

## Siegel-Tukey Test for Differences in Scale

There are a few different reasons we explore differences in scale.

Keep in mind that the scale of a dataset is basically the spread of the data. For most datasets, we're examining the variance.

Hypothesis tests comparing means vary depending on the assumption of equal variances. Thus testing that assumption requires methods to adequately test the homogeneity of variances. The F-test should come to mind as it is a common approach.

Some datasets do not lend themselves to using the F-test, which is applicable using real numbers. Some datasets gather information that is ordinal or interval data, thus we need another approach to test for differences in scale.

### A Non-Parametric Difference in Scale Test

The Siegal Turkey test is a non-parametric test for the comparison of data that is at least ordinal. It also provides a nonparametric method for interval, integer, and real data.

There are 4 steps:

Define the null and alternate hypotheses

Combine the data and arrange from smallest to largest

Assign rank order alternating smallest to largest then largest to smallest

Apply the Wilcoxon rank sum test

The key here is the ranking assignments of the data. If the spread of the two groups of data are the same then both groups will have values at the extremes of the combine set in equal proportions.

If not, then one group will end up with a small set of rank values than the other group, thus allowing the Wilcoxon rank sum test to detect the difference.

### Step 1 Define the hypothesis test

Let's examine some temperature data gathered over a week of operation.

There are two lines in our factory, line A and line B. The chamber temperature at one stage of the process is critical to the reliability of the final product.

We suspect line A has less variation in temperature than line B.

The null hypothesis is the two lines and the variation of the temperatures are the same.

$$H_0: \sigma_A^2 = \sigma_B^2$$

The alternative hypothesis is line A has a less spread of temperatures than line B.

$$H_1: \sigma_A^2 < \sigma_B^2$$

Underlying the test, we are also assuming the two groups have equivalent medians.

## Step 2 Combine and Sort the Data

The n temperature readings in Fahrenheit for line A are: 44, 71, 95, 69, 67, 82, 86

The m temperature readings for line B are: 24, 27, 79, 62, 75, 87

The median of line A is 71 and for line B it is 68.5.

The median of line A is 71 and for line B it is 68.5.

The combined 13 reading in smallest to largest order is: 24, 27, 44, 62, 67, 69, 71, 75, 79, 82, 86, 87, 95

Maintain which reading is from which line. The ordered readings are from lines A or B as follows (here a table may be helpful):

Line	B	B	A	B	A	A	A	B	B	A	A	B	A
Temp	24	27	44	62	67	69	71	75	79	82	86	87	95

## Step 3 Assign Rank Using Alternating Method

Next assign the rank of 1 to the smallest value, in this case, 24. Then assign the rank of 2 and 3 to the two largest values, in this case, 95, and 87, respectively.

Line	B	B	A	B	A	A	A	B	B	A	A	B	A
Temp	24	27	44	62	67	69	71	75	79	82	86	87	95
Rank	1											3	2

Continue assigning ranks alternating the next two smallest, then next two largest till all values have a rank assigned.

Line	B	B	A	B	A	A	A	B	B	A	A	B	A
Temp	24	27	44	62	67	69	71	75	79	82	86	87	95
Rank	1	4	5	8	9	12	13	11	10	7	6	3	2

## Step 4 Apply the Wilcoxon rank sum test

The Wilcoxon rank-sum test uses the sum of the ranks for the two groups. The rank sum for line A and B readings are:

$$W_A = 5 + 9 + 12 + 13 + 7 + 6 + 2 = 54$$

$$W_B = 1 + 4 + 8 + 11 + 10 + 3 = 37$$

If the null hypothesis is true the two sums should be about the same given the sizes of the two groups of data.

Part of the Wilcoxon rank sum test is subtracting the minimum possible rank sum from each total. The minimum value is found using

$$n(n+1)/2$$

The U statistics for this set of data are:

$$U_A = 54 - 7(7+1)/2 = 26$$

$$U_B = 37 - 6(6+1)/2 = 16$$

The critical value is found with a U-table given the  $n_1$

$$n_1 = 6$$

n2	0.1%	0.5%	1%	2.5%	5%	10%
6	—	23	24	26	28	30
7	21	24	25	27	29	32
8	22	25	27	29	31	34
9	23	26	28	31	33	36
10	24	27	29	32	35	38
11	25	28	30	34	37	40
12	25	30	32	35	38	42
13	26	31	33	37	40	44
14	26	32	34	38	42	46
15	28	33	36	40	44	48
16	29	34	37	42	46	50
17	30	36	39	43	47	52
18	31	37	40	45	49	55

19	32	38	41	46	51	57
20	33	39	43	48	53	59

The critical value with a one-sided 5% significance is 29.

Note: the paper has other tables for  $n_1$  from 1 through 20. With more than 10 samples per group, there is a normal approximation approach available.

Since the test statistic, 16 in this example, is less than the critical value, 29, we do have convincing evidence to conclude the two groups of readings have different variances.