STA 104 Applied Nonparametric Statistics

Chapter 4: One-Way Layout Problems: Nonparametric One-Way Analysis of Variance

Xiner Zhou

Department of Statistics, University of California, Davis

Table of contents

Two-Sided All-Treatments Multiple Comparisons for General Alternative

After rejection of

$$H_0 : \underbrace{\tau_1 = \ldots = \tau_k}_{F_1 = F_2 \ldots = F_k \equiv F}$$

$$H_1 : \qquad \underbrace{\tau_1 \ldots \tau_k \text{ not all equal}}_{\text{at least two of the treatment effects are not equal}}$$

with the Kruskal-Wallis test, it is important to reach conclusions about exactly which treatment is different from which treatment, that is, all $\binom{k}{2} = k(k-1)/2$ individual differences between pairs of treatment effects (τ_i, τ_j), for $i < j$, and these conclusions are naturally two-sided in nature.

Hypothesis

$$\left.\begin{cases} H_0: \tau_1 = \tau_2 & H_1: \tau_1 \neq \tau_2 \\ H_0: \tau_1 = \tau_3 & H_1: \tau_1 \neq \tau_3 \\ \ldots & \\ H_0: \tau_{k-1} = \tau_k & H_1: \tau_{k-1} \neq \tau_k \end{cases}\right\} \frac{k(k-1)}{2} \text{ simultaneous tests/multiple}$$

comparisons

Motivation

To test for each : $\underbrace{H_0: \tau_i = \tau_j \quad H_1: \tau_i \neq \tau_j}_{\text{Two-sample problem : Wilcoxon rank sum test}}$:

$\Rightarrow$ For each pair of treatments (i, j), for $1 \leq i < j \leq k$, let

$$W_{ij} = \sum_{b=1}^{n_j} R_{jb}$$

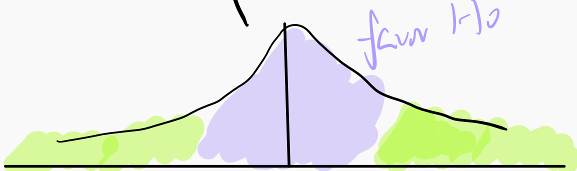
where R_{jb} are the ranks of X_{jb} among the combined i th and j th samples; that is, W_{ij} is the Wilcoxon rank sum of the j th sample ranks in the joint two-sample ranking of the i th and j th sample observations.

$\Rightarrow$ standardized (under H_0) version of W_{ij} multiplied by $\sqrt{2}$

$$W_{ij}^* = \boxed{\sqrt{2}}\left[\frac{W_{ij} - E_0\left(W_{ij}\right)}{\left\{\operatorname{var}_0\left(W_{ij}\right)\right\}^{1/2}}\right] = \frac{W_{ij} - \frac{n_i\left(n_i + n_j + 1\right)}{2}}{\left\{n_i n_j\left(n_i + n_j + 1\right)/24\right\}^{1/2}}, \quad \text{for } 1 \leq i < j \leq k.$$

(Wilcoxon rank sum statistic standardized : $W^* = \dfrac{W - \frac{n(N+1)}{2}}{\sqrt{\frac{mn(N+1)}{12}}}$)

$m = n_i$
$n = n_j$
$N = m+n = n_i + n_j$



favor H_1 $\longrightarrow$ take absolute value $|W_{ij}^*|$

Optimal :

$\Rightarrow$

- When H_0 is true, the $[k(k-1)/2]$-component vector $\left(W_{12}^*, W_{13}^*, \ldots, W_{k-1,k}^*\right)$ has, as $\min(n_1, \ldots, n_k)$ tends to infinity, an asymptotic multivariate normal distribution with mean vector $\mathbf{0}$.
- and when $n_1 = n_2 = \cdots = n_k$,

$$\left(W_{12}^*, W_{13}^*, \ldots, W_{k-1,k}^*\right) \sim (\ldots Z_i - Z_j \ldots)$$

the $[k(k-1)/2]$-component vector of differences where $Z_1 \ldots Z_k$ are indpendent $N(0,1)$

$\Rightarrow$

$$\max_{1 \le i < j \le k} \left| W_{ij}^* \right| \sim range(Z_1 \ldots Z_k)$$

$\Rightarrow$ To get the null distribution for the simultaneous tests, it is equivalent to know the null distribution of $\max_{1 \le i < j \le k} \left| W_{ij}^* \right|$

$\Rightarrow$ It is then equivalent to the distribution of the range when we draw k independent $N(0,1)$

Procedure

critical value of range $(z_1 \ldots z_k)$

For each pair of treatments (i, j), for $1 \leq i < j \leq k$,

Decide $\tau_i \neq \tau_j$ if $\left| W_{ij}^* \right| \geq q_\alpha$; otherwise decide $\tau_u = \tau_v$.

q_α is the upper α quantile of the range of k normal variates.

Example: Length of YOY Gizzard Shad

Site I	Site II	Site III
29(5)	60(15)	33(8)
46(13)	32(7)	26(2)
37(9)	42(10)	25(1)
31(6)	45(12)	28(4)
44(11)	52(14)	27(3)

Simultaneous test :

$H_0: \tau_1 = \tau_2 \quad H_1: \tau_1 \neq \tau_2$

$H_0: \tau_1 = \tau_3 \quad H_1: \tau_1 \neq \tau_3$

$H_0: \tau_2 = \tau_3 \quad H_1: \tau_2 \neq \tau_3$

Control experimentwise error rate to be 10%

```
> library(NSM3)
> cRangeNor(0.1,k=3)
[1] 2.903
```

To compute critical value of 0.1

Decide $\tau_u \neq \tau_v$ if $|W^*_{uv}| \geq 2.903$.

$$W^*_{12} = \frac{[34 - 5(11)/2]}{\sqrt{5 \times 5 \times 11/24}} = 1.92$$

$$W^*_{13} = \frac{[17 - 5(11)/2]}{\sqrt{5 \times 5 \times 11/24}} = -3.10$$

$$W^*_{23} = \frac{[16 - 5(11)/2]}{\sqrt{5 \times 5 \times 11/24}} = -3.397$$

$$\Rightarrow |W^*_{12}| = 1.92 < 2.903 \implies \text{decide } \tau_1 = \tau_2$$

$$|W^*_{13}| = 3.10 > 2.903 \implies \text{decide } \tau_1 \neq \tau_3$$

$$|W^*_{23}| = 3.397 > 2.903 \implies \text{decide } \tau_2 \neq \tau_3$$

10% = 0.1

Thus, at an experimentwise error rate of .05, the multiple comparison decisions can be summarized by the statement $(\tau_1 = \tau_2) \neq (\tau_3)$.

This multiple comparison procedure provides more detailed information about the lengths of the YOY gizzard shad population in Kokosing Lake. We now know that sites I and II may be viewed as providing similar living environments for gizzard shad. However, we also know that the common living environment at sites I and II is significantly different from the common living environment at sites III.