

BIST P8130: Biostatistics Methods I

Recitation 06 – Wilcoxon Rank Sum test, Wilcoxon Signed-Rank test in SAS

Zilan Chai

Fall 2017

This recitation's big ideas:

- Use PROC NPAR1WAY to perform a Rank Sum test
- Use PROC UNIVARIATE to perform a Signed-Rank test

Wilcoxon Rank Sum Test

Wilcoxon Rank-Sum (or Mann-Whitney U) test is the non-parametric equivalent of two-sample independent T-test. It is a test of location, i.e., compares the medians: whether there is (or not) a significance difference in medians.

SAS provides both one- and two-sided p-values. One should be careful because the test is not symmetric, so you can't divide the two-sided p-value by 2.

SAS syntax:

```
proc npar1way data = data_name Wilcoxon;  
  class grouping_var;  
  var outcome_var;  
run;
```

Example: (Rosner 9.8) Suppose we want to compare the length of hospital stay for patients with the same diagnosis at two different hospitals.

Obs	hospital	length_of_stay
1	hosp1	21
2	hosp1	10
...
23	hosp2	44
24	hosp2	238

[Please see the SAS code for checking normality assumption.]

```

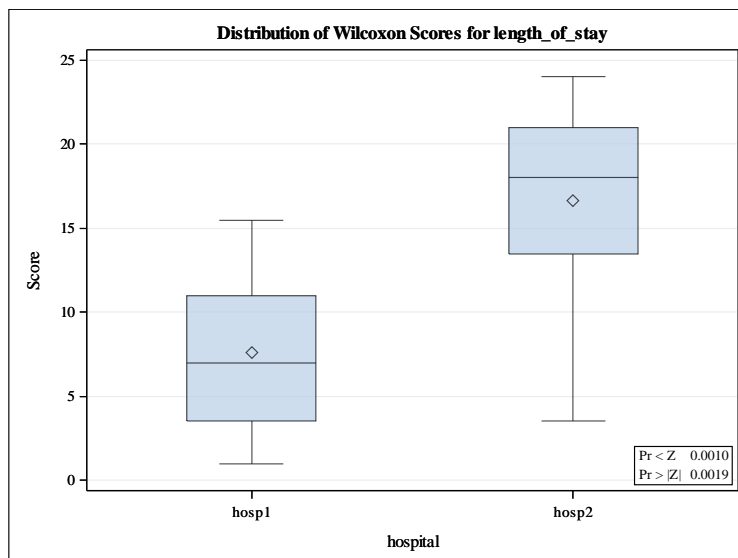
proc npar1way data = hospital_stay Wilcoxon;
  class hospital;
  var length_of_stay;
run;

```

Wilcoxon Scores (Rank Sums) for Variable length_of_stay Classified by Variable hospital					
hospital	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
hosp1	11	83.50	137.50	17.249002	7.590909
hosp2	13	216.50	162.50	17.249002	16.653846
Average scores were used for ties.					

Wilcoxon Two-Sample Test	
Statistic	83.5000
Normal Approximation	
Z	-3.1016
One-Sided Pr < Z	0.0010
Two-Sided Pr > Z	0.0019
t Approximation	
One-Sided Pr < Z	0.0025
Two-Sided Pr > Z	0.0050
Z includes a continuity correction of 0.5.	

Kruskal-Wallis Test	
Chi-Square	9.8008
DF	1
Pr > Chi-Square	0.0017



Wilcoxon Signed-Rank Test

By default, SAS calculates the exact Wilcoxon Signed-Rank test in PROC UNIVARIATE.
Example (Rosner 9.20)

Obs	id	sbp_baseline	sbp_1mo	diff
1	1	119.67	117.33	2.34
2	2	100.00	98.78	1.22
...
16	16	134.44	126.67	7.77
17	17	108.67	108.67	0.00

```
proc univariate data=sbp NORMAL;
  var diff;
  qqplot diff;
run;
```

Basic Statistical Measures			
Location		Variability	
Mean	2.950588	Std Deviation	3.27001
Median	2.560000	Variance	10.69297
Mode	.	Range	12.44000
		Interquartile Range	4.23000

Tests for Location: $\mu_0=0$				
Test	Statistic		p Value	
Student's t	t	3.720351	Pr > t	0.0019
Sign	M	5	Pr >= M	0.0213
Signed Rank	S	56	Pr >= S 	0.0021

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.982278	Pr < W	0.9750
Kolmogorov-Smirnov	D	0.106262	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.024329	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.168225	Pr > A-Sq	>0.2500