Biostatistics

Chapter 8 Nonparametric Statistics

Jing Li

jing.li@sjtu.edu.cn

http://cbb.sjtu.edu.cn/~jingli/courses/2017fall/bi372/

Dept of Bioinformatics & Biostatistics, SJTU

Continuous outcome (means)

Parametric statistics

Outcome Variable	Are the observations correlated?		Alternatives if the normality assumption is violated (and small n):
	independent	correlated	
Continuous (e.g. blood pressure, age, pain score)	**Ttest:** compares means between two independent groups **ANOVA:** compares means between more than two independent groups **Pearson's correlation coefficient** (linear correlation): shows linear correlation between two continuous variables **Linear regression:** multivariate regression technique when the outcome is continuous; gives slopes or adjusted means	**Paired ttest:** compares means between two related groups (e.g., the same subjects before and after) **Repeated-measures ANOVA:** compares changes over time in the means of two or more groups (repeated measurements) **Mixed models/GEE modeling**: multivariate regression techniques to compare changes over time between two or more groups	

Parametric Test Procedures

1. Involve Population Parameters (Mean)

2. Have Stringent Assumptions
   (Normality)

3. Examples: Z Test, t Test, F test

Example

- You want to see if the success rates for two protocols is the same. For protocol 1, the rates (% of capacity) are **71**, **82**, **77**, **92**, **88**. For protocol 2, the rates are **85**, **82**, **94** & **97**. Do the rates have the same **probability distributions** at the **.05** level?

Continuous outcome (means)

Parametric statistics | Non-parametric statistics

Outcome Variable	Are the observations correlated? independent	correlated	Alternatives if the normality assumption is violated (and small n):
Continuous (e.g. blood pressure, age, pain score)	**Ttest:** compares means between two independent groups **ANOVA:** compares means between more than two independent groups **Pearson's correlation coefficient** (linear correlation): shows linear correlation between two continuous variables **Linear regression:** multivariate regression technique when the outcome is continuous; gives slopes or adjusted means	**Paired ttest:** compares means between two related groups (e.g., the same subjects before and after) **Repeated-measures ANOVA:** compares changes over time in the means of two or more groups (repeated measurements) **Mixed models/GEE modeling**: multivariate regression techniques to compare changes over time between two or more groups	**Wilcoxon sign-rank test:** non-parametric alternative to paired ttest **Wilcoxon sum-rank test** (=Mann-Whitney U test): non-parametric alternative to the ttest **Kruskal-Wallis test:** non-parametric alternative to ANOVA **Spearman rank correlation coefficient:** non-parametric alternative to Pearson's correlation coefficient

Nonparametric Test Procedures

1. Do Not Involve Population Parameters

   Example: Probability Distributions

2. Data Measured on Any Scale (Ratio or Interval, Ordinal)

3. Example: Wilcoxon Rank Sum Test

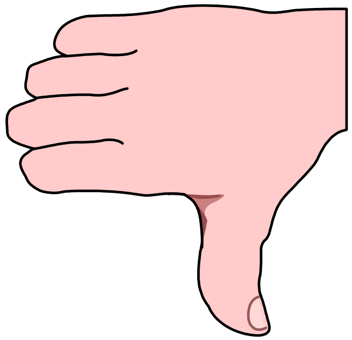
Advantages of Nonparametric Tests

1. Used With All Scales
2. Easier to Compute
3. Make Fewer Assumptions
4. Need Not Involve Population Parameters
5. Results May Be as Exact as Parametric Procedures

Disadvantages of Nonparametric Tests

1. May Waste Information

2. Difficult to Compute by hand for Large Samples
3. Tables Not Widely Available



Wilcoxon Rank Sum Test

Parametric	Nonparametric
t-test $t = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	Wilcoxon Rank Sum test (秩和检验) (Also known as the Mann–Whitney U-test or the Mann–Whitney–Wilcoxon test

Wilcoxon Rank Sum Test

1. Tests Two Independent Population Probability Distributions
2. Corresponds to t-Test for 2 Independent Means
3. Assumptions
   - Independent, Random Samples

Wilcoxon Rank Sum Test Procedure

1. Assign Ranks, R_i, to the $n_1 + n_2$ Sample Observations

   If Unequal Sample Sizes, Let n_1 Refer to Smaller-Sized Sample

2. Sum the Ranks, T_i, for Each Sample

3. Test Statistic Is T_A (Smallest Sample)

Null hypothesis: both samples come from the same underlying distribution

Wilcoxon Rank Sum Test

- You want to see if the success rates for two protocols is the same. For protocol 1, the rates (% of capacity) are **71**, **82**, **77**, **92**, **88**. For protocol 2, the rates are **85**, **82**, **94** & **97**. Do the rates have the same **probability distributions** at the **.05** level?

