Chapter 12 Non-parametric Test

Daxiang Na

2022-10-25

Contents

1 Questions:

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

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 - psignrank(75.5,n=14))` = 0.135 but `2*psignrank(29.5,n=14)` != 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 - psignrank(T,n))` and `2*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?

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 =` 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 $T = min(T^+, T^-)$ 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 (n1, n2 > 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