

Numeric Recipes for Econometrics(1)

回帰分析編

SHIMURA Masato

jcd02773@nifty.com

2011年1月24日

目次

1	単純回帰モデルと重回帰モデル	3
1.1	最少自乗法	3
1.2	単回帰モデル	4
1.3	重回帰モデル	10
1.4	対数回帰/The Log Linear Model	14
2	回帰モデルの検定	18
2.1	t 検定	19
2.2	p-value	21
2.3	AIC によるモデルの選択	22
2.4	AIC による変数選択の方法	23
3	トレンド項とダミー変数	26
3.1	トレンド項	26
3.2	ダミー変数を用いたモデル	28
4	不均一分散・系列相関	32
4.1	系列相関 (1)・ダービン・ワトソン検定	32
4.2	系列相関 (2)・コ克蘭・オーカット法 Cochrane-Orcutt(C-O) Method	34
4.3	不均一分散 (1)・一般化最少自乗法	36
4.4	不均一分散 (2)・加重最少自乗法 WLS	38
4.5	関数一覧	39
5	種々の回帰分析	41

5.1	多項式モデル	41
5.2	乗法モデル	46
5.3	ロジスチック回帰	51
5.4	動学モデル	52
5.5	ハウ・タッカー モデル	55
5.6	ブラウン型消費関数	58
6	Newton 法による非線形回帰	61
6.1	Newton 法	61
7	ロレンツ曲線とジニ係数	69
7.1	関数一覧	72
8	Logit と Probit	74
8.1	Group Logit	74
8.2	Probit	84
付録 A	SNA のデータで	89
A.1	SNA データの取り出しとグラフ	89
A.2	Jでの読み込み	89
A.3	グラフの作成	89

最少自乗法を用いてモデルを作成する方法を中心に、多変数の最少自乗法や対数変換の方法、トレンド項、ダミー変数などを付加する方法を紹介する。多変数の最少自乗法は変数を増やしていくと相関係数が上がるが、多重共線性が生じるといった厄介な性質があるが、モデル選択に AIC を用いて最適モデルを選択する。また、不均一分散が生じた場合のコクラン・オーカット法や一般化最少自乗法を紹介する。

本稿は数式からアルゴリズムを構想し Script を作成することを主眼においているので、例題に用いたデータ資料の多くは出典を記載した上で Script を作成時に用いたものを掲載した。原典も参照いただければ幸甚である。

最少自乗法は誰が ルジャンドル (Adrean-Marie Legendre 1752-1833 Paris) が 1806 年に最少自乗法を記した書を刊行した。後にガウス (1777-1855) が 1795 年から使っていたと主張したことから大論争が起ったが、状況証拠からガウスの方が早かったと証明され、ガウスは 1809 年に刊行した。^{*1}

ラプラスは 1812 年に最少自乗法に最初の証明を与えた。フランス革命の混乱の後のナポレオンの皇帝就任や、ベートーベンの交響曲 3 番「英雄」の作曲は 1804 年であった。

1 単純回帰モデルと重回帰モデル

1.1 最少自乗法

概要

回帰モデルは線形数学を用いると簡潔に表すことができる

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \frac{\mathbf{X}'\mathbf{y}}{\mathbf{X}'\mathbf{X}} = \frac{y}{X}$$

定数項も求めるときは 1 X と 1 のみで構成した (縦) ベクトルを X に付加する。

これを Script で表すと次のように 1 行で記述でき、単回帰、重回帰共に同じスクリプトで足りる。(1 からなるベクトルは自動付加)

```
reg_ols=: %. 1&,.@]
```

データを加工すれば指数や対数にもこれをそのまま用いることができる。

^{*1} ジャンケンの後出しのような執拗な論争にルジャンドルは大いに傷ついた。(Wikipedia に怒った似顔絵が拳がっている)他にルジャンドルとガウスの双方の名を冠する理論も多い

1.1.1 最少自乗法の正規方程式

多変数の重回帰モデルの正規方程式は行列で表すと次のようになる。(独立変数が2変数の例、単回帰の拡張にすぎない)

$$\begin{pmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_2 x_1 & \sum x_2^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \sum Y \\ \sum x_1 Y \\ \sum x_2 Y \end{pmatrix}$$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \quad \beta = \begin{pmatrix} a \\ b \end{pmatrix} \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

1.1.2 最少自乗法の性質

最少自乗法による推定値は最良線形不偏推定値即ち、線形不偏推定値の中で尤も分散の小さい推定値である。(ガウス・マルコフの定理)

1. 残差 (e_i) の総和は 0

$$\sum e_i = 0$$

2. 説明変数 (X_i) と残差 (e_i) の積和は 0 である。

$$\sum X_i e_i = 0$$

3. 回帰式の理論値 (\hat{Y}) と残差 (e_i) の積和は 0

$$\sum \hat{Y}_i e_i = 0$$

このことは残差が正規分布であることが前提となっている。

$u_t \sim N(1, \sigma^2)$ u_t は平均 0, 分散 σ の正規分布に従う。

$E(u_t) = 0$ t_t の平均は 0

$E(u_t^2) = \sigma$ 均一分散の仮定

$E(u_t u_s) = 0$ 系列相関はない

などが残差 (誤差項 (u_t)) に仮定されているが、現実には聞き分けのよい優等生ばかりではない。

1.2 単回帰モデル

1.2.1 matrix-divide

配列計算言語は行列除算機能を持っている。正方行列は当然のこと、縦長の行列の逆行列が即座に得られる。

これは K.E.Iverson が APL で最初に実装したもので、J では `matrix-divide(%.)` である。この縦長の

matix-divide で回帰分析は飛躍的に簡略化された。^{*2}

単回帰の正規方程式と重回帰モデルの正規方程式は次のように表すことができ、正規方程式を経由する必要はない。

$$(\mathbf{X}'\mathbf{X})\hat{\beta} = \mathbf{X}'y$$

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y$$

これらは、更に次のように簡約される。

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y = \frac{\mathbf{X}'y}{\mathbf{X}'\mathbf{X}} = \frac{y}{\mathbf{X}}$$

線型数学で表現すれば単回帰も重回帰も同じであり、配列計算言語では、単回帰も重回帰も同じスクリプトで記述でき、入力データにより、単回帰、重回帰のどちらにも適用できる。

定数項を求める場合には

$$\frac{y}{1, \mathbf{X}}$$

と \mathbf{X} の左に 1 を付加する。

単回帰

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdots \\ y_n \end{bmatrix}, \quad \begin{bmatrix} 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{13} \\ \vdots & \vdots \\ 1 & x_{1n} \end{bmatrix}$$

重回帰

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdots \\ y_n \end{bmatrix}, \quad \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ 1 & x_{13} & x_{23} & \cdots & x_{k3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}$$

これを I で最も簡潔に定義すれば次のとおりであり、単回帰にも重回帰にも用いることができる

^{*2} Iverson が W.Leontief の逆行列 (正方行列) をサポートしたときに思いついた

reg_ols=: %. 1&,.@]

J Grammar .

tacit 関数型で定義する

単項と両項 %. は最少自乗法で用いる場合は両項の *Hook* を構成する

```
reg_ols=: %. 1&,.@]
                                     /      \
                                     y      1&,.
                                     |
                                     X
```

入力データの構成 次の GDP データの場合は *X* は次のように構成する

*3

- *X* と *y* の統合

本稿では $1, .X, .y$ と、*y* を一番右の列に付加する (単項の) データフォームを原則としている。後は *Script* が両項として計算してくれる

```
]A=: |: 4 3 9 8 6,1 2 0 2 5,: 0 1 5 6 8
```

```
x0x1|y
```

```
4 1 |0
```

```
3 2 |1
```

```
9 0 |5
```

```
8 2 |6
```

```
6 5 |8
```

- *X* に *I* を付加する。

```
1&,.@]
```

stich(,.) で *I* を付加するとこれが定数項になる。

@] は右引数に作用させる

```
1, . }. "1 A
```

```
X1X2
```

```
1 1 0
```

```
1 2 1
```

```
1 0 5
```

```
1 2 6
```

*3 最少自乗法は従属変数 (被説明変数) を *y*、説明変数を *X* とする。J は左引数を *x*、右引数を *y* として用いる。

1 5 8

この定数項の I は *Script* で自動的に付加されるようにしている。従って入力データは A の様に y を右端の列に持ってくればよい。

配列計算言語 J の逆行列計算機能 % で最小自乗法の回帰係数は単回帰、重回帰共いとも簡単に求められる。

```
({"1 A) reg_ols 0 1 {"1 A
      ( y)          (x1 x2)
```

```
_5.4023 1.12644 1.32184
```

$$f = -5.4023 + 1.12644x_0 + 1.32184x_1$$

Script .

```
reg0=:3 : 0
NB. select trend data or multi data
if. 1= +/ * $ y do. reg_t y
elseif. do. regx y end.
)

regx=:3 : '({"1 y ) %. 1,.}:"1 y ' NB. (1)
reg_t=:3 : ' y %. 1,. >: i. # y ' NB. (2)
```

reg0 は

1. y を右に付加する *DAT* の形で回帰する、
2. データが y のみで X は順序数の場合に順序数を自動で付加する

```
reg0 A
_5.4023 1.12644 1.32184
```

1.2.2 散布図

Worked Example

Appendix の *SNA* データ (*gdp_consumer.csv*) から実質国内総生産と実質民間最終消費支出の 1980-2008 (単位=10 億円) を取り出す。

```
require 'files csv'
a=. readcsv dir=. '/data/sna/esri/gdp/2010/gdp_consumer.csv'
a1=. |: ; ("1) ". L:0 }. "1 a
```

```
DAT= . 1 37{"1 a1
```

```
axis= . {"1 a1
```

時系列に2つを並べる

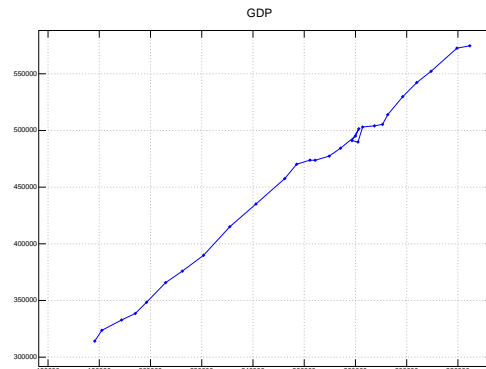
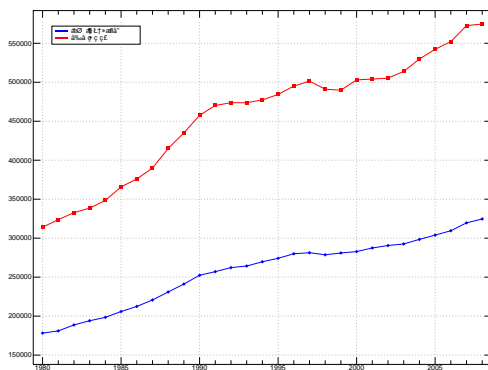
```
plot axis; |:DAT
```

;により x の $axis$ と y の |: DAT を連結し $plot$ 用 'type line,marker;title GDP' plot { |:DAT のデータを構成する

x	;	Y
-----	---	-----

散布図

転置 (|:) と単項の catalogue({) により 2 列のデータは x, y に分かれる



1.2.3 Line-fit と推計値

Worked Example

Appendix の GDP(1080-2008) の作図と回帰

X 軸に西暦をそのまま用いて見よう。図から見ると 2 ないし 3 のモードに分かれるが、先ず単回帰とラインフィットを試みよう。ダミー変数を用いる方法は §(3.2) で紹介する。

```
reg0 0 37{"1 a1
```

```
_1.70498e7 8779.72
```

```
linefit_reg0 0 _1{"1 a1
```

$f = -1.70498e7 + 8779.72x$ (単位 10 億円)

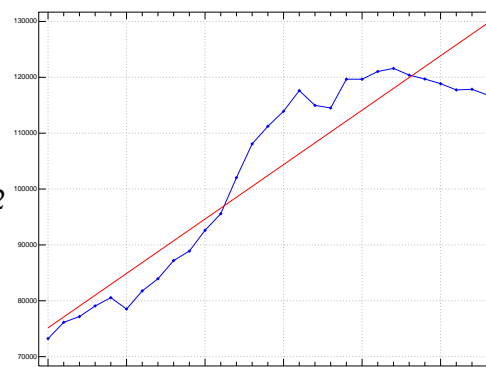
この場合 X に西暦を用いたが単に順序数 (0 1 2 …) としても勾配は変わらない

```
reg0 37{"1 a1
```

```
325266 8779.72
```

$f = 325266 + 8779.72x$

$f = a + bx$ のフォームとなる



*4

推計値の計算方法は 1.3.1 参照

reg0 は単項型の OLS を求める関数であり、 X_0, X_1, \dots, X_n, y の様に y を最右列として縦に変数を一塊りに並べる。

OLS は $f = a + bx$ となる。

```
reg0 |."1 DAT NB. |. lotate (change x,y)
1459.37 0.559894
```

$$f = 1459.37 + 0.559894x$$

実質民間消費支出 = 1459.37 + 0.559894(実質国内総生産)

reg_exam_ad は OLS の検査エンジンで OLS の後ろに付加して用いる。J では動詞 (OLS) をパワーアップする副詞として定義されている。

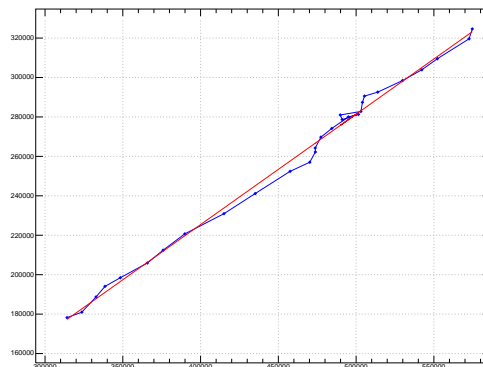
```
reg0 reg_exam_ad |."1 DAT
```

```

+-----+-----+
| f=      |1459.37 0.559894|
+-----+-----+
| R_sqr:  |99.4909          |
+-----+-----+
| AIC:    |469.199          |
+-----+-----+
| DW:     |0.707257         |
+-----+-----+
| t=:     |0.408705 72.6804|
+-----+-----+

```

```
linefit_reg0 |."1 DAT
```



linefit_reg0 は観測値 y を回帰して理論値 \hat{y} を求めて図に表している。

cortable は相関行列を求める関数であり、相関行列の 2 乗が決定係数である

```

cortable DAT
      1 0.997451
0.997451      1
民間消費支出 国民総生産

```

*4 弾性値とも言われる

1.3 重回帰モデル

線形数学では単回帰と重回帰（多変量回帰）は同じ方程式で表すことができ、 J では、単回帰と全く同じプログラムで計算できる。

*5

即ち、重相関はデータ作成のみに帰するものであり、データを $X_0, .X_1, \dots, .X_n, .y$ の様に y を最右列として縦に変数を一塊りに並べれば単相関と同一の関数を用いることができる。

図は独立変数 X が 2 変数で有れば 3D で表すことができる。これとは別に y と推計値（理論値） \hat{y} を *linefit* として図示することができる。*linefit_reg0* は年次のデータを X_0 （最初の列）に入れておけば年を X 軸に表示してくれる。

Worked Example 年を独立変数に用いた重回帰
先の実質民間消費支出と実質国民総生産に年を加える

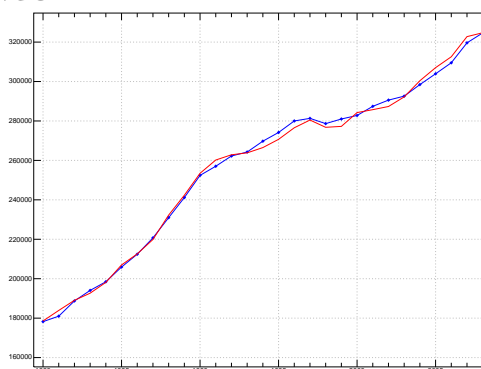
```
DAT2=. 0 37 1{"1 a1
```

```
reg0 DAT2  
_1.9976e6 1027.61 0.450498
```

$$f = -1.9976e6 + 1027.61year + 0.450498GDP$$

評価用の副詞 *reg_exam_ad* を付加して確認する

```
reg0 reg_exam_ad DAT2  
+-----+-----+  
|f=      |_1.9976e6 1027.61 0.450498|  
+-----+-----+  
|R_sqr:  |99.7563                    |  
+-----+-----+  
|AIC:    |449.826                      |  
+-----+-----+  
|DW:     |0.938842                     |  
+-----+-----+  
|t=:     |_7.2972 0.170789 0.108582 |  
+-----+-----+
```



