重回帰分析

青木繁伸

1　目的

重回帰分析を行う。

2　使用法

```
from mreg import mreg
mreg(dat, tolerance = False, verbose=True)
```

2.1　引数

dat	従属変数を最後の列に置いたデータフレーム
tolerance	torelance を出力する場合に True にする デフォルトでは VIF を出力する
verbose	必要最小限のプリント出力をする（デフォルトは True）

2.2　戻り値の名前

"result"	偏回帰係数，標準誤差，t 値，p 値, 標準化偏回帰係数，VIF
"anova"	回帰の分散分析
"stderr"	残差標準誤差
"dfe"	残差標準誤差の自由度
"R"	重相関係数
"R2"	決定係数（重相関係数の 2 乗）
"R2s"	自由度調整済み重相関係数の二乗
"F"	回帰の分散分析の F 値
"dfr"	第 1 自由度（第 2 自由度は "dfe"
"p"	F の p 値
"loglik"	対数尤度
"AIC"	AIC

3　使用例

```
import pandas as pd

dat = pd.read_csv("data/mreg.csv")
print(dat.head())
```

```
     X1    X2    X3    X4    X5    X6    X7    X8    X9   X10   X11   X12  \
0  71.5  39.5  71.6  46.1  41.9  65.6  36.7  50.5  46.4  63.4  68.5  42.7
1  38.4  40.9  47.2  50.9  53.9  31.8  33.1  47.4  44.6  58.4  55.7  43.0
2  49.8  54.2  41.7  27.0  72.8  50.3  43.9  36.3  28.6  29.6  48.0  40.2
3  53.0  42.3  71.9  64.7  65.2  59.9  48.2  48.2  43.2  61.0  66.2  81.5
4  31.2  26.9  33.9  54.0  68.3  38.4  45.1  71.0  54.5  51.0  43.8  33.7

      Y
0  75.7
1  32.5
2  12.4
3  72.9
4  11.4
```

3.1 VIF を出力する

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

a = mreg(dat)
```

```
Coefficients
           Estimate  Std. Error   t value Pr(> |t|)      beta      VIF
X1         0.412004    0.054423  7.570410  < 0.0001   0.26804  3.44236
X2        -0.111885    0.041141 -2.719546    0.0079  -0.07279  1.96704
X3         0.319443    0.050622  6.310382  < 0.0001   0.20786  2.97944
X4         0.433128    0.058125  7.451685  < 0.0001   0.28190  3.92989
X5        -0.394926    0.043465 -9.086134  < 0.0001  -0.25706  2.19797
X6         0.471159    0.055647  8.466993  < 0.0001   0.30660  3.60081
X7         0.164192    0.043884  3.741505    0.0003   0.10684  2.23915
X8        -0.150200    0.037658 -3.988524    0.0001  -0.09773  1.64871
X9         0.215244    0.041273  5.215104  < 0.0001   0.14004  1.98013
X10        0.081436    0.053246  1.529427    0.1298   0.05298  3.29532
X11        0.368669    0.061393  6.005041  < 0.0001   0.23980  4.37881
X12        0.235303    0.033327  7.060501  < 0.0001   0.15306  1.29054
constant -57.178256    7.045523 -8.115545  < 0.0001

Residual standard error: 2.93338 on 87 degrees of freedom
Multiple R: 0.98403, Multiple R-squared: 0.96832, Adjusted R-squared: 0.96395
F-statistic: 221.584 on 12 and 87 DF,  p-value: < 0.0001
loglik = -242.54617  AIC = 513.09235

Anova Table
            Df        Sum Sq      Mean Sq    F value    Pr(>F)
regression  12  22879.990759  1906.665897  221.58426  < 0.0001
```

```
residuals   87    748.608841     8.604699
total       99  23628.599600   238.672723
```

3.2 tolerance を出力する

