Non-parametric Tests for Two Samples

45.2

Introduction

In Section 45.1 we look at the sign test and the Wilcoxon signed-rank test. Each of these is a one-sample test which is used for hypotheses about the location (or "average" of some sort) of a single distribution. When we looked at t-tests in HELM 41 we saw how hypotheses concerning the mean of a single normal distribution could be tested using a one-sample t-test and the means of two normal populations could be compared using a two-sample t-test. In the same way we can have a two-sample nonparametric test to compare the locations of two distributions when we are unwilling to assume that the distribution is normal or belongs to some other particular type. In this Section we will look at one such test, the Wilcoxon rank-sum test.

Prerequisites

Before starting this Section you should . . .

- be familiar with the general ideas and terms of significance tests
- be familiar with the ideas of a nonparametric test and rank-based tests as explained in Section 45.1
- be familiar with t-tests
- be familiar with the general ideas of continuous distributions

Learning Outcomes

On completion you should be able to . . .

- decide when a Wilcoxon rank-sum test may be used
- use and interpret the results of a Wilcoxon rank-sum test

1. The Wilcoxon rank-sum test

Sometimes called the Mann-Whitney test, the Wilcoxon rank-sum test may be applied to continuous distributions which have the same shape and spread but may have different means. If we take the distributions as X_1 with mean μ_1 and X_2 with mean μ_2 then the Wilcoxon rank-sum test may be used to test the null hypothesis

$$H_0 : \mu_1 = \mu_2$$

Against the alternatives

$$H_1 : \mu_1 \neq \mu_2$$

$$H_1 : \mu_1 > \mu_2$$

$$H_1 : \mu_1 < \mu_2$$

Now assume that a random sample of size n_1 is taken from population X_1 and a random sample of size n_2 is taken from population X_2. As with the Wilcoxon signed-rank test, the theory is demanding but the application is straightforward. The test procedure is as follows:

1. Arrange all of the n_1+n_2 sample members in ascending order and assign ranks to them. Equal ranks are dealt with in the usual way.

2. Find the sum of the ranks assigned to members of the smaller of the two samples and call this S_1.

3. Find the sum of the ranks assigned to members of the larger of the two samples and call this S_2. Normally, this is **not** done directly. It may be shown that

$$S_2 = \frac{(n_1+n_2)(n_1+n_2+1)}{2} - S_1$$

   and it is usual to use this relationship to find S_2 rather than do the direct calculation to save both time and effort.

4. When testing $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$, Tables 2 and 3 given at the end of this Workbook may be used directly to test at both the 5% and 1% levels of significance. Rejection of the null hypothesis occurs when *either* rank sum is *less* than the tabulated critical value.

5. In the case of one-tailed tests the same tables may be used but with these tables the levels of significance are restricted to 2.5% (from the 5% table) and 0.5% (from the 1% table). Examples given here will normally use a two-tailed test and the 5% level of significance.

6. The tables gives critical values for sample sizes $n \leq 25$. For $n > 25$ we use a normal distribution as an approximation to the distribution of the rank sum.

Example 6

The properties of a new alloy for potential use in aircraft wing construction are being investigated. If the new alloy is to replace the one in current use, it must be established that the mean axial twisting resistance of the two alloys does not differ significantly. 10 samples of each alloy are tested and the mean axial twisting resistance is measure. The results are given in the table below.

Mean Axial Twisting Resistance			
Current Alloy		New Alloy	
2224	2306	2247	2387
2340	2356	2302	2407
2410	2367	2405	2409
2389	2380	2399	2388
2402	2401	2378	2397

Use the Wilcoxon rank-sum test to decide, at the 5% level of significance, whether there is evidence of a significant difference in the mean axial twisting resistance of the two alloys.

Solution

Denoting the mean axial twisting resistance of the current alloy by μ_1 and the mean axial twisting resistance of the new alloy by μ_2, we will test the hypothesis

$$H_0 : \mu_1 = \mu_2$$

against the alternative

$$H_1 : \mu \neq \mu_2.$$

Note that in the following table the use of -c and -n to denote current and new alloys is simply a device to enable use to distinguish between the two samples for the purposes of calculation.

