

総当たり法による正準判別分析

青木繁伸

2020年3月17日

1 目的

総当たり法による正準判別分析を行う。

2 使用法

```
import sys
sys.path.append("statlib")
from multi import APSS_candis
APSS_candis(dat, sort_by="Correct rate")
```

2.1 引数

<code>dat</code>	群を表す変数が最右列にあるようなデータフレーム
<code>sort_by</code>	デフォルト ("Correct rate") では正判別率の大きい順に並べ替える。 "Wilks' lambda" を指定すると、Wilks' の λ の小さい順に並べ替える。 "Can. corr. coeff." を指定すると、正準相関係数の大きい順に並べ替える。 なお、並べ替えに使う Wilks' の λ と正準相関係数は第1正準判別解のものである。

2.2 戻り値

結果の総括表を返す。

3 使用例

```
import pandas as pd
import sys
sys.path.append("statlib")
from multi import APSS_candis

data = pd.read_csv("data/iris.csv")
```

以下のようなデータフレームにおいて、`sl`、`sw`、`pl`、`pw`の4変数を用いて`sp`を予測(判別)する。

```
data.head()
```

```
   sl  sw  pl  pw  sp
0  5.1  3.5  1.4  0.2  setosa
1  4.9  3.0  1.4  0.2  setosa
2  4.7  3.2  1.3  0.2  setosa
3  4.6  3.1  1.5  0.2  setosa
4  5.0  3.6  1.4  0.2  setosa
```

結果の表示

```
APSS_candis(data)
```

	Correct rate	Wilks' lambda	Can. corr. coeff.	variables
6	98.000000	0.024976	0.983818	[sw, pl, pw]
14	98.000000	0.023439	0.984821	[sl, sw, pl, pw]
10	97.333333	0.030580	0.981906	[sl, pl, pw]
4	96.666667	0.038316	0.976028	[sw, pw]
9	96.666667	0.039878	0.979237	[sl, pl]
13	96.666667	0.031546	0.980800	[sl, sw, pl]
0	96.000000	0.071117	0.963786	[pw]
2	96.000000	0.043777	0.975519	[pl, pw]
8	96.000000	0.071044	0.963792	[sl, pw]
12	96.000000	0.035025	0.977679	[sl, sw, pw]
5	95.333333	0.036884	0.977690	[sw, pl]
1	94.666667	0.058628	0.970243	[pl]
11	80.000000	0.166544	0.898133	[sl, sw]
7	74.666667	0.381294	0.786578	[sl]
3	55.333333	0.599217	0.633074	[sw]

```
APSS_candis(data, sort_by="Wilks' lambda")
```

	Correct rate	Wilks' lambda	Can. corr. coeff.	variables
14	98.000000	0.023439	0.984821	[sl, sw, pl, pw]
6	98.000000	0.024976	0.983818	[sw, pl, pw]
10	97.333333	0.030580	0.981906	[sl, pl, pw]
13	96.666667	0.031546	0.980800	[sl, sw, pl]
12	96.000000	0.035025	0.977679	[sl, sw, pw]
5	95.333333	0.036884	0.977690	[sw, pl]
4	96.666667	0.038316	0.976028	[sw, pw]
9	96.666667	0.039878	0.979237	[sl, pl]
2	96.000000	0.043777	0.975519	[pl, pw]
1	94.666667	0.058628	0.970243	[pl]

8	96.000000	0.071044	0.963792	[sl, pw]
0	96.000000	0.071117	0.963786	[pw]
11	80.000000	0.166544	0.898133	[sl, sw]
7	74.666667	0.381294	0.786578	[sl]
3	55.333333	0.599217	0.633074	[sw]

```
APSS_candis(data, sort_by="Can. corr. coeff.")
```

	Correct rate	Wilks' lambda	Can. corr. coeff.	variables
14	98.000000	0.023439	0.984821	[sl, sw, pl, pw]
6	98.000000	0.024976	0.983818	[sw, pl, pw]
10	97.333333	0.030580	0.981906	[sl, pl, pw]
13	96.666667	0.031546	0.980800	[sl, sw, pl]
9	96.666667	0.039878	0.979237	[sl, pl]
5	95.333333	0.036884	0.977690	[sw, pl]
12	96.000000	0.035025	0.977679	[sl, sw, pw]
4	96.666667	0.038316	0.976028	[sw, pw]
2	96.000000	0.043777	0.975519	[pl, pw]
1	94.666667	0.058628	0.970243	[pl]
8	96.000000	0.071044	0.963792	[sl, pw]
0	96.000000	0.071117	0.963786	[pw]
11	80.000000	0.166544	0.898133	[sl, sw]
7	74.666667	0.381294	0.786578	[sl]
3	55.333333	0.599217	0.633074	[sw]