

マクネマー検定（拡張を含む）

青木繁伸

2020年3月17日

1 目的

マクネマー検定を行う。

Python でのマクネマー検定としては `statsmodels.stats.contingency_tables.mcnemar()` がある。しかし、 2×2 より大きな分割表での検定には対処していない。

R では、マクネマー検定を拡張した `mcnemar.test()` を提供している。このような拡張は他に **SPSS** が提供している方法もあるので、それも含めた以下のプログラムを提示する。

2 使用法

```
import sys
sys.path.append("statlib")
from xtest import McNemar_test
McNemar_test(x, y=None, correct=True, exact=True, extension="R", verbose=True)
```

2.1 引数

x データベクトルまたは分割表（2次元配列）
y データベクトル（**x** が分割表の場合は、省略）
correct 2×2 分割表の場合でイエーツの補正を行う場合に **True**（デフォルト）
exact 2×2 分割表の場合に、**exact=True**（デフォルト）の場合は、二項検定による正確な検定を行う
extension
 2×2 より大きな分割表に適用する拡張法。デフォルトでは **R** による拡張。**SPSS** による拡張法を採用するときは "**SPSS**" を指定する。
verbose 必要最小限のプリント出力をする（デフォルトは **True**）

2.2 戻り値の名前

"chisq" χ^2 近似検定の場合に χ^2 値
"df" χ^2 近似検定の場合に自由度
"pvalue" p 値

"table" 分割表 (SPSS の拡張マクネマー検定の場合は 2×2 に縮小された分割表)
"method" 検定の名前

3 使用例

3.1 2×2 分割表の場合

```
x = [[79, 15], [8, 57]]

import sys
sys.path.append("statlib")
from xtest import McNemar_test

a = McNemar_test(x, exact=False)
```

McNemar's chi-squared test with continuity correction
chisq = 1.5652, df = 1, p value = 0.21090

```
a = McNemar_test(x, exact=False, correct = False)
```

McNemar's chi-squared test
chisq = 2.1304, df = 1, p value = 0.14440

```
a = McNemar_test(x, exact=True)
```

Exact binomial test
x = 8, n = 23, p value = 0.21004

3.2 2×2 より大きな分割表の場合

3.2.1 R の場合の拡張法

`extension="R"` で指定する (デフォルトなので省略可)。

```
a = McNemar_test([[11, 3, 2], [4, 9, 5], [1, 2, 13]])
```

McNemar's chi-squared test
chisq = 1.7619, df = 3, p value = 0.62326

次の例のように、 $x_{ij} = x_{ji} = 0$ となるような場合は、割り算の分子・分母が共に 0 になるので注意が必要。 χ^2 統計量には算入しないが、自由度は減らさない (0/NaN = NaN にして、`sp.nansum()` を使う)。

```
a = McNemar_test([[7, 11, 4, 3], [2, 2, 0, 15], [5, 0, 5, 14], [3, 3, 9, 10]])
```

McNemar's chi-squared test
chisq = 15.429, df = 6, p value = 0.01717

カイ二乗値は 15.42884, 自由度は 6, p 値は 0.01717 である。

3.2.2 SPSS の場合の拡張法

定義により、行と列の 2 変数は順序尺度でなければならない。

`extension="SPSS"` で指定する。

```
a = McNemar_test([[11, 3, 2], [4, 9, 5], [1, 2, 13]],
                  extension="SPSS")
```

Exact binomial test

x = 7, n = 17, p value = 0.62906

R での拡張と違い、 $x_{ij} = x_{ji} = 0$ となるような部分を含んでもかまわない。

```
a = McNemar_test([[7, 11, 4, 3], [2, 2, 0, 15], [5, 0, 5, 14], [3, 3, 9, 10]],
                  extension="SPSS")
```

Exact binomial test

x = 22, n = 69, p value = 0.00354

4 既存の Python 関数との比較 `mcnemar()`

4.1 2×2 分割表の場合

`statsmodels.stats.contingency_tables.mcnemar(table, exact=True, correction=True)`

`exact=True` の場合は、二項検定による正確な方法を使う。

`exact=False` の場合は、 χ^2 分布による近似検定である。また、このとき `correction=True` ならば連続性の補正を行う。

2×2 より大きな分割表には対応していない（左上隅の 4 つのセルを対象として計算する）。

```
from statsmodels.stats.contingency_tables import mcnemar
x = [[79, 15], [8, 57]]
print(mcnemar(x, exact=False))
```

pvalue 0.21090292605660677

statistic 1.565217391304348

```
print(mcnemar(x, exact=False, correction=False))
```

pvalue 0.1443997922130672

statistic 2.130434782608696

```
print(mcnemar(x, exact=True))
```

pvalue 0.21003961563110343

statistic 8.0

4.2 2×2 より大きな分割表の場合 `SquareTable` クラス

```
import numpy as np
from statsmodels.stats.contingency_tables import SquareTable

tbl = np.array([[1520, 266, 124, 66],
                [ 234, 1512, 432, 78],
                [ 117, 362, 1772, 205],
                [ 36, 82, 179, 492]])

inst = SquareTable(tbl)
print(inst.symmetry())
```

```
df          6.0
pvalue      0.003987419840428568
statistic   19.106550215266772
```

```
a = McNemar_test(tbl, extension="R")
```

```
McNemar's chi-squared test
chisq = 19.107, df = 6, p value = 0.00399
```

```
print(a["chisq"])
```

```
19.106550215266772
```

```
print(a["pvalue"])
```

```
0.003987419840428568
```

分母が 0 になる場合を含む場合には注意が必要である。

```
tbl2 = np.array([[7, 11, 4, 3],
                 [2, 2, 0, 15],
                 [5, 0, 5, 14],
                 [3, 3, 9, 10]])
```

`symmetry()` は全てのセルに 0.5 を加えてから検定統計量を計算する。

統計量は $9^2/14 + (-1)^2/10 + 0^2/1 + 0^2/7 + 12^2/19 + 5^2/24 = 14.506328320802$ になる。

0 による割り算が起きないように (?), 全ての項の分母に 1 が加えられていることになる。

このため、`McNemar_test()` と結果が異なることになる。

```
a = McNemar_test(tbl2, extension="R")
```

```
McNemar's chi-squared test
chisq = 15.429, df = 6, p value = 0.01717
```

```
print(a["chisq"])
```

```
15.428836863619471
```

```
print(a["pvalue"])
```

```
0.017171133904320426
```

統計量は $9^2/13 + (-1)^2/9 + 0^2/0 + 0^2/6 + 12^2/18 + 5^2/23 = 15.4288368636195$ になる (途中にある $0^2/0$ は計算しない)。