

Non-parametric tests I

Objectives

- Mann-Whitney
- Wilcoxon Signed Rank
- Relation of Parametric to Non-parametric tests

the problem

Our testing procedures thus far have relied on assumptions of independence, equal variance, and normality of the data. Although the tests used are robust to departures from these assumptions (especially with larger sample sizes and balanced data), we may find some assumptions are too badly violated to use these procedures. Here we begin studying tests that do not require the assumption of normality.

Non-parametric tests

- Parametric: involving an assumption about the underlying distribution of the data
- Non-parametric: (a.k.a. “distribution-free”) not requiring an assumption of the data’s distribution. *NOT* however, assumption free in general.

What will we learn?

Equivalent to:

Mann-Whitney	Two-sample t-test
--------------	----------------------

Wilcoxon Signed Rank	Paired sample t-test
-------------------------	-------------------------

Kruskall- Wallis	One-way ANOVA
---------------------	------------------

Two-sample tests

Recall the null and alternative hypothesis for a two-tailed test in a two-sample case:

- $H_0 : \mu_1 = \mu_2$
- $H_A : \mu_1 \neq \mu_2$

The two-sample t-test requires independence, equal variance and normality. What if our data are not normally distributed?

Mann-Whitney test

The Mann-Whitney test (a.k.a the Wilcoxon Rank Sum Test) requires independent data with equal variances. To conduct this (or any other statistical test) we need:

- test statistic
- critical value from null distribution

M.W. test statistic

To find the M.W. test statistics U and U' :

1. Rank data and separate into two groups

$$2. U = n_1n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$3. U' = n_1n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

Where: $R_i = \sum$ ranks in group i and $n_1 \leq n_2$

NB: $U = n_1n_2 - U'$ and $U' = n_1n_2 - U$ Also:
 $R_1 + R_2 = \frac{N(N+1)}{2}$ where $N = n_1 + n_2$ is a useful check.

M.W. critical value

When you've found U and U' , take the larger of the two and compare it to U_{α, n_1, n_2} (this is where $n_1 < n_2$ is relevant) found in many introductory statistics texts (e.g. Zar Table B.11).

If the larger value of U and U' is greater than U_{α, n_1, n_2} , reject the null hypothesis at the α -level. If not, fail to reject the null hypothesis.

Example

Goal: determine whether the feather density (# feathers/sq. inch) of the Adelie penguin is the same as the Chinstrap penguin.

We have data from 7 and 8 penguins randomly selected from each species.

Adelie	Rank	Chinstrap	Rank
87.0		87.2	
82.5		92.5	
78.7		81.5	
81.3		78.8	
80.5		86.2	
74.5		85.3	
77.8		87.1	
		89.0	

Our hypotheses are:

H_0 : Adelie density = Chinstrap density

H_A : Adelie density \neq Chinstrap density

Why the bizarre hypothesis language?

The hypotheses above have been framed in words rather than in terms of parameters because we're considering a shift in the location of the distributions, rather than inference particularly about the means.

This does not mean we cannot say that one population will be larger than the other more often than not, however, which is often what we want to know.

Example

Adelie	Rank	Chinstrap	Rank
87.0	5	87.2	3
82.5	8	92.5	1
78.7	13	81.5	9
81.3	10	78.8	12
80.5	11	86.2	6
74.5	15	85.3	7
77.8	14	87.1	4
		89.0	2
sum	76		44

$$U = 7 * 8 + \frac{7(7 + 1)}{2} - 76 = 8$$

$$U' = 8 * 7 + \frac{8(8 + 1)}{2} - 44 = 48$$

$$U_{.05(2),7,8} = 45$$

R code for critical value

```
qwilcox(.975,7,8,lower.tail = T,log.p=F)
```

Example

Because the larger of our test statistics: $U' = 48$ is larger than our critical value $U_{.05(2),7,8} = 45$ we reject the null hypothesis.

This leads to a few questions:

- Why does this work?
- How do we deal with tied ranks?
- What about one-tailed hypotheses?
- Does it matter if ranking is large to small or small to large?

Why does Mann-Whitney work?

If neither distribution tends to produce greater values than the other, i.e. if the null hypothesis is true, *and* the variances are equal, then we would expect the rankings to be randomly intermingled between the two samples.

If one distribution *does* tend to produce greater values than the other, then one sample will usually have higher rankings than the other, in which case one of U or U' will be a large value.

How do we deal with tied ranks?

When multiple observations have the same rank, give them the appropriate average.