```
a = mreg(dat, tolerance=True)
```

```
Coefficients
           Estimate  Std. Error   t value Pr(> |t|)      beta tolerance
X1         0.412004    0.054423  7.570410  < 0.0001   0.26804   0.29050
X2        -0.111885    0.041141 -2.719546    0.0079  -0.07279   0.50838
X3         0.319443    0.050622  6.310382  < 0.0001   0.20786   0.33563
X4         0.433128    0.058125  7.451685  < 0.0001   0.28190   0.25446
X5        -0.394926    0.043465 -9.086134  < 0.0001  -0.25706   0.45497
X6         0.471159    0.055647  8.466993  < 0.0001   0.30660   0.27772
X7         0.164192    0.043884  3.741505    0.0003   0.10684   0.44660
X8        -0.150200    0.037658 -3.988524    0.0001  -0.09773   0.60653
X9         0.215244    0.041273  5.215104  < 0.0001   0.14004   0.50502
X10        0.081436    0.053246  1.529427    0.1298   0.05298   0.30346
X11        0.368669    0.061393  6.005041  < 0.0001   0.23980   0.22837
X12        0.235303    0.033327  7.060501  < 0.0001   0.15306   0.77487
constant -57.178256    7.045523 -8.115545  < 0.0001

Residual standard error: 2.93338 on 87 degrees of freedom
Multiple R: 0.98403, Multiple R-squared: 0.96832, Adjusted R-squared: 0.96395
F-statistic: 221.584 on 12 and 87 DF,  p-value: < 0.0001
loglik = -242.54617  AIC = 513.09235

Anova Table
            Df        Sum Sq      Mean Sq    F value    Pr(>F)
regression  12  22879.990759  1906.665897  221.58426  < 0.0001
residuals   87    748.608841     8.604699
total       99  23628.599600   238.672723
```

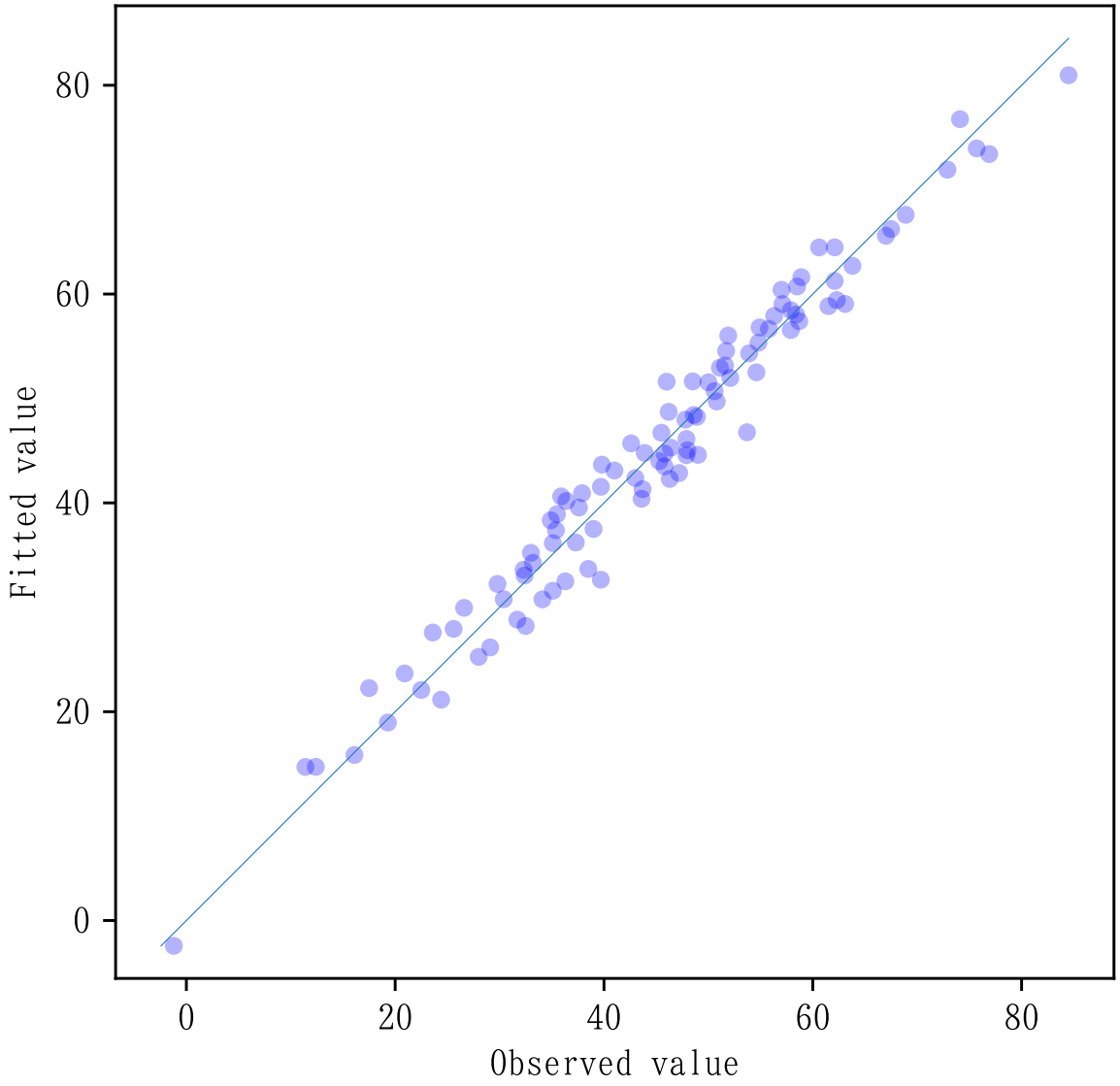
3.3 実測値と予測値のプロット

