

重回帰分析におけるAICを用いた変数選択

帝京平成大学 鈴木義一郎

まず、回帰分析に必要な「関数群」を次のように用意しておく。

```
stat_reg=:3 :0
regb=:[%.1:,.] NB. regression coefficient
regp=:(1:,.)+/. *regb NB. predicted value
regq=:[:+/[:*:[-regp NB. sum of residueale
regcd=:100*_1:-regq%[:+/[:*:(-+/%#)@[ NB. coefficient of determination(%)
mat=:[:%. (|:+/ .*)@(1:,.) NB. inverse of data matrix
resvar=:regq%[:-/[:$1:,.] NB. residual variance
regt=:regb%[:%:resvar*[:(<1 0)&|:mat@] NB. t-value of regression coefficients
mll=:>:@^.@((o.2)^*_regq%#@[)*#@[%_2: NB. MLL(Maximum Log Likelyhood
regaic=:+:@(1:+#@(1:,:)))-2:*mll
'program set of regression model'
)
s tat_reg ''
program set of regression model
これで、関数群が導入された。
```

そこで、竹内氏のJAPLA 2005夏合宿のTest Dataを定義する。

```
Y=:23.6 26.5 24 19 21.7 22.9 26 28.2 25.5 23 21.6 21 26.2 27 24
X1=:20.6 22.8 18.3 16.5 17 22.1 18.9 19.3 19.3 15.4 18.8 18.3 18.1 23.9 22.4
X2=:11.5 15.8 13 15.9 11.1 16.1 15 11.8 15.2 15.6 17.2 14.5 16.1 16.6 13
X3=:2.9 6.2 3.7 5.6 2.1 5 6 2.1 7 4.1 6.8 4.4 4.6 6 6
X4=:7 8.7 6.1 9.7 7 7 7.3 6.1 5.2 5.2 8.7 4.2 7.9 7.9 7.3
X5=:1.7 4.3 3.4 5.1 1.7 3.5 3.5 1.7 3.5 3.6 2.6 3.3 2.6 4.4 3.5
X6=:8.73 8.6 7.62 8.68 7.83 9.65 7.27 6.84 7.57 6.7 8.7 8.71 6.91 8.85 9.33 %
10
X7=:2.97 3.28 2.54 5.11 3.23 3.06 2.81 2.16 2.04 2.26 4.03 2 3.02 2.93 3.04 %
10
X8=:0.83 1.89 1.86 3.09 1 1.58 1.85 0.88 1.81 2.34 1.38 1.8 1.44 1.84 1.56 % 10
X9=:5.86 6.94 9.19 9.11 8.1 7 5.83 8.1 5 8.78 3.82 7.5 5.65 7.33 5.83 % 10
```

```
X10=:6.09 5.51 4.69 6.1 6.31 4.35 4.87 5.17 3.42 3.33 5.06 2.9 4.91 4.76 5.62 %
10
```

```
X=: |:X1, X2, X3, X4, X5, X6, X7, X8, X9, :X10
X,.Y
20.6 11.5 2.9 7 1.7 0.873 0.297 0.083 0.586 0.609 23.6
22.8 15.8 6.2 8.7 4.3 0.86 0.328 0.189 0.694 0.551 26.5
18.3 13 3.7 6.1 3.4 0.762 0.254 0.186 0.919 0.469 24
16.5 15.9 5.6 9.7 5.1 0.868 0.511 0.309 0.911 0.61 19
17 11.1 2.1 7 1.7 0.783 0.323 0.1 0.81 0.631 21.7
22.1 16.1 5 7 3.5 0.965 0.306 0.158 0.7 0.435 22.9
18.9 15 6 7.3 3.5 0.727 0.281 0.185 0.583 0.487 26
19.3 11.8 2.1 6.1 1.7 0.684 0.216 0.088 0.81 0.517 28.2
19.3 15.2 7 5.2 3.5 0.757 0.204 0.181 0.5 0.342 25.5
15.4 15.6 4.1 5.2 3.6 0.67 0.226 0.234 0.878 0.333 23
18.8 17.2 6.8 8.7 2.6 0.87 0.403 0.138 0.382 0.506 21.6
18.3 14.5 4.4 4.2 3.3 0.871 0.2 0.18 0.75 0.29 21
18.1 16.1 4.6 7.9 2.6 0.691 0.302 0.144 0.565 0.491 26.2
23.9 16.6 6 7.9 4.4 0.885 0.293 0.184 0.733 0.476 27
22.4 13 6 7.3 3.5 0.933 0.304 0.156 0.583 0.562 24
```

```
>:/: Y regaic"1 |:X
7 1 6 8 9 5 4 10 2 3
```

