

# Nonparametric Methods

## STA 102: Introduction to Biostatistics

Yue Jiang

The following material was used by Yue Jiang during a live lecture.

Without the accompanying oral comments, the text is incomplete as a record of the presentation.

# Why nonparametric methods?

For the methods we have studied so far, we have assumed the populations from which the data were drawn were either normally distributed or approximately so. This is a necessary property for the tests to be valid. Because the distributional form is assumed known, with only values of  $\mu$  and  $\sigma$  unknown, these methods are known as *parametric* methods.

*Nonparametric methods* make fewer assumptions about the nature of the underlying distributions and may be appropriate when some assumptions of parametric methods are not satisfied.

# Sign test

The **sign test** is a nonparametric alternative to the paired t-test. It does not assume normality but just requires that observations are independent. The null hypothesis of the sign test is that the median difference among pairs in the underlying population is 0.

## Sign test: cold data

A medical researcher claims that a new vaccine will decrease the number of colds in adults. You randomly select 14 adults and record the number of colds each has in a one-year period. After giving the vaccine to each adult, you again record the number of colds each has in a one-year period. At  $\alpha = 0.05$ , do the data support the researcher's claim?

Year	Number of Colds													
Before Vaccine	3	4	2	1	3	6	4	5	2	0	2	5	3	3
After Vaccine	2	1	0	3	1	3	3	2	2	2	3	4	3	2

# Sign test

The sign test has just a few simple steps:

1. Take the difference of each pair of observations (e.g., before-after)
2. Record the sign of each difference (+, -, or 0)
3. Count the number of + signs
4. Test using the Binomial distribution with  $n$  = number of pairs with nonzero differences and  $p = 0.5$ , as we would expect the same number of positive and negative signs under  $H_0$

# Sign test

The differences do NOT look very normal and we have a relatively small sample size. So, we cannot use previous methods to solve this question (and must rely on the sign test)!

Year	Number of Colds													
Before	3	4	2	1	3	6	4	5	2	0	2	5	3	3
After	2	1	0	3	1	3	3	2	2	2	3	4	3	2
Difference	1	3	2	-2	2	3	1	3	0	-2	-1	1	0	1
Sign	+	+	+	-	+	+	+	+	0	-	-	+	0	+

$n = 12$  nonzero differences, with 9 + signs

## Sign test for cold data

Remember, we had 9 successes out of 12 trials. What is the probability we see 9 successes out of 12 trials in a binomial distribution (or more extreme), if the null hypothesis that  $p = 0.5$  is true?

## Conducting the sign test

We can calculate, for  $X \sim \text{Binom}(12, 0.5)$ ,

$$P(X = 9) + P(X = 10) + P(X = 11) + P(X = 12)$$

and then double it (for the two-sided test).

## Conducting the sign test

We fail to reject  $H_0$  and hence conclude that there is no evidence to suggest there is a difference in the median number of colds in the years before and after vaccine administration.

# Wilcoxon signed-rank test

The sign test is appealing because it avoids all distributional assumptions. However, it ignores the magnitude of the differences.

If we are willing to assume that the differences are symmetrically distributed around the median, we can incorporate the magnitude of the differences and gain considerable power using the **Wilcoxon Signed-Rank Test**.

The Wilcoxon signed-rank test is an alternative to the paired t-test. The  $H_0$  is that the median difference in the underlying population is 0.

# Rank-based methods

Many non-parametric methods have a similar flavor to parametric methods but are computed on the ranks instead of the observed data.

The *rank* of an observation, among a set of observations, is its position when the observations are ordered from smallest to largest. The smallest observation has rank 1, next smallest has rank 2, and so forth. If observations are tied, the rank assigned to each is the average of the ranks appropriate to the equal numbers.

