Chapter 9

—

Ⓐ Wilcoxon Rank Sum test

many non-P procedures involve ranks.

$Y_1, \ldots, Y_{n_y} \overset{iid}{\sim} F_Y$ & $X_1, \ldots, X_{n_x} \overset{iid}{\sim} F_X$

wtk is $F_Y = F_X$?

but, the test targets means: $H_0 : \mathbb{E}\{Y\} = \mathbb{E}\{X\}$

$H_1 : \mathbb{E}\{Y\} \neq \mathbb{E}\{X\}$

Like the permutation test, the Wilcoxon test works on the combined sample of size $n = n_y + n_x$

Let $\boxed{r_i^{(y)} \text{ \& } r_j^{(x)}}$ denote the rank assigned to y_i & x_j in the combined sample, for all $i = 1, \ldots, n_y$
$j = 1, \ldots, n_x$

There are several common <u>test stats</u>, but most involve

$$\bar{r} \equiv \bar{r}^{(y)} = \sum_{i=1}^{n_y} r_i^{(y)}$$

Simulating from the null distribution is easy by randomly permuting ranks $\{1, \ldots, n\}$

The distribution of $\bar{R}$ may be obtained by considering the distribution of n_y integers selected w/o replacement from $\{1, \ldots, n\}$ & summed.

- the number of ways to sum n_y from n things is $\binom{n_y}{n}$

- counting how many of those sum up to $\bar{r}$ (72) is harder

$$\frac{SSP(\bar{r}, n_y, \{1, \ldots, n\})}{\binom{n_y}{n}}$$

Subset sum problem (NP-hard)

<u>see App B.3 in book</u>

In R there is a CDF called <u>pwilcox</u> & it works with a shifted test stat

$$\bar{r}' = \bar{r} - \frac{n_y(n_y+1)}{2}$$

Ch9 also has <u>CLT approx</u> & one more example.

Chapter 9

(B) Wilcoxon <u>signed ranks</u> test

is like a non-p <u>paired</u> t-test.

same hypotheses as (A) but paired data.

$(x_1, y_1), \ldots, (x_{n'}, y_{n'})$

let $d_i = y_i - x_i \quad i = 1, \ldots, n$

& discard any $d_i = \emptyset$ & re-label to

$d_1, \ldots, d_n$

- rank the absolute differences $|d_1|, \ldots, |d_n|$
  $\rightarrow r_1, \ldots, r_n$

- put the sign of each d_i back on its rank

$$r_i^{\pm} = \begin{cases} r_i & \text{if } y_i > x_i \\ -r_i & \text{otherwise} \end{cases} \quad i=1,\dots,n$$

Again, several stats are used for testing, but most are a variation on

$$\bar{r}^+ = \sum_{i=1}^{n} r_i^{\pm} \mathbb{1}_{\{d_i > 0\}}$$

As usual simulation is straightforward.

The math is harder. A SSP is involved, R has psignrank for the CDF

<u>One-sample non-P location test</u>

$Y_1, \dots, Y_{n_y} \overset{iid}{\sim} f_Y$ & wtk

$$H_0 : \mathbb{E}\{Y\} = \mu \quad \text{for some } \mu$$

$$\text{or } H_1 : \mathbb{E}\{Y\} \neq \mu$$

Can use <u>signed-ranks test</u> by replacing

$Y_i = \mu$ everywhere.