この結果は、単回帰の場合の情報量規準(AIC)の小さい順の変数Xの番号を示している。

次に、目的変数(Y)を左引数に、説明変数(X)と用いる変数の個数(k)を「k;X」の形で右引数に入力して、採用した変数と情報量規準(AIC)をボックスの形で出力する両側関数を次のように定義する：

```
compare=:4 :0
s=. (>{.y.)=+/"1 t)#t=. #:i.2^#y.>{:y.
r=.,:(k{s);x.regaic|:(k=.0){u=.s#y
while. k<<:#s
do. r=.r,(k{s);x.regaic|:(k=.k+1){u
end.
)
```

まず、全部の変数を用いた場合の情報量規準

```
Y regaic X
```

_18.2885

のように算出される。

そこで、説明変数の個数を9個、8個と順次減少せしめた時の「AIC」を最小にする変数の組合せとAICの値を示してみると、以下のようになる。

((=<./>){:"1 t)#t=:Y compare 9; |:X

1 1 1 1 1 1 1 0 1	_17.4898
1	

((=<./>){:"1 t)#t=:Y compare 8; |:X

1 1 1 1 0 1 1 0 1	_17.1509
1	

((=<./>){:"1 t)#t=:Y compare 7; |:X

1 1 0 1 1 1 1 0 0	_12.9917
1	

((=<./>){:"1 t)#t=:Y compare 6; |:X

1 1 0 0 1 1 1 0 0	_9.22404
1	

((=<./>){:"1 t)#t=:Y compare 5; |:X

1 1 0 1 0 1 0 0 0	_0.890523
1	

((=<./>){:"1 t)#t=:Y compare 4; |:X

1 0 0 0 1 1 0 1 0	1.07253
0	

((=<./>){:"1 t)#t=:Y compare 3; |:X

1 0 0 0 1 1 0 0 0	2.20332
0	

((=<./>){:"1 t)#t=:Y compare 2; |:X

1 0 0 0 0 1 0 0 0	3.09146
0	

((=<./>){:"1 t)#t=:Y compare 1; |:X

0 0 0 0 0 0 1 0 0	71.1598
0	

7個の変数までの「決定係数」を算出してみると

Y regcd X

99. 8097

Y regcd | :X1, X2, X3, X4, X5, X6, X7, X9, :X10

99. 7993

Y regcd | :X1, X2, X3, X4, X6, X7, X9, :X10

99. 7947

Y regcd | :X1, X2, X4, X5, X6, X7, :X10

99. 7291

といった結果になるから、当てはまりはかなり良好である。

最後に、4つのモデルの「t-値」をチェックしてみると

```
Y regt X
5. 7979 8. 53216 _2. 64812 _1. 21633 1. 01017 0. 531599 _8. 73857 1. 61007 _0. 467718
_1. 19988 _2. 85622
```

```
Y regt | :X1, X2, X3, X4, X5, X6, X7, X9, :X10
8. 78653 10. 2421 _2. 83758 _1. 32083 1. 31097 0. 338029 _11. 1566 1. 82217 _1. 27949
_3. 06918
```

```
Y regt | :X1, X2, X3, X4, X6, X7, X9, :X10
10. 0292 11. 0977 _3. 05993 _2. 85687 1. 4502 _12. 194 1. 94324 _3. 26486 _3. 31798
```

```
Y regt | :X1, X2, X4, X5, X6, X7, :X10
9. 49097 10. 4314 _2. 75603 1. 41377 _2. 92771 _11. 4584 1. 54838 _2. 95442
```

といった結果になる。

ところで、全部の変数を使ったモデルと9個の変数を使ったモデルにはt-値の絶対値の小さな箇所が幾つか散見される。これに対して「X1, X2, X3, X4, X6, X7, X9, :X10」という8個の変数を使ったモデルでは、X4のt-値が「1.4502」、X7のt-値が「1.94324」で、他の変数のt-値の絶対値はかなり大きいことから、このモデルを採択すべきであるという結論が得られる。

