

ステップワイズ重回帰分析

青木繁伸

1 目的

ステップワイズ重回帰分析を行う。

2 使用法

```
from sreg import sreg
sreg(data, stepwise=True, P_in=0.05, P_out=0.05, predict=True, verbose=True)
```

2.1 引数

data	従属変数が最右列にあるようなデータフレーム
stepwise	ステップワイズ変数選択をする (デフォルトは True)
P_in	変数編入基準
P_out	変数除去基準
predict	予測値を求める (デフォルトは True)
verbose	必要最小限のプリント出力をする

2.2 戻り値の名前

"Results of analysis"	偏回帰係数などの結果をまとめたデータフレーム
"analysis of variance table"	回帰の分散分析表のデータフレーム
"coefficient of determination"	決定係数など
"prediction"	予測値

3 使用例

```
import pandas as pd

data = pd.DataFrame({
    "x1": [1.2, 1.6, 3.5, 4., 5.6, 5.7, 6.7, 7.5, 8.5, 9.7],
    "x2": [1.9, 2.7, 3.7, 3.1, 3.5, 7.5, 1.2, 3.7, 0.6, 5.1],
    "y": [0.9, 1.3, 2., 1.8, 2.2, 3.5, 1.9, 2.7, 2.1, 3.6]})

import sys
sys.path.append("statlib")
from sreg import sreg
```

```
ans = sreg(data)
```

```
Valid cases :      10
```

```
Dependent variable : y
```

```
***** Basic Statistics *****
```

	Mean	Variance	S.D.
x1	5.4	7.997778	2.828034
x2	3.3	3.922222	1.980460
y	2.2	0.744444	0.862812

```
***** Correlation Coefficients Matrix *****
```

	x1	x2	y
x1	1.000000	0.128156	0.754990
x2	0.128156	1.000000	0.743877
y	0.754990	0.743877	1.000000

```
Threshold to enter Pin: 0.05
```

```
Threshold to remove Pout: 0.05
```

```
Candidate variable to enter: x1 P : 0.0116 ***** entered
```

```
***** STEP 1 *****
```

```
entered variable: x1
```

```
***** Estimated Partial Regression Coefficients *****
```

	Estimate	Std. err.	t value	p value	Std. coef.	Tolerance	VIF
x1	0.230342	0.070732	3.256544	0.0116	0.75499	1.0	1.0
constant	0.956154	0.426497	2.241880	0.0553			

```
Residual standard error: 0.6001 on 8 degrees of freedom
```

```
Multiple R: 0.7550, Multiple R-squared: 0.5700, Adjusted R-squared: 0.5163
```

```
ANOVA F-statistic: 10.605 on 1 and 8 degrees of freedom, p-value: 0.0116
```

```
Log-likelihood: -7.96703, AIC: 21.93406
```

```
***** Analysis of Variance Table for Multiple Regression Analysis *****
```

	SS	df	MS	F value	p value
Regression	3.819066	1	3.81907	10.6051	0.0116
Residual	2.880934	8	0.36012		
Total	6.700000	9			

```
increased R2: 0.57001, as F value: 10.60508, d.f.: (1, 8)
```

p value of F: 0.01159

Candidate variable to remove: x1 P : 0.0116 ***** not removed

Candidate variable to enter: x2 P : < 0.0001 ***** entered

***** STEP 2 *****

entered variable: x2

***** Estimated Partial Regression Coefficients *****

	Estimate	Std. err.	t value	p value	Std. coef.	Tolerance \
x1	0.204617	0.007564	27.050556	< 0.0001	0.67067	0.98358
x2	0.286634	0.010802	26.536454	< 0.0001	0.65793	0.98358
constant	0.149176	0.054506	2.736849	0.029		

	VIF
x1	1.0167
x2	1.0167
constant	

Residual standard error: 0.0636 on 7 degrees of freedom

Multiple R: 0.9979, Multiple R-squared: 0.9958, Adjusted R-squared: 0.9946

ANOVA F-statistic: 823.477 on 2 and 7 degrees of freedom, p-value: < 0.0001

Log-likelihood: 15.13807, AIC: -22.27614

***** Analysis of Variance Table for Multiple Regression Analysis *****

	SS	df	MS	F value	p value
Regression	6.671644	2	3.33582	823.477	< 0.0001
Residual	0.028356	7	0.00405		
Total	6.700000	9			

increased R2: 0.42576, as F value: 704.18341, d.f.: (1, 7)

p value of F: 0.00000

Candidate variable to remove: x2 P : < 0.0001 ***** not removed

***** Prediction *****

	Observed	Predicted	Residuals	Std. Resid.
0	0.9	0.939320	-0.039320	-0.781587
1	1.3	1.250474	0.049526	0.931741
2	2.0	1.925881	0.074119	1.270261
3	1.8	1.856209	-0.056209	-0.945412
4	2.2	2.298250	-0.098250	-1.628550

5	3.5	3.465247	0.034753	0.866405
6	1.9	1.864071	0.035929	0.658447
7	2.7	2.744350	-0.044350	-0.761441
8	2.1	2.060402	0.039598	0.870312
9	3.6	3.595795	0.004205	0.086345

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

***** Estimated Partial Regression Coefficients *****

	Estimate	Std. err.	t value	p value	Std. coef.	Tolerance \
x1	0.204617	0.007564	27.050556	< 0.0001	0.67067	0.98358
x2	0.286634	0.010802	26.536454	< 0.0001	0.65793	0.98358
constant	0.149176	0.054506	2.736849	0.029		

	VIF
x1	1.0167
x2	1.0167
constant	

Residual standard error: 0.0636 on 7 degrees of freedom
Multiple R: 0.9979, Multiple R-squared: 0.9958, Adjusted R-squared: 0.9946
ANOVA F-statistic: 823.477 on 2 and 7 degrees of freedom, p-value: < 0.0001
Log-likelihood: 15.13807, AIC: -22.27614

***** Analysis of Variance Table for Multiple Regression Analysis *****

	SS	df	MS	F value	p value
Regression	6.671644	2	3.33582	823.477	< 0.0001
Residual	0.028356	7	0.00405		
Total	6.700000	9			