```
print(a["pred"].head(20))
```

```
    Observed values  Fitted values  Residuals  Std. Residuals
0              75.7      73.952919   1.747081        0.700449
1              32.5      28.210814   4.289186        1.540464
2              12.4      14.716876  -2.316876       -0.863695
3              72.9      71.913482   0.986518        0.375458
4              11.4      14.705985  -3.305985       -1.260960
5              52.1      51.966178   0.133822        0.048331
```

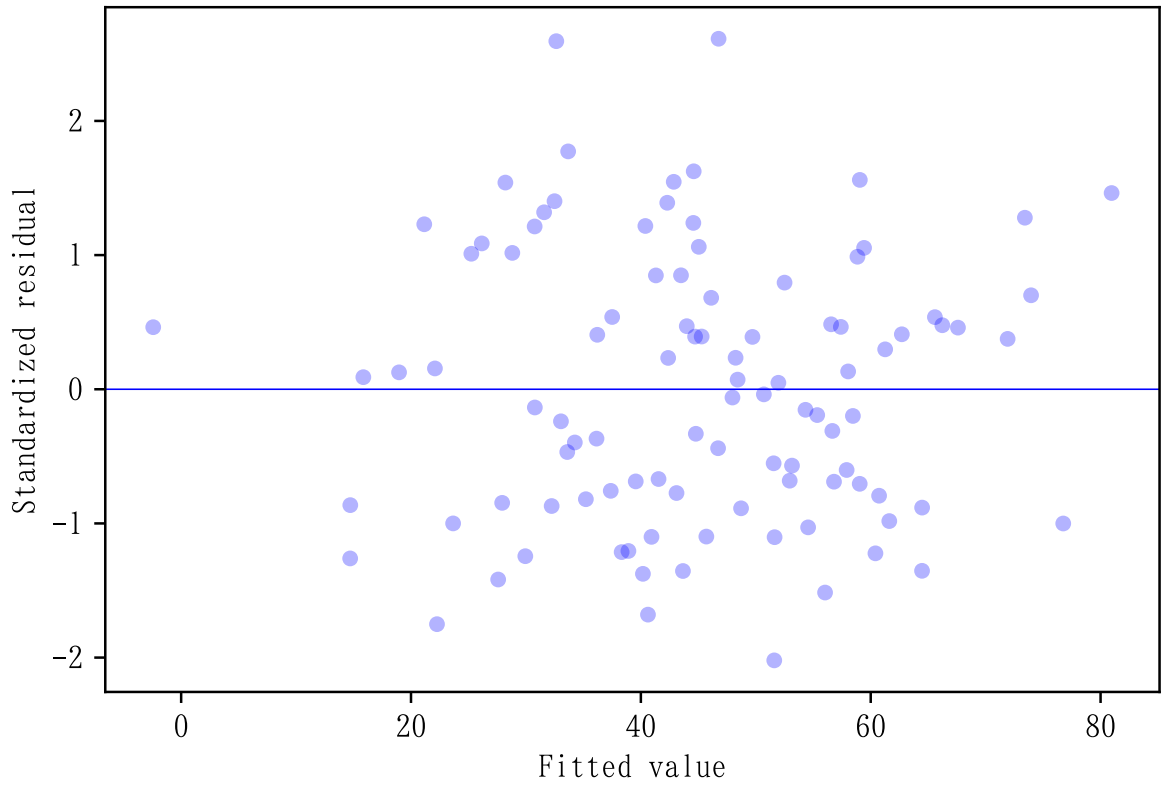
6	67.5	66.229810	1.270190	0.477150
7	45.8	43.491259	2.308741	0.849349
8	32.4	33.046478	-0.646478	-0.238944
9	54.8	55.344161	-0.544161	-0.192383
10	37.3	36.207497	1.092503	0.405899
11	33.2	34.252977	-1.052977	-0.396641
12	57.9	56.553729	1.346271	0.484167
13	34.1	30.758051	3.341949	1.213360
14	47.9	46.119264	1.780736	0.681590
15	32.3	33.595222	-1.295222	-0.467855
16	-1.2	-2.430090	1.230090	0.462647
17	57.0	60.410464	-3.410464	-1.222941
18	37.9	40.925912	-3.025912	-1.100175
19	48.5	51.635974	-3.135974	-1.102556

