

非参数检验算法

目录

SPSSAU 数据格式..... 1

计算公式..... 2

 1. 基本说明..... 2

 2. MannWhitney 2

 3. Kruskal-Wallis 2

 4. 多重比较..... 3

参考文献..... 3

SPSSAU 中，非参数检验指 X 对于 Y 的差异关系研究，其包括 MannWhitney 和 Kruskal-Wallis 检验，并且 SPSSAU 自动判断选择。在 SPSSAU 中支持：

- ✓ 批量对‘标题’进行分析；
- ✓ 自动选择 MannWhitney 和 Kruskal-Wallis 检验；
- ✓ 支持多重比较。



SPSSAU 数据格式

X	Title1	Title2	Title3
Male	7.859206	7.699458	8.171911
Male	8.27663	9.567024	3.670163
Male	2.098827	3.968077	5.659362
Male	6.024826	0.623412	4.920464
Male	3.102632	4.343761	5.748209
Female	2.963518	7.800297	3.122878
Female	8.232136	0.099224	0.112001
Female	1.798314	8.374911	9.011049
Female	2.194865	0.566183	9.472328
Female	2.442311	7.202591	6.13348
Female	5.312099	1.245504	7.203394
Female	4.655058	5.38763	4.949512

比如上图中 X 为性别，其仅包括两个类别分别是 Male 和 Female，需要研究不同性别对于 3 个 title 的差异关系。

计算公式

1. 基本说明

SPSSAU 算法会自动判断 X 的组别个数，当 X 组别个数为 2 时则使用 MannWhitney 检验，当 X 的组别个数>2 时，则使用 Kruskal-Wallis 检验，分别如下：

2. MannWhitney

- ✓ 将两个样本合并,对所有观测值按大小排序,并赋予相应的秩次,如果有并列值(即存在相同的对比数字),则给予平均秩次。
- ✓ 计算每个样本的秩次和,分别记为 R_1 和 R_2 。
- ✓ 计算 U_1 和 U_2 :

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

- ✓ 计算 MannWhitney U

$$U = \min(U_1, U_2)$$

其中：

n_1 和 n_2 分别是两个组别的样本量

R_1 和 R_2 分别是两个组别的秩次和

- ✓ 计算 z 值

$$T_i = \frac{t^3 - t}{12}$$

$$z = \frac{(U - n_1 n_2 / 2)}{\sqrt{\frac{n_1 n_2}{N(N-1)} \left(\frac{N^3 - N}{12} - \sum_{i=1}^m T_i \right)}}$$

其中：

t ：某个并列值的出现次数

m ：并列值的个数

N ：所有样本的总样本数

3. Kruskal-Wallis

- ✓ 将所有样本合并,对所有观测值按大小排序,并赋予相应的秩次,如果有并列值,则给予平均秩次。
- ✓ 计算统计量 H

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

如果存在并列值(即存在相同的对比数字),需要进行校正

$$H_{\text{corrected}} = \frac{H}{1 - \frac{\sum_{j=1}^m (t_j^3 - t_j)}{N^3 - N}}$$

其中:

t_j : 第 j 个并列值的出现次数

m : 并列值的个数

N : 所有样本的总样本数

k : 组别个数

n_i : 第 i 组的样本量

R_i : 第 i 组的秩次和

4. 多重比较

SPSSAU 中提供三种事后比较,分别是 Dunn's t 检验、Dunn's t 检验(校正 p 值)和 Nemenyi 法,其中校正 p 值使用的具体方法为 'bonferroni'。SPSSAU 借助 Python 中 scikit_posthocs 包实现,可参考链接: <https://scikit-posthocs.rtfd.io>。

参考文献

【1】 The SPSSAU project (2024). SPSSAU. (Version 24.0) [Online Application Software]. Retrieved from <https://www.spssau.com>.

【2】 scikit_posthocs: Statistical Post-hoc Tests for Multiple Comparisons (Version 0.7.0). Available at: <https://scikit-posthocs.rtfd.io>.

【3】 周俊,马世澎. SPSSAU 科研数据分析方法与应用.第 1 版[M]. 电子工业出版社,2024.