*5 縦の逆行列 % を用いる

1.3.1 推計値の求め方

```
· A,. reg_sub0 A
x0x1y estimate-y
4 1 0 0.425287
3 2 1 0.62069
9 0 5 4.73563
8 2 6 6.25287
6 5 8 7.96552
```

```
reg_sub0=:3 : 0
if. 1 = +/*$ y do.
  (1,. i. # y ) mp reg0 y
else. (1,. }:"1 y ) mp reg0 y
end.
)
```

1. X の左にに I を付加する. $1, X$

```
] a3=.1,. 0 1{"1 A
1 4 1
1 3 2
1 9 0
1 8 2
1 6 5
```

2. $reg0$ で回帰係数を求める

```
] f=. reg0 A
_5.4023 1.12644 1.32184
```

3. 内積を用いて推計値を求める。 $a \times f$

$mp=:+/ . *$ は J の内積演算

```
a3 +/ . * f
0.425287 0.62069 4.73563 6.25287 7.96552
```

第 0 行は次の計算を行っている

```
1 4 1 * f
_5.4023 4.50575 1.32184
+/ 1 4 1 * f
0.425287
```

*6

*6 回帰係数 f はベクトルであるので右から転置して内積演算をする事も可能である。

```
f +/ . * |: a3
0.425287 0.62069 4.73563 6.25287 7.96552
```

1.3.2 正規方程式で計算する

多変数の重回帰モデルの正規方程式は行列で表すと次のようになる。(独立変数が2変数の例、単回帰の拡張にすぎない)

$$\begin{pmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_2 x_1 & \sum x_2^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \sum Y \\ \sum x_1 Y \\ \sum x_2 Y \end{pmatrix}$$

配列計算言語では回帰係数は $\frac{Y}{X}$ で直ちに求められるので正規方程式を用いることは希であるが、些かクラシックに正規方程式を作成して連立方程式を解いて求めてみよう。

正規方程式への当てはめは *Script* を作成し、連立方程式の解法はクラメール法を用いた。

```
cr=%:}:"1
```

```

composition_neq A                                clean cr  composition_neq A
5  30 10  20                                     1  0  0 _5.4023
30 206 56 144                                    0  1  0 1.12644
10  56 34  54                                    0  0  1 1.32184
```

右の列が連立方程式の解で、完全に解ければ左は単位行列なる。

最右の列が回帰係数 左は単位行列

```
cr=: %: }:"1 NB. Cramer method
```

```
combi=:[:{ [;~ i.
```

```
neqx_sub=: 3 : 0
```

```
NR=. {: $ y
```

```
MAT0=. (combi NR){ "1 L:0 y
```

```
MAT=. ;("1)+/ L:0 */"1 L:0 MAT0
```

```
((# y),+/ y),.(+/ y),MAT
```

```
)
```

```
composition_neq=: 3 : 0
```

```

a3 +/ . * f
0.425287 0.62069 4.73563 6.25287 7.96552
```

```

NB. u a1
'X0 Y0'=: (}: "1 y) ; {: "1 y
MATX=. neqx_sub X0
MATY=.(+/ Y0) , ; +/ L:0 ({ |: X0) * L:0 Y0
MATX,.MATY
)
reg_neq=: cr@composition_neq

```

J Grammar .

X の組み合わせ 正規方程式の 0 行、0 列を除いた行構成

```

combi 3
+-----+
|0 0|0 1|0 2|
+-----+
|1 0|1 1|1 2|
+-----+
|2 0|2 1|2 2|
+-----+

```

計算 各 Box 内で行 (横) 方向に掛けてから合計する

X に襟巻きを付ける 0 行と 0 列を付加する。(#y), +/ X0
y を構成 y, X × y

```

MATY=.(+/ Y0) , ; +/ L:0 ({ |: X0) * L:0 Y0

```

拡大係数行列にする y を右の列に付けて正規方程式の拡大係数行列を作成する

クラメール法で解く 連立方程式をクラメール法で解く

```

reg_neq=: cr@composition_neq

```

1.4 対数回帰/The Log Linear Model

線形とは、原因と結果が何らかの意味で比例的であるということである（山口昌哉）

経済データでは対数線型を取り扱うことが多い。これは、

1. 対数にすると変化の量でなく率で表され、非線形データでも多くが対数線型関係になる。。
2. 係数が、*elastically* を表す。*elastically* (弾性値)とは、 x が 1% 増加したときに、 y が何 % 増加するかを示し、単位を取捨した係数で表されるので、都合がよい。(単位の差は、定数項にあらわれる。)

半対数線型は、 Y, X のどちらかが、比率や % になっている場合に、その対数を取らずに回帰する場合などに用いられる。

<i>Model</i>	<i>Form</i>	$Slope = \frac{dY}{dX}$	$Elasticity = \frac{dY}{dX} \cdot \frac{X}{Y}$ 弾性
<i>Linear</i>	$Y = B_1 + B_2X$	B_2	$B_2 \left(\frac{X}{Y} \right)$
<i>Log-linear</i> 対数線形	$\ln Y = B_1 + B_2 \ln X$ ($Y = B_1 X^{B_2}$)	$B_2 \left(\frac{Y}{X} \right)$	B_2
<i>Log-linear</i> 半対数線形	$\ln Y = B_1 + B_2 X$ ($Y = e^{B_1 + B_2 X}$)	$B_2(Y)$	$B_2(X)$
<i>Linear-log</i> 対数線形	$Y = B_1 + B_2 \ln X$	$B_2 \left(\frac{1}{X} \right)$	$B_2 \left(\frac{1}{Y} \right)$
<i>Reciprocal</i> 逆数	$Y = B_1 + B_2 \left(\frac{1}{X} \right)$	$-B_2 \left(\frac{1}{X^2} \right)$	$-B_2 \left(\frac{1}{XY} \right)$

このような各々の対数に対応したスクリプトを用意するのではなくデータを変形して、単純な OLS で線形回帰することとする。

データは縦型で $X_0, .X_1, \dots, .X_n, .Y$ と並べる。たての列の取り出しは `m{"1)`, 縦の列の連結は `stich(, .)`

*7

J Grammar J と対数

*7 *stich* とは詩の行のことで *K.E.Iverson* の命名

J の対数機能 .

- 底を指定した対数は両項で用いる

```
10 ^ . 10 100 1000 NB. log(10)
1 2 3
2 ^ . 2 4 8 16 NB. log(2)
1 2 3 4
```

- 自然対数 ($e \ln$) は単項で用いる

```
^ . 10 NB. ln 10
2.30259
^ 2.30259 NB. inverse
10
```

1x1 e の数値は $1x1$ で表すことができる

```
1x1 2x1 2x2
2.71828 5.43656 14.7781
e 2e 2e^2
```

1.4.1 Worked Example

製造業の生産関数 (コブ・ダグラス生産関数)

1980-2002 の製造業のデータ。総労働時間 (LAB) は就業者数 (単位 1000 万人) に月平均労働時間を、資本稼働率 (CAP) は民間資本ストックに稼働率を掛ける。

```
DM0=. ".@> readcsv jpath '~user/classes/numeric_recipe/data/csv/morimune_0.csv'
```

$DM0$ から目的のデータを作成する。

```
X0=.* / ("1) 2 3 {"1 DM0 NB. LAB
X1=.* / ("1) 4 5 {"1 DM0 NB. CAP
Y=1 {"1 DM0 NB. GDP
DM27=:X0, .X1, .Y NB. X,Y
```

(一行では)

```
DM27=: (* / ("1) 2 3 {"1 DM0), .(* / ("1) 4 5 {"1 DM0), .1 {"1 DM0
```

<i>X0</i> 民間資本ストック × 稼働率 <i>LAB</i>	対数変換 (^.)	
<i>X1</i> 就業者 × 労働時間 <i>CAP</i>	^. DM27	
<i>Y</i> 製造業 <i>GDP</i>		
DM27	<i>X0</i>	<i>X1</i> <i>Y</i>
	5.49609	11.9897 11.1407
<i>X0</i>	5.50523	11.9977 11.1846
161085	243.736	68921
162379	245.976	72010
166040	244.122	75123
176496	250.127	77606
197495	259.415	83306
...		

```
reg0 ^. DM27 NB. ln
1.27786 0.7074 0.25785
```

$$\ln Y = 1.27786 + 0.7074 \ln CAP + 0.25785 \ln LAB$$

```
reg0 reg_exam_ad DM27
+-----+-----+
|f=      |1.27786 0.7074 0.25785 |
+-----+-----+
|R  sqr: |99.3993          |
+-----+-----+
|AIC:   |_188.238         |
+-----+-----+
|DW:    |1.25405          |
+-----+-----+
|t=:    |3.46684 53.2258 5.43342|
+-----+-----+
```

製造業の *GDP* は資本財を 1% 増加させれば 0.707% 増え、労働力を 1% 増加させれば 0.25% 増加する。相関係数は非常に高く *t* 値も十分である。

図が示すように、バブル期の過剰設備投資が経済の変調で過剰設備となっており、その調整を労働面に求めていった軌跡がよく示されている。

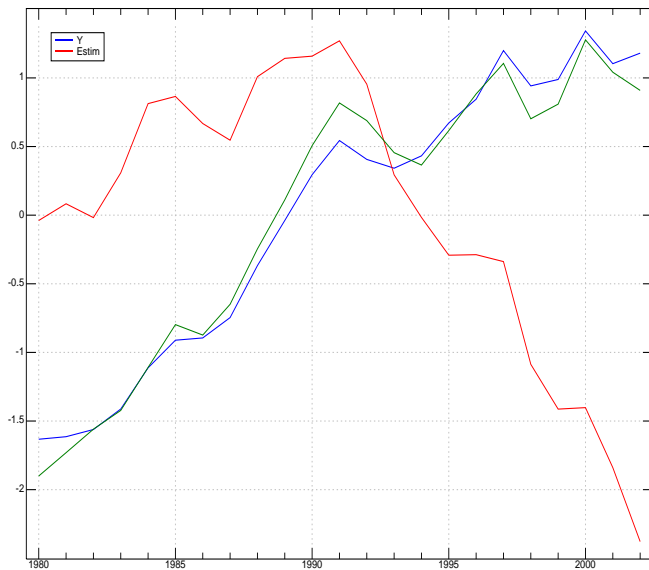


図1 製造業の時系列データ

多変数回帰はこのグラフのように足並みが揃っていないデータのモデル化が可能である。グラフは多様な数値を一枚に表せるように規準化している。

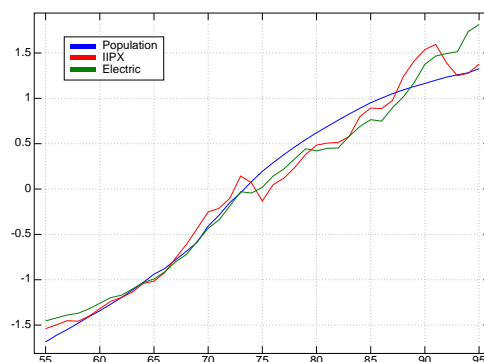
Worked Example 指数を用いた重相関の例 発電量と経済指数 (1969-1994)

重相関回帰は多入力 / 出力の回帰である。使用するデータを全て初年度の数値で割って指数化することで単位を取ってしまう手法である。

ここではデータが大きいのので、csv で読み込んでいる。

```
read_data0 '' NB. adjust your data folder
DS0      NB. J data
DN28     NB. 3 4 7 2 {"1 DS0 // pickup
DAT=: DN28 %("1) {. DN28 NB. divide by first raw
```

y 発電量
 x_1 人口
 x_2 鉱工業生産指数
 x_3 原油輸入量



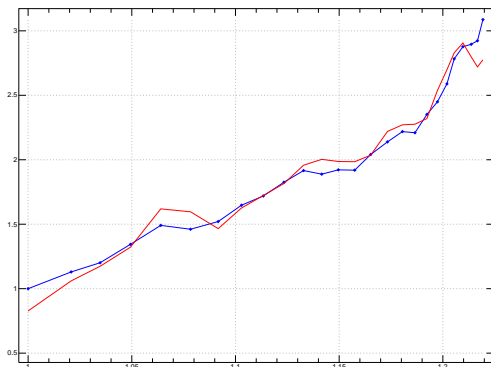
```
reg0 DAT
```

_3.10519 2.74842 0.860231 0.323955

$$f = -3.10519 + 2.74842x_1 + 0.860231x_2 + 0.323955x_3$$

reg0 reg_exam_ad DAT

```
+-----+-----+
|f=      |_3.10519 2.74842 0.860231 0.323955|
+-----+-----+
|R  sqr: |96.8057      |
+-----+-----+
|AIC:    |_109.527     |
+-----+-----+
|DW:     |0.582066    |
+-----+-----+
|t=:     |_3.25926 2.56923 5.78009 2.88985 |
+-----+-----+
```



Source: 杉原 敏夫「適応的モデルによる経済時系列分析」1996 工学図書)

2 回帰モデルの検定

t 検定は回帰係数毎の当て嵌まり具合を見ることができる。

重回帰モデルは変数を増やしていくと多重共線性という厄介な性質が生じる。 AIC を持ちいて変数を選択する事が望ましい。

2.1 t 検定

(スチューデントの t 検定) *Guinness brewery* の化学, 数学の技師であった *William Sealey Gossett*(1876-1937) が少量のサンプルで出来るビールの品質管理の研究を *Student* のペンネームで発表したものが *t - distribution* t - 分布 である。

正規分布を使用する平均値の差の検定は, 母分散が既知という条件下でしかも, 大標本のときだけ信頼できることは知られている. 小さな標本から平均値の差の有意性を判定するために考案された平均値の検定がスチューデントの t 検定である。

n 個の観測値の標本平均 m と母平均 μ の差 (距離) を不偏標本分散の平方根 u で割った統計量

$$t = \frac{(n - u)}{u / \sqrt{n}}$$

の分布が自由度 $n - 1$ の t 分布に従うことはゴセット (筆名 : スチューデント) が最初に発見し, フィッシャーが厳密に証明したことは歴史的事実として有名である。

William Sealey Gosset (1876-1937)



Oxford で化学と数学を学び, 1899 年ダブリンのギネス社の技師となる。1905 年にロンドンに出ての *Pearson* に学ぶ。確率論と結びついた近代統計学の幕開けは 1908 年にギネスビールの技師ゴセットによるペンネーム・スチューデントの t 検定で始まった。ギネス社は数年前に醸造責任者の論文により醸造の機密が漏れたことから論文発表を禁じていた。このスチューデントは終生身分を明らかにせず, 多くの論文を自宅で執筆し発表した。ギネス社が知ったのはゴセットの急死後, 友人がゴセットの論文集を刊行するにあたりギネス家に寄付を求めたときであった。

<http://www-history.mcs.st-andrews.ac.uk/Biographies/Gosset.html>

x が $(0, 1^2)$ の分布に従い, y が自由度 n の χ 分布に従うとき

$$\frac{x}{\sqrt{\frac{y}{n}}}$$

は自由度 n の t 分布に従う。

$N(\mu, \sigma^2)$ の分布から大きさ n の標本を繰り返し取って得られる分布の平均値を \bar{x} とし, 分散を s^2 とすると

$$\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}}$$

は自由度 $n - 1$ の t 分布に従う。

$$\frac{\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{ns^2}{\sigma^2}}}{n-1} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \cdot \frac{1}{\frac{\sqrt{n} \frac{s}{\sigma}}{\sqrt{n-1}}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}}$$

自由度 n の t 分布関数は

$$f_n(x) = \frac{1}{\sqrt{n}B\left(\frac{1}{2}, \frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

