Non-parametric Tests

Why non-parametric methods

- Certain **statistical tests** like the t-test require **assumptions** of the **distribution** of the study variables in the population
  - **t-test** requires the underlying assumption of a **normal distribution**
  - Such tests are known as **parametric tests**

- There are situations when it is obvious that the study variable cannot be normally distributed, e.g.,
  - # of hospital admissions per person per year
  - # of surgical operations per person

Parametric and Non-Parametric Tests

- **Parametric Tests**: Relies on theoretical distributions of the test statistic under the null hypothesis and **assumptions** about the **distribution** of the sample data (i.e., **normality**)

- **Non-Parametric Tests**: Referred to as "Distribution Free" as they do **not assume** that data are drawn from **any particular distribut**ion

Why Use Non-parametric Methods

- The study variable generates data which are scores and so should be treated as a categorical variable with data measured on ordinal scale
  - E.g., severity of symptoms after taking headache pill:
    - 1: feeling worse
    - 2: feeling better
    - 3: no change

- For such type of data, the assumption required for parametric tests seem invalid => non-parametric methods should be used

- Aka **distribution-free tests**, because they make no assumption about the underlying distribution of the study variables

Wilcoxon rank sum test (aka Mann-Whitney U test)

- **Non-parametric** equivalent of parametric **t-test** for **2 independent samples** (unpaired t-test)

- Suppose the waiting time (in days) for cataract surgery at two eye clinics are as follows:

Patients at clinic A (n_A=18)	1, 5, 15, 7, 42, 13, 8, 35, 21, 12, 12, 22, 3, 14, 4, 2, 7, 2
Patients at clinic B (n_B=15)	4, 9, 6, 2, 10, 11, 16, 18, 6, 0, 9, 11, 7, 11, 10

Wilcoxon rank sum test (aka Mann-Whitney U test)

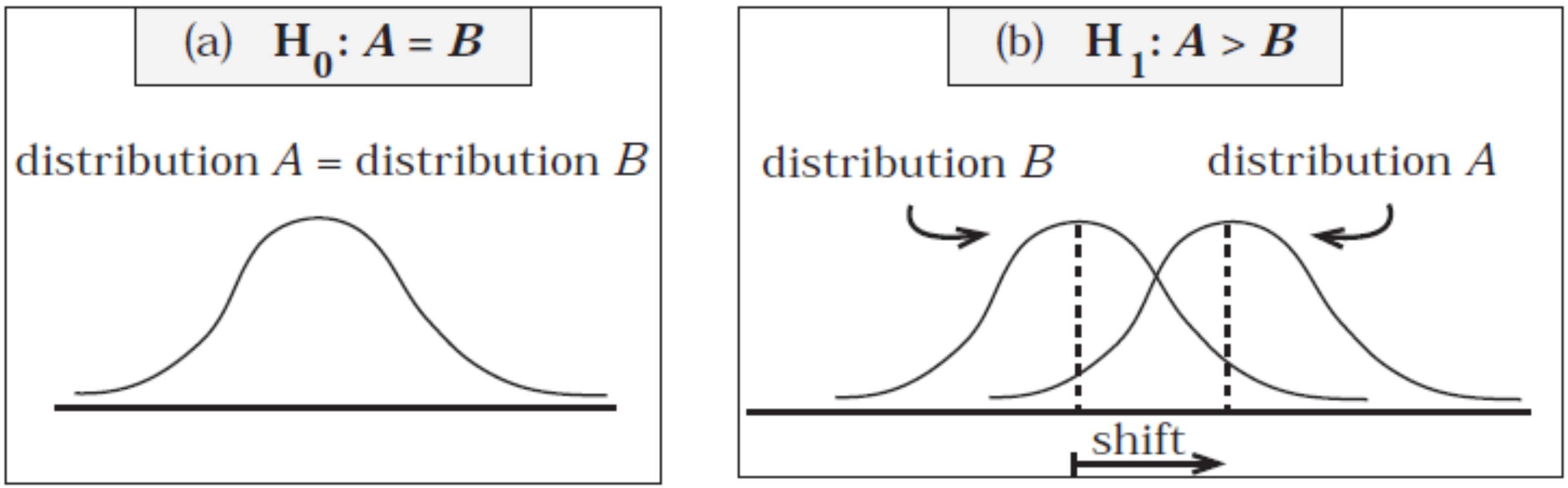

Illustration of $H_0 : A = B$ versus $H_1 : A > B$.