```
import sys
sys.path.append("statlib")
from mreg import mreg_plot

mreg_plot(a)
```



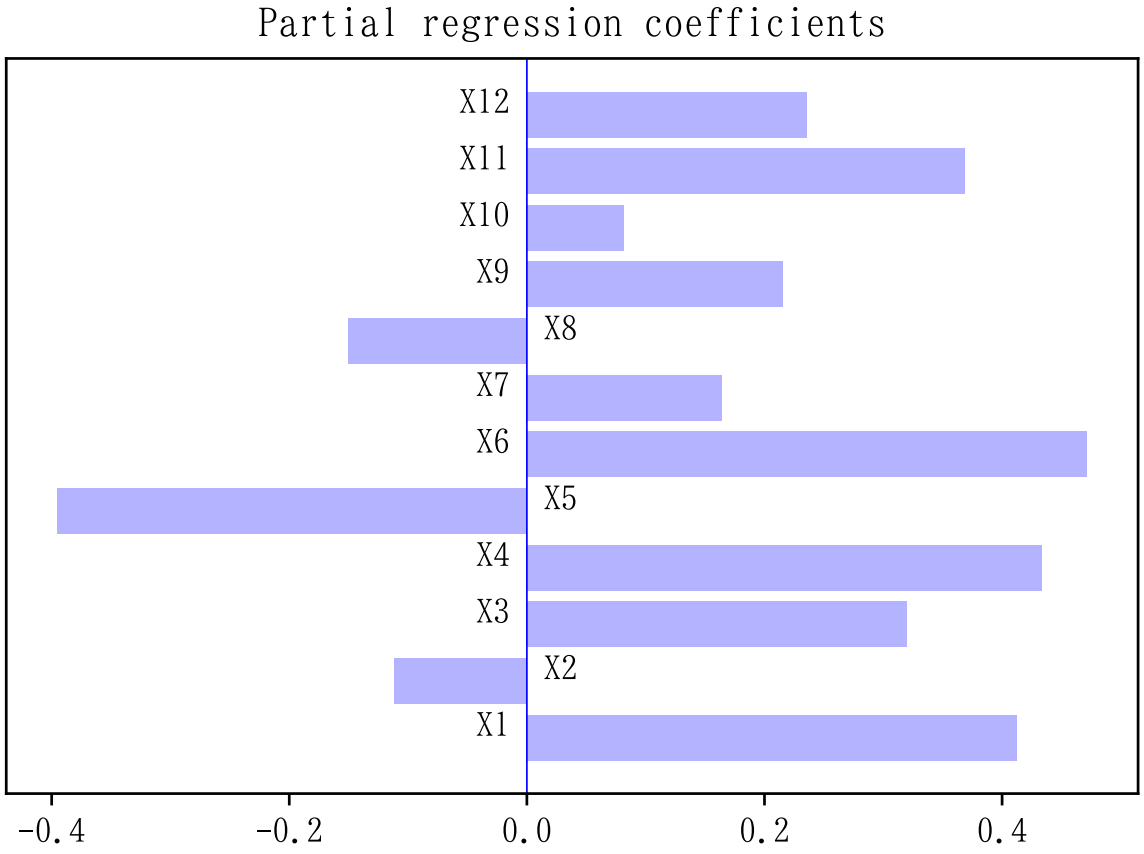
3.4 標準化残差のプロット

```
import sys
sys.path.append("statlib")
from mreg import mreg_plot

mreg_plot(a, type="e")
```



3.5 偏回帰係数の大きさのプロット

```
mreg_plot(a, type="c")
```




```
mreg_plot(a, type="b")
```