平均値 $E[x] = 0$

分散 $V[x] = \frac{n}{n-2}, (n \geq 3)$

この関数は正規分布より少し背が低く裾野が広い分布となる。

t 分布は個々の回帰係数に関しても検定でき、有意でない説明変数を示すことができる。 t 値は自由度 $n - 2$ の t 分布に従う。

*8

t 値はサンプルが大凡 30 個程度あれば、両側検定で 2.0 以上、片側検定で 1.7 以上あれば概ね回帰係数は有意であると見なせる。(有意水準 5%)

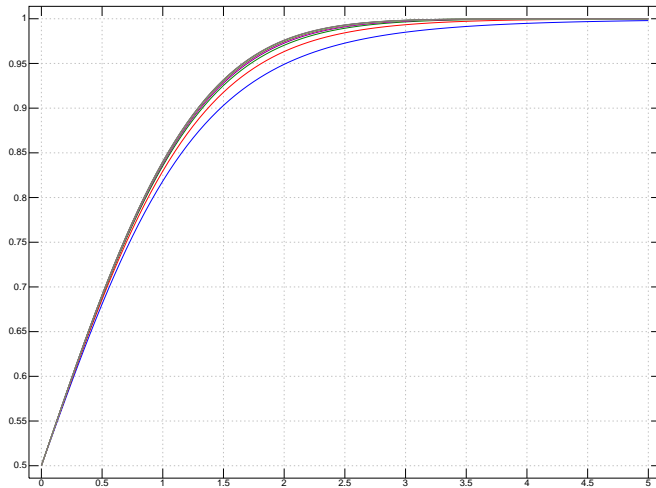
t 値とパーセント点の計算 自由度 1 5 10 20 30 40 50 70 80 90 100 120 のグラフ

$tcdf$ は *E.Show* が作成した芸術的な関数である。詳細は数学付録参照

```
a=. steps 0 5 100 NB. from 0 to 5 divide by 100 steps
```

```
plot a; ;("1"),.({@>1 5 10 20 30 40 50 70 80 90 100 120) tcdf L:0 a
```

$steps$ は *numeric.ijs* に入っている関数で、この例では 0 と 5 の間を 100 の等間隔刻みの数列を生成する。



X 軸に自由度、Y 軸にパーセント点をとる。Y 軸の 97.5% の値は自由度が 20 程度で概ね 2 に近い値になり大きくばらつくことはない。

自由度 50 t 値 2.009 のパーセント点は 97.5 で 両側検定ならば 95%、片側検定ならば 97.5% の値である。 $tcdf$ からはパーセント点が計算できる。

*8 サンプル数から説明変数 + 1 を引いた数

```
50 tcdf 2.009
0.975024
```

左引数に自由度 (19) 右引数に α が 0.1 0.05 0.025 0.01 0.005 ののポイントの t 値を指定してパーセント点を求める。ここで得られるのは片側検定の値で 0.95 0.975 0.985 0.995 0.9975 が得られる。両側検定のポイントは 0.90 0.95 0.975 0.99 0.995 ($1 - \alpha$ の値) となる。

```
19 tcdf 1.729 2.093 2.433 2.861 3.174
0.949988 0.974999 0.987488 0.995001 0.997502
```

t 分布は原点に対して対象な分布なので、 $t_{2.5} = -t_{2.5}$ である。 t 値の 95% 点 (95% 信頼区間) は 97.5% の t 値を用いる。

観測値が多いデータならば t 分布を正規分布で置き換えることもできる。正規分布の 97.5% 点は 1.96 であるので、この値を近似的に 2 とすれば t 値と概ね合致する。

2.2 p-value

最近の経済学などでは、 t 値と併せて、 p 値が記載されることが多い。

p value (probability value) は t -値から求められる。

p -value は帰無仮説のもとで検定統計量の値を超える確率を示し、帰無仮説を棄却する最少の有意水準を示す。

例えば、自由度 (df) 8 の t -value が 5.86 であった場合、 t -table の α の値は 0.0010 では 5.041 である。 $tcdf$ で求めた値は 5.041 よりも大きく α は 0.001 より小さくなる。 t 値から計算すると、 p 値は 0.000189233 でエラーの確率は概ね 0.02 %、 $\frac{2}{10000}$ であるといえる。

```
8 tcdf 5.86
0.999811
```

```
1- 8 tcdf 5.86
0.000189233 NB. p-value
```

2.3 AIC によるモデルの選択



赤池弘次 (1927-2009) 元統計数理研究所長
秩父セメントのエンジニア中川東一郎に協力してセメントキル
ンの自動制御技術を研究したとき、AIC を考案した。
二人の共著「ダイナミックシステムの統計的解析と制御」
1972/2000 (新版) (サイエンス社) に詳細が紹介されている。

2.3.1 AIC

AIC(Akaike Information Criteria)(赤池) 情報量規準はモデル選択の重要な指標である。

2003 年の暮れに横浜で、AIC 誕生 30 周年のシンポジウムが開催された。AIC は統計数理研究
所の赤池弘次により創られた理論である。セメントキルンのオートメーション化を図っていた秩父
セメントの中川東一郎に協力する過程で、数理理論として最適な制御方程式を曖昧さを残さないで
決定するという現実の課題の下で開発され、多くの現場で使い込まれた頑強な手法である。

最尤推定量を求めるには、尤度関数

$$LL = \log L = \sum_{i=1}^n \log f(x_i, \theta)$$

を最大にすることである。L の対数を取ったものを対数尤度 LL と、最大対数尤度を MLL とあ
らわす。

対数尤度を θ に関して微分したものを最尤方程式とする。

$$\frac{\partial}{\partial \theta} = \sum_{i=1}^n \log f(x_i, \theta) = 0$$

を解いて得られる解の中から LL を最大にする θ を選ばばよい。

正規分布の確率密度関数の対数尤度は

$$LL = \sum_{i=1}^n \frac{(x_i - u)^2}{2\sigma^2} + (u \text{ を含まない項})$$

最尤方程式は

$$\frac{\partial}{\partial u} LL = \sum_{i=1}^n \frac{x_i - u}{\sigma^2} = -\frac{1}{\sigma^2} [\sum x_i - nu] = 0$$

解き進めて u に対する最尤推定を求めると

$$u = \frac{1}{n} \sum_{i=1}^n x_i = M(x)$$

σ^2 に対する最尤推定量は

$$\sigma^2 = \frac{1}{n} \left[\sum_{i=1}^n (x_i - u)^2 \right] = V(x)$$

2.3.2 情報量規準 AIC のアルゴリズム

残差平方和

$$Q = (y - Xb)'(y - Xb) = y'y - (y'X)b$$

AIC 検定は、AIC の値が小さいほうのモデルの方がより良く当てはまるので、重相関のモデル選択に有用である。AIC はクライテリアであり、数値の比較で、小さい方のモデルの方が優れているという決定を行う。

数値自体には意味はないので、モデル選択時には、AIC の値を添えればよい。reg_exam_ad は回帰係数、決定係数、t-value、AIC を同時に計算する。

$$MLL = -(2/n) \log(Q/n) AIC =: n \log(Q/n) + 2(k+1)$$

重相関回帰は変数を増やしていくと決定係数が上がって一見両構想に見えるがデータの派閥とも言うべき多重共線性が生じると言う厄介な問題があるが AIC の小さい方のモデルを選択することで解決でき、DW 検定 よりも使いやすい。

2.4 AIC による変数選択の方法

重相関では AIC による変数選択が有用である。

変数を選択する場合に、AIC のみを用いる簡単な 2 の関数を用いて最適指標を選択する方法を考える。

このような多くの変数からふさわしい変数を AIC で選択する方法を紹介する。SCRIPT は竹内・鈴木による。

2.4.1 竹内・鈴木関数のテスト

```
R=, sample_data '' NB. sample data
```

サンプルデータ (R) から 9 変数を選択する場合の 9 個の組み合わせと AIC 値 (ソート後・最初の 10 個)

*9

*10

*9 サンプルデータは numeric_recipe_data.ijs に入っている。numeric_recipe_main.ijs でプリロードの指定がされていないときは手動でロードする。

ここではデータは Y,X は分離された型となっている。

*10 sample_data'' で Y1,Y2 Y3 R が利用できる。

竹内・鈴木関数はデータ X は横長の形で受け付けるようになっている。縦長のデータを用いるときは (|:) で *rotate* する。

find_best0 は AIC の小さい順に並べる。

```

10{. find_best0 Y1 compare 9;R
+-----+
|1 1 0 1 0 1 1 1 0 0 1 1 1|209.464|
+-----+
|1 1 0 1 0 1 1 1 1 0 1 0 1|209.542|
+-----+
|1 1 0 1 1 1 1 1 0 0 1 0 1|209.602|
+-----+
|1 1 0 1 0 1 1 1 0 1 1 0 1|209.603|
+-----+
|1 1 1 1 0 1 1 1 0 0 1 0 1|209.614|
+-----+
|1 1 0 1 0 1 0 1 1 0 1 1 1|209.667|
+-----+
|1 1 1 1 0 1 0 1 1 0 1 0 1|209.691|
+-----+
|1 1 0 1 1 1 0 1 1 0 1 0 1|209.705|
+-----+
|1 1 0 1 0 1 0 1 0 1 1 1 1|209.714|
+-----+
|1 1 0 1 0 1 0 1 1 1 1 0 1|209.718|
+-----+

```

find_best0 を見ても、AIC の値は多少うつろうが、並べてみて、0 の谷の深いところを外していけば、大綱では概ね妥当な選択が出来ている。

find_best は組み合わせの範囲を指定して、各組毎の、AIC の最少値を表示する。組み合わせは、 ${}_m C_n$ の n が大きいと *out_of_memory* を頻発する。

次は 7 変数から 12 変数の場合の AIC 最少の組み合わせを選ぶ方法である。

```

(7;12) find_best Y1; R
+---+-----+-----+
|7 |1 1 0 1 0 1 0 1 0 0 1 0 1|209.881|
+---+-----+-----+

```



```

|8 |1 1 0 1 0 1 1 1 0 0 1 0 1|209.615|
+---+-----+-----+
|9 |1 1 0 1 0 1 1 1 0 0 1 1 1|209.464|
+---+-----+-----+
|10|1 1 0 1 0 1 1 1 0 1 1 1 1|209.395|
+---+-----+-----+
|11|1 1 0 1 1 1 1 1 0 1 1 1 1|209.36 |
+---+-----+-----+
|12|1 1 0 1 1 1 1 1 1 1 1 1 1|209.357|
+---+-----+-----+

```

ESRI(内閣府経済社会総合研究所) が毎月発表する景気動向指数 (*Leading Coincident Lagged*) の月次指標を用いる方法を考えると景気動向指数の項目は 32 あり、これから重相関に用いる指標を数項目選びだそうとすると、

$${}_{32}C_{10}$$

$10!32 = 64512240$ の組み合わせが出来る。大綱を選択しなければ収拾がつかない。

景気動向指数のデータを用いる場合は予選リーグを行い、*LC LG* 毎に相関の少ないものを先に外していくと計算量がずっと少なくなって、妥当な結果が得られる。

理論と関係なくマシンに変数選択を委ねることはないが重相関の「多重共線性と *AIC* の関係」を体験することは実際に変数選択を行う場合に生かされる

3 トレンド項とダミー変数

3.1 トレンド項

豌豆がのびるとき支柱を立てる。時系列データを相互に回帰しようとするとき、トレンド項を入れると、トレンドがトレンド項に吸収され、目的のデータ相互は、トレンドが除去されて回帰できる。

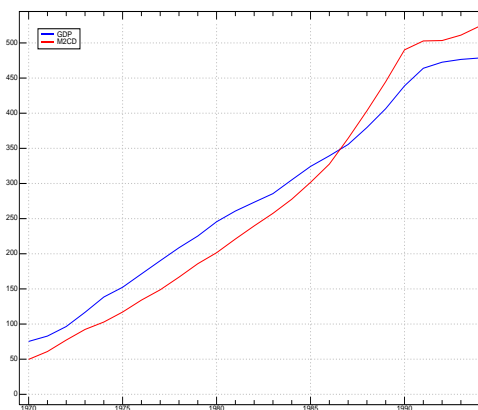
Worked Example トレンドの除去 (1) *GDP* と *M2CD*

```
read_data0 '' --> DMN0
DN29=: (i. # DMN0),. ^ . 2 1{"1 DMN0
```

出典：箕谷「計量経済学（第3版）」東洋経済新報社 1997
DN29

```
trend lnM2CD lnGDP(y)
```

0	3.90657	4.32146
1	4.10935	4.41763
2	4.34661	4.5694
3	4.52536	4.75974
4	4.6334	4.93052
5	4.76311	5.02626
.....		
23	6.23655	6.16655



```
24 6.26131 6.17084
```

マネーサプライと *GDP* の関係は *GDP, M2CD* を対数に変換して、トレンド項を入れることによって求められる。トレンドの係数は β_0 にでる。

$$\ln NGDP = \beta_0 + \beta_1 TIME + \beta_2 M2CDAV$$

$$\ln NGDP = 0.493964 - 0.017084 TIME + 0.967544 \ln M2CDAV$$

マネーサプライ 1% の増加は *GDP* を 0.97% 押し上げる。

```
DN29=: (i. # DMN0),. ^ . 2 1{"1 DMN0
```

```
reg0 DN29
```

```
0.493964 _0.017084 0.967544
```

*11

G

*11 トレンドのオリジンは0とした

為替レート EXR
 Worked Example 消費者物価 CPI
 賃金率 Yr

read_data0 '' --> DN30

出典:浅野・中村「計量経済学」有斐閣 2000

トレンド項を入れる	<pre>reg0 (1{"1 DN30}),. (>:i. # DN30) ,. 2{"1 DN30 26.4252 0.0381768 3.5465 CPI = 26.4252 + 0.0381768EXR + 3.5465Yr >:i.# は.トレンド項を作成(>:は I オリジン)</pre>
EXR の逆数とトレンド項	<pre>reg0 (% 1{"1 DN30}),. (>:i. # DN30) ,. 2{"1 DN30 51.4814 _5487.67 4.42389 y = 51.4814 - 5487.67(1/EXR) + 4.42389Yr %は逆数を求める</pre>
対数とトレンド項	<pre>reg0 (^ .1{"1 DN30}),.(>:i. # DN30),. ^ . 2{"1 DN30 1.50601 0.377514 0.0639988 y = ln1.50601 + ln0.377514 + 0.0639988Yr</pre>

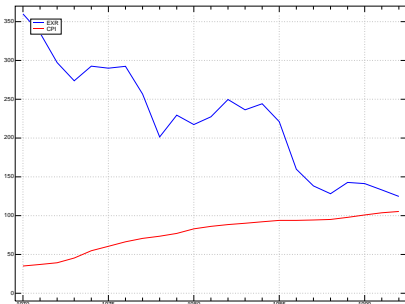
トレンド項は対数にしない。

AIC で見てもこの例では2番目の EXR の逆数とトレンド項を用いたモデルの信頼性が一番高い。

```
reg0 reg_exam_ad (% 1 {"1 DN30}),.(>:i.# DN30),.2{"1 DN30
+-----+-----+
|f=      |51.4814 _5487.67 4.42389|
+-----+-----+
|R_sqr: |95.478                      |
+-----+-----+
|AIC:   |76.8481                      |
+-----+-----+
|DW:    |0.364896                      |
+-----+-----+
```

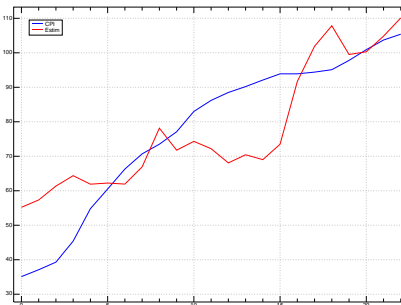
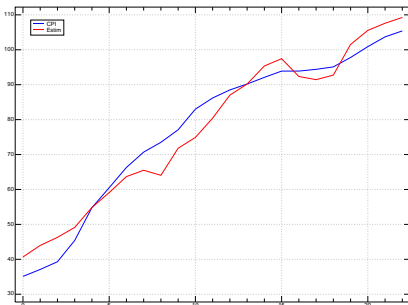
|t=: |13.2416 _3.71547 11.9488|

+-----+-----+-----+-----+-----+



トレンド項はX軸が時間でない変数の時に挿入するとトレンドデータがトレンド項に集まり分析が容易となる。

この例で有り無しをグラフで比較してみよう。無い場合の相関係数は0.65程度である。



3.2 ダミー変数を用いたモデル

ダミー変数は

1. パラメーターの構造変化（バブル, リセッション）や数量化できない質的属性（災害, 戦争）などを含んだデータの構造変化や特性を考慮し, ダミーを付けることにより, 回帰方程式を何本かに分けることができる。
2. 季節変動のある四季や月次データを, 四季や繁忙期を分けるのにも用いられる。

ダミーデータを用いたて, 重相関と同じように回帰できる。

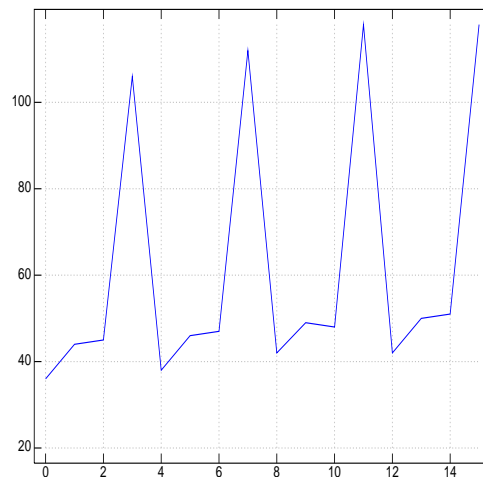
3.2.1 季節ダミー

季節変動はダミー項を入れることによりダミー項にウエイトが吸収される。

四季のダミーを入れる	<i>mk_dummy4</i>	<i>x mk_dummy4 y</i> 左パラメータ <i>0 Spring</i> <i>1 Summer</i> <i>2 Autumn</i> <i>3 Winter</i>
月のダミーを入れる (複数指定可能)	<i>mk_dummy12</i>	<i>8 1 2 mk_dummy12 y.</i>

Worked Example *Canada* のある店の宝石売り上げの 4 年分のシーズンデータ。
クリスマスシーズンが突出している。

```
4 4 $ DN31
36 44 45 106
38 46 47 112
42 49 48 118
42 50 51 118
I II III Xmas
```



経過と解説 .

