無相関検定

青木繁伸

2020 年 3 月 17 日

1 目的

ピアソンの積率相関係数，スピアマンの順位相関係数，ケンドールの順位相関係数について，無相関検定を行う。

2 使用法

```
import sys
sys.path.append("statlib")
from xtest import cor_test
cor_test(x, y, alternative="two_sided", method="Pearson", exact=True,
         conflevel=0.95, continuity=False, verbose=True)
```

2.1 引数

x, y	同じ長さのデータベクトル
alternative	対立仮説: "two_sided", "less", "greater" のいすれか（デフォルトは "two_sided"）
method	求める相関係数の指定: "Pearson", "Spearman", "Kendall" のいずれか（デフォルトは "Pearson"）
exact	method が "Spearman" または "Kendall" のとき，正確な p 値を求めるかどうか デフォルトは True で，条件を満たす場合は正確な p 値が求められる method="Spearman" の場合，$n \leqq 9$ なら正確な p 値，$n \leqq 1290$ ならエッジワース近似による p 値，$n \geqq 1291$ なら t 分布により漸近近似による p 値 method="Kendall" の場合，$n \leqq 100$ なら正確な p 値，$n \geqq 101$ なら正規分布により漸近近似による p 値 exact=False の場合は，漸近近似による p 値が求められる
conflevel	信頼区間を求めるときの信頼率（デフォルトは 0.95）
continuity	method が "spearman" または "kendall" のとき，検定において連続性の補正を行うかどうか（デフォルトは False）
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

"statistic"	検定統計量
"names"	相関係数の種別
"parameter"	自由度
"pvalue"	p 値
"estimate"	標本相関係数
"nullvalue"	母相関係数
"alternative"	対立仮説の種別
"conflevel"	信頼率
"confint"	信頼区間
"method"	検定手法名

3　使用例

テストデータ

```
x = [2, 4, 3, 2, 1, 2, 3, 4, 5, 6]
y = [3, 4, 2, 6, 5, 4, 7, 6, 5, 7]
```

3.1　ピアソンの積率相関係数，両側検定

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

a = cor_test(x, y)
```

```
Pearson's product-moment correlation
t = 1.0691,  df = 8,  p value = 0.31620
alternative hypothesis: true r is not equal to 0
95 percent confidence interval = [-0.35510, 0.80418]
sample estimates r = 0.35358
```

3.2　ピアソンの積率相関係数，片側検定

```
a = cor_test(x, y, alternative="less")
```

```
Pearson's product-moment correlation
t = 1.0691,  df = 8,  p value = 0.84190
alternative hypothesis: true r is less than 0
```

```
95 percent confidence interval = [-1.00000, 0.75788]
sample estimates r = 0.35358
```

3.3 ピアソンの積率相関係数，片側検定

```
a = cor_test(x, y, alternative="greater")
```

```
Pearson's product-moment correlation
t = 1.0691,  df = 8,  p value = 0.15810
alternative hypothesis: true r is greater than 0
95 percent confidence interval = [-0.24695, 1.00000]
sample estimates r = 0.35358
```

3.4 スピアマンの順位相関係数，両側検定

```
a = cor_test(x, y, method="Spearman", continuity=True)
```

```
Spearman's rank correlation rho
t = 1.0171,  df = 8,  p value = 0.33887 (asymptotic, using t distribution)
alternative hypothesis: true rho is not equal to 0
sample estimates rho = 0.33438
```

```
a = cor_test(x, y, method="Spearman")
```

```
Spearman's rank correlation rho
t = 1.0035,  df = 8,  p value = 0.34499 (asymptotic, using t distribution)
alternative hypothesis: true rho is not equal to 0
sample estimates rho = 0.33438
```

3.5 スピアマンの順位相関係数，片側検定

```
a = cor_test(x, y, method="Spearman", alternative="less")
```

```
Spearman's rank correlation rho
t = 1.0035,  df = 8,  p value = 0.82751 (asymptotic, using t distribution)
alternative hypothesis: true rho is less than 0
sample estimates rho = 0.33438
```

3.6 スピアマンの順位相関係数，片側検定

