use a W table like the one shown earlier in this document. Count the number of rows (differences) you have and cross reference to the alpha value established to find W-critical.

Here's an online version of the table: https://real-statistics.com/statistics-tables/wilcoxon-signed-ranks-table/

Since the W-test value of 3 is less or equal to the W-critical value of 3, we will reject the null hypothesis in favor of the alternative hypothesis. This indicates that there is a difference in reaction times between morning and evening.

Here is the R-code for the “reaction timing” data set.

```
x1 <- c(25,31,36,52,27,40,44,22,39,48) #Before

x2 <- c(24,30,36,52,29,43,47,26,45,59) #After

wilcox.test(x2, x1, alternative = "two.sided", mu = 0, paired = TRUE, exact = FALSE, correct = TRUE)
```

Here are the results:

```
Wilcoxon signed rank test with continuity correction

data:  x2 and x1

V = 33, p-value = 0.04181

alternative hypothesis: true location shift is not equal to 0
```

(a statement about the Ha, not the result!)

Interpretation:

Since p-value < alpha of .05, then we will reject the null hypothesis in favor of the alternative hypothesis. The V value is the sum of the differences of the negative ranks.