

Chapter 9

① Wilcoxon Rank Sum test

many non-P procedures involve ranks.

$$Y_1, \dots, Y_{n_Y} \stackrel{\text{iid}}{\sim} F_Y \quad \& \quad X_1, \dots, X_{n_X} \stackrel{\text{iid}}{\sim} F_X$$

$$\text{WtK is } F_Y = F_X ?$$

but, the

test targets means: $H_0: E\{Y\} = E\{X\}$

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

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

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

There are several common test stats, but most involve

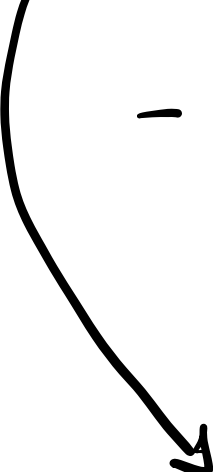
$$\bar{r} \equiv \bar{r}^{(y)} = \sum_{i=1}^{n_y} r_i^{(y)}$$

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

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


$$\text{SSP}(\bar{r}, n_y, \{1, \dots, n\})$$



$$\binom{n_y}{n}$$

Subset sum problem (NP-hard)

see App B.3 in book

In R there is a CDF called pnwilcox
& it works with a shifted test stat

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

Ch 9 also has CLT approx & one more example.

Chapter 9

⑧ Wilcoxon signed rank test

is like a non-p paired t-test.

Same hypothesis as ⑦ but paired data.

$$(x_1, y_1), \dots, (x_n, y_n)$$

$$\text{let } d_i = y_i - x_i \quad i=1, \dots, n$$

discard any $d_i = 0$ & re-label to

$$d_1, \dots, d_n$$

— rank the absolute differences $|d_1|, \dots, |d_n|$
 $\rightarrow r_1, \dots, r_n$

— put the sign of each d_i back on
its rank

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

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

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

As usual simulation is straightforward.

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

One-sample non-P location test

$$Y_1, \dots, Y_n \stackrel{\text{id}}{\sim} f_Y \quad \& \quad \text{with}$$

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

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

Can use signed-ranks test by replacing

$\chi_i = \mu$ everywhere.