Data	Sorted	Ranked
2224-c	2224-c	1
2340-c	2247-n	2
2410-c	2302-n	3
2389-c	2306-c	4
2402-c	2340-c	5
2306-c	2356-c	6
2356-c	2367-c	7
2367-c	2378-n	8
2380-c	2380-c	9
2401-c	2387-n	10
2247-n	2388-n	11
2302-n	2389-c	12
2405-n	2397-n	13
2399-n	2399-n	14
2378-n	2401-c	15
2387-n	2402-c	16
2407-n	2405-n	17
2409-n	2407-n	18
2388-n	2409-n	19
2397-n	2410-c	20

Note that a spreadsheet such as Excel will sort quickly and accurately when this notation is used.

Solution (contd.)

We now calculate the sum of the ranks assigned to the current (-c) alloy. Note that in this case the choice of which sum to calculate is arbitrary since both samples are the same size. We have

$$S_C = (1 + 4 + 5 + 6 + 7 + 9 + 12 + 15 + 16 + 20) = 95$$

The sum S_N of the ranks assigned to the new alloy is calculated as follows:

$$S_N = \frac{(10 + 10)(10 + 10 + 1)}{2} - S_C = \frac{20 \times 21}{2} - 95 = 115$$

From Table 2, the critical value at the 5% level of significance corresponding to two samples each of size 10 is 78. As neither rank sum is less than (or equal to) this value we conclude that on the basis of the available evidence we cannot reject the null hypothesis at the 5% level of significance.

Now do this Task.

Task

A motorcycle engineer is investigating the resistance to stretching of two alloy steels for potential use in chains. The engineer wishes to establish in the first instance whether there is any difference in the mean resistance to stretch of the two alloys. 10 samples of one alloy and 12 samples of the second alloy are tested under the same conditions and the actual stretch is measured. All samples are the same length. The results are given in the table below.

Actual Stretch Found (mm)			
Steel-Alloy 1		Steel-Alloy 2	
2.22	2.30	2.24	2.38
2.34	2.35	2.31	2.43
2.41	2.36	2.42	2.25
2.38	2.39	2.45	2.43
2.40	2.41	2.37	2.29
		2.28	2.46

Use the Wilcoxon rank-sum test to decide, at the 5% level of significance, whether there is evidence of a significant difference in the mean resistance to stretching of the two alloys.

Your solution

Work the problem on a separate piece of paper. Record the important stages of the work here together with your conclusions.

Answer

Denoting the mean resistance to stretching of alloy 1 by μ_1 and the mean resistance to stretching of alloy 2 by μ_2, we will test the hypothesis

$$H_0 : \mu_1 = \mu_2$$

against the alternative

$$H_1 : \mu_1 \neq \mu_2.$$

Note that the use of -1 and -2 to denote the two alloys is simply a device to enable us to distinguish between the two samples for the purposes of calculation.

Data	Sorted	Ranked
2.22-1	2.22-1	1
2.34-1	2.24-2	2
2.41-1	2.25-2	3
2.38-1	2.28-2	4
2.40-1	2.29-2	5
2.30-1	2.30-1	6
2.35-1	2.31-2	7
2.36-1	2.34-1	8
2.39-1	2.35-1	9
2.41-1	2.36-1	10
2.24-2	2.37-2	11
2.31-2	2.38-1	12.5
2.42-2	2.38-2	12.5
2.45-2	2.39-1	14
2.37-2	2.40-1	15
2.28-2	2.41-1	16.5
2.38-2	2.41-1	16.5
2.43-2	2.42-2	18
2.25-2	2.43-2	19.5
2.43-2	2.43-2	19.5
2.29-2	2.45-2	21
2.46-2	2.46-2	22