Wilcoxon rank sum test

1. **Rank all observations** (n_A+n_B) in **ascending** order (least time to longest) along with the group identity each observation belongs

2. **Resolve tied ranks** by dividing sum of the ranks by the number of entries for a particular set of ties, i.e. average the ranks

time	rank	clinic	time	rank	clinic
0	1	B	8	15	A
1	2	A	9	16.5	B
2	4	A	9	16.5	B
2	4	B	10	18.5	B
2	4	A	10	18.5	B
3	6	A	11	21	B
4	7.5	A	11	21	B
4	7.5	B	11	21	B
5	9	A	12	23.5	A
6	10.5	B	12	23.5	A
6	10.5	B	13	25	A
7	13	A	etc	etc	etc
7	13	A			
7	13	B			

Wilcoxon rank sum test

1. **Sum up ranks separately** for the two groups.
   - If the two populations from which the samples have been drawn have **similar distributions**, we would expect the **sum of ranks** to be **close**.
   - If not, we would expect the **group** with the **smaller median** to have the **smaller sum of ranks**

2. If the **group sizes** in both groups are the **same**, **take** the group with the **smaller** sum of ranks. If both groups have **unique sample sizes**, then **use** the sum of ranks of the **smaller group**

3. **Test for statistical significance**

Wilcoxon rank sum test

- In this example
  - sum of group A ranks = 324.5
  - sum of group B ranks = 236.5
- T= 236.5 (sum of ranks of the smaller group)
- If $n=n_A+n_B$ <=25, then looking up table giving critical values of T for various size of n_A and n_B
- If n>25, we assume that T is practically normally distributed with

$$\mu_T = \frac{n_A(n_A+n_B+1)}{2}, \text{where } n_A < n_B$$

$$SE_T = \sqrt{\frac{n_B \mu}{6}}$$

Wilcoxon rank sum test

- For our problem, T=236.5, n_A=18, n_B=15

$$z = \frac{T - \mu_T}{SE_T} = \frac{236.5 - 255}{27.66} = 0.67$$

- Result is not statistically significant at 5% (P=0.05) level

- **No strong evidence** to show that the difference in waiting time for the two clinics are statistically significant

Example

Samples of individuals from **several ethnic groups** were taken. Blood samples were collected from each individual and several variables measured. We shall compare the groups labeled "Native American“ and " Caucasian" with respect to the variable.
The data is as follows:

Native American (n_A=7)	8.50, 9.48, 8.65, 8.16, 8.83, 7.76, 8.63
Caucasian (n_B=9)	8.27, 8.20, 8.25, 8.14, 9.00, 8.10, 7.20 8.32, 7.70

TABLE 5

Critical values of T_L and T_U for the Wilcoxon rank sum test: independent samples. Test statistic is rank sum associated with smaller sample (if equal sample sizes, either rank sum can be used).

a. $\alpha = .025$ one-tailed; $\alpha = .05$ two-tailed

n_2 \ n_1	3		4		5		6		7		8		9		10	
	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	5	16	6	18	6	21	7	23	7	26	8	28	8	31	9	33
4	6	18	11	25	12	28	12	32	13	35	14	38	15	41	16	44
5	6	21	12	28	18	37	19	41	20	45	21	49	22	53	24	56
6	7	23	12	32	19	41	26	52	28	56	29	61	31	65	32	70
7	7	26	13	35	20	45	28	56	37	68	39	73	41	78	43	83
8	8	28	14	38	21	49	29	61	39	73	49	87	51	93	54	98
9	8	31	15	41	22	53	31	65	41	78	51	93	63	108	66	114
10	9	33	16	44	24	56	32	70	43	83	54	98	66	114	79	131

b. $\alpha = .05$ one-tailed; $\alpha = .10$ two-tailed

n_2 \ n_1	3		4		5		6		7		8		9		10	
	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	6	15	7	17	7	20	8	22	9	24	9	27	10	29	11	31
4	7	17	12	24	13	27	14	30	15	33	16	36	17	39	18	42
5	7	20	13	27	19	36	20	40	22	43	24	46	25	50	26	54
6	8	22	14	30	20	40	28	50	30	54	32	58	33	63	35	67
7	9	24	15	33	22	43	30	54	39	66	41	71	43	76	46	80
8	9	27	16	36	24	46	32	58	41	71	52	84	54	90	57	95
9	10	29	17	39	25	50	33	63	43	76	54	90	66	105	69	111
10	11	31	18	42	26	54	35	67	46	80	57	95	69	111	83	127

Source: From F. Wilcoxon and R. A. Wilcox, *Some Rapid Approximate Statistical Procedures* (Pearl River, N.Y. Lederle Laboratories, 1964), pp. 20–23. Reproduced with the permission of American Cyanamid Company.