- 指定した 4 半期を分けて、フラグを立てる。
四半期は会計区分や四季（春夏秋冬）に対応する。時間（季節）は 0 オリジンの順序数とする

```
|: 1 2 3 =/ 0 1 2 3 NB. rotate identity matrix
0 0 0
1 0 0
0 1 0
0 0 1
```

- X* を作成する
I 期を除き *II, III, Xmas* 期にダミーを入れる（0 オリジン。自動でトレンド項が左に入る）

```

1 2 3 mk_dummy4 DN31
0 0 0 0 36
1 0 0 1 44
2 0 1 0 45
3 1 0 0 106 NB. Xmas
4 0 0 0 38
5 0 0 1 46
6 0 1 0 47
7 1 0 0 112 NB. Xmax
8 0 0 0 42
9 0 0 1 49
10 0 1 0 48
11 1 0 0 118 NB. Xmas
12 0 0 0 42
13 0 0 1 50
14 0 1 0 51
15 1 0 0 118 NB. Xmas

```

- 回帰係数を求める

```

reg0 1 2 3 mk_dummy4 DN31
35.6 0.65 72.05 6.95 7.1

```

```

linefit_reg0 1 2 3 mk_dummy4 DN31

```

$$f = 35.6 + 0.65Q_0 + 72.05Q_1 + 6.95Q_2 + 7.1Q_3$$

- クリスマスシーズンのみが突出しているので、3のみにタグを付けても良い。

```

|: 3 mk_dummy0 DN31
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1

```

```

reg0 3 mk_dummy4 DN31
39.69 0.734756 67.1972

```

```

linefit_reg0 3 mk_dummy4 DN31

```

$$f = 39.69 + 0.734756Q_3 + 67.1972Q$$

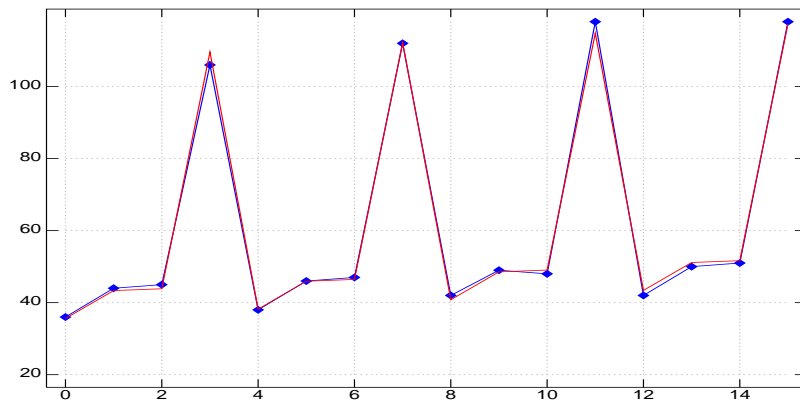


図2 宝石の売り上げと推計値

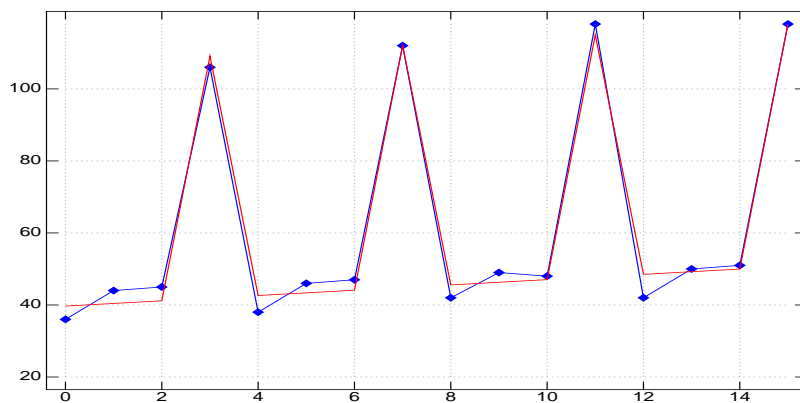


図3 宝石の売り上げと推計値

3.2.2 重相関とダミー変数

多変数の重回帰に対してもダミー変数をつけることができる。経済変動により何種類かのダミー変数をつけて回帰することも可能である。

4 不均一分散・系列相関

4.1 系列相関 (1)・ダービン・ワトソン検定

Durbin は 2000 年に来日した。London School of Economics で教えており、時系列分野でも活躍している

自己相関のテストとして、ダービンとワトソンにより創られたダービン・ワトソンテストがある。系列相関のテスト。系列相関がない場合は、値が 2 に近くなる。

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

Worked Example 消費支出と過分所得

```
DN32=. ".@> readcsv DIRCSV,'shirasago_0.csv'
```

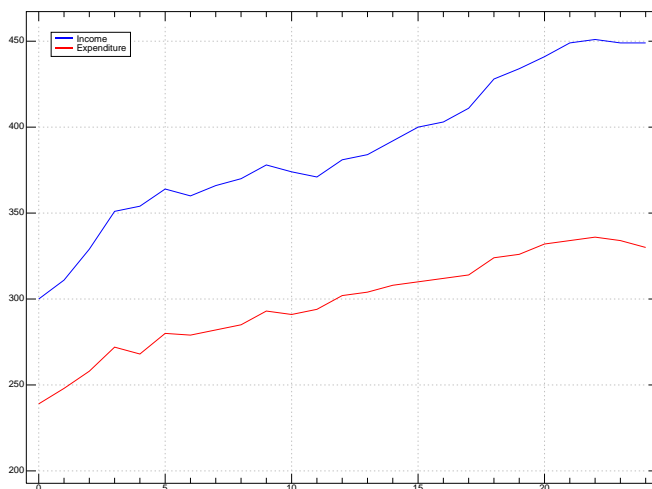


図 4 家計の所得と支出 (実質)

	Y	X	Y	X		
1970	239	300	83	304	384	Y = 実質家計可処分所得
71	248	311	84	308	392	X = 実質家計消費支出
72	258	329	85	310	400	
73	272	351	86	312	403	(出典) 白砂「初歩からの計量経済学」
74	268	354	87	314	411	
75	280	364	88	324	428	


```

| 76 279 360|89 326 434|
| 77 282 366|90 332 441|
| 78 285 370|91 334 449|
| 79 293 378|92 336 451|
| 80 291 374|93 334 449|
| 81 294 371|94 330 449|
| 82 302 381|          |
+-----+-----+

```

```

OLS      reg0 2 1{"1 DN32
50.8745 0.637437
y = 50.8745 + 0.637437x

```

```

      reg0 reg_exam_ad 2 1{"1 DN32
+-----+-----+
|f=      |50.8745 0.637437|
+-----+-----+
|R sqr: |97.5095          |
+-----+-----+
|AIC:   |77.2247          |
+-----+-----+
|DW:    |0.352762          |
+-----+-----+
|t=:    |6.13607 30.0085 |
+-----+-----+

```

相関係数、 t 値は申し分ないが DW 値が低すぎ、系列相関が存在する。

```

DW 検定  dw 2 1 {"1 DN32
0.352762

```

```

      dw=: 3 : 0
NB. usage: dw y same reg0
TMP0=. estim_reg0 y
TMP1=. ({"1 y)- TMP0 NB. ut
TMP2=. ; -/&|. (L:0) 2<\ TMP1 NB. ut-ut_1
(+/ ^&2 TMP2)% +/ ^&2 TMP1
)

```

多重共線性はデータの派閥で、厄介なことに重相関ではデータを増やすと多重共線性が生じ相関係数が上がって、もっともらしくなる。AIC や DW 検定で整理しなければならない。

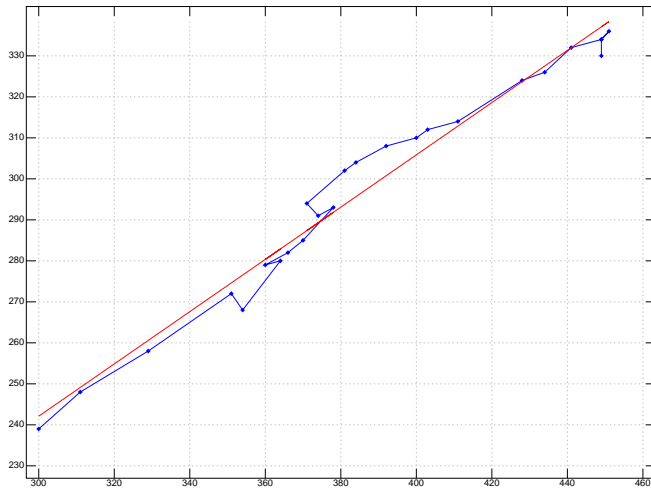


図5 家計の所得 (X) と支出 (実質 Y)

4.2 系列相関 (2)・コ克蘭・オーカット法 Cochrane-Orcutt(C-O) Method

階差モデルにより, 系列相関を取り去る方法に, *Cochrane-Orcutt Method* がある.

回帰モデル

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

誤差項

$$u_t = \rho u_{t-1} + \epsilon_t$$

t 時点前の式に ρ をかける

$$\rho Y_{t-1} = \rho \beta_1 + \rho \beta_2 X_{t-1} + \rho u_{t-1}$$

この式を回帰モデルから引く

$$(Y_t - \rho Y_{t-1}) = (1 - \rho)\beta_1 + \beta_2(X_t - \rho X_{t-1}) + \epsilon_t$$

置き換えの項

$$Y_t^* = Y_t - \rho Y_{t-1}$$

$$X_t^* = X_t - \rho X_{t-1}$$

$$\beta_1^* = (1 - \rho)\beta_1$$

$$\beta_2^* = \beta_2$$

系列相関のないモデル

$$Y_t^* = \beta_1^* + \beta_2^* X_t^* + \epsilon_t$$

YI 消費と XI 可処分所得、消費関数のモデルであるが、消費が遅れ、系列相関が出ている。(曲線が原データの散布図)

C-O Method

4 はリピート回数 4 回でほぼ収束している. $cor|f = \beta_1 + \beta_2 X$

```

4 co_method 2 1 {"1 DN32
+-----+-----+
|0.850961|93.7563 0.535125|
+-----+-----+
|0.832879|92.0731 0.538302|
+-----+-----+
|0.828025|91.5287 0.539413|
+-----+-----+
|0.826887|91.3968 0.539686|
+-----+-----+

```

$$\rho = 0.826887$$

$$Y_t = 91.3968 + 0.539686X_t$$

4.2.1 Script

```

co_method=:4 : 0
NB. Usage: 4 co_method y
NB. Cochrane-Orcutt Method by Giichiro Suzuki
NB. slightly modified M.Shimura
YY=. {|. |: y NB. slightly modified
cor=: ([: +/ }. *}:) % [: +/ [: *: }:
mdy=: }.@] - [*}:@]
q=: ,. ''
b=: (x0=>{.YY) %. y1=: 1,.y0=: >{: YY
while. x >#q
do. r=: cor u0=: x0-y1 +/ . * b

```

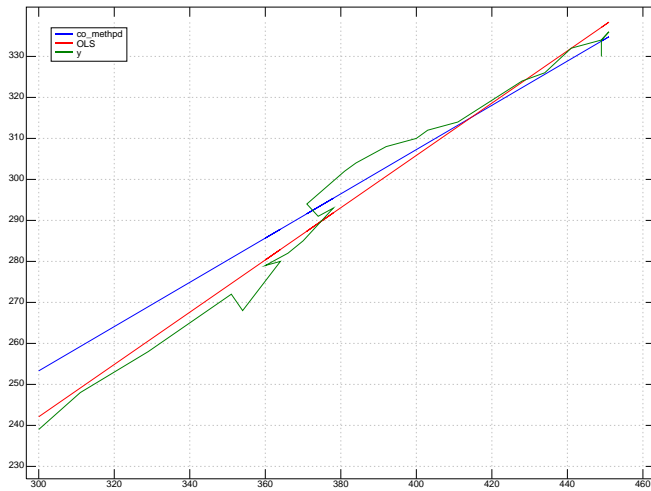


図 6 家計の所得 (X) と支出 (Co OLS)

```

b=: ( r mdy x0) %. 1, .r mdy y0
q=: q,r;b=: (({. b)%1-r),}. b
end.
)

```

4.3 不均一分散 (1)・一般化最少自乗法

最少自乗法ではモデルの誤差が一定であると仮定しているが、不均一なケースもある。不均一分散 (*heteroscedasticity*) はトレンドや国別統計の大小格差などにより分散が乱れた場合に生じる。

プレイス・ウインスティン変換による一般化最少自乗法

最初の $t = 1$

$$\begin{aligned}
 Y_1^* &= \sqrt{1 - \hat{\rho}^2} Y_1 \\
 C_1^* &= \sqrt{1 - \hat{\rho}^2} \\
 X_1^* &= \sqrt{1 - \hat{\rho}^2} X_1
 \end{aligned}$$

$t = 2, \dots, n$ のとき

$$\begin{aligned}
 Y_t^* &= Y_t - \hat{\rho} Y_{t-1} \\
 C_t^* &= 1 - \hat{\rho} \\
 X_t^* &= X_t - \hat{\rho} X_{t-1}
 \end{aligned}$$

定数項なしの OLS で推定し、 α と β を求める。

```
pw_method 2 1 {"1 DN32
```

55.4676 0.621306

$$Y = 55.4676 + 0.621306X$$

	pw_sub0	a1				
Year	Ct*	Xi*	Yt*			
				82	0.149039	65.2935 51.8175
				83	0.149039	59.7839 47.0098
1970	0.525229	157.569	125.53	84	0.149039	65.231 49.3079
71	0.149039	55.7117	44.6203	85	0.149039	66.4233 47.904
72	0.149039	64.3511	46.9617	86	0.149039	62.6156 48.2021
73	0.149039	71.0338	52.4521	87	0.149039	68.0627 48.5002
74	0.149039	55.3127	36.5386	88	0.149039	78.255 56.7982
75	0.149039	62.7598	51.9425	89	0.149039	69.7887 50.2886
76	0.149039	50.2502	40.7309	90	0.149039	71.6829 54.5867
77	0.149039	59.654	44.5819	91	0.149039	73.7262 51.4809
78	0.149039	58.5483	45.029	92	0.149039	68.9185 51.779
79	0.149039	63.1444	50.4761	93	0.149039	65.2166 48.0771
80	0.149039	52.3367	41.6684	94	0.149039	66.9185 45.779
81	0.149039	52.7406	46.3703			