```
a = cor_test(x, y, method="Spearman", alternative="greater")
```

```
Spearman's rank correlation rho
t = 1.0035,  df = 8,  p value = 0.17249 (asymptotic, using t distribution)
alternative hypothesis: true rho is greater than 0
sample estimates rho = 0.33438
```

3.7 ケンドールの順位相関係数，両側検定

```
a = cor_test(x, y, method="Kendall")
```

```
Kendall's rank correlation tau
Z = 0.92937, p value = 0.35270 (asymptotic, using normal distribution)
alternative hypothesis: true tau is not equal to 0
sample estimates tau = 0.24693
```

3.8 ケンドールの順位相関係数，片側検定

```
a = cor_test(x, y, method="Kendall", alternative="less")
```

```
Kendall's rank correlation tau
Z = 0.92937, p value = 0.82365 (asymptotic, using normal distribution)
alternative hypothesis: true tau is less than 0
sample estimates tau = 0.24693
```

3.9 ケンドールの順位相関係数，片側検定

```
a = cor_test(x, y, method="Kendall", alternative="greater")
```

```
Kendall's rank correlation tau
Z = 0.92937, p value = 0.17635 (asymptotic, using normal distribution)
alternative hypothesis: true tau is greater than 0
sample estimates tau = 0.24693
```

3.10 正確な p 値

順位相関係数の場合には，同順位（タイ）がなく，サンプルサイズが小さい場合には正確な p 値が計算される。

3.10.1 スピアマンの順位相関係数の場合

exact=True（デフォルト）が指定され，$n \leqq 9$ で同順位がない場合には正確な p 値が計算される。

```
x = [45.8, 50.5, 56.8, 60.8, 64.8]
```

```
y = [42.5, 45.8, 64.9, 53.6, 57.4]
b = cor_test(x, y, method="Spearman", exact=False)
```

```
Spearman's rank correlation rho
t = 1.6977,  df = 3,  p value = 0.18812 (asymptotic, using t distribution)
alternative hypothesis: true rho is not equal to 0
sample estimates rho = 0.70000
```

```
b = cor_test(x, y, method="Spearman")
```

```
Spearman's rank correlation rho
S = 6.00000, p value = 0.23333 (exact p value)
alternative hypothesis: true rho is not equal to 0
sample estimates rho = 0.70000
```

```
c = cor_test(x, y, method="Spearman", alternative="less")
```

```
Spearman's rank correlation rho
S = 6.00000, p value = 0.93333 (exact p value)
alternative hypothesis: true rho is less than 0
sample estimates rho = 0.70000
```

```
d = cor_test(x, y, method="Spearman", alternative="greater")
```

```
Spearman's rank correlation rho
S = 6.00000, p value = 0.11667 (exact p value)
alternative hypothesis: true rho is greater than 0
sample estimates rho = 0.70000
```

exact=True（デフォルト）が指定され，$10 \leqq n \leqq 1290$ で同順位がない場合には Edgeworth 近似による p 値が計算される。

```
import numpy as np
np.random.seed(123)
x = np.random.randn(1290)
y = np.random.randn(1290)
b = cor_test(x, y, method="Spearman")
```

```
Spearman's rank correlation rho
S = 337505304.00000, p value = 0.04185 (Edgeworth approximation)
alternative hypothesis: true rho is not equal to 0
sample estimates rho = 0.05667
```

```
c = cor_test(x, y, method="Spearman", alternative="less")
```

```
Spearman's rank correlation rho
S = 337505304.00000, p value = 0.97907 (Edgeworth approximation)
alternative hypothesis: true rho is less than 0
sample estimates rho = 0.05667
```

```
d = cor_test(x, y, method="Spearman", alternative="greater")
```

```
Spearman's rank correlation rho
S = 337505304.00000, p value = 0.02093 (Edgeworth approximation)
alternative hypothesis: true rho is greater than 0
sample estimates rho = 0.05667
```

exact=False が指定された場合，同順位がある場合または $n \geqq 1291$ の場合には t 分布で近似した p 値が計算される。