Wilcoxon matched pairs signed ranks test

- **Non-parametric equiv**. of parametric **paired t-test**
- Suppose the anxiety scores recorded for 10 patients receiving a new drug and a placebo in random order in a cross-over clinical trial are:

Patients	1	2	3	4	5	6	7	8	9	10
Drug score	19	11	14	17	23	11	15	19	11	8
Placebo score	22	18	17	19	22	12	14	11	19	7

- Question: Is there any statistical evidence to show that the new drug can significantly lower anxiety scores when compared with the placebo?

Wilcoxon matched pairs signed ranks test

1. Take the difference for each pair of readings

Patients	1	2	3	4	5	6	7	8	9	10
Drug score	19	11	14	17	23	11	15	19	11	8
Placebo score	22	18	17	19	22	12	14	11	19	7
difference	-3	-7	-3	-2	1	-1	1	8	-8	1

2. Rank the differences from the smallest to the largest, ignoring signs and omitting 0 differences

Patients	1	2	3	4	5	6	7	8	9	10
Drug score	19	11	14	17	23	11	15	19	11	8
Placebo score	22	18	17	19	22	12	14	11	19	7
difference	-3	-7	-3	-2	1	-1	1	8	-8	1
rank	6.5	8	6.5	5	2.5	2.5	2.5	9.5	9.5	2.5

Wilcoxon matched pairs signed ranks test

- Add up ranks of positive differences and ranks of negative differences. Call the sum of the smaller group T

Patients	1	2	3	4	5	6	7	8	9	10
Drug score	19	11	14	17	23	11	15	19	11	8
Placebo score	22	18	17	19	22	12	14	11	19	7
difference	-3	-7	-3	-2	1	-1	1	8	-8	1
Rank -	6.5	8	6.5	5		2.5			9.5	
Rank +					2.5		2.5	9.5		2.5

- Sum of + ranks: 17 (n_+ =4)
- Sum of – ranks: 38 (n_- = 6)
- T (sum of ranks of smaller group) = 17

Wilcoxon matched pairs signed ranks test

- Test for statistical significance
  - If n<25, then look up table giving critical values of T for various size of n
  - If n>25, we can assume that T is practically normally distributed with

$$\mu_T = \frac{n(n+1)}{4}$$

$$SE_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{\mu_T(2n+1)}{6}}$$

- For our problem, T=17 and n=10, hence we look up table

Wilcoxon matched pairs signed ranks test

critical value for P=0.05 at N=10 is 8 (for 2-tailed test)

Note that critical values go progressively smaller as P gets smaller

Table B **Table of Critical Values of T in the Wilcoxon's Matched-Pairs Signed-Ranks Test**

	Level of significance for one-tailed test		
	0.025	0.01	0.005
N	Level of significance for two-tailed test		
	0.05	0.02	0.01
6	0	-	-
7	2	0	-
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61
25	89	77	68

Wilcoxon matched pairs signed ranks test

- For our problem, we found that T value of 17 is **higher** than the critical value for **statistical significance** at the **5%** level

- **There is insufficient evidence** to show that the new drug can significantly lower anxiety scores than the placebo.

- Therefore, we **cannot rule out** the possibility that the observed differences among scores are **due to sampling error.**

Non-parametric vs. parametric methods

Advantages:

- **Do not require the assumption** needed for parametric tests.
  - Therefore useful for data which are markedly skewed
- **Good for data** generated from **small samples**.
  - For such small samples, parametric tests are not recommended unless the nature of population distribution is known
- **Good for observations** which are scores,
  - i.e. measured on ordinal scale
- **Quick and easy to apply** and yet compare quite well with parametric methods

Non-parametric vs. parametric methods

Disadvantages

- **Not suitable for estimation** purposes as confidence intervals are difficult to construct
- **No equivalent methods** for more complicated parametric methods like testing for interactions in **ANOVA models**
- **Not quite as statistically efficient** as **parametric** methods if the assumptions needed for the parametric methods have been met