# Why ranks?

Using the ranks provides robustness against outliers:

- ▶ Weights: 175, 169, 190
  - ▶ Ranks: 2, 1, 3
- ▶ Weights: 175, 169, 490
  - ▶ Ranks: 2, 1, 3 (robust to outlier)
- ▶ Weights: 175, 175, 190
  - ▶ Ranks: 1.5, 1.5, 3 (take sum of ranks and divide by number tied)

# Wilcoxon signed-rank test

Like the sign test, the **Wilcoxon signed-rank test** is for *paired* differences and has just a few simple steps:

1. Calculate the difference for each pair of observations
2. Rank the absolute values of these differences from smallest to largest (drop 0's; assign an average rank to ties); let  $n$  be the number of non-zero differences
3. Assign each rank a  $+$  or  $-$  depending on the original sign of the difference
4. Add up all positive ranks; add up all negative ranks; let  $T$  be equal to the smaller sum

# Wilcoxon signed-rank test

Calculate the z-score

$$z_T = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

and obtain a p-value from the normal distribution.

Don't do this by hand – we will use R.

## Example: middle ear effusion

A common symptom of otitis media in young children is the prolonged presence of fluid in the middle ear, known as middle-ear effusion. The presence of fluid may result in temporary hearing loss and interfere with normal learning skills in the first 2 years of life.

One hypothesis is that babies who are breastfed for at least one month build some immunity and have less prolonged effusion than their bottle-fed counterparts. A small study of 14 babies is set up, with babies matched one-to-one by age, gender, socioeconomic status, and health condition. One member of the pair is a breastfed baby, and the other member is a bottle-fed baby. The outcome is the duration (days) of middle-ear effusion after the first episode of otitis media.

## Example: middle ear effusion

Duration of Effusion for Each Pair (Days):

Pair Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Breastfed	26	3	12	28	7	39	12	30	7	15	65	10	19	11
Bottle-fed	18	7	6	33	7	57	29	28	8	27	78	17	16	35

What is the  $H_0$ ?

## Example: middle ear effusion

### Duration of Effusion for Each Pair (Days):

Pair Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Breastfed	26	3	12	28	7	39	12	30	7	15	65	10	19	11
Bottle-fed	18	7	6	33	7	57	29	28	8	27	78	17	16	35
Difference	8	-4	6	-5	0	-18	-17	2	-1	-12	-13	-7	3	-24
—Difference—	8	4	6	5	0	18	17	2	1	12	13	7	3	24
Rank	8	4	6	5	NA	12	11	2	1	9	10	7	3	13
Signed Rank	8	-4	6	-5	NA	-12	-11	2	-1	-9	-10	-7	3	-13

$n = 13$  non-zero ranks

Sum of positive ranks is smaller than sum of negative ranks

Sum of positive ranks  $= 8 + 6 + 2 + 3 = 19 = T$

$$z_T = \frac{19 - \frac{13(14)}{4}}{\sqrt{\frac{13(14)(27)}{24}}} = \frac{19 - 45.5}{14.309} = -1.85; p > 0.05$$

What do we conclude?

# Wilcoxon rank sum test

Don't be confused with the Wilcoxon Signed-Rank Test! The *Wilcoxon Rank Sum Test*:

- ▶ Is the non-parametric counterpart of the two-sample t-test
- ▶ Assumes two samples are independent but does not require normality or equal variance of groups
- ▶ Assumes two distributions have "roughly the same shape"
- ▶ Also called the *Mann-Whitney U test* or *Wilcoxon-Mann-Whitney test*

# Wilcoxon rank sum test

The **Wilcoxon Rank Sum Test** is an extension of the Wilcoxon signed rank test for independent samples. Instead of taking differences, the observations are lumped together and ranked, with then the sum of the ranks calculated by group.

We'll use software to perform this test.

## Example: Exercise capacity and coronary artery disease

Consider data from a two-group study of exercise capacity. Two groups of men, one with diagnosed three-vessel coronary artery disease (3VD), and the other group of men with suspected disease (SD) in one or more vessels. The total time participants could exercise on a treadmill set to increase in speed and slope according to a set schedule, is below.