X軸には可処分所得を取り、散布図 (Blue) と OLS (Green) PW による一般化最少自乗法 (Red) を落としてみると、GLSの方が素直な回帰になっている。

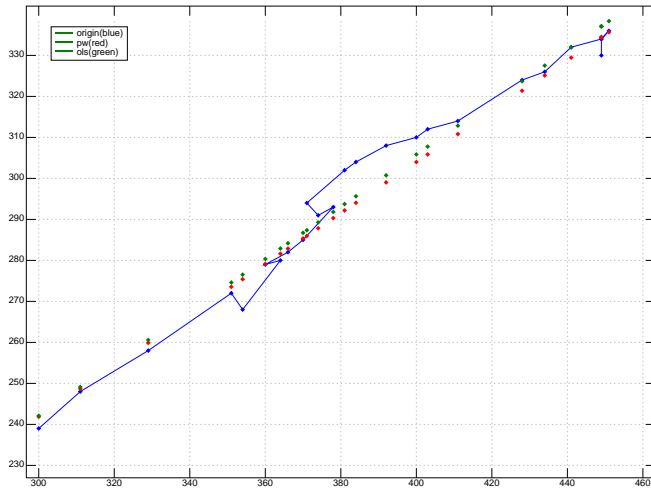


図7 家計の所得と支出 (実質)

4.4 不均一分散 (2)・加重最少自乗法 WLS

4.4.1 σ_i^2 is known/Weighted Least Squares(WLS)

不均一分散が観察できる場合に、回帰式 $Y_i = B_1 + B_2 X_i + u_i$ において、 σ_i^2 が計算できる場合は、左右の項を、 σ_i で、割るかまたは、収縮 (deflate) した次のモデルが適用できる。

$$\frac{Y_i}{\sigma_i} = B_1 \frac{1}{\sigma_i} + B_2 \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i}$$

$$v_i = \frac{u_i}{\sigma_i}$$

$$v_i^2 = \frac{u_i^2}{\sigma_i^2}$$

4.4.2 $\sqrt{X_i}$ transformation

誤差の分散が X_i に比例する場合には、

$\sqrt{X_i}$ (square root transformation) を適用できる。

$$\frac{Y_i}{\sqrt{X_i}} = B_1 \frac{1}{\sqrt{X_i}} + B_2 \frac{X_i}{\sqrt{X_i}} + \frac{u_i}{\sqrt{X_i}} = B_1 \frac{1}{\sqrt{X_i}} + B_2 \sqrt{X_i} + v_i$$

```
wls 2 1 {"1 DN32
48.3814 0.643862
```

```
y = 48.3814 + 0.643862X
wls
```

```
wls reg_exam_ad 2 1 {"1 DN32
+-----+-----+
|f=      |48.3814 0.643862|
+-----+-----+
|R_sqr: |97.4996          |
+-----+-----+
|AIC:   |77.3239          |
+-----+-----+
|DW:    |0.352839         |
+-----+-----+
|t=:    |5.8238 30.2508  |
+-----+-----+
```

4.5 関数一覧

<i>OLS</i>	<i>reg0</i> <i>reg_exam_ad</i>	<i>reg0 y</i> <i>reg0 reg_exam_ad y</i>
正規方程式	<i>test_neq</i>	<i>test_neq y</i>
クラメル法	<i>cr</i>	<i>cr test_neq y</i>
<i>t</i> 検定、 <i>p</i> 値	<i>tcdf</i>	<i>m tcdf n</i>
季節ダミー	<i>mk_dummy4</i> <i>mk_dummy12</i>	<i>m mk_dummy4 n</i> <i>m mk_dummy12 n</i>
<i>reg_select</i>	<i>find_best0</i> <i>find_best</i>	<i>find_best0 y cmpare n;x</i> <i>(m1;m2)find_best y;x</i>

系列相關、不均一分散	<i>dw</i> <i>co_method</i> <i>pw_method</i> <i>wls</i>	<i>dw y</i> <i>co_method y</i> <i>pw_method y</i> <i>wls y</i>
------------	---	---

5 種々の回帰分析

多項式、乗法モデルやニュートン法を用いた非線形モデルなど、曲線による柔軟な回帰モデルや、質的データをグループロジットモデル、自己の過去のデータを織り込む動学モデルやハウ・タッカーモデルを紹介する。また、ジニ係数による所得配分の推移の比較を行ってみる。

5.1 多項式モデル

5.1.1 多項式のモデル選択

多項式の最適次数は AIC の最少の値で求める。

東京都の3月の最低気温のデータでモデル選択を行う。4次の AIC が最少である。

DATA is DN12 in numeric_recipe_data.ijs

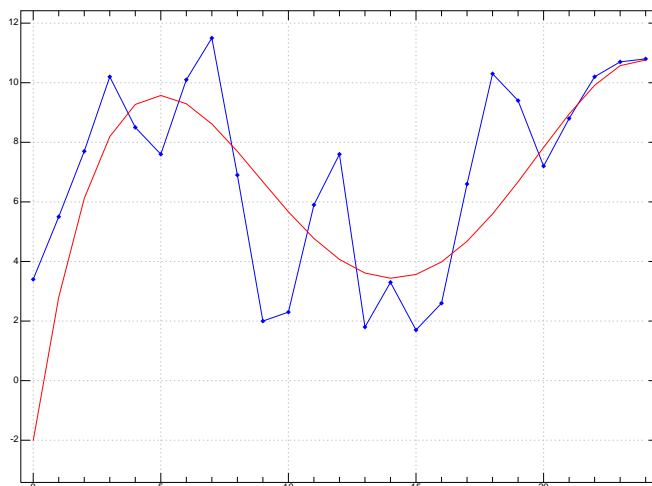


図8 東京の冬の最低気温(青)と推計値(赤)4次

```
3 poly0 DN12
4.89554 1.13288 _0.141456 0.00438882

y = 4.89554 + 1.13288x - 0.141456x2 + 0.00438882x3

3 poly_exam DN12
co. of det.(%): 39.09
value of AIC : 53.52
Value of MLE: 4.896 1.133 _0.141 0.004
```

4 poly_exam DN12
 co. of det.(%): 62.51
 value of AIC : 43.38
 Value of MLE: _2.015 5.656 _0.889 0.049 _0.001

5 poly_exam DN12
 co. of det.(%): 62.51
 value of AIC : 45.38
 Value of MLE: _1.936 5.585 _0.871 0.047 _0.001 0.000

AIC が最少の 4 次のモデルを採用する。

$$y = -2.01542 + 5.65633x - 0.889141x^2 + 0.0485201x^3 - 0.000848678x^4$$

linefit linefit と散布図を描く

4 linefit_poly0 DN12

推計値を求める estim_poly0

4 estim_poly0 DN12	7.6	4.06875
3.4 _2.01542	1.8	3.61157
5.5 2.79945	3.3	3.43786
7.7 6.11526	1.7	3.56376
10.2 8.19261	2.6	3.98505
8.5 9.27167	6.6	4.67716
7.6 9.5723	10.3	5.59512
10.1 9.29394	9.4	6.67361
11.5 8.6157	7.2	7.82693
6.9 7.6963	8.8	8.94903
2 6.6741	10.2	9.91348
2.3 5.66709	10.7	10.5735
5.9 4.77289	10.8	10.7618

Linefit *linefit_poly*

linefit_poly でグラフを描いて比較してみる。(図 8)

```
4 linefit_poly0 DN12
```

J Grammar

p. *J* の多項式の記述法は、多項式 (*p.*) に渡す関数定義では、高次の項が右に来る。

steps *steps* は *nimeric* のファイルに入っている関数で、グラフを描くときは間隔が滑らかになる。

i:n *i : n* は主にグラフ用に 0 の両側に *n* 個ずつ展開するので、*X* 軸の指定に有用である。

```
require 'plot numeric trig'
```

```
    i:5  
  _5 _4 _3 _2 _1 0 1 2 3 4 5
```

5.1.2 2 変数の多項式

速度 *v* の時のブレーキを踏んでから車が停止するまでの距離 *y* が *v* との 2 次式の関係にある。

$$y = av^2 + bv + c$$

5 回の測定データから多項式で回帰を行う。

DN40

V Y

9 3

17 9

17 5

25 14

35 23

V_i	走行速度	kmh		9	17	17	25	35
y	制動距離	m		3	9	5	14	23

5.1.3 J の方法

正規方程式を経由しないで直接に回帰係数 (ベクトル) を求める

```
2 poly0 DN40
```

$$y = -0.524753 + 0.259012x + 0.0119353x^2$$

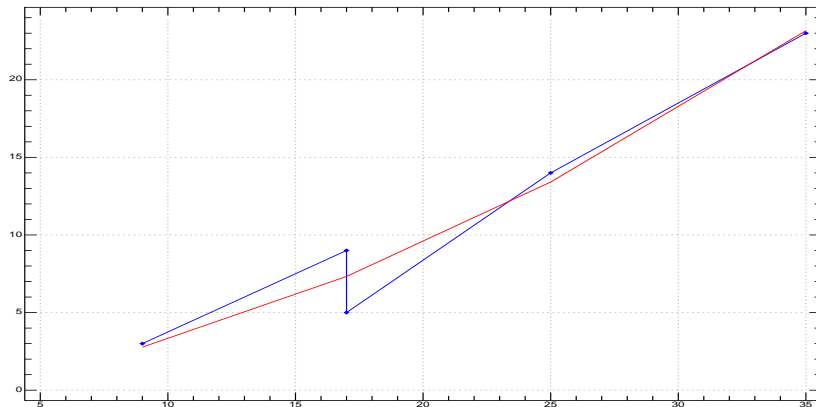


図9 車の速度と停止距離

<pre>({. "1 DN40")^/i.3 1 9 81 1 17 289 1 17 289 1 25 625 1 35 1225</pre>	<pre>({: "1 DN40) %. ({. "1 DN40)^/i.3 _0.524759 0.259012 0.0119352</pre>
---	---

5.1.4 最適次数の選択

最適次数はAICにより定める。

```
2 poly_exam DN40
co. of det.(%): 96.64
value of AIC : 8.74
Value of MLE: _0.525 0.259 0.012
```

```
3 poly_exam DN40
co. of det.(%): 96.88
value of AIC : 10.35
Value of MLE: 5.330 _0.790 0.067 _0.001
```

```
4 poly_exam DN40
co. of det.(%): 95.29
value of AIC : 14.42
```

Value of MLE:*****

この例では 2 次の AIC が最少である。2 linefit_poly0 DN40 によりグラフを描いてみよう。

5.1.5 多項式の正規方程式

多項式の正規方程式を用いると

$$\begin{bmatrix} \sum v_i^4 & \sum v_i^3 & \sum v_i^2 \\ \sum v_i^3 & \sum v_i^2 & \sum v_i \\ \sum v_i^2 & \sum v_i & \sum 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum v_i^2 y_i \\ \sum v_i y_i \\ \sum y_i \end{bmatrix}$$

$\begin{bmatrix} \sum v_i^4 & \sum v_i^3 & \sum v_i^2 \\ \sum v_i^3 & \sum v_i^2 & \sum v_i \\ \sum v_i^2 & \sum v_i & \sum 1 \end{bmatrix}$	<pre>a=. ; ("1) ,.3<\ +/ V ^/ 4 3 2 1 0 2.06485e6 69055 2509 69055 2509 103 2509 103 5</pre>
$\begin{bmatrix} \sum v_i^2 y_i \\ \sum v_i y_i \\ \sum y_i \end{bmatrix}$	<pre>b=. +/ Y * V ^/ 2 1 0 41214 1420 54</pre>
	<pre>b %. a 0.0119352 0.259012 _0.524761 y = 0.0119352x^2 + 0.259012x - 0.524761</pre>

*12

多項式の計算の留意事項 多項式の計算に際しての留意事項

64,32 ビットなど有限の桁数の CPU で計算するので、数値計算を専門とする言語でも桁数が多くなると分数演算や浮動小数点に移行する。多項式の計算を行う場合には次のような現象が現れることがあるので注意を要する。

*12 J の多項式関数の統一記述方式で、逆に出る

1. 正規方程式を経由する場合と、データを直接計算する場合とでは、次数が高くなると回帰係数に差が生じる。
2. 多項式回帰では、正規方程式を作る方法は、乗数が $2n+1$ になり、発散が早い。
3. 高次の多項式は次数を増やすと、高次の部分が高周波部分に相当し、微少な調整を行う。
AIC は微少な差も検出して値を出す、最高次の数を見ながら、適当な次数を選択する。(30 次数などむやみに高次を求めない)
4. 次数が上がっていくと、AIC は小さくなるが計算がオーバーフローして、回帰係数が崩れてしまう場合がある。図で確認すると全く外れている。

5.2 乗法モデル

指数を用いた回帰は乗法弾力性モデルと呼ばれる。

$$Y_t = a_0 \times x_{t1}^{\alpha_1} \times x_{t2}^{\alpha_2} \times \dots \times x_{tm}^{\alpha_m} \times \epsilon_t$$

$$\ln Y_t = \ln a_0 + \alpha_1 \ln x_{t1} + \alpha_2 \ln x_{t2} + \dots + \alpha_m \ln x_{tm} + \ln \epsilon_t$$

乗法モデルは、従属変数の変動を m 個の独立変数の乗法結合で表す。回帰は対数で行い、モデルの記述と推計の時には乗法を用いる。この対数回帰の時に、両対数の乗法弾力性モデルと従属変数に対数を用いる乗法指数モデルがある。

Worked Example データは DN41 で、季節性のある、アメリカのある都市の大きな本屋の売り上げである。まず、季節のタグを付けてトレンドと季節変動を区分する。

*13

これで直ちに多変量回帰が行える。売り上げに対数を取ることもしできる。対数回帰も同時に行ってみよう。

重回帰モデル	<pre>reg0 1 2 3 mk_dummy0 DN41 64.8625 1.98125 113.056 48.2875 40.0188 f_t = 64.8625 + 1.98125t + 113.056Q_{4t} + 48.2875Q_{3t} + 40.0188Q_{2t}</pre>
--------	--

*13 トレンド項のオリジンを 0 としている。定数項の値は 1 オリジンと異なるが推計結果は同じである。

1 2 3 mk_dummy4 DN41

DN41=:1 2 3 mk_dummy4 ^.DN41

NB. 对数

0 0 0 0 72
1 0 0 1 110
2 0 1 0 117
3 1 0 0 172
4 0 0 0 76
5 0 0 1 112
6 0 1 0 130
7 1 0 0 194
8 0 0 0 78
9 0 0 1 119
10 0 1 0 128
11 1 0 0 201
12 0 0 0 81
13 0 0 1 134
14 0 1 0 141
15 1 0 0 216

0 0 0 0 4.27667
1 0 0 1 4.70048
2 0 1 0 4.76217
3 1 0 0 5.14749
4 0 0 0 4.33073
5 0 0 1 4.7185
6 0 1 0 4.86753
7 1 0 0 5.26786
8 0 0 0 4.35671
9 0 0 1 4.77912
10 0 1 0 4.85203
11 1 0 0 5.3033
12 0 0 0 4.39445
13 0 0 1 4.89784
14 0 1 0 4.94876
15 1 0 0 5.37528

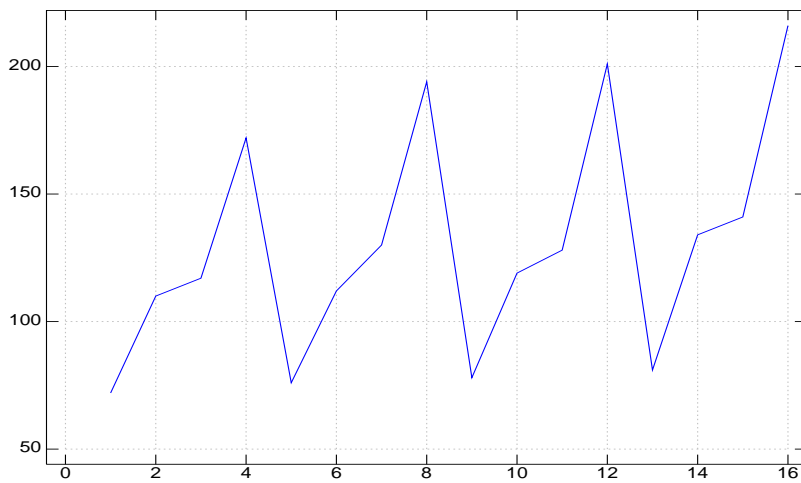


图 10 exp-reg

対数重回帰モデル	<p>対数モデル</p> $\ln(D_t) = a' + b'_1 t + b'_2 Q_{2t} + b'_3 Q_{3t} + b'_4 Q_{4t} + e'_t$ <pre>reg0 1 2 3 mk_dummy0 ^ . DN41 4.25357 0.0143442 0.890812 0.489297 0.420002</pre> $\ln \hat{f}_t = 4.25357 + 0.0143442t + 0.890812Q_{4t} + 0.489297Q_{3t} + 0.420002Q_{2t}$
指数乗法モデル	<p>対数回帰の値を e のべき乗とする</p> $Y_t = a_0 \times \alpha_1^{xt1} \times \alpha_2^{xt2} \times \dots \times \alpha_m^{xtm} \times \epsilon_t, (t = 1, \dots, N)$ $\hat{D}_t = (e^{a'}) (e^{b'_1 t}) (e^{b'_2 Q_{2t}}) (e^{b'_3 Q_{3t}}) (e^{b'_4 Q_{4t}})$ $\hat{D}_t = (e^{4.25357}) (e^{+0.0143442t}) (e^{0.890812Q_{4t}}) (e^{0.489297Q_{3t}}) (e^{0.420002Q_{2t}})$

5.2.1 推計値を求める

書けば 1 行のスクリプトだが、計算過程は複雑である

```
exp_reg=: 3 : 0
NB. y. is beforehand make ln
(^{: "1 y),. */("1) 1x1 ^ (1,. }:"1 y) * "1 reg0 y
)
```

経過と説明 .

