

## 同じ値 (タイ) がある場合も考慮した順位検定統計量の明解な表現

$(X_1, \dots, X_{n_1})$  を平均  $\mu_1$  の連続型の分布関数  $F_\sigma(x - \mu_1) = F(\frac{x-\mu_1}{\sigma})$  をもつ母集団からの大きさ  $n_1$  の無作為標本,  $(Y_1, \dots, Y_{n_2})$  を平均  $\mu_2$  の分布関数  $F_\sigma(x - \mu_2) = F(\frac{x-\mu_2}{\sigma})$  をもつ母集団からの大きさ  $n_2$  の無作為標本とする. すなわち,  $X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}$  は互いに独立な確率変数の実現値で, 各  $X_i$  は同一の分布関数  $F_\sigma(x - \mu_1)$  をもち, 各  $Y_j$  は同一の分布関数  $F_\sigma(x - \mu_2)$  をもつとする.  $n \equiv n_1 + n_2$  と置く.

帰無仮説  $H_0 : \mu_1 = \mu_2$  v.s. 対立仮説  $H_1 : \mu_1 \neq \mu_2$  の有意水準  $\alpha$  の検定を考える. 実際のデータには  $X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}$  の中に同じ値がある場合が多いため, 中間順位を採用している.

$$I(A) = \begin{cases} 1 & (\text{不等式 } A \text{ が真のとき}) \\ 0 & (\text{不等式 } A \text{ が偽のとき}) \end{cases}$$

とし,

$$R_i \equiv \frac{\sum_{k=1}^{n_1} I(X_k \leq X_i) + \sum_{k=1}^{n_2} I(Y_k \leq X_i)}{2} + \frac{\sum_{k=1}^{n_1} I(X_k < X_i) + \sum_{k=1}^{n_2} I(Y_k < X_i) + 1}{2}, \quad (1)$$

$$R_{n_1+j} \equiv \frac{\sum_{k=1}^{n_1} I(X_k \leq Y_j) + \sum_{k=1}^{n_2} I(Y_k \leq Y_j)}{2} + \frac{\sum_{k=1}^{n_1} I(X_k < Y_j) + \sum_{k=1}^{n_2} I(Y_k < Y_j) + 1}{2} \quad (2)$$

で定義すれば, 同じ値がない場合のデータに対して, 通常の  $X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}$  を小さい方から並べたときの  $X_i$  の順位が  $R_i$ ,  $Y_j$  の順位が  $R_{n_1+j}$  となっている. 同じ値がある場合のデータに対しても (1), (2) は中間順位を与えている. このとき,

$$T_R \equiv \frac{1}{n_1} \sum_{i=1}^{n_1} (R_i - \frac{n+1}{2})$$

がウィルコクソン (マン・フィットニー) の順位検定統計量で,  $|T_R|$  が大きいとき  $H_0$  を棄却する.

$$Z_R \equiv \sqrt{\frac{n(n-1)}{n_1 n_2}} \frac{n_1 T_R}{\sqrt{\sum_{k=1}^n (R_k - \frac{n+1}{2})^2}} \quad (3)$$

の分布は,  $n_1, n_2$  がともに大きいとき, 標準正規分布  $N(0, 1)$  で近似できる.  $m$  を同じ値 (同順位) の組の数,  $t_i$  を  $i$  番目の組の同順位の観測値の個数とすれば,  $Z_R$  は

$$Z_R = \frac{n_1 T_R}{\sqrt{\frac{n_1 n_2}{n(n-1)} [\frac{n^3 - n^2}{12} - \sum_{i=1}^m \frac{t_i^3 - t_i^2}{12}]}} \quad (4)$$

とも表現できる. 順位検定法が紹介された本では, 同じ値 (同順位) がある場合の補正として (4) の表現がなされているが, (3) の表現の方が明解である. 2 標本のウィルコクソン検定を

例に取っているが，1 標本のウィルコクソンの符号付順位検定についても，(4) に対応した表現が，順位検定法が紹介された本に掲載されている．符号付順位検定も (3) に対応した明解な表現が可能である．Esoft のプログラミングでは (1), (2), (3) を使ったアルゴリズムが採用されている．

[著書と発展のページにリンク](#)