

polychoric 相関係数

青木繁伸

2020年3月17日

1 目的

$k \times m$ 分割表として集計されている 2 変数が、潜在的には正規分布に従うと仮定できる場合について polychoric 相関係数を求める。tetrachoric 相関係数を求めるために別の関数が用意されることがあるが、 2×2 分割表に対して polychoric 相関係数を求めるのと同じであるので、区別しない。

R の **polycor** パッケージにある `polychor()` を **Python** に翻訳・修正したものである。

R の **polycor** の情報

```
Package:                polycor
Version:                0.7-9
Date:                  2016-08-26
Title:                 Polychoric and Polyserial Correlations
Authors@R:             person("John", "Fox", role = c("aut", "cre"), email =
                        "jfox@mcmaster.ca")
Depends:               R (>= 3.3.0)
Imports:               stats, mvtnorm, Matrix
ByteCompile:          yes
LazyLoad:              yes
Description:           Computes polychoric and polyserial correlations by quick
                        "two-step" methods or ML, optionally with standard errors;
                        tetrachoric and biserial correlations are special cases.
License:               GPL (>= 2)
URL:                   https://r-forge.r-project.org/projects/polycor/,
                        http://CRAN.R-project.org/package=polycor
Author:                John Fox [aut, cre]
Maintainer:            John Fox <jfox@mcmaster.ca>
Repository:            CRAN
Repository/R-Forge/Project:  polycor
Repository/R-Forge/Revision:  13
Repository/R-Forge/DateTimeStamp:  2016-08-26 18:25:37
Date/Publication:      2016-08-27 00:22:11
NeedsCompilation:      no
```

Packaged: 2016-08-26 18:45:33 UTC; rforge
Built: R 3.5.0; ; 2018-04-23 16:48:15 UTC; unix

参考文献

Dragow, F. (1986) Polychoric and polyserial correlations. Pp. 68 – 74 in S. Kotz and N. Johnson, eds., *The Encyclopedia of Statistics, Volume 7*. Wiley.

Olsson, U. (1979) Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika* **44**, 443-460.

2 使用法

```
import sys
sys.path.append("statlib")
from multi import polychoric
polychoric(x, y=None, two_step=False, initval=0.9, verbose=True)
```

2.1 引数

<code>x</code>	変数ベクトル, または二重リスト・二次元配列 (二重クロス集計表)
<code>y</code>	<code>x</code> が変数ベクトルの場合は変数ベクトル
<code>two_step</code>	相関係数だけを推定するときは <code>False</code> , 変数のカットポイントも含めて推定するときは <code>True</code> 。
<code>initval</code>	相関係数の初期値 (デフォルトでは 0.9)
<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

<code>"method"</code>	分析手法
<code>"rho"</code>	相関係数の推定値
<code>"SE"</code>	相関係数の標準誤差
<code>"chisq"</code>	元の変数が二変量正規分布に従うかどうかの検定統計量 (χ^2 分布にしたがう)
<code>"df"</code>	自由度
<code>"pvalue"</code>	p 値
<code>"rowCuts"</code>	行のカットポイント
<code>"colCuts"</code>	列のカットポイント

3 使用例

```
x = [[155, 437, 162, 21],
      [7, 88, 84, 46]]
```

```
import sys
sys.path.append("statlib")
from multi import polychoric

a = polychoric(x)
```

Polychoric Correlation: ML estimator = 0.52314, Std.Err. = 0.038468
chisq = 2.7387, df = 2, p = 0.25427

Row Treshold

	Threshold	Std.Err.
0	0.753609	0.02552

Column Treshold

	Threshold	Std.Err.
0	-0.984497	0.064519
1	0.483815	0.053894
2	1.500813	0.120374

```
b = polychoric(x, two_step=True)
```

Polychoric Correlation: two step ML estimator = 0.52306, Std.Err. = 0.037285
chisq = 2.7578, df = 2, p = 0.25186

Row Treshold

[0.75541503]

Column Treshold

[-0.9862713 0.48736457 1.49851307]

polyserial 相関係数の場合

```
y = [[103, 89], [265, 218]]

c = polychoric(y)
```

Polychoric Correlation: -0.018451

Row Treshold: -0.56969

Column Treshold: 0.11351

```
import numpy as np
import pandas as pd
from multi import gendat # 相関係数を指定してテストデータを発生
np.random.seed(123)
d = gendat(300, 0.5)
```

```
x = [sum(w > np.array([-np.Inf, -0.5, 0, 0.5, np.Inf])) for w in d[:,
0]]
y = [sum(w > np.array([-np.Inf, -0.7, 0, 0.5, 1.2, np.Inf])) for w in
d[:, 1]]
d = pd.DataFrame({"x": x, "y": y})
tbl = np.array(pd.crosstab(d["x"], d["y"]))
print(tbl)
```

```
[[46 17 20 10  2]
 [12 20 12 10  3]
 [12 14 16 13  5]
 [ 5 16 19 22 26]]
```

```
a1 = polychoric(x, y)
```

```
Polychoric Correlation: ML estimator = 0.52827, Std.Err. = 0.051394
chisq = 8.8677, df = 11, p = 0.63411
```

Row Treshold

	Threshold	Std.Err.
0	-0.475275	0.091031
1	0.011840	0.092689
2	0.538377	0.103639

Column Treshold

	Threshold	Std.Err.
0	-0.666354	0.111374
1	-0.064110	0.097286
2	0.504842	0.103553
3	1.171224	0.143286

```
a2 = polychoric(tbl)
```

```
Polychoric Correlation: ML estimator = 0.52827, Std.Err. = 0.051394
chisq = 8.8677, df = 11, p = 0.63411
```

Row Treshold

	Threshold	Std.Err.
0	-0.475275	0.091031
1	0.011840	0.092689
2	0.538377	0.103639

Column Treshold

	Threshold	Std.Err.
0	-0.666354	0.111374
1	-0.064110	0.097286

2	0.504842	0.103553
3	1.171224	0.143286