Wilcoxon Rank Sum Test Solution

- H_0: Identical Distrib.
- H_a: Shifted Left or Right
- **= .05**
- n_1 **= 4** n_2 **= 5**

Wilcoxon Rank Sum Test Computation Table

Protocol 1		Protocol 2	
Rate	Rank	Rate	Rank
71		85	
82		82	
77		94	
92		97	
88			
Rank Sum			

Wilcoxon Rank Sum Test Computation Table

Protocol 1		Protocol 2	
Rate	Rank	Rate	Rank
71	1	85	
82		82	
77	2	94	
92		97	
88			
Rank Sum			

Wilcoxon Rank Sum Test Computation Table

Protocol 1		Protocol 2	
Rate	Rank	Rate	Rank
71	1	85	
82	~~4~~ 3.5	82	~~5~~ 3.5
77	2	94	
92		97	
88			
Rank Sum			

Wilcoxon Rank Sum Test Computation Table

Protocol 1		Protocol 2	
Rate	Rank	Rate	Rank
71	1	85	5
82	~~4~~ 3.5	82	~~5~~ 3.5
77	2	94	8
92	7	97	9
88	6		
Rank Sum	19.5		25.5

Wilcoxon Rank Sum Test Solution

- **H0: Identical Distrib.**
- **Ha: Shifted Left or Right**
- **= .05**
- $n_1 = 4$ $n_2 = 5$
- **Critical Value(s):**

Test Statistic:

T = 25.5(Smallest Sample**)**

Wilcoxon Rank Sum Table (Portion)

α = .05 two-tailed

		n_1						
		4		5		6		..
		T_L	T_U	T_L	T_U	T_L	T_U	..
	4	10	26	16	34	23	43	..
n_2	5	11	29	17	38	24	48	..
	6	12	32	18	42	26	52	..
	:	:	:	:	:	:	:	:

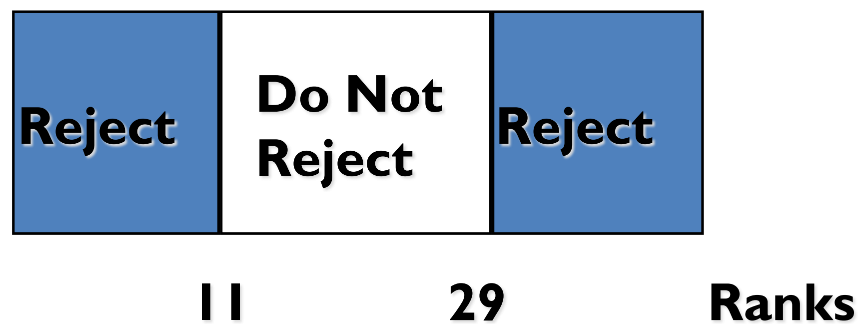
Wilcoxon Rank Sum Test Solution

- **H0: Identical Distrib.**
- **Ha: Shifted Left or Right**
- **= .05**
- $n_1 = 4 \quad n_2 = 5$

Test Statistic:

T = 25.5(Smallest Sample)

Critical Value(s):



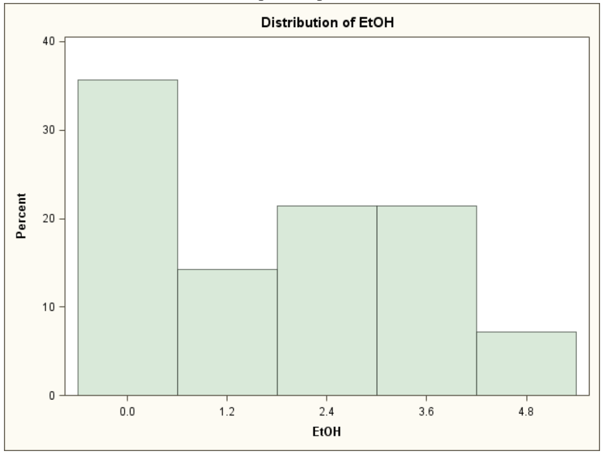
Decision: **Do Not Reject at = .05**

Conclusion: **There is No evidence for unequal distribution**

Practice

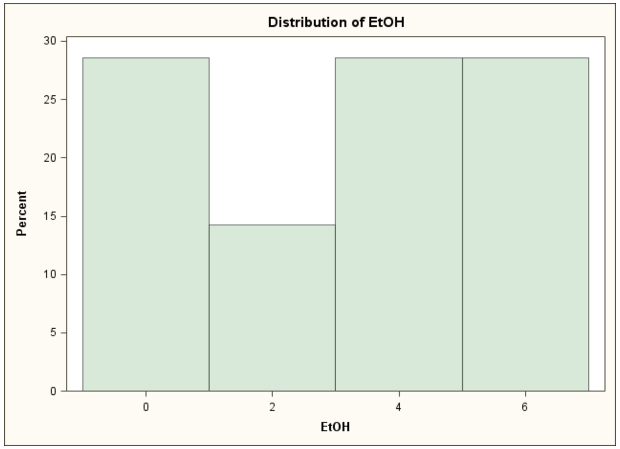
- Street smart students drink more alcohol than book smart students?
- What is the outcome variable? Weekly alcohol intake (drinks/week)
- What type of variable is it? Continuous