1. 片対数回帰の係数を求める。X は t と季節ダミー

```
reg0 DN410=. 1 2 3 mk_dummy4 ^ . DN41
4.25357 0.0143442 0.890812 0.489297 0.420002
```

2. x の前に定数項の 1 を付け、Y を落とす

```
1,. }:"1 DN410
1 0 0 0 0
1 1 0 0 1
1 2 0 1 0
1 3 1 0 0
```


3. 1, X に回帰係数を掛ける

```
reg0 DN410 NB. 回帰係数
4.25357 0.0143442 0.890812 0.489297 0.420002
```

```
(1, .}: "1 DN410) * "1 reg0 DN410
4.25357      0      0      0      0
4.25357 0.0143442      0      0 0.420002
4.25357 0.0286885      0 0.489297      0
4.25357 0.0430327 0.890812      0      0
```

4. $e^n e$ のべき乗化

```
1x1^ (1, .}: "1 DN410) * "1 reg0 DN410
70.3564      1      1      1      1
70.3564 1.01445      1      1 1.52196
70.3564 1.0291      1 1.63117      1
70.3564 1.04397 2.43711      1      1
```

5. t 各行毎 (横に) に足し合わせる

```
+/( "1) 1x1^ (1, .}: "1 DN410) * "1 reg0 DN410
74.3564 74.8928 75.0167 75.8375
```

6. exp_reg NB. exp_reg で計算する

exp_reg DN410	194 189.577
72 70.3564	78 78.9116
110 108.627	119 121.836
117 118.103	128 132.464
172 179.006	201 200.773
76 74.5113	81 83.5717
112 115.042	134 129.031
130 125.078	141 140.287
	216 212.629

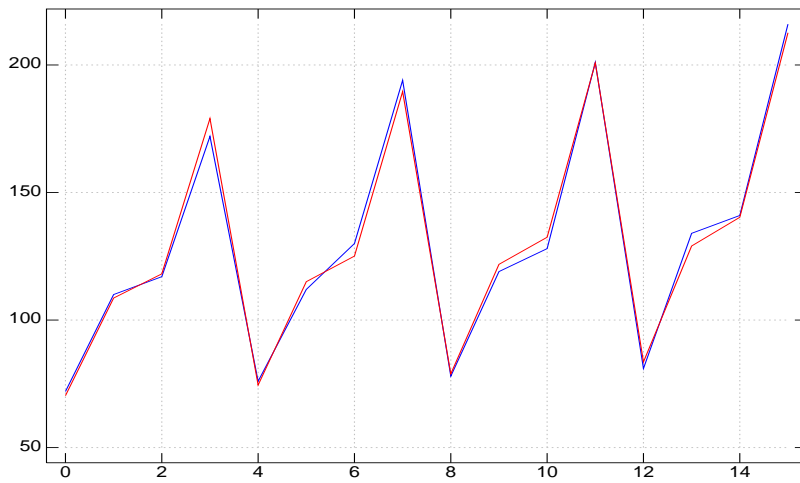


図 11 exp_reg

回帰係数の意味 需要の増加

$$\ln(D_t) = a' + b'_1 t + b'_2 Q_{2t} + b'_3 Q_{3t} + b'_4 Q_{4t} + e'_t$$

$$b'^1, b'^2, b'^3, b'^4$$

```
reg0 1 2 3 mk_dummy4 ^ . DN41
4.25357 0.0143442 0.890812 0.489297 0.420002
```

```
1-~1x1^ }. reg0 1 2 3 mk_dummy4 ^ . DN41
0.0144476 1.43711 0.631169 0.521965
```

$$e^{b'^1} - 1 = e^{0.1434} - 1 = 0.144$$

1.445 % increase in demand each quarter 各四半期単位で 1.445% づつ需要が伸びている。

e^{b^2, b^3, b^4} は第 1 シーズンに対する季節需要の倍率を示す。

(Q_2 52 % , Q_3 63 % 、 Q_4 143 %)

Reference

Stephen A.DeLurgio [Forecasting Principles and Applications] McGraw-Hill 1998

5.3 ロジスティック回帰

ロジスティック曲線は個体群生物学で考案された微分方程式のモデリングである。その後 1976 年にロバート・メイにより離散データの複雑な挙動が明らかにされ、リー・ヨークによりカオス理論の形成にも貢献した。逆 S 字の形が成長や製品の普及率を表し、医学、薬学以外にもマーケティングや金融工学での応用も盛んである。

ロジスティック式の解は

$$Y = \frac{L}{1 + e^{-b(x-a)}}$$

DN42

```

+-----+-----+
| 1 3 5 7 9 | 1 2 5 7 8 |
+-----+-----+
      X           Y
  
```

10 logis DN42

5.5744 0.470032

$$Y = \frac{10}{1 + e^{-5.5744(x-0.470032)}}$$

$L=10$ とした。

'line marker' plot DN42

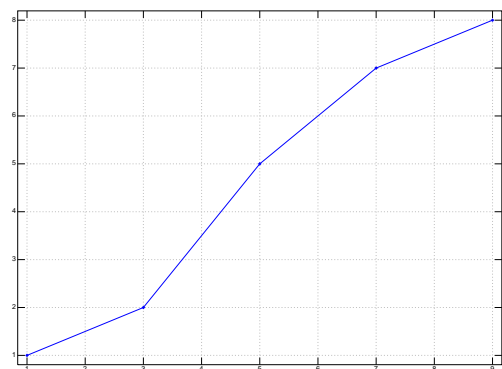


図 12 logistic

Script .

NB. Logistic regression

```
logis=:4 : 0
```

NB. usage L logis x;y (L = maximum upper value)

```
'X0 Y0'=: y
```

```
nr=: # Y0
```

```
log_y=: ^.(Y0 % x - Y0)
```

```
tmp=: log_y %. 1&,. X0
```

```
k2=:(( +/ log_y ) - ((1{tmp) * +/ X0)) % (1{tmp) * nr  
(- k2), 1{tmp  
)
```

Reference

石村 貞夫 「グラフ統計のはなし」東京図書 1995

5.4 動学モデル

従属変数 Y の時差変数を独立変数として含むモデル。

動学モデルでは、従属変数の時差変数 T_{t-k} が Y_t に影響を与えていると見ている。販売予測や需要予測の分野でも動学モデルは活躍している。

耐久財の消費関数に関するハウタッカーモデルも、複雑な動学モデルである。

3.1 はトレンド項 *index* とれんどこう@トレンド項を挿入するモデルでいわば竹の支柱を差すようなものであるが動学モデルは網のメッシュネットに蔓を絡ませる柔軟な構造を採用する。

時変 matrix の作成と時差変数を含むモデル . 本多の電力需要と人口、鉱工業生産高の例 (1955-1995) によって、動学モデルを作成してみよう。

このモデルは、従属変数の時差変数を確率変数として含み、他の独立変数も、差分変数である。

データの件数が多い場合は、このような動学モデルも実用になり、 Y の実績値を変数に含むので推計値の信頼度は高い。

動学モデルは回帰に用いる時系列を含むデータの作成が肝要であり、最終データを作れば後は単純な多変数回帰モデルと同一の計算である。

マトリクスの操作は多少複雑なので、簡単に時差変数を作成する関数 *mk_dougaku* を作成した。

作成するモデルは

$$\hat{Y}_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 \nabla x_{t1} + \alpha_3 \nabla x_{t2}$$

Y_{t-1} は Y (電力需要) の時差変数

$\alpha_2 \nabla x_{t1} + \alpha_3 \nabla x_{t2}$ は人口と鉱工業生産指数の差分である

```
DN45=. ".@> readcsv DIRCSV,'honda_311.csv'
```

最初の 10 個のサンプル (10{"1 DN45)

DN45				(2 3;1) mk_dougaku DN45			
年電力需要	人口	鉱工業生産指数		Y-1	d-P	d-IPX	Y
Y	P	IPX					
55	442.37	0.8928	6	442.37	0.0089	1.3	515.81
56	515.81	0.9017	7.3	515.81	0.0076	1.4	587.98
57	587.98	0.9093	8.7	587.98	0.0084	1.2	630.76
58	630.76	0.9177	8.5	630.76	0.0087	1.7	740.81
59	740.81	0.9264	10.2	740.81	0.0078	2.5	877.46
60	877.46	0.9342	12.7	877.46	0.0087	2.5	1015.54
61	1015.54	0.9429	15.2	1015.54	0.0089	1.3	1076.81
62	1076.81	0.9518	16.5	1076.81	0.0098	1.8	1229.44
63	1229.44	0.9616	18.3	1229.44	0.0102	2.9	1387.92
64	1387.92	0.9718	21.2				

(2 3;1) は、2,3 列の差分、第 1 列の Y-1 を作成し、最終行に Y を置く 4 列のマトリクスを作成する。
0 オリジン

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 x_{t2} + \alpha_3 x_{t3} + \epsilon_t$$

```
reg0 (2 3 ; 1) mk_dougaku DN45
13.571 1.02176 4088.13 24.3003
```

$$\hat{Y}_t = 13.571 + 1.02176Y_{t-1} + 4088.13\nabla x_{t1} + 24.3003\nabla x_{t2}$$

```
reg0 reg_exam_ad (2 3 ; 1) mk_dougaku DN45
+-----+-----+
|f=      |13.571 1.02176 4088.13 24.3003 |
+-----+-----+
|corr=:  |99.7973 |
+-----+-----+
|AIC:    |375.935 |
+-----+-----+
```

```
|DW= |2.16315 |
+-----+-----+
|t=: |0.191998 114.896 0.796813 4.44903|
+-----+-----+
```

人口 x_1 と鉱工業生産指数 X_2 は差分 $\nabla x_t = x_t - x_{t-1}$ に変換する。
 このモデルは、決定係数は十分であるが、人口 ∇t_1 の t 値が小さい。(5%でも有効でない)

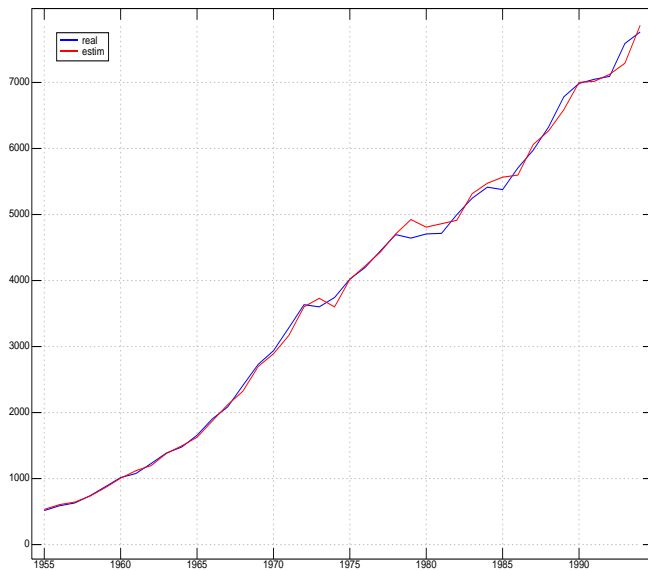


図 13 電力需要

Script .

```
sabun_sub=: 3 : ' ;-/& |.(L:0) 2<\ y'
sabun=: 4 : 0
if. 1=# x do. ANS=. sabun_sub x{"1 y
else.
ANS=. ,./>sabun_sub L:0 { |: x{"1 y
end.
ANS
)
mk_t_1=: 4 : '({: tmp),. }. tmp=: x {"1 y'
mk_dougaku=: 4 : 0
NB. x is e.g. 2 0 1 3;1 2;0
'X0 X1 X2'=: x
```

```
X0{"1 (X1 sabun y),. X2 mk_t_1 y
)
```

5.5 ハウ・タッカー モデル

いささかクラシックだが頑強な耐久消費財の消費の分析法であるハウタッカーモデルである。

$$q_t = A_0 + A_1 q_{t-1} + A_2 \Delta x_t + A_3 x_{t-1} + u_t$$

Δx_t は支出（消費支出や可処分所得）の差分であり、 t_1 は 1 期前のデータである。

この様にデータを並べたマトリクスを作り、単に *OLS* で回帰すればよい。

A_0, A_1, A_2, A_3 から、ハウタッカー関数のパラメータを $\alpha, \beta, \delta, \gamma$ を計算する。

$$\alpha = \frac{2A_0(A_2 - 0.5A_3)}{A_3(A_1 + 1)}$$

$$\beta = \frac{2(A_1 - 1)}{(A_1 + 1)} + \frac{A_3}{A_2 - 0.5A_3}$$

$$\gamma = \frac{2(A_2 - 0.5A_3)}{(A_1 + 1)}$$

$$\delta = \frac{A_3}{A_2 - 0.5A_3}$$

冷蔵庫の需要の例

```
DN46=. ".@> readcsv DIRCSV,'minotani_91.csv'
```

```
reg0 mk_ht 1 2 {"1 DN46
218.707 0.55144 7.45617 1.08877
```

$$q_t = 218.707 + 0.55144q_{t-1} + 7.45617\Delta x_t + 1.08877x_{t-1}$$

```
reg0 reg_exam_ad mk_ht 1 2 {"1 DN46
+-----+-----+
|f=      |218.707 0.55144 7.45617 1.08877 |
+-----+-----+
|corr=:  |55.6471          |
+-----+-----+
|AIC:    |153.876          |
+-----+-----+
|DW=     |1.69933         |
+-----+-----+
|t=:     |0.847684 2.36688 2.25759 2.07664|
```

DN46

冷蔵庫出荷 最終消費支出 変化率
台/1万世帯 万円/1世帯 消費支出

	q	x	dx
40	1093.3	166.084	0
41	1080.7	176.579	10.495
42	1201	187.534	10.955
43	1174	198.203	10.669
44	1103.5	210.204	12.001
45	1038.8	218.286	8.082
46	1015.2	224.765	6.479
47	1113.1	239.983	15.218
48	1293.5	254.226	14.243
49	1212.9	246.952	-7.274
50	1080.8	252.041	5.089
51	1127.3	256.084	4.043
52	1186.7	261.896	5.812
53	1293.5	270.653	8.757
54	1315.5	282.799	12.146
55	1226.4	282.258	-0.541
56	1202.8	279.892	-2.366
57	1230.9	288.201	8.309
58	1242.7	293.472	5.271

mk_ht 1 2 {"1 DN46

qt-1	dx	x	q
1093.3	10.495	176.579	1080.7
1080.7	10.955	187.534	1201
1201	10.669	198.203	1174
1174	12.001	210.204	1103.5
1103.5	8.082	218.286	1038.8
1038.8	6.479	224.765	1015.2
1015.2	15.218	239.983	1113.1
1113.1	14.243	254.226	1293.5
1293.5	-7.274	246.952	1212.9
1212.9	5.089	252.041	1080.8
1080.8	4.043	256.084	1127.3
1127.3	5.812	261.896	1186.7
1186.7	8.757	270.653	1293.5
1293.5	12.146	282.799	1315.5
1315.5	-0.541	282.258	1226.4
1226.4	-2.366	279.892	1202.8
1202.8	8.309	288.201	1230.9
1230.9	5.271	293.472	1242.7

