

# ステップワイズ判別分析

青木繁伸

2020年3月17日

## 1 目的

ステップワイズ変数選択による判別分析を行う。

## 2 使用法

```
import sys
sys.path.append("statlib")
from multi import sdis
sdis(data, stepwise=True, P_in=0.05, P_out=0.05, predict=False, verbose=True)
```

### 2.1 引数

<code>data</code>	判別される変数が最右列にあるようなデータフレーム
<code>stepwise</code>	ステップワイズ変数選択をするかどうか (デフォルトは <code>True</code> )
<code>P_in</code>	変数編入基準 (デフォルトは 0.05)
<code>P_out</code>	編入除去基準 (デフォルトは 0.05)
<code>predict</code>	判別値を出力するかどうか (デフォルトは <code>False</code> であるが、判別値は必ず計算される)
<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>"function"</code>	分類関数
<code>"classification"</code>	分類結果詳細
<code>"result"</code>	分類結果集計表

## 3 使用例

```
import pandas as pd
```

```

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

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

ans = sdis(data, stepwise=True, predict=False)

```

Valid cases : 150

Group variable: sp

\*\*\*\*\* Means \*\*\*\*\*

	setosa	versicolor	virginica	Grand Mean
sl	5.006	5.936	6.588	5.843333
sw	3.428	2.770	2.974	3.057333
pl	1.462	4.260	5.552	3.758000
pw	0.246	1.326	2.026	1.199333

\*\*\*\*\* Pooled Within Correlation Coefficient matrix \*\*\*\*\*

	sl	sw	pl	pw
sl	1.000000	0.530236	0.756164	0.364506
sw	0.530236	1.000000	0.377916	0.470535
pl	0.756164	0.377916	1.000000	0.484459
pw	0.364506	0.470535	0.484459	1.000000

Threshold to enter Pin: 0.050000

Threshold to remove Pout: 0.050000

\*\*\*\*\* Candidate to enter: pl P : < 0.0001 entered

\*\*\*\*\* STEP 1: entered variable: pl

\*\*\*\*\* Classification Function \*\*\*\*\*

	setosa	versicolor	virginica	F value	p value
pl	-15.789381	-46.007362	-59.960768	1180.16118	< 0.0001
constant	11.542037	97.995680	166.451092		

Wilks' Lambda: 0.05863

Equivalent F value: 1180.16118

Degrees of Freedom: (2, 147.000)

P value: < 0.0001

\*\*\*\*\* Candidate to remove: pl P : < 0.0001 not removed

\*\*\*\*\* Candidate to enter: sw P : < 0.0001 entered

\*\*\*\*\* STEP 2: entered variable: sw

\*\*\*\*\* Classification Function \*\*\*\*\*

	setosa	versicolor	virginica	F value	p value
pl	2.257843	-36.964081	-52.011816	1112.95382	< 0.0001
sw	-60.498025	-30.314945	-26.646530	43.03545	< 0.0001
constant	102.043131	120.719691	184.008191		

Wilks' Lambda: 0.03688

Equivalent F value: 307.10467

Degrees of Freedom: (4, 292.000)

P value: < 0.0001

\*\*\*\*\* Candidate to remove: sw P : < 0.0001 not removed

\*\*\*\*\* Candidate to enter: pw P : < 0.0001 entered

\*\*\*\*\* STEP 3: entered variable: pw

\*\*\*\*\* Classification Function \*\*\*\*\*

	setosa	versicolor	virginica	F value	p value
pl	-6.186436	-36.458190	-46.174379	38.72447	< 0.0001
sw	-70.526325	-29.714156	-19.714077	54.57694	< 0.0001
pw	49.636953	-2.973715	-34.313479	34.56869	< 0.0001
constant	119.299060	120.781625	192.254462		

Wilks' Lambda: 0.02498

Equivalent F value: 257.50317

Degrees of Freedom: (6, 290.000)

P value: < 0.0001

\*\*\*\*\* Candidate to remove: pw P : < 0.0001 not removed

\*\*\*\*\* Candidate to enter: sl P : 0.0103 entered

\*\*\*\*\* STEP 4: entered variable: sl

\*\*\*\*\* Classification Function \*\*\*\*\*

	setosa	versicolor	virginica	F value	p value
pl	32.861278	-10.422902	-25.533090	35.59017	< 0.0001
sw	-47.175741	-14.145020	-7.370559	21.93593	< 0.0001
pw	34.796822	-12.868458	-42.158226	24.90433	< 0.0001
sl	-47.088333	-31.396418	-24.891698	4.72115	0.0103
constant	170.419715	143.507990	206.539415		

Wilks' Lambda: 0.02344  
 Equivalent F value: 199.14534  
 Degrees of Freedom: (8, 288.000)  
 P value: < 0.0001

\*\*\*\*\* Candidate to remove: sl P : 0.0103 not removed

===== Final Results =====

\*\*\*\*\* Classification Function \*\*\*\*\*

	setosa	versicolor	virginica	F value	p value
pl	32.861278	-10.422902	-25.533090	35.59017	< 0.0001
sw	-47.175741	-14.145020	-7.370559	21.93593	< 0.0001
pw	34.796822	-12.868458	-42.158226	24.90433	< 0.0001
sl	-47.088333	-31.396418	-24.891698	4.72115	0.0103
constant	170.419715	143.507990	206.539415		

Wilks' Lambda: 0.02344  
 Equivalent F value: 199.14534  
 Degrees of Freedom: (8, 288.000)  
 P value: 4

\*\*\*\*\* Discriminant Function \*\*\*\*\*

\*\*\*\*\* Discrimination between setosa and versicolor

Mahalanobis' Distance: 9.47967  
 Probability of Misclassification: < 0.0001

	Disc. Coef.	Std. Disc. Coef.
pl	-21.642090	-38.07718
sw	16.515361	7.17445
pw	-23.832640	-18.10548

sl	7.845958	6.47528
constant	-13.455863	

\*\*\*\*\* Discrimination between setosa and virginica

Mahalanobis' Distance: 13.39346

Prpbability of Misclassification: < 0.0001

	Disc. Coef.	Std. Disc. Coef.
pl	-29.197184	-51.36964
sw	19.902591	8.6459
pw	-38.477524	-29.23109
sl	11.098318	9.15946
constant	18.059850	

\*\*\*\*\* Discrimination between versicolor and virginica

Mahalanobis' Distance: 4.14742

Prpbability of Misclassification: 0.0191

	Disc. Coef.	Std. Disc. Coef.
pl	-7.555094	-13.29246
sw	3.387230	1.47145
pw	-14.644884	-11.12561
sl	3.252360	2.68418
constant	31.515713	

\*\*\*\*\* Classification Result \*\*\*\*\*

	Classified Group		
	setosa	versicolor	virginica
setosa	50	0	0
versicolor	0	48	2
virginica	0	1	49

Correct rate = 98.0