Example: The 3rd, 4th and 5th observations are all equal. They get assigned $(3+4+5)/3 = 4$. The *next* observation gets assigned rank 6.

Example: The 7th and 8th observations are equal. They get assigned $(7 + 8)/2 = 7.5$. The next observation gets assigned rank 9.

See Zar example 8.14 (p.150) for a full example of this.

One-tailed hypotheses

In a two-tailed test, we need to consider both U and U' . In a one-tailed test, only one of these will be relevant. Which one we pick, however, depends on whether we rank the largest observation as 1 or the smallest as 1. Zar's table 8.2 (p.149) gives us a helpful summary of how to decide which one we want. (Here, G1 refers to Group 1.)

	$H_0: G1 \geq G2$	$H_0: G1 \leq G2$
	$H_A: G1 < G2$	$H_A: G1 > G2$
Rank 1 small	U	U'
Rank 1 large	U'	U

Note that this still assumes $n_1 \leq n_2$!

MW: Direction of ranking

For a two-tailed test, it doesn't matter if ranking is smallest to largest or largest to smallest, because you consider both U and U' .

For a one-tailed test, you need to pay attention to which test statistic you want to consider, which will depend on the direction of ranking.

What about paired samples?

Recall the paired-sample t-test:

- run a 1-sample test on the differences
- requires a natural pairing of the data

Wilcoxon Signed Rank test

The Wilcoxon signed rank test requires independent data with equal variances. We also have an assumption that the data are symmetric about the median. To conduct this we need:

- test statistic
- critical value from null distribution
- paired sampling of the data

WSR test statistic

1. calculate $d_j = x_j - y_j$
2. rank $|d_j|$ (1=smallest, ties are handled as in Mann-Whitney)
3. gives signs to the ranks (“+” if d_j is positive, “-” if d_j is negative)
4. Calculate

$$T_+ = \sum (+) \text{ ranks}$$

$$T_- = \sum (-) \text{ ranks}$$

NB: T_- is the sum of the ranks, not the signed ranks, so it is a positive number.

WSR critical value

For two-tailed tests we reject H_0 that the measurements from the two populations are the same if *either* T_+ or T_- is *smaller* than the critical $T_{\alpha(2),n}$ value found in Zar table B.12.

As with the Mann-Whitney, only use one of T_+ or T_- in one-tailed tests. (NB: ranking is small to large.):

H_0 : Obs in pop 1 \leq Obs in pop 2

H_A : Obs in pop 1 $>$ Obs in pop 2

Reject H_0 if $T_- \leq T_{\alpha(1),n}$.

H_0 : Obs in pop 1 \geq Obs in pop 2

H_A : Obs in pop 1 $<$ Obs in pop 2

Reject H_0 if $T_+ \leq T_{\alpha(1),n}$.

Why does WSR work?

If the two populations are the same, we would expect the differences to be roughly split between positive and negative, in which case $T_+ \approx T_-$ and neither one will be very small.

When one population is different from the other, this will show up as either many positive or many negative differences, leading to very few of the other, and hence a small value for T_+ or T_- .

Wilcoxon Signed Rank - Example Let's revisit the boys shoe data from homework 4 and use the Wilcoxon Signed Rank test.

A	B	Diff	Rank $ d $	Sign
13.2	14	-0.8	9	-
8.2	8.8	-0.6	8	-
10.9	11.2	-0.3	4	-
14.3	14.2	0.1	1	+
10.7	11.8	-1.1	10	-
6.6	6.4	0.2	2	+
9.5	9.8	-0.3	4	-
10.8	11.3	-0.5	6.5	-
8.8	9.3	-0.5	6.5	-
13.3	13.6	-0.3	4	-

$$T_+ = 1 + 2 = 3$$

$$T_- = 9 + 8 + 4 + 10 + 4 + 6.5 + 6.5 + 4 = 52$$

Example

For a two-tailed test, we compare the smaller of these values to the critical $T_{\alpha(2),10}$ value. (Reject if *smaller* than the critical value in this case!)

For a one-tailed null that thicknesses from A are $\leq B$, we compare T_- to $T_{\alpha(1),10}$ and reject if T_- is smaller.

For a one-tailed null that thicknesses from A are $\geq B$, we compare T_+ to $T_{\alpha(1),10}$ and reject if T_+ is smaller.

For $\alpha = .05$, $T_{.05(2),10} = 8$ and $T_{.05(1),10} = 10$. We would reject the two-tailed null, as well as the second null given above.