Y1=:4 6 5 5 5 6 6 6 6 3 4 6 5 4 5 5 6 6 6 4 6 5 6 4 6 5 5 6 4 6 5 1 6 6 5 6
5 4 3 5 5 5 6 6 6 1 6 6 2 5 3 1 5 5 5 1 2 5 4 5 5 5 6 3 2 4 6 5 6 6 4 6 6 5 6 6
5 5 6 5 5 2 6 4 3 3 1

Y2=:4 5 5 6 5 6 5 6 6 4 3 6 4 5 5 5 6 6 5 2 6 5 6 4 6 5 5 6 4 6 4 1 6 6 5 5
5 4 3 5 6 5 6 6 6 1 6 6 1 5 2 1 5 5 4 1 2 5 4 6 6 5 6 1 2 5 6 5 6 6 5 6 6 5 6 6
5 5 6 6 5 1 6 4 2 2 2

Y3=:4 5 5 6 5 6 6 6 6 3 4 6 5 5 5 5 6 6 5 2 6 5 6 3 6 5 5 6 4 6 4 1 6 6 6 5
5 4 3 5 6 5 6 6 6 1 6 5 2 5 2 1 5 5 3 1 2 5 4 6 6 6 6 1 2 4 6 5 6 6 5 6 6 4 6 5
5 5 6 6 5 2 6 3 2 3 1

A=:5 5 5 5 4 5 5 4 3 2 4 5 3 3 3 5 5 5 5 5 5 3 3 5 3 5 5 5 5 2 5 5 4 5 5
1 3 5 5 5 5 5 5 1 5 5 2 5 5 1 2 5 3 2 3 3 1 5 5 3 5 2 2 2 4 5 5 5 3 5 3 4 5 5 2
5 5 5 5 2 5 5 2 1 2

B=:5 5 5 4 4 5 5 4 3 2 2 5 3 3 2 5 5 5 5 5 5 3 3 5 3 5 5 5 5 2 5 5 4 5 5
1 2 5 4 5 5 5 5 1 5 5 2 5 5 1 2 5 2 2 4 2 1 5 5 3 5 2 1 2 4 5 5 5 2 5 3 4 5 5 2
5 5 5 5 2 5 5 2 1 1

C=:4 3 5 4 5 5 5 3 3 1 1 5 1 2 1 3 5 5 5 5 5 4 3 3 5 2 4 4 3 5 5 2 4 5 3 3 5
1 1 5 4 3 5 5 5 1 5 4 1 5 3 1 1 3 1 2 2 1 1 5 5 2 4 2 1 1 4 5 5 5 1 5 3 4 5 5 1
4 4 5 3 1 5 3 1 1 1

D=:5 5 5 5 5 5 5 5 3 2 3 5 5 3 3 5 5 5 5 5 5 4 3 5 3 5 5 5 5 2 5 5 5 5 5
2 2 5 5 5 5 5 5 1 5 5 2 5 4 1 3 5 2 2 4 3 1 5 5 3 5 2 2 2 4 5 5 5 2 5 5 4 5 5 3
5 5 5 5 1 5 5 2 2 2

E=:5 5 5 4 5 5 5 5 3 2 3 5 3 3 2 5 5 5 5 5 5 4 3 5 3 3 5 5 5 2 5 5 3 5 5
2 2 5 3 5 5 5 5 2 5 5 2 5 4 2 2 5 2 2 2 2 2 5 5 3 3 2 2 2 4 5 5 5 2 5 3 5 5 5 2
3 5 5 5 2 5 5 2 2 2

F=:5 5 5 5 5 5 5 4 4 3 5 5 5 4 5 5 5 5 5 5 5 3 5 5 5 5 5 3 5 5 5 5 5 1
3 3 5 5 5 5 5 5 1 5 5 3 5 5 1 3 5 3 2 4 3 1 5 5 5 5 2 2 3 5 5 5 5 3 5 5 5 5 5 4
5 5 5 5 1 5 5 3 2 2