We now calculate the sum S_1 of the ranks assigned to alloy 1 since this is the smaller sample. We have:

$$S_1 = (1 + 6 + 8 + 9 + 10 + 12.5 + 14 + 15 + 16.5 + 16.5) = 108.5$$

The sum S_2 of the ranks assigned to the second alloy is calculated as follows:

$$S_2 = \frac{(10 + 12)(10 + 12 + 1)}{2} - S_1 = \frac{22 \times 23}{2} - 108.5 = 144.5$$

From Table 2, the critical value at the 5% level of significance corresponding to samples of sizes 10 and 12 is 85. As neither rank sum is less than (or equal to) this value we conclude that on the basis of the available evidence we cannot reject the null hypothesis at that 5% level of significance.

General comments about the Wilcoxon rank-sum test

1. It can be shown that in cases where the underlying distribution is normal, the t-test is preferable to the Wilcoxon rank-sum test.

2. In cases where the underlying distribution in non-normal and the conditions for the t-test cannot reasonably be met, it may well be preferable to use the Wilcoxon rank-sum test.

3. In cases where the underlying distribution is symmetric but non-normal and exhibits substantially larger tails then the normal distribution, it is often preferable to use the Wilcoxon rank-sum test since the mean of such distributions is often unstable.

Example 7

A civil engineer is investigating the compressive strength of a new type of insulating block for potential use in the building of new houses.

The engineer wishes to establish whether there is any difference in the mean compressive strengths of the blocks in current usage and the proposed new block.

Ten samples of the current block and 14 samples of the new block are tested under the same conditions and their compressive strength in pounds per square inch (psi) is measured. All samples are of the standard size used in the building industry.

The results are given in the table below.

Compressive Strength (mm)			
Current Block		New Block	
2228	2301	2243	2389
2342	2354	2311	2436
2413	2366	2425	2258
2387	2398	2456	2437
2408	2417	2371	2293
		2284	2467
		2313	2324

Use the Wilcoxon rank-sum test to decide, at the 5% level of significance, whether there is evidence of a significant difference in the mean compressive strengths of the two types of insulating blocks.

Solution

Denoting the mean compressive strength of the current blocks by μ_1 and the mean compressive strength of the new blocks by μ_2, we will test the hypothesis

$H_0 : \mu_1 = \mu_2$ against the alternative $H_1 : \mu_1 \neq \mu_2$.

Note that the use of -c and -n to denote the current and new blocks is simply to device to enable us to distinguish between the two samples for the purposes of calculation.

Data	Sorted	Ranked
2228-c	2228-c	1
2342-c	2243-n	2
2413-c	2258-n	3
2387-c	2284-n	4
2408-c	2293-n	5
2301-c	2301-c	6
2354-c	2311-n	7
2366-c	2313-n	8
2398-c	2324-n	9
2417-c	2342-c	10
2243-n	2354-c	11
2311-n	2366-c	12
2425-n	2371-n	13
2456-n	2387-c	14
2371-n	2389-n	15
2284-n	2398-c	16
2313-n	2408-c	17
2389-n	2413-c	18
2436-n	2417-c	19
2258-n	2425-n	20
2437-n	2436-n	21
2293-n	2437-n	22
2467-n	2456-n	23
2324-n	2467-n	24

We now calculate the sum S_C of the ranks assigned to the blocks in current usage since this is the smallest sample. We have:

$$S_C = (1 + 6 + 10 + 11 + 12 + 14 + 16 + 17 + 18 + 19) = 124$$

The sum S_N of the ranks assigned to the new type of block is calculated as follows:

$$S_N = \frac{(10 + 14)(10 + 14 + 1)}{2} - S_C = \frac{24 \times 25}{2} - 124 = 176$$

From Table 2, the critical value at the 5% level of significance corresponding to samples of sizes 10 and 14 is 91. As neither rank sum is less than (or equal to) this value we conclude that on the basis of the available evidence we cannot reject the null hypothesis at the 5% level of significance.

Task

The breaking strengths of cables made with two different compounds are compared. Standard lengths of ten samples using compound A and twelve using compound B are tested. The breaking strengths in newtons are as follows.