+-----+-----+-----+-----+-----+

相関係数は高くない。 x_{t-1} の t 値はほんの少し足りないがこのモデルを採用する。

hau_takker mk_ht 1 2 {"1 DN46

ALPHA	BETA	GAMMA	DELTA	1/DELTA
1789.82	-0.420726	8.91015	0.157524	6.34822

α は定数項、 β はマイナスならばストック調整が働いている。

γ は需要に対する消費支出の短期的効果で l 単位あたりの需要の増加分を示し、 δ は減価率でそ

の逆数 ($1/\delta$) は平均耐用年数を表す。

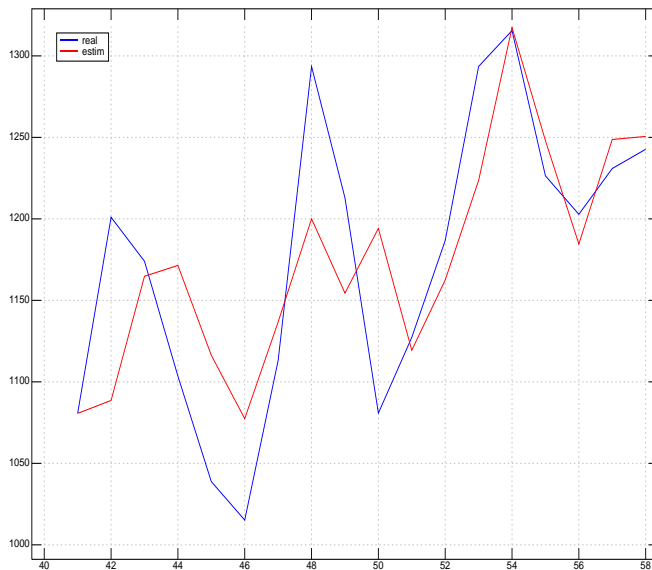


図 14 冷蔵庫の需要 HT-Model

Worked Example $\dot{y}(t) = \alpha\delta + (\beta - \delta)y(t) + \gamma\dot{x}(t) + \gamma\delta x(t)$

$$y(t) = \frac{1}{2}(y_t + y_{t-1})$$

$$\dot{y}(t) = \Delta y_t = (y_t - y_{t-1})$$

$$x(t) = \frac{1}{2}(x_t + x_{t-1})$$

$$\dot{x}(t) = \Delta x_t = (x_t - x_{t-1})$$

$y_t = A_0 + A_1 y_{t-1} + A_2 x_t + A_3 x_{t-1}$ 型モデル

(差分を取らない $\alpha, \beta, \gamma, \delta$ の計算式が異なる)

$$\alpha = \frac{A_0(A_2 - A_3)}{(A_1 + 1)(A_2 + A_3)}$$

$$\beta = \frac{2(A_1 - 1)}{A_1 + 1} + \frac{2(A_2 + A_3)}{A_2 - A_3}$$

$$\gamma = \frac{A_2 - A_3}{A_1 + 1}$$

$$\delta = \frac{2(A_2 + A_3)}{A_2 - A_3}$$

```
DM189=. ".@> readcsv DIRCSV,'minotani_189.csv'
```

```
reg0 0 3 1 2{"1 mk_t_1 1 2{"1 DM189
```

```
_0.206106 0.836571 0.242656 _0.218758
```

$$CDS D = -0.206106 + 0.836571 CDS D_{-1} + 0.242656 YDHPCDS D - 0.218758 YDHPCDS D_{-1}$$

CDS D: 耐久財 + 半耐久財

YDHPCDSD: $\frac{\text{名目家計可分所得}}{\text{CDS D のデフレーター}}$

```

    hau_takker2 0 3 1 2{"1 mk_t_1 1 2{"1 DM189
+-----+-----+-----+-----+-----+
|ALPHA  |BETA      |GAMMA   |DELTA   |1/DELTA|
+-----+-----+-----+-----+-----+
|_2.1667|_0.0743826|0.251237|0.103589|9.65351|
+-----+-----+-----+-----+-----+

```

```

    mk_ht=: 3 : 0
NB. usage: u m1 m2 {"1 y
NB. column of m1 m2 is q x
NB. q qt-1 delta-x xt-1
SABUN=: ; -/&|. (L:0) 2<\ {"1 y
    (}: {"1 y) ,.SABUN,. (}: {"1 y),. (}. {"1 y)
)

```

```

    hau_takker=: 3 : 0
NB. Usage: u mk_t m1 m2 {"1 y
'A0 A1 A2 A3'=: reg0 y
ALPHA=: ((+: A0) *A2- -: A3)% A3* >: A1
BETA=: +/((+: <: A1)% >: A1), A3 % A2- -: A3
GAMMA=: (+: A2- -: A3)% >: A1
DELTA=: A3%A2- -: A3
CAL=: 'ALPHA'; 'BETA'; 'GAMMA'; 'DELTA'; '1/DELTA'
CAL, : ANS=: ALPHA; BETA; GAMMA; DELTA; % DELTA
)

```

5.6 ブラウン型消費関数

T.M. Brown が 1952 年に習慣形成効果を入れたディューゼンペリとは異なるタイプの消費関数を定式化した。ブラウンの言葉で「過去に享有された実際の消費は、人間の生理学的、心理学的組織に、印象づけられて、習慣、風習、標準、および水準を形成し、これは消費者行動における慣性あるいは履歴的效果をもたらす。」と述べられている。

$C_t = \alpha(1 - \lambda) + \beta(1 - \lambda)Y_t + \lambda C_{t-1}$ ブラウン型消費関数は

$$C_t = \alpha + \beta YD_t + \gamma C_{t-1}$$

のように C_{t-1} のデータを作成して重相関のデータテーブルに入れて、OLS で計算すれば求め

られる。 $t-1$ は最新のデータを一つ落とし残りの項は古い先頭のデータを一つ落とす。

```
DN48=. ".@> readcsv DIRCSV,'minotani_20.csv'
```

```
reg0  (}. 2{"1 DN48),. mk_t_1 1{"1 DN48  
8.20431 _0.00712897 1.00022
```

```
reg0 reg_exam_ad  (}. 2{"1 DN48),. mk_t_1 1{"1 DN48
```

```
+-----+-----+  
|f=      |8.20431 _0.00712897 1.00022|  
+-----+-----+  
|corr=: |99.7752                      |  
+-----+-----+  
|AIC:   |62.5429                      |  
+-----+-----+  
|DW=    |1.21154                      |  
+-----+-----+  
|t=:    |2.8677 _0.0768726 9.87364  |  
+-----+-----+
```

$$C = 8.20701 - 0.007268YD + 1.00038C_{-1}$$

year	YD	Ct-1	Ct				
1966	89.2311	69.1303	76.4408	1981	212.753	171.145	174.605
1967	99.5354	76.4408	84.1501	1982	217.719	174.605	182.758
1968	109.297	84.1501	92.29	1983	223.195	182.758	187.981
1969	121.983	92.29	101.664	1984	228.76	187.981	192.628
1970	133.525	101.664	108.885	1985	237.088	192.628	199.461
1971	139.053	108.885	115.524	1986	244.9	199.461	207.29
1972	154.528	115.524	127.133	1987	250.547	207.29	215.918
1973	168.379	127.133	134.987	1988	261.553	215.918	228.024
1974	177.955	134.987	136.942	1989	270.513	228.024	237.825
1975	184.558	136.942	142.001	1990	280.107	237.825	247.726
1976	189.055	142.001	146.828	1991	292.364	247.726	254.435
1977	194.784	146.828	152.479	1992	295.818	254.435	257.237
1978	201.008	152.479	161.586	1993	300.989	257.237	261.618
1979	205.983	161.586	170.055	1994	305.79	261.618	265.364
1980	208.52	170.055	171.145				

issue 養谷 1 p20

C: 名目家計最終支出/PC

PC: Cのデフレーター (1990=1.0)

YD: 名目家計可処分所得/PC

単位:兆円

```
a =. (}. 2{"1 DN48),. mk_t_1 1{"1 DN48
'key native brown' plot |: ({"1 a),. estim_reg0 a
pd 'eps temp/brown0.eps'
```

Reference

本多正久「経営のための需要の予測と分析」産能大学出版部 2000

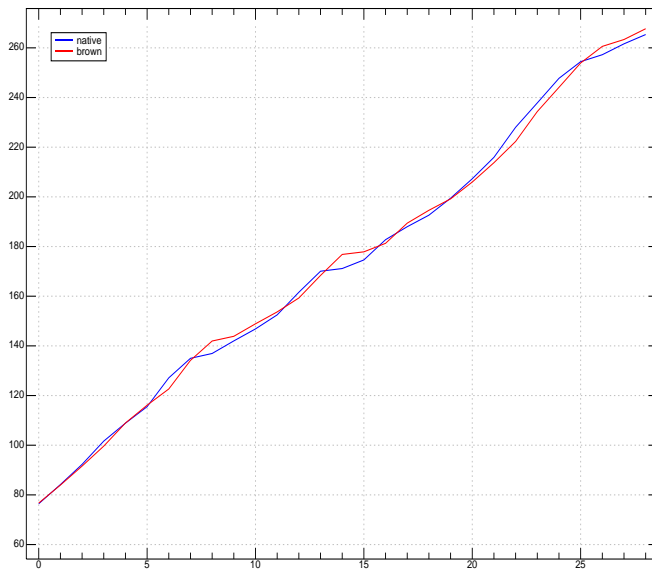


図 15 名目家計可分所得と耐久消費財

6 Newton 法による非線形回帰

6.1 Newton 法

J は微分のプリミティブ ($D.$) を備えている。N.Thomson(former IBM) が 1994 年にこの $D.$ を用いる簡潔なニュートン法を発表している。

ニュートン法は、

$$x - \frac{f(x)}{f'(x)}$$

の反復計算を行うものである。 J は微分演算子 $D.$ を備える。あくまで、数式処理でなく数値計算を行う。

ニュートン法を用いると微分演算子を用いて、数式をシンプルに定義できる。

```
new_1=: 1 : ' ] - x % x D.1' (^:_)("0)
```

J Grammar

1. 動詞を左パラメーターに、初期値を右パラメーターに取るので、`1 : 0` と副詞で定義する。
2. J 右パラメータから引く。
3. $D.1$ 一階微分関数
4. `"0` ランク指定。ベクトルに作用させる。
5. `(^:_)` は収束まで計算する。`(^:100)` 程度で打ち切っても良い。

Worked Example $x^2 - x - 2$ を定義する。多項式 p. を用いる方法による。多項式 p. は回帰係数をそのまま左引数に用いるので $a + bx$ のように表記され、高次の項ほど右に来る。ff2 は動詞 (関数) である。

```
ff2=. _2 _1 1&p.
```

```
ff2      0 1 2 3
_2 _2 0 4
```

ニュートン法の反復式	マトリクスの場合 <i>F is vector-valued</i>
$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$	$x^{(n+1)} = x^{(n)} - \frac{F(x^{(n)})}{F'(x^{(n)})}$

```
ff2 newton
```

```
2
```

```
(i:5),.(ff2 i:5),. ff2 newton i:5 NB.Newton 法の解
```

```
i:5/ff2/newton ff2
```

```
-----
```

```
_5 28 _1
_4 18 _1
_3 10 _1
_2 4 _1
_1 0 _1
0 _2 _1
1 _2 2
2 0 2
3 4 2
4 10 2
5 18 2
```

$y = x^3 + x - 1$ をニュートン法で解いてみよう。

1. fn で $x^3 + x - 1$ を定義する。

```
fn=. _1 1 0 1 &p.
```

2. 初期値 (複数=ベクトル) の定義

初期値を-5 から 5 までの 11 個とした。 i : は i と異なり、マイナス側にも鏡映の数列を取る。

```
i: 5
_5 _4 _3 _2 _1 0 1 2 3 4 5
```

3. `fn i:5` を計算

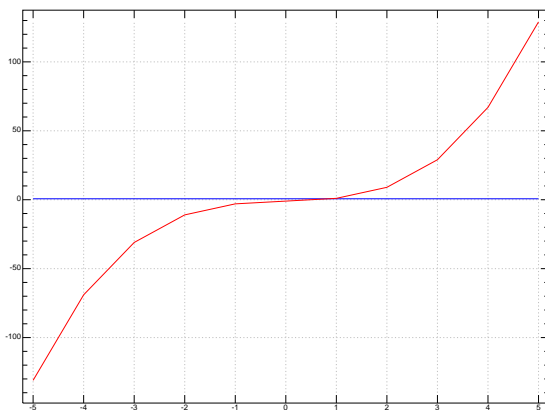
```
fn i:5
_131 _69 _31 _11 _3 _1 1 9 29 67 129
```

4. `plot`

X軸を正しく生成させるため、`Newton` に与えた引数と同じとした。

```
plot (i:5); fn i: 5
```

5. `graph`



6. `newton` 法

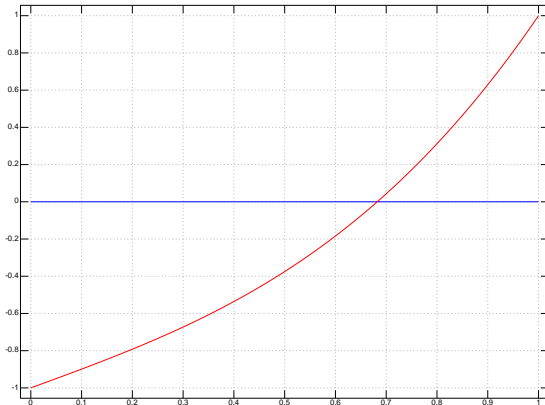
```
fn newton i: 5
```

```
0.682328 0.682328 0.682328 0.682328 0.682328 0.682328 0.682328 0.682328 0.682328 0.682328 0.682328
```

初期値を $i:5$ とした範囲では解は 0.68 (同一) である

7. 解の検証もかねて、詳細のグラフを書いてみよう

```
plot (steps 0 1 100);(100#0),:fn steps 0 1 100
```



8. J の多項式解法 ($p.$) で求めた解

`p. _1 1 0 1`

```

++++-----+
|1|_0.341164j1.16154 _0.341164j_1.16154 0.682328|
++++-----+

```

実数は 0.68

左の Box の 0.1 は反復階数で数字が多いときは苦戦している

6.1.1 Newton 法による非線形回帰

この Newton 法を用いた Thomson のアルゴリズムによる非線形回帰である。

Working Example *sample data is DN43 (in nonlinear_regression.ijs)*
from N.Thomson APL94 paper