```
x = np.random.randn(1291)
y = np.random.randn(1291)
b = cor_test(x, y, method="Spearman")
```

```
Spearman's rank correlation rho
t = -0.89477,  df = 1289,  p value = 0.37108 (asymptotic, using t distribution)
alternative hypothesis: true rho is not equal to 0
sample estimates rho = -0.02491
```

```
c = cor_test(x, y, method="Spearman", alternative="less")
```

```
Spearman's rank correlation rho
t = -0.89477,  df = 1289,  p value = 0.18554 (asymptotic, using t distribution)
alternative hypothesis: true rho is less than 0
sample estimates rho = -0.02491
```

```
d = cor_test(x, y, method="Spearman", alternative="greater")
```

```
Spearman's rank correlation rho
t = -0.89477,  df = 1289,  p value = 0.81446 (asymptotic, using t distribution)
alternative hypothesis: true rho is greater than 0
sample estimates rho = -0.02491
```

3.10.2 ケンドールの順位相関係数の場合

exact=True（デフォルト）が指定され，$n \leqq 50$ で同順位がない場合には正確な p 値が計算される。

```
x = [45.8, 50.5, 56.8, 60.8, 64.8]
y = [42.5, 45.8, 64.9, 53.6, 57.4]
b = cor_test(x, y, method="Kendall", exact=False)
```

```
Kendall's rank correlation tau
Z = 1.46969, p value = 0.14164 (asymptotic, using normal distribution)
alternative hypothesis: true tau is not equal to 0
sample estimates tau = 0.60000
```

```
b = cor_test(x, y, method="Kendall")
```

```
Kendall's rank correlation tau
T = 8.00000, p value = 0.23333 (exact p value)
alternative hypothesis: true tau is not equal to 0
sample estimates tau = 0.60000
```

```
c = cor_test(x, y, method="Kendall", alternative="less")
```

```
Kendall's rank correlation tau
T = 8.00000, p value = 0.95833 (exact p value)
alternative hypothesis: true tau is less than 0
sample estimates tau = 0.60000
```

```
d = cor_test(x, y, method="Kendall", alternative="greater")
```

```
Kendall's rank correlation tau
T = 8.00000, p value = 0.11667 (exact p value)
alternative hypothesis: true tau is greater than 0
sample estimates tau = 0.60000
```

```
x = np.random.randn(50)
y = np.random.randn(50)
b = cor_test(x, y, method="Kendall")
```

```
Kendall's rank correlation tau
T = 626.00000, p value = 0.82873 (exact p value)
alternative hypothesis: true tau is not equal to 0
sample estimates tau = 0.02204
```

```
c = cor_test(x, y, method="Kendall", alternative="less")
```

```
Kendall's rank correlation tau
T = 626.00000, p value = 0.59211 (exact p value)
alternative hypothesis: true tau is less than 0
sample estimates tau = 0.02204
```

```
d = cor_test(x, y, method="Kendall", alternative="greater")
```

```
Kendall's rank correlation tau
T = 626.00000, p value = 0.41436 (exact p value)
alternative hypothesis: true tau is greater than 0
sample estimates tau = 0.02204
```

```
x = np.random.randn(51)
y = np.random.randn(51)
b = cor_test(x, y, method="Kendall")
```

```
Kendall's rank correlation tau
Z = 0.43048, p value = 0.66685 (asymptotic, using normal distribution)
```

```
alternative hypothesis: true tau is not equal to 0
sample estimates tau = 0.04157
```

```
c = cor_test(x, y, method="Kendall", alternative="less")
```

```
Kendall's rank correlation tau
Z = 0.43048, p value = 0.66658 (asymptotic, using normal distribution)
alternative hypothesis: true tau is less than 0
sample estimates tau = 0.04157
```

```
d = cor_test(x, y, method="Kendall", alternative="greater")
```

```
Kendall's rank correlation tau
Z = 0.43048, p value = 0.33342 (asymptotic, using normal distribution)
alternative hypothesis: true tau is greater than 0
sample estimates tau = 0.04157
```