

Chapter 12 Non-parametric Test

Daxiang Na

2022-10-25

Contents

1 Questions:	1
1.1 What is V in the returned result of <code>wilcox.test</code>	1
1.2 What is the definition of T in <code>psignrank</code> ?	1
1.3 How can we decide which probability to calculate in one-sided Wilcoxon Rank Sum test?	2
1.4 Question about CTL	2
1.5 is it okay to share the Inference cheat sheet raw file?	2
2 Wilcoxon Signed Rank Test	2
3 Wilcoxon Rank-Sum test (also known as Mann-Whitney U test)	3

1 Questions:

1.1 What is V in the returned result of `wilcox.test`

```
## Wilcoxon signed rank exact test data: data$air and data$sulf.diox
## V = 21, p-value = 0.006653 alternative hypothesis: true location
## shift is not equal to 0
```

1.2 What is the definition of T in `psignrank`?

In the `psignrank` document(<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/SignRank>), the the Wilcoxon signed rank statistic is "the sum of the ranks of the absolute values $x[i]$ for which $x[i]$ is positive". I am wondering if $T = T^+$ instead of $T = \min(T^+, T^-)$? This appears to be correct when solving the example problem, where $2*(1 - \text{psignrank}(75.5, n=14)) = 0.135$ but $2*\text{psignrank}(29.5, n=14) \neq 0.135$

I therefore wonder if $T = T^+$ should be the definition of `psignrank` in R.

Another question regarding this topic: when testing the two-tailed hypothesis, when should

we use $2*(1 - \text{psignrank}(T,n))$ and $2*\text{psignrank}(T, n)$? i.e. what is the mean for the wilcoxon signed rank distribution? Is it $n(n+1)/4$ as stated in the document?

This is right.

So when we calculate the p value, we determine whether to use $2*(1 - \text{psignrank}(T,n))$ or $2*\text{psignrank}(T, n)$ based on if $T > n(n+1)/4$ or $T < n(n+1)/4$, right?

1.3 How can we decide which probability to calculate in one-sided Wilcoxon Rank Sum test?

In the Wilcoxon Rank Sum test, we always use $W = \min(W_1, W_2)$ and $\mu_W = n_{\min} * (n_{\min} + n_{\text{large}} + 1)/2$ no matter if we want to know if $H_1 : \mu_1 - \mu_2 < 0$ or $H_1 : \mu_2 - \mu_1 < 0$, how to we determine if p-value = $1 - \Pr(z < z_W)$ or p-value = $\Pr(z < z_W)$?

My understanding: it depends on which group has W_{\min} . In the Wilcoxon Rank Sum test we are calculating the probability of getting W_{\min} in the distribution of $N(n_{\min} * (n_{\min} + n_{\text{large}} + 1)/2, \sigma)$. In the example problem in the slides, H1 is about group 1 has less values, and group 1 turns out to have W_{\min} , therefore p-value = $\Pr(z < 0)$.

1.4 Question about CTL

For sampling distribution, it can have sampling size n and sampling time m , what determines if it follows CLT? n or m ? If we sample for 1 time and 100 times, each time with same size n , does that makes a difference? Another way to ask this question: if we sample for 3 times ($m = 3$), but each time with sample size $n = 1000$, does that follow CLT?

1.5 is it okay to share the Inference cheat sheet raw file?

Willing to extend it.

2 Wilcoxon Signed Rank Test

- Only for paired sample.
- Evaluate the null hypothesis: $Z_T = (T - \mu_T)/\sigma_T$

- Note:

$$\mu_T = 0$$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{6}}$$

- When n is large enough ($n > 12$), we get

$$Z_T \sim N(0, 1)$$

- calculate the probability of getting Z_T when $\mu = 0$ is true.
- For two-sided test, follow what we do in the sampling distribution:
 - $2 * p$ when $z < 0$
 - $2 * (1 - p)$ when $z > 0$
- if $n > 12$, you can just apply CLT, the R code is: `wilcox.test(before, after, paired = T, exact = F, correct = F)`. `exact = F` determines if the statistics follow normal distribution (`exact = F`) or exact distribution (`exact = T`).
- If $n \leq 12$, we cannot use the normal approximation. In that case, we use `psignrank(T,n)` in R to calculate the exact distribution.
 - R requires $T = T^+$ for this to work correctly!

3 Wilcoxon Rank-Sum test (also known as Mann-Whitney U test)

- nonparametric analog to the two-sample t-test
- get W_1 and W_2
- $W = \min(W_1, W_2)$
- n_1 = sample size with the smaller sum of ranks.
- n_2 = sample size with the larger sum of ranks.

$$\mu_W = \frac{n_1(n_1 + n_2 + 1)}{2} \text{ and } \sigma_W = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

-
- $z_W = \frac{W - \mu_W}{\sigma_W}$
- $z_W \sim N(0, 1)$ when n_1 and n_2 are large enough ($n_1, n_2 > 10$).

- in R: `wilcox.test(..., exact = F, correct = F, paired = F, alt = "")`

- When n_1 and n_2 are very small (i.e. either is less than or equal to 10), we can use the exact distribution to calculate the p-values. In R: `pwilcox(Wobs, n1, n2)`
 - in this case, $W_{obs} = W - n_1(n_1 + 1)/2$
 - `wilcox.test` also works when `exact = T`
- correct: correct the data with continuity correction