More Nonparametric Methods

December 4, 2019

Wilcoxon Signed-Rank Test

The **Wilcoxon signed-rank test** is the nonparametric alternative to the paired t-test.

- Requires numeric data.
- The t-test requires that the differences are normally distributed.
- The Wilcoxon signed-rank test does not have these same requirements.

Example

A manufacturing firm is examining task completion times for two production methods.

Worker	Method 1	Method 2	Difference
1	10.2	9.5	0.7
2	9.6	9.8	-0.2
3	9.2	8.8	0.4
4	10.6	10.1	0.5
5	9.9	10.3	-0.4
6	10.2	9.3	0.9
7	10.6	10.5	0.1
8	10.0	10.0	0
9	11.2	10.6	0.6
10	10.7	10.2	0.5
11	10.6	9.8	0.8

Wilcoxon Signed-Rank Test

1. Take the difference.
2. Take the absolute value of the difference.
3. Rank the absolute values of the differences.
4. Reapply the signs from the differences to these ranks.
5. Sum the signed ranks.

Example

Find the sum of the signed ranks for the manufacturing firm data.

Wilcoxon Signed-Rank Test

Let T denote the sum of the signed-rank values.

For data with at least 10 pairs, T is well-approximated by

$$N\left(\mu = 0, \sigma = \sqrt{\frac{n(n+1)(2n+1)}{6}}\right)$$

where n is the number of pairs with a nonzero difference.

Example

Find the normal distribution that approximates T for the manufacturer data.

Wilcoxon Signed-Rank Test

Then the test statistic z is

$$z = \frac{T - \mu}{\sigma}$$

Find the test statistic for the manufacturer data.

The Mann-Whitney-Wilcoxon Test

- Now we want to examine the difference between two *unpaired* samples.
- All of these nonparametric tests require only independent, random samples.
- (No distributional assumptions.)
- This is the nonparametric analog to the two-sample t-test.

The Mann-Whitney-Wilcoxon Test

The MWW Test examines

$$H_0: \text{ The two populations have the same distribution.}$$
$$H_A: \text{ The two populations do not have the same distribution.}$$

Example

- Most of the students attending a particular high school came from one of two middle schools.
- The administration wants to know if the academic potential is the same between the students from these two schools.

Example

Class standings of a random sample of 9 students are

Garfield Students

Student	Class Standing
Fields	8
Clark	52
Jones	112
Tibbs	21

Mulberry Students

Student	Class Standing
Hart	70
Phipps	202
Kirkwood	144
Abbott	175
Guest	146

To perform a MWW test,

1. Rank the combined data.
2. Split the data back up and sum the ranks.

The MWW Test

For small samples ($n_1, n_2 < 10$), critical values for the MWW Test come from a table.

- T_L, the lower tail of the rejection region, comes directly from the table.
- T_U, the upper tail, is calculated as

$$T_U = n_1(n_1 + n_2 + 1) - T_L$$

  - n_1 is the sample whose rank sum is being used in the test.

The MWW Test: Tables

		n_2								
		2	3	4	5	6	7	8	9	10
	2	3	3	3	3	3	3	4	4	4
	3	6	6	6	7	8	8	9	9	10
	4	10	10	11	12	13	14	15	15	16
	5	15	16	17	18	19	21	22	23	24
	6	21	23	24	25	27	28	30	32	33
n_1	7	28	30	32	34	35	37	39	41	43
	8	37	39	41	43	45	47	50	52	54
	9	46	48	50	53	56	58	61	63	66
	10	56	59	61	64	67	70	73	76	79

We reject H_0 if $T < T_L$ or $T > T_U$.

Large Sample Case

...but we prefer not to use tables or unfamiliar distributions!

If $n_1, n_2 \geq 10$, we use a normal approximation with

$$\mu = \frac{1}{2} n_1 (n_1 + n_2 + 1)$$

$$\sigma = \sqrt{\frac{1}{12} n_1 n_2 (n_1 + n_2 + 1)}$$

Example

- The Third National Bank has two branch offices.
- Random samples of account balances taken from each branch.
- Do the data indicate that the account balances are distributed differentially between the two branches?

Example

Branch 1	
Account	Balance
1	1095
2	995
3	1200
4	1195
5	925
6	950
7	805
8	945
9	875
10	1055
11	1025
12	975

Branch 2	
Account	Balance
1	885
2	850
3	915
4	950
5	800
6	750
7	865
8	1000
9	1050
10	935

Example

Branch 1			Branch 2		
Account	Balance	Rank	Account	Balance	Rank
1	1095	20	1	885	7
2	995	14	2	850	4
3	1200	22	3	915	8
4	1195	21	4	950	12.5
5	925	9	5	800	2
6	950	12.5	6	750	1
7	805	3	7	865	5
8	945	11	8	1000	16
9	875	6	9	1050	18
10	1055	19	10	935	10
11	1025	17			
12	975	15			