

クラスカル・ウォリス検定と多重比較

青木繁伸

2020年3月17日

1 目的

クラスカル・ウォリス検定を行い、引き続き多重比較を行う。

2 使用法

原データを用いる場合

```
import sys
sys.path.append("statlib")
from xtest import Kruskal_Wallis_test
Kruskal_Wallis_test(x, g=None, multiple_comparison=False, verbose)
```

二次データを用いる場合

```
import sys
sys.path.append("statlib")
from xtest import Kruskal_Wallis_test2
Kruskal_Wallis_test2(d, multiple_comparison=False, verbose)
```

2.1 引数

<code>x</code>	データベクトル (リスト) または群ごとのデータの二重リスト
<code>g</code>	<code>x</code> がデータベクトルの場合はそれぞれのデータの所属群を表すベクトル (リスト) <code>x</code> が二重リストの場合は省略される
<code>multiple_comparison</code>	多重比較 (デフォルトは多重比較をしない:False)
<code>d</code>	$k \times m$ 行列
<code>verbose</code>	必要最小限のプリント出力をする

2.2 戻り値の名前

"chisq"	検定統計量 (χ^2 分布にしたがう)
"df"	自由度

"pvalue" p 値
"comparison" 多重比較の結果のデータフレーム
"method" 検定手法名

3 使用例

3.1 使用例 1

データと所属群の2つのベクトル (リスト) で与える。

```
import sys
sys.path.append("statlib")
from xtest import Kruskal_Wallis_test

a = Kruskal_Wallis_test([3, 2, 4, 2, 3, 5, 4, 3, 3, 1, 3, 4, 5, 4, 5],
                        [1, 1, 2, 1, 3, 3, 1, 3, 1, 2, 2, 2, 1, 2, 2])
```

```
Kruskal-Wallis test with multiple comparison
chisq = 0.62667, df = 2, p value = 0.73101
```

群の指定は数値でも文字列 (1文字に限らない) でもよい。

```
b = Kruskal_Wallis_test([3, 2, 4, 2, 3, 5, 4, 3, 3, 1, 3, 4, 5, 4, 5],
                        ["a", "a", "b", "a", "c", "c", "a", "c", "a", "b",
                         "b", "b", "a", "b", "b"])
```

```
Kruskal-Wallis test with multiple comparison
chisq = 0.62667, df = 2, p value = 0.73101
```

3.2 使用例 2

複数の群のデータを二重リストで与える。

使用例 1 と同じデータを群ごとに整理して与える。また、多重比較を行う。

```
c = Kruskal_Wallis_test([[3, 2, 2, 4, 3, 5],
                          [4, 1, 3, 4, 4, 5],
                          [3, 5, 3]], multiple_comparison=True)
```

```
Kruskal-Wallis test with multiple comparison
chisq = 0.62667, df = 2, p value = 0.73101
```

```
Multiple comparison
      chisq  p value
0:1  0.537778 0.764228
0:2  0.326667 0.849308
1:2  0.000741 0.999630
```

3.3 使用例 3

二次データ (クロス集計表) を `Kruskal_Wallis_test2()` に与えることで、検定することができる。
使用例 1, 2 と同じデータをクロス集計して与える。

```
from xtest import Kruskal_Wallis_test2
d = [[0, 2, 2, 1, 1],
      [1, 0, 1, 3, 1],
      [0, 0, 2, 0, 1]]
e = Kruskal_Wallis_test2(d)
```

```
Kruskal-Wallis test with multiple comparison
chisq = 0.62667, df = 2, p value = 0.73101
```

4 既存の Python 関数との比較 `scipy.stats.kruskal()`

`scipy.stats.kruskal()` は複数のリストで指定するが、`statlib.Kruskal_Wallis_test()` では二重リストで指定するという違いはあるが、同じ結果になる。

```
x = [3, 2, 1, 2, 3, 2, 2, 2, 1, 2, 3]
y = [4, 3, 3, 2, 2, 3, 4, 3, 2]
z = [3, 2, 2, 1, 1, 1, 2, 3, 2, 4, 1, 5, 3]
from scipy.stats import kruskal
kruskal(x, y, z)
```

```
KruskalResult(statistic=3.929213737190095, pvalue=0.14021099712855986)
```

```
c = Kruskal_Wallis_test([x, y, z])
```

```
Kruskal-Wallis test with multiple comparison
chisq = 3.9292, df = 2, p value = 0.14021
```

```
print("chisq =", c["chisq"])
```

```
chisq = 3.929213737190111
```

```
print("p value =", c["pvalue"])
```

```
p value = 0.14021099712855875
```