G=:5 5 4 4 3 5 5 3 3 1 1 5 3 2 3 3 5 5 5 5 5 4 3 2 5 1 5 4 3 5 5 2 4 4 3 4 5
1 1 5 5 4 5 5 5 1 5 4 1 3 1 1 2 4 1 2 1 1 1 5 5 3 3 1 1 1 4 5 5 5 1 5 3 3 5 5 1
3 3 5 3 1 5 4 1 1 1

H=:5 5 5 5 5 5 5 5 3 3 5 5 3 3 5 5 5 5 5 3 5 5 5 3 5 5 5 5 5 5 3 5 5 5 5 5
2 2 5 5 5 5 5 5 1 5 5 4 5 5 1 3 5 1 2 4 2 2 5 5 5 5 1 2 3 4 5 5 5 3 5 3 5 5 5 3
5 5 5 5 1 5 5 3 1 3

I=:5 5 5 5 5 5 5 5 3 3 5 5 5 3 3 5 5 5 5 5 5 3 5 5 5 5 5 5 5 2 5 5 5 5 5
2 2 5 5 5 5 5 5 1 5 5 2 5 4 1 3 5 2 2 4 3 2 5 5 4 5 3 2 2 4 5 5 5 2 5 5 5 5 5 3
5 5 5 5 1 5 5 3 2 2

```
J=:5 5 5 5 5 5 5 5 3 2 3 5 3 3 2 5 5 5 5 5 5 5 3 5 3 5 4 5 5 5 4 5 5 5 5 5
1 2 5 5 5 5 5 5 1 5 5 2 5 5 1 3 5 2 2 3 3 2 5 5 3 5 2 2 2 3 5 5 5 2 5 3 5 5 5 3
5 5 5 5 2 5 5 2 2 2
```

```
K=:3 5 5 4 3 5 5 4 3 2 2 5 3 2 2 5 5 5 5 3 5 5 3 3 5 3 4 5 5 5 5 2 4 5 3 5 5
2 2 5 3 3 5 5 5 1 5 5 2 5 3 1 2 4 2 2 2 2 1 5 5 3 3 2 2 2 3 5 5 5 2 5 3 3 5 5 2
3 3 5 3 1 5 4 2 2 2
```

```
L=:5 5 5 5 5 5 5 5 3 2 2 5 3 3 2 5 5 5 5 5 5 5 4 2 5 2 2 5 5 5 5 2 5 5 4 5 5
2 2 5 4 5 5 5 5 2 5 5 2 5 5 2 2 5 2 2 2 2 2 5 5 3 4 2 2 2 4 5 5 5 2 5 3 4 5 5 3
5 5 5 5 2 5 5 2 2 2
```

```
M=:5 5 5 4 3 5 5 4 3 1 3 5 3 2 3 4 5 5 5 5 5 5 3 3 5 3 4 5 3 5 5 1 5 5 4 5 5
1 1 5 4 4 5 5 5 1 5 4 1 4 3 1 1 5 2 1 2 1 1 5 5 4 5 1 1 1 4 5 5 5 3 5 3 4 5 5 3
5 5 5 4 1 5 5 1 1 1
```

```
$ R=:A,B,C,D,E,F,G,H,I,J,K,L,:M
```

```
13 87
```

```
>:/: Y1 regaic"1 R
```

```
13 4 9 6 11 7 1 2 5 10 12 8 3
```

```
>:/: Y2 regaic"1 R
```

```
13 4 6 8 9 11 7 1 10 3 2 12 5
```

```
>:/: Y3 regaic"1 R
```

```
13 4 6 8 9 7 11 1 10 12 2 3 5
```

```
>:/: Y regaic"1 R
```

```
13 4 6 9 7 11 8 1 10 2 12 3 5
```

```
$X=:M,D,F,I,G,:H
```

```
6 87
```

```
compare=:4 :0
```

```
s=. ((>{. y.)=+/"1 t)#t=. #:i.2^#y=.>{:y.
```

```
r=.,:(k{s);x.regaic|:(k=.0){u=.s#y
```

```
while. k<<:#s
```

```
do. r=.r,(k{s);x.regaic|:(k=.k+1){u
```

```
end.
```

```
)
```