Compound A				Compound B			
10854	11627	10000	11632	11000	10856	10245	9157
9106	10051	13720	11222	11072	9540	11000	10959
10325	10001			8851	11513	10030	11197

Use a Wilcoxon rank-sum test to test the null hypothesis that the mean breaking strengths for the two compounds are the same against the two-sided alternative. Use the 5% level of significance.

Your solution

Answer

The data and their ranks are as follows.

Data		Sorted		
Strength	Compound	Strength	Compound	Rank
10854	A	8851	B	1
11627	A	9106	A	2
10000	A	9157	B	3
11632	A	9540	B	4
9106	A	10000	A	5
10051	A	10001	A	6
13720	A	10030	B	7
11222	A	10051	A	8
10325	A	10245	B	9
10001	A	10325	A	10
11000	B	10854	A	11
10856	B	10856	B	12
10245	B	10959	B	13
9157	B	11000	B	14.5
11072	B	11000	B	14.5
9540	B	11072	B	16
11000	B	11197	B	17
10959	B	11222	A	18
8851	B	11513	B	19
11513	B	11627	A	20
10030	B	11632	A	21
11197	B	13720	A	22

The sum of the ranks for Compound A is 123. The sum of the ranks for Compound B is

$$\frac{22 \times 23}{2} - 123 = 130.$$

From Table 2 we see that the critical value at the 5% level for a two-tailed test is 85. Neither rank sum is less than this so we do not reject the null hypothesis. There is no significant evidence of a difference in mean breaking strength between cables made with the two compounds.

Exercises

1. The lifetimes of plastic clips with two different designs are compared by subjecting clips to continuous flexing until they break. Twelve of each design are tested. The lifetimes in hours are as follows.

Design A				Design B			
36.1	16.6	24.6	38.5	62.5	28.2	19.9	33.9
15.6	28.3	16.0	44.7	13.3	39.4	19.3	23.7
14.3	10.8	0.7	6.5	12.7	122.0	168.0	55.0

   Use a Wilcoxon rank-sum test to test the null hypothesis that the mean lifetimes are equal for the two designs against the alternative that they are not. Use the 5% level of significance. Comment on any assumptions which are necessary.

2. An experiment is conducted to test whether the installation of cavity-wall insulation reduces the amount of energy consumed in houses. Out of twenty otherwise similar houses on a housing estate, ten are selected at random for insulation. The total energy consumption over a winter is measured for each house. The data, in mwh, are as follows.

Without insulation					With insulation				
12.6	11.8	12.8	11.4	14.4	10.8	9.9	9.5	10.0	10.4
12.3	11.5	13.2	11.0	11.8	10.7	11.8	7.5	11.8	10.1

   Use a Wilcoxon rank-sum test to test the null hypothesis that the insulation has no effect against the alternative that it reduces energy consumption. Use the 1% level of significance.

Answers

1. The data, sorted into ascending order within each design, and their ranks are as follows.

	Design A				Design B			
Obs.	0.7	6.5	10.8	14.3	12.7	13.3	19.3	19.9
Rank	1	2	3	6	4	5	10	11
Obs.	15.6	16.0	16.6	24.6	23.7	28.2	33.9	39.4
Rank	7	8	9	13	12	14	16	19
Obs.	28.3	36.1	38.5	44.7	55.0	62.5	122.0	168.0
Rank	15	17	18	20	21	22	23	24

The rank sum for design A is 119 and the rank sum for design B is

$$\frac{24 \times 25}{2} - 119 = 181.$$

Table 2 shows that the critical value for a two-sided test at the 5% level of significance is 115. Neither rank sum is less than 115 so we do not reject the null hypothesis. There is no significant evidence of a difference in the mean lifetimes between the designs.

Comment: We assume that the two distributions have the same shape and spread. It may be that the spread in this case would increase with the mean but this could be corrected by application of a transformation such as taking logs and this would not affect the ranks and so would have no effect on the test outcome. In fact it is sufficient to assume that the two distributions would be the same under the null hypothesis and this seems reasonable in this case.