Book smart (13):



Mean=1.6 drinks/week; median = 1.5

Street smart (7):



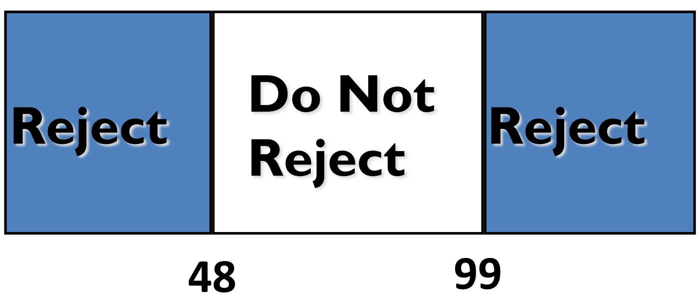
Mean=2.7 drinks/week; median = 3.0

- Is it normally distributed? **No** (and small n)
- Are the observations correlated? No
- Are groups being compared, and if so, how many? Two

Run Rank Sum Test

- Book smart values (n=13): 0 0 0 0 1 1 2 2 2 3 3 4 5
- Street Smart values (n=7): 0 0 2 3 3 5 6

n1,n2=7,13; two-sided P-value=0.05



Another example:

You work in the biostatistics department. Is the **new** R package **faster** (**.05** level)? You collect the following data entry times:

User	Current	New
Donna	9.98	9.88
Santosha	9.88	9.86
Sam	9.90	9.83
Tamika	9.99	9.80
Brian	9.94	9.87
Jorge	9.84	9.84

Are the observations correlated? **No**

Signed Rank Test

Parametric	Nonparametric
Paired t-test $$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$ $$s_d = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}}$$	Wilcoxon Signed Rank test (符号秩检验)

Wilcoxon Signed Rank Test

1. Tests Probability Distributions of Two Related Populations
2. Corresponds to t-test for Paired Means
3. Assumptions

   Random samples; Both populations are continuous; paired samples.

Signed Rank Test Procedure

1. Obtain Difference Scores, $D_i = X_{1i} - X_{2i}$
2. Take Absolute Value $|D_i|$ and *rank them* (Do not count $D_i = 0$)
3. Assign Ranks, R_i, with Smallest = 1
4. Calculate range and mean rank for $|D_i|$
5. Sum '+' Ranks (T_+) & '-' Ranks (T_-)

Signed Rank Test Computation Table

X_{1i}	X_{2i}	D_i	$\|D_i\|$	R_i	Sign	Sign R_i
9.98	9.88	+0.10	0.10	4	+	+4
9.88	9.86	+0.02	0.02	1	+	+1
9.90	9.83	+0.07	0.07	2 2.5	+	+2.5
9.99	9.80	+0.19	0.19	5	+	+5
9.94	9.87	+0.07	0.07	3 2.5	+	+2.5
9.84	9.84	0.00	0.00	...	...	Discard
Total					$T_+ = 15$, $T_- = 0$	

Signed Rank Test Solution

- H0: Identical Distrib.
- Ha: Different Distrib
- α = .05
- n' = 5 (not 6; 1 elim.)
- Critical Value(s):

Reject	**Do Not Reject**

Significant if T < critical value

Test Statistic:

T= smaller of T_+ and T_-

T=0, α = .1

Decision: **not reject H0**

Conclusion:

Border line at α = .1, no significant difference

Signed Rank Test Table (two-sides)

Critical Values of the Wilcoxon Signed Ranks Test

n	Two-Tailed Test		One-Tailed Test	
	α = .05	α = .01	α = .05	α = .01
5	--	--	0	--
6	0	--	2	--
7	2	--	3	0
8	3	0	5	1
9	5	1	8	3
10	8	3	10	5
11	10	5	13	7
12	13	7	17	9
13	17	9	21	12
14	21	12	25	15
15	25	15	30	19
16	29	19	35	23
17	34	23	41	27
18	40	27	47	32
19	46	32	53	37
20	52	37	60	43
21	58	42	67	49
22	65	48	75	55
23	73	54	83	62
24	81	61	91	69
25	89	68	100	76
26	98	75	110	84
27	107	83	119	92
28	116	91	130	101
29	126	100	140	110
30	137	109	151	120

Practice

A study of early childhood education asked kindergarten students to retell two fairy tales that had been read to them earlier in the week. Each child told two stories. The first had been read to them, and the second had been read but also illustrated with pictures. An expert listened to a recording of the children and assigned a score for certain uses of language.