Y compare 1;M, D, :F

0 0	449.194
1	
0 1	443.763
0	
1 0	439.743
0	

Y compare 2;M, D, :F

0 1	437.466
1	
1 0	430.914
1	
1 1	435.859
0	

Y compare 3;M, D, :F

1 1	430.545
1	

Y compare 1;X

0 0 0 0 0 1	456.151
0 0 0 0 1 0	456.074
0 0 0 1 0 0	451.814
0 0 1 0 0 0	449.194
0 1 0 0 0 0	443.763
1 0 0 0 0 0	439.743

Y compare 2;X

0 0 0 0 1 1	442.669
0 0 0 1 0 1	447.024
0 0 0 1 1 0	442.582
0 0 1 0 0 1	444.481
0 0 1 0 1 0	434.071
0 0 1 1 0 0	442.071
0 1 0 0 0 1	441.005
0 1 0 0 1 0	441.07
0 1 0 1 0 0	443.457
0 1 1 0 0 0	437.466
1 0 0 0 0 1	437.176
1 0 0 0 1 0	439.422
1 0 0 1 0 0	437.688
1 0 1 0 0 0	430.914
1 1 0 0 0 0	435.859

Y compare 3;X

0 0 0 1 1 1	439.717
0 0 1 0 1 1	433.492
0 0 1 1 0 1	441.115
0 0 1 1 1 0	433.324
0 1 0 0 1 1	438.349
0 1 0 1 0 1	441.004
0 1 0 1 1 0	440.423
0 1 1 0 0 1	437.032
0 1 1 0 1 0	433.002
0 1 1 1 0 0	437.464
1 0 0 0 1 1	436.698
1 0 0 1 0 1	436.554
1 0 0 1 1 0	437.29
1 0 1 0 0 1	430.912
1 0 1 0 1 0	430.096
1 0 1 1 0 0	430.913
1 1 0 0 0 1	435.046
1 1 0 0 1 0	435.843
1 1 0 1 0 0	435.858
1 1 1 0 0 0	430.545

Y compare 4;X

0 0 1 1 1 1	433.108
----------------	---------

0 1 0 1 1 1	438.259
0 1 1 0 1 1	432.782
0 1 1 1 0 1	436.987
0 1 1 1 1 0	432.968
1 0 0 1 1 1	436.064
1 0 1 0 1 1	430.093
1 0 1 1 0 1	430.911
1 0 1 1 1 0	430.096
1 1 0 0 1 1	434.957
1 1 0 1 0 1	434.996
1 1 0 1 1 0	435.837
1 1 1 0 0 1	430.539
1 1 1 0 1 0	429.984
1 1 1 1 0 0	430.282

Y compare 5:X

0 1 1 1 1 1	432.777
1 0 1 1 1 1	430.091
1 1 0 1 1 1	434.935
1 1 1 0 1 1	429.984
1 1 1 1 0 1	430.281
1 1 1 1 1 0	429.874

Y compare 6:X

1 1 1 1 1 1	429.872
----------------	---------

Y regcd |:X

61.7886

Y1 regcd |:X

65.3528

Y2 regcd |:X

56.8329

Y3 regcd |:X

57.7126

Y1 regaic |:X

222.431

Y2 regaic |:X

258.784

Y1 regt |:X

3.76077 2.21832 0.201419 2.65222 0.476612 0.432455 _1.39699

Y regt |:X

2.97485 1.64825 0.449666 2.18957 _0.321256 0.614117 0.0487277

Y regt |:M, D, I, :G

4.19895 2.10677 1.17501 0.0726581 0.139235

Y regt |:M, D, :G

4.61451 2.26709 1.8671 0.123791

Y regcd |:M, D, :G

59.074

Y regcd M, .D

59.0664

Y regt M, .D

4.68096 2.82659 1.95832

Y regcd M, .F

61.3282

Y regt M, .F

3.3709 4.43191 2.99535

Y1 compare 2;M, D, :F

0	1	1	233.315
1	0	1	225.134
1	1	0	230.342

Y1 regcd M,.F

64. 2594

Y1 regt M,.F

4. 12399 4. 85388 3. 04765

Y1 regcd |:M,D,:F

64. 4023

Y1 regt |:M,D,:F

3. 82062 2. 92391 0. 577103 2. 33965