2. The data, sorted into ascending order within each group, and their ranks are as follows.

	Without insulation					With insulation				
Obs.	11.0	11.4	11.5	11.8	11.8	7.5	9.5	9.9	10.0	10.1
Rank	9.0	10.0	11.0	13.5	13.5	1.0	2.0	3.0	4.0	5.0
Obs.	12.3	12.6	12.8	13.2	14.4	10.4	10.7	10.8	11.8	11.8
Rank	16.0	17.0	18.0	19.0	20.0	6.0	7.0	8.0	13.5	13.5

The rank sum for houses without insulation is 147. The rank sum for houses with insulation is

$$\frac{20 \times 21}{2} - 147 = 63.$$

From Table 3 we see that the critical value for a two-sided test at the 1% level is 71. The rank sum for the houses with insulation is 63. This is less than 71 so our result is significant at the 1% level in a two-tailed test and therefore significant at the 0.5% level in a one-tailed test. The table given does not give one-sided 1% critical values but, because the result is significant at the 0.5% level, we can deduce that it is significant at the 1% level. Therefore we reject the null hypothesis and conclude that the insulation does reduce energy consumption.

Table 1

Critical values for the Wilcoxon signed-rank test

n \ α	0.10	0.05	0.02	0.01	Two − tailed tests
	0.05	0.025	0.01	0.005	One − tailed tests
4					
5	0				
6	2	0			
7	3	2	0		
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	

For $n > 25$ the rank sum has an approximately normal distribution with mean $M = \dfrac{1}{4}n(n+1)$ and standard deviation $s = \sqrt{n(n+1)(2n+1)/24}$.

Table 2

Critical Values for the Wilcoxon Rank-Sum Test (5% Two-tail Values)

$n_2 \backslash n_1$	4	5	6	7	8	9	10	11	12	13	14	15
4	10											
5	11	17										
6	12	18	26									
7	13	20	27	36								
8	14	21	29	38	49							
9	15	22	31	40	51	63						
10	15	23	32	42	53	65	78					
11	16	24	34	44	55	68	81	96				
12	17	26	35	46	58	71	85	99	115			
13	18	27	37	48	60	73	88	103	119	137		
14	19	28	38	50	63	76	91	106	123	141	160	
15	20	29	40	52	65	79	94	110	127	145	164	185
16	21	31	42	54	67	82	97	114	131	150	169	
17	21	32	43	56	70	84	100	117	135	154		
18	22	33	45	58	72	87	103	121	139			
19	23	34	46	60	74	90	107	124				
20	24	35	48	62	77	93	110					
21	25	37	50	64	79	95						
22	26	38	51	66	82							
23	27	39	53	68								
24	28	40	55									
25	28	42										
26	29											
27												
28												

Table 3

Critical Values for the Wilcoxon Rank-Sum Test (1% Two-tail Values)

$n_2 \backslash n_1$	4	5	6	7	8	9	10	11	12	13	14	15
5		15										
6	10	16	23									
7	10	17	24	32								
8	11	17	25	34	43							
9	11	18	26	35	45	56						
10	12	19	27	37	47	68	71					
11	12	20	28	38	49	61	74	87				
12	13	21	30	40	51	63	76	90	106			
13	14	22	31	41	53	65	79	93	109	125		
14	14	22	32	43	54	67	81	96	112	129	147	
15	15	23	33	44	56	70	84	99	115	133	151	171
16	15	24	34	46	58	72	86	102	119	137	155	
17	16	25	36	47	60	74	89	105	122	140		
18	16	26	37	49	62	76	92	108	125			
19	17	27	38	50	64	78	94	111				
20	18	28	39	52	66	81	97					
21	18	29	40	53	68	83						
22	19	29	42	55	70							
23	19	30	43	57								
24	20	31	44									
25	20	32										
26	21											
27												
28												