3VD times: 864, 636, 638, 708, 786, 600, 1320, 750, 594, 750

SD times: 1014, 684, 810, 990, 840, 978, 1002, 1110

3VD median: 729

SD median: 984

## Example: exercise capacity and coronary artery disease

Value	Rank	Group	Value	Rank	Group
594	1	3VD	810	10	SD
600	2	3VD	840	11	SD
636	3	3VD	864	12	3VD
638	4	3VD	978	13	SD
684	5	SD	990	14	SD
708	6	3VD	1002	15	SD
750	7.5	3VD	1014	16	SD
750	7.5	3VD	1110	17	SD
786	9	3VD	1320	18	3VD

Sum of ranks in SD group ( $n = 8$ ): 101

Sum of ranks in 3VD group ( $n = 10$ ): 70

## Example: exercise capacity and coronary artery disease

The sum of the ranks in the SD group ( $n = 8$ ) was 101, and the 3VD group had 10 subjects with sum of ranks 70. Is this enough evidence to reject the null hypothesis? (Ask software).

As it turns out, we reject  $H_0$  and conclude the two groups do not have the same median. The median in the SD group is higher.

# Kruskal-Wallis test

The **Kruskal-Wallis Test** is the nonparametric version of ANOVA, generalizing the Wilcoxon rank sum test to more than 2 groups.

## Kruskal-Wallis test for pet data

Consider the study of pulse rate from the ANOVA lecture. Remember that pulse rates were not necessarily normally distributed within groups.

## Kruskal-Wallis test for pet data

Calculating the sum of the ranks, we find that in the pet group, out of  $n = 15$ , the sum of the ranks was 190; out of the  $n = 15$  in the friend group, the sum of the ranks was 495, and out of the  $n = 15$  where neither were present, the sum of the ranks was 350. Is this enough evidence to reject the null hypothesis? (Ask software).

The p-value is  $< 0.0001$ . What may we conclude?

## Pairwise comparisons: pet vs. friend

Again, we can step-down and look at pairwise differences. In this case, we would use the Wilcoxon Rank-Sum test (why?).

# Contingency tables

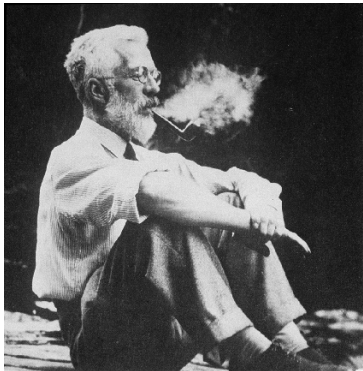
Typical questions of interest:

- ▶ Is there an association between the row variable (indexed by  $r$ ) and the column variable (indexed by  $c$ )?
- ▶ How strong is any association?

# Hypothesis test for contingency tables

- ▶  $H_0$ : pneumococcal serotype is unrelated to mortality (they are not associated)
- ▶  $H_1$ : pneumococcal serotype is related to mortality (they are associated)

# Fisher's exact test



Fisher's exact test is a great choice for testing a relationship between two variables in a contingency table. While it has been around for almost 100 years, it was originally used only for very small samples due to the computational burden involved.

"Sir Ronald Fisher could be regarded as Darwin's greatest twentieth-century successor." Richard Dawkins,  
*River out of Eden*

## Fisher's exact test

Fisher's exact test is fairly intuitive. We assume the column and row totals are fixed (so for our pneumococcus example, we assume we have 38 deaths and 218 survivors and that we have 34 in serogroup 31, 44 in serogroup 10, 72 in serogroup 15, and 106 in serogroup 20).

Next, we construct *all possible contingency tables* with the same margins, and then sum up the probabilities of all tables as extreme or more extreme than our own table to get the p-value.

## Fisher's exact test

Our Table

Serogroup	Died	Survived	Total
31	10	24	34
10	7	37	44
15	12	60	72
20	9	97	106
Total	38	218	256

Another Table With Same Margins

Serogroup	Died	Survived	Total
31	9	25	34
10	8	36	44
15	12	60	72
20	9	97	106
Total	38	218	256

These probabilities had to be calculated using the *hypergeometric* distribution. Obviously, this was no fun before modern computing.

The  $p$ -value using Fisher's exact test for the IPD data is 0.027.  
What can we conclude?

# Why not always go nonparametric?

Non-parametric methods are desirable because they do not require as many restrictive assumptions as parametric ones. However, this flexibility comes at a price – *if* the assumptions underlying a parametric test are satisfied, the nonparametric test is less powerful than the comparable parametric technique.