Is there any difference between these two education ways ?

Child	1	2	3	4	5	6	7	8
Story 2	77	49	66	28	38	56	68	42
Story 1	40	72	0	36	55	45	51	40
Difference	37	-23	66	-8	-17	11	17	2

ANOVA

- Assumption:

    - Response variables are normally distributed (or approximately normally distributed)
    - Variances of populations are equal
    - Groups are independent

Kruskal-Wallis test

Use it,

if the data are not normally distributed;

if the variances for the different conditions are markedly different;

if the data are measurements on an ordinal scale.

Parametric	Nonparametric
ANOVA $$F_{k-1,n-k} = \frac{MS_B}{MS_W} = \frac{SS_B/(k-1)}{SS_W/(n-k)}$$	**Kruskal-Wallis test**

Kruskal-Wallis test

- A k-sample non-parametric test on the means (k > 2).
- Pool observations together $N = \sum n_i$ and assign ranks to individuals.
- Compute the rank sum R_i for each sample.

$$H = H^* = \frac{12}{N(N+1)} \times \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

- Test statistic Chi-squares with *df* = (k-1)

  one tail prob. Compare with χ^2_{k-1},

Step by step example of the Kruskal-Wallis test

Does physical exercise alleviate depression? We find some depressed people and check that they are all equivalently depressed to begin with. Then we allocate each person randomly to one of three groups: no exercise; 20 minutes of jogging per day; or 60 minutes of jogging per day. At the end of a month, we ask each participant to rate how depressed they now feel, on a Likert scale that runs from 1 ("totally miserable") through to 100 (ecstatically happy").

- We have three separate groups of participants, each of whom gives us a single score on a rating scale.
- Ratings are examples of an ordinal scale of measurement
- So, the data are not suitable for a parametric test.

Kruskal-Wallis test

- Here are the data

Rating on depression scale:

	No exercise	Jogging for 20 minutes	Jogging for 60 minutes
	23	22	59
	26	27	66
	51	39	38
	49	29	49
	58	46	56
	37	48	60
	29	49	56
	44	65	62
mean rating (SD):	39.63 (12.85)	40.63 (14.23)	55.75 (8.73)

Kruskal-Wallis test

- Step 1, Rank all of scores

	C1 (No exercise)	C2 (Jogging for 20 minutes)	C3 (Jogging for 60 minutes)
	23 (2)	22 (1)	59 (20)
	26 (3)	27 (4)	66 (24)
	51 (16)	39 (9)	38 (8)
	49 (14)	29 (5.5)	49 (14)
	58 (19)	46 (11)	56 (17.5)
	37 (7)	48 (12)	60 (21)
	29 (5.5)	49 (14)	56 (17.5)
	44 (10)	65 (23)	62 (22)
mean rank (SD)	**9.56 (6.25)**	**9.94 (6.84)**	**18.00 (5.09)**
sum of ranks (Tc)	**76.5**	**79.5**	**144**

If two or more scores are the same then they are "tied". "Tied" scores get the average of the ranks

Kruskal-Wallis test

- Step 2, find "H"

$$H = H^* = \frac{12}{N(N+1)} \times \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

N is the total number of participants (all groups combined). We have 24 participants (3 groups of 8).

R_i is the rank total for each group. R_1 = 76.5, R_2 = 79.5, and R_3 = 144.

n_i is the number of participants in each group. Here, n_1 = 8, n_2 = 8 and n_3 3 = 8.

Kruskal-Wallis test

Four our data

$$H = \left[\frac{12}{24 * (24+1)} * \sum \frac{R_i^{\;2}}{n_i} \right] - 3 * (24+1)$$

$$\frac{76.5^2}{8} + \frac{79.5^2}{8} + \frac{144^2}{8}$$

$$= 731.5313 + 790.0313 + 2592.0000 = \mathbf{4113.5625}$$

$$H = \left[\frac{12}{600} * 4113.5625 \right] - 75 \quad = \mathbf{7.27}$$

Kruskal-Wallis test

- Step 3, the degree of freedom

  the number of groups minus one. Here we have three groups, and so we have 2 d.f.

- Step 4, assessing the significance

  H is 7.27, with 2 d.f.

  P<0.05

Table of critical Chi-Square values:

df	$p = .05$	$p = .01$	$p = .001$
1	3.84	6.64	10.83
2	**5.99**	**9.21**	**13.82**
3	7.82	11.35	16.27

- Step 5, conclusion

  There is a difference of some kind between our three groups

Summary

Nonparametric	**Parametric**
Wilcoxon Rank – Sum test	Two sample t-test
Wilcoxon Signed-Rank test	Two paired sample t-test
Kruskal-Wallis test	ANOVA