

polyserial 相関係数

青木繁伸

2020年3月17日

1 目的

2変数のうちの片方が連続変数, もう一方が順序カテゴリー変数の場合の相関係数を求める。順序カテゴリー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 polyserial 相関係数を求める。

R の **polycor** パッケージにある `polyserial()` 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.

2 使用法

```
import sys
sys.path.append("statlib")
from multi import polyserial
polyserial(x, y, two_step=False, bins=4, verbose=True)
```

2.1 引数

x	連続変数ベクトル
y	順序カテゴリー変数ベクトル
two_step	相関係数だけを推定するときは False, 順序カテゴリー変数のカットポイントも含めて推定するときは True。
bins	連続変数の分割数
verbose	必要最小限のプリント出力をする

2.2 戻り値の名前

"method"	分析手法
"rho"	相関係数の推定値
"SE"	相関係数の標準誤差
"chisq"	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)
"df"	自由度
"pvalue"	p 値
"colCuts"	順序カテゴリー変数のカットポイント

3 使用例

```
x = [0.7492, -0.22295, 0.11472, 0.53715, -0.51242, 0.35807, 0.49264,
-0.51353, -0.94197, 1.15984, 1.12984, -1.02436, -1.07608, -0.30467,
0.54926, 1.35279, 2.40187, 0.37274, -0.74321, 1.71418, 0.47398,
-0.78159, 1.2034, -1.21809, 0.22509, -0.01787, 0.14278, -0.57617,
0.88083, 1.46454, -0.20298, 0.94596, -1.04114, 1.2703, 0.1639,
-0.15033, -0.41042, 1.59887, 0.58806, 0.72434, 0.89338, 0.34713,
1.42137, 0.56944, 0.73173, -1.15237, 1.72124, -0.76932, 0.07062,
-0.75912, -0.08296, 0.06515, -0.00246, 0.10176, -0.34965, -1.53095,
```

```

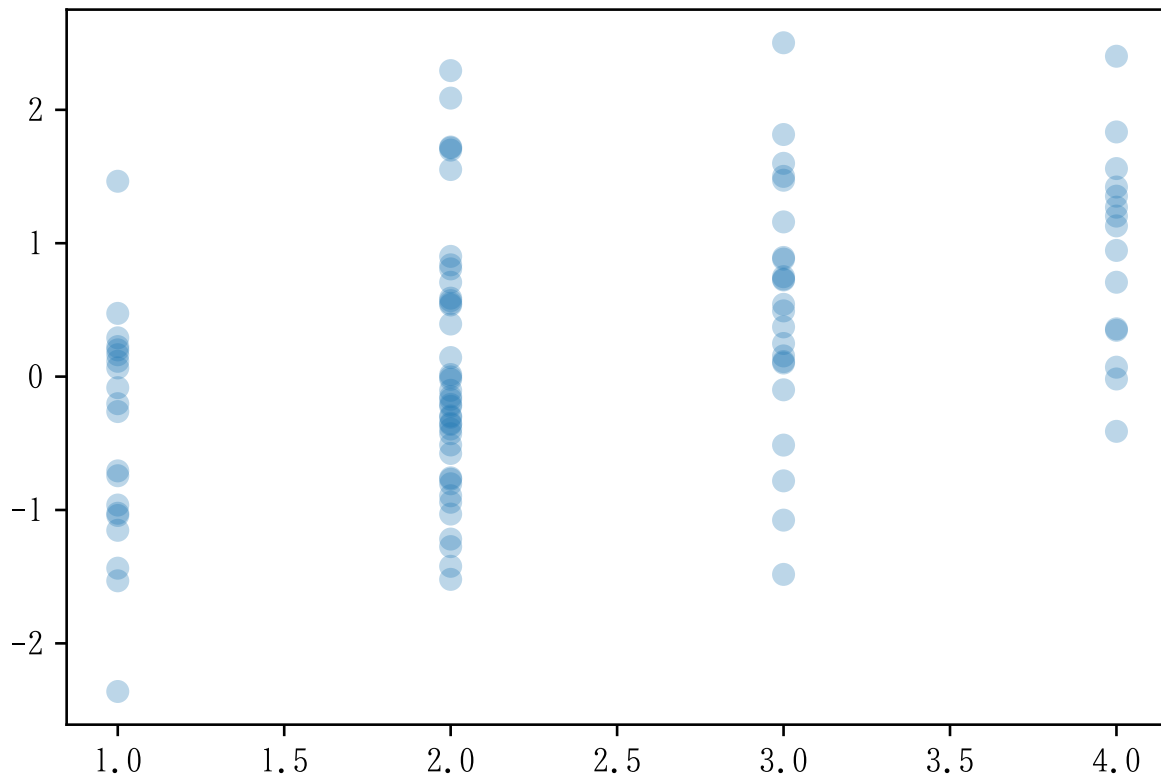
1.55971, -0.20873, 0.11238, 2.08786, 1.83424, 0.70741, -0.70681,
-1.48295, -2.36075, 1.81494, 0.20038, 2.29407, -0.80354, 0.0154,
-1.43625, -0.96287, 0.15512, -1.27466, 0.90138, -1.42222, -0.02011,
-0.16987, 1.55233, -0.10292, -0.39256, -0.35841, -1.03065, 0.29056,
1.69807, 0.54179, 0.24826, -1.52018, 2.50281, -0.2621, -0.89337,
-0.09901, 0.80746, 1.50104, -0.29506, 0.39425, 0.83788, 1.47223,
0.70706, -0.42811]
y = [3, 2, 1, 2, 2, 4, 3, 3, 2, 3, 4, 1, 3, 2, 2, 4, 4, 3, 1, 2,
1, 3, 4, 2, 1, 4, 2, 2, 3, 1, 1, 4, 1, 4, 1, 2, 4, 3, 2, 3, 3,
4, 4, 2, 3, 1, 2, 2, 4, 2, 1, 1, 2, 3, 2, 1, 4, 2, 3, 2, 4, 4,
1, 3, 1, 3, 1, 2, 2, 2, 1, 1, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
1, 2, 3, 3, 2, 3, 1, 2, 3, 2, 3, 2, 2, 2, 3, 2, 2]

```

```

import matplotlib.pyplot as plt
plt.scatter(y, x, alpha=0.3)

```



```

import sys
sys.path.append("statlib")
from multi import polyserial

a = polyserial(x, y, two_step = True)

```

Polyserial Correlation: two step ML estimator = 0.467, Std.Err. = 0.079169
Test of bivariate normality: chisq = 5.4889, df = 11, p value= 0.9052

```
a = polyserial(x, y)
```

Polyserial Correlation: ML estimator = 0.46619, Std.Err. = 0.080643

Test of bivariate normality: chisq = 5.4841, df = 11, p value= 0.90548

	Threshold	Std.Err.
0	-0.845491	0.133969
1	0.305882	0.118274
2	1.042709	0.144710

```
a = polyserial(x, y, bins=6)
```

Polyserial Correlation: ML estimator = 0.46619, Std.Err. = 0.080643

Test of bivariate normality: chisq = 12.914, df = 19, p value= 0.84293

	Threshold	Std.Err.
0	-0.845491	0.133969
1	0.305882	0.118274
2	1.042709	0.144710