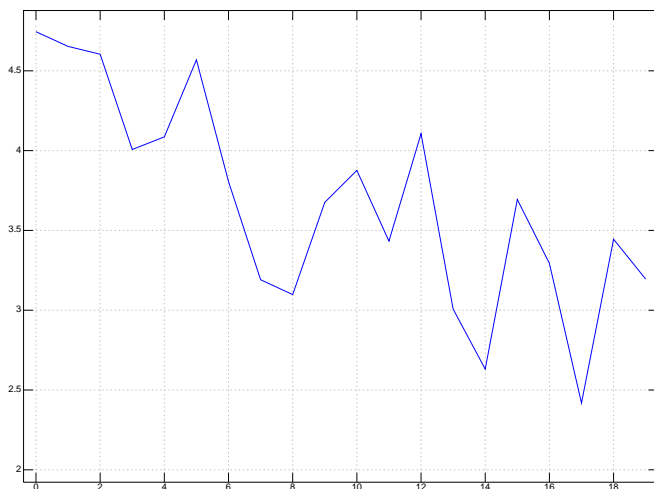


図 16 DN43

求める関数は動詞で定義する。非線形回帰は鑄型を前もって用意する。
 データは *plot* の X;Y の逆の Y;X の形のため、(| . X;Y) として用いる。

$$\begin{bmatrix} \Sigma \left(\frac{\partial f_i}{\partial a_0} \right)^2 & \Sigma \left(\frac{\partial f_i \partial f_i}{\partial a_0 \partial a_1} \right) & \cdots & \Sigma \left(\frac{\partial f_i \partial f_i}{\partial a_0 \partial a_m} \right) \\ \Sigma \frac{\partial f_i \partial f_i}{\partial a_1 \partial a_0} & \Sigma \left(\frac{\partial f_i}{\partial a_1} \right)^2 & \cdots & \Sigma \frac{\partial f_i \partial f_i}{\partial a_1 \partial a_m} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma \frac{\partial f_i \partial f_i}{\partial a_m \partial a_0} & \Sigma \frac{\partial f_i \partial f_i}{\partial a_m \partial a_1} & \cdots & \Sigma \left(\frac{\partial f_i}{\partial a_m} \right)^2 \end{bmatrix} \begin{bmatrix} \Delta a_0^{(v)} \\ \Delta a_1^{(v)} \\ \vdots \\ \Delta a_m^{(v)} \end{bmatrix} = \begin{bmatrix} - \Sigma \frac{\partial f_i}{\partial a_0} r_i^{(v)} \\ - \Sigma \frac{\partial f_i}{\partial a_1} r_i^{(v)} \\ \vdots \\ - \Sigma \frac{\partial f_i}{\partial a_m} r_i^{(v)} \end{bmatrix}$$

求める関数	$y = c_0 + c_1 \exp^{-c_2 x}$ 右肩下がりなのでマイナスとする
Newton 法での定義	<code>fn0=: 4 : ' (0{y)+(1{y}) * ^-x * 2{y}'</code>
初期値	<code>C0=:3 4 0.4</code>
	<pre> 5 4 \$ DN43 4.745 4.653 4.604 4.007 4.086 4.569 3.806 3.191 3.098 3.676 3.876 3.433 4.106 3.007 2.631 3.694 3.293 2.418 3.445 3.195 DT2=: (i.20);DN43 sample data DN43 is N.Thomson APL94 paper </pre>

```
(|. DT2)&(fn0 sub_a2) newton_fit C0=. 3 4 0.4
2.89046 1.93904 0.118559
```

$$f = 2.89046 + 1.93904e^{-0.118559x}$$

```
plot DN43 ,:2.89046+ 1.93904* 1x1 ^ -0.118559*i.20
```

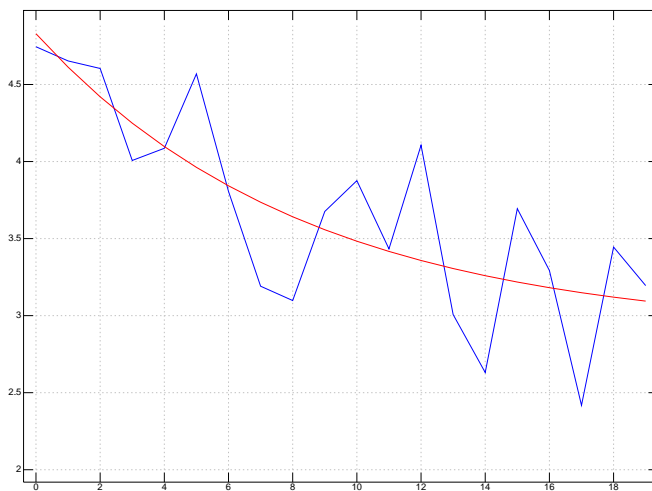


図 17 Newton_fit

6.1.2 指数回帰の場合

```
reg0 ^.DN43
1.53001 _0.0232633
```

$$f = 1.53001e^{-0.0232633x}$$

Working Example 小国「新数値計算法」サイエンス社 1997
p149 の例 data is DO_0

求める関数	$y = c_0 e^{c_1 x}$
-------	---------------------

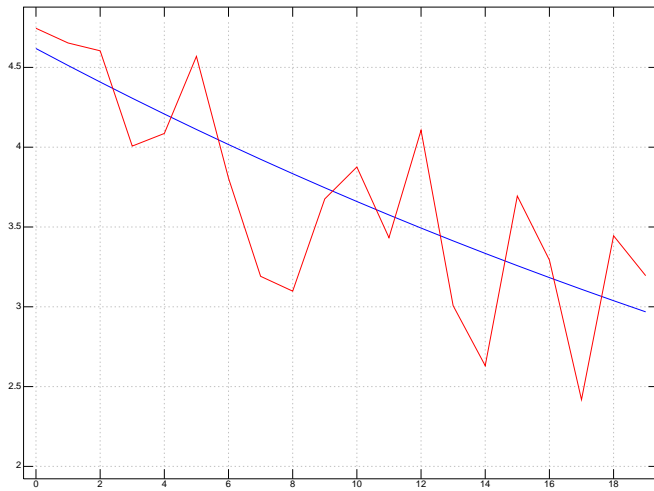


図 18 regression with exponential

Newton 法での定 義	fn2=: 4 : ' (0{y})* ^x * 1{y}'
初期値	C2=:1.3 -1
	<p>DN44</p> <pre> +-----+-----+ 0.1 0.2 0.3 0.4 0.5 1.1112 0.8232 0.6099 0.4518 0.3347 +-----+-----+ </pre>

```
(|. DN44)&(fn2 sub_a2) newton_fit C2=. 1.3 _1
1.49993 _2.9998
```

$1.49993e^{-2.9998x}$ NB. true value is $1.5e^{-3}$

6.1.3 Newton 法の Script

```
NB. -----newton_curve_fit-----
NB. Norman Thomson 1994
NB. slightly modified by M.Shimura
NB. x is |. plot x;y data
NB. y is syokiti
newton_fit=: 1 : ']' + (>@{. @x)(%.|:)(>@{:@x) D.1'(^:5)
```

```
sub_a0=: 1 : ' (>@{: @[) x ]'  
sub_a1=: 1 : ' ((>@{. @[)- x);x '  
sub_a2=: 1 : '(x sub_a0) sub_a1'
```

Reference

Norman Thomson [Applying Matrix Divide in APL and J] APL Quote Quad 9/1994

小国「新数値計算法」サイエンス社 1997

7 ロレンツ曲線とジニ係数

ロレンツ曲線は 1905 年にアメリカのマック・ロレンツにより、ジニ係数は 1936 年にイタリアのコッラド・ジニにより考案された。所得や資産の分布などの格差や不平等の度合いや集中度を明らかにするための手法で、ロレンツ曲線は、X 軸に累積世帯比率を、Y 軸に累積所得比率を取ってプロットする。ジニ係数はロレンツ曲線を計測可能な指数にしたもので、0 から 1 の間の数値を取り、数値が大きくなるほど所得分配の不平等の度合いが大きくなるとされる。

$$G = 1 - ((X_1 - X_0)(Y_1 - Y_0) + ((X_2 - X_1)(Y_2 - Y_1) + \dots + ((X_n - X_{n-1})(Y_n - Y_{n-1})))$$

$$= 1 - \sum((X_n - X_{n-1})(Y_n - Y_{n-1}))$$

X_i : 累積世帯比率

Y_i : 累積所得比率

Working Example アメリカの 5 分位階級別の所得

DN52					
分位累積	1966	1980	1990	2005	(%)
1:	0.2	5.6	5.3	4.6	4
2:	0.4	12.4	11.6	10.7	9.6
3:	0.6	17.7	17.6	16.6	15.3
4:	0.8	23.8	24.4	23.8	23
5:	1	40.4	41.1	44.3	48.1

Total	99.9	100	100	100	

出典) 白砂表 2 - 2

5 分位階級別 (20% ずつ) の所得シェア (%) の推移を示す。第一分位が低所得層

7.0.4 経過と解説

累積世帯・所得比率 累積世帯比率と累積所得比率を計算する。

```

jini_sub DN52
+-----+-----+
|0.2 0.4 0.6 0.8 1|0.056 0.053 0.046 0.04|
|                   | 0.18 0.169 0.153 0.136|
|                   |0.357 0.345 0.319 0.289|
|                   |0.595 0.589 0.557 0.519|
|                   |0.999      1      1      1|
  
```

```

calc_jini DN52
0.325 0.3376 0.37 0.4064
-----
1966 1980 1990 2005
0.325 0.3376 0.37 0.4064

```

アメリカではジニ係数が上がっており、所得格差は拡大している。

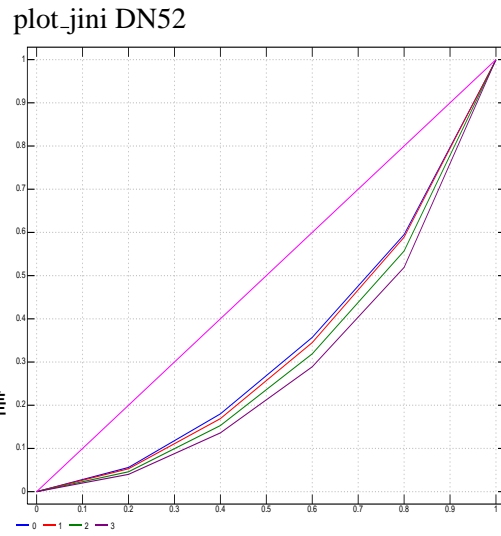


図 19 jini

+-----+-----+

一見 2005 年に向け第一分位が減っているが、第 3、第 4 分位も減っている。富の集中が進んでいる。

階層比較 累積所得比率を 2 個ずつ組み合わせ

```

,. |.(L:0) 2<\ 0,CUM0
+-----+
|0.056 0.053 0.046 0.04 |
| 0 0 0 0 |
+-----+
| 0.18 0.169 0.153 0.136|
|0.056 0.053 0.046 0.04|
+-----+
|0.357 0.345 0.319 0.289|
| 0.18 0.169 0.153 0.136|
+-----+
|0.595 0.589 0.557 0.519|
|0.357 0.345 0.319 0.289|
+-----+
|0.999 1 1 1|
|0.595 0.589 0.557 0.519|
+-----+

```

加えて ボックス内で縦に合計し

```

;("1), .+/(L:0) 2<\0,CUM0
0.056 0.053 0.046 0.04
0.236 0.222 0.199 0.176
0.537 0.514 0.472 0.425
0.952 0.934 0.876 0.808
1.594 1.589 1.557 1.519

```

$X_1 - X_{i-1}$ 累積世帯比率 $X_i - X_{i-1}$ をかける

```

;("1) ,. 0.2 */ L:0 +/(L:0) 2<\0,CUM0
0.0112 0.0106 0.0092 0.008
0.0472 0.0444 0.0398 0.0352
0.1074 0.1028 0.0944 0.085
0.1904 0.1868 0.1752 0.1616
0.3188 0.3178 0.3114 0.3038

```

Σ 横に Σ

```

+;("1) ,. 0.2 */ L:0 +/(L:0) 2<\0,CUM0
0.675 0.6624 0.63 0.5936

```

$1 - \Sigma$ 1 から引く

```

calc_jini DN52
0.325 0.3376 0.37 0.4064

```

Working Example アジア各国の 5 分位階級と所得分布

DN53

階級比率	BG	PK	IND	PH	CH	TH	ML	TW	SKR	JP
1: 0.2	9	9.3	8.1	5.4	4.7	6.3	4.4	6.7	7.9	8.6
2: 0.4	12.5	13	11.6	8.8	9	9.9	8.1	12.5	13.6	13
3: 0.6	15.9	16.3	15	13.1	14.2	14	12.9	17.4	18	17.4
4: 0.8	21.2	21.1	19.3	20.5	22.1	20.8	20.3	23.2	23.1	23.4
5: 1	41.4	40.3	46	52.2	50	49	54.3	40.2	37.4	37.6

出典) 白砂表 2 - 3

アジア各国のジニ係数。過剰労働力の国と高率の税金での所得再分配が進んだ国ではジニ係数は低くでる。

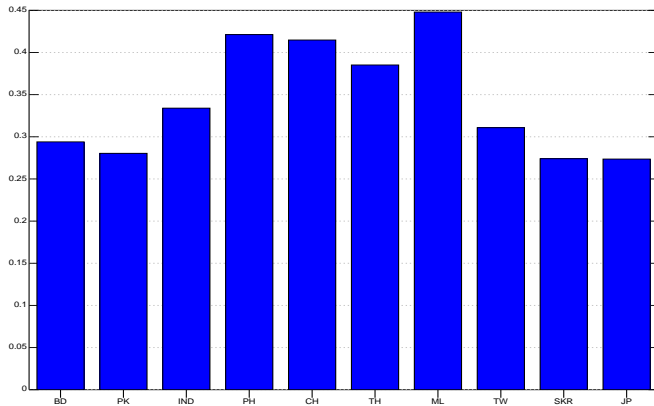


図 20 Logistic curve of \hat{P}

|: K, .{@>calc_jini DN53

```

+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|BangD|PaKis |IND  |Phili |China |Thai  |Malay|Tiwan |SKrea|JP   |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|0.294|0.2804|0.334|0.4212|0.4148|0.3852|0.448|0.3108|0.274|0.2736|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

7.1 関数一覧

多項式	<i>poly0</i> <i>poly_exam</i> <i>line_fit_poly0</i>	
乗法モデル	<i>mk_dummy4</i> <i>mk_dummy12</i>	
ロジスティック曲線	<i>logis</i>	
ニュートン法	<i>newton</i> <i>newton_fit</i>	
動学モデル	<i>mk_dougaku</i> <i>hau_takker</i> <i>hau_takker2</i>	

ジニ係数	<i>jini_sub</i> <i>calc_jini</i>	

8 Logit と Probit

可能な選択肢から答えを選ぶ場合、従属変数は連続とならない。YES - NO や購入する、しないなどの結果が2の選択肢から1つを選ぶ場合と、これを拡張して、数項の選択肢を用意する場合がある。最初は、2の選択肢の場合とする。

線形関数であらわす場合は、不均一分散が生じたり、確率の推定値が1を超えるといった問題が生じる。

このような場合の解法として *probit* と *Logit* がある。*probit* は正規分布、*Logit* はロジスティック分布を用いる。

*14

8.1 Group Logit

保有確率	$P_i = E(Y_i = 1 X_i) = \frac{1}{1 + \exp(-(\alpha + \beta X_i))} = \frac{1}{1 + \exp(-Z_i)}$ $Z_i = \alpha + \beta X_i$
非保有確率	$1 - P_i = \frac{\exp(-Z_i)}{1 + \exp(-Z_i)}$
保有と非保有の確率比・オッズ	$\frac{P_i}{1 - P_i} = \exp(Z_i)$
対数	$\ln \frac{P_i}{1 - P_i} = Z_i = \alpha + \beta X_i$
攪乱項を加える	$\ln \frac{P_i}{1 - P_i} = Z_i = \alpha + \beta X_i + \epsilon_i$

グループデータの標本数が大きければ、攪乱項は近似的に正規分布に従い、分散は、 $\frac{1}{N_i P_i (1 - P_i)}$ となる。攪乱項の分散は不均一となり、最少自乗法は使えない。

このような場合に分散の平方根の逆数で加重した加重最少自乗法をもちいる。

分散は、 \hat{P}_i から $\hat{\sigma}_i^2 = \frac{1}{N_i \hat{P}_i (1 - \hat{P}_i)}$ で推定値が求められる。

Working Example 所得と自動車の保有割合（グループデータ）

8.1.1 経過と説明

1. $\hat{P} = n/N$ NB. 各所得レベルに応じた保有割合

P=: (2{"1 DN50}) % 1{"1 DN50

0.4 0.4 0.35 0.5 0.7 0.833333 0.75 0.925

*14 Logit は少々複雑だが必ず計算できる。Probit は *erf* 関数の逆関数がない場合は数式を眺めるだけにした方がよい

DN50

所得	標本数	保有者
200	50	20
400	100	40
600	200	70
800	300	150
1000	200	140
1200	120	100
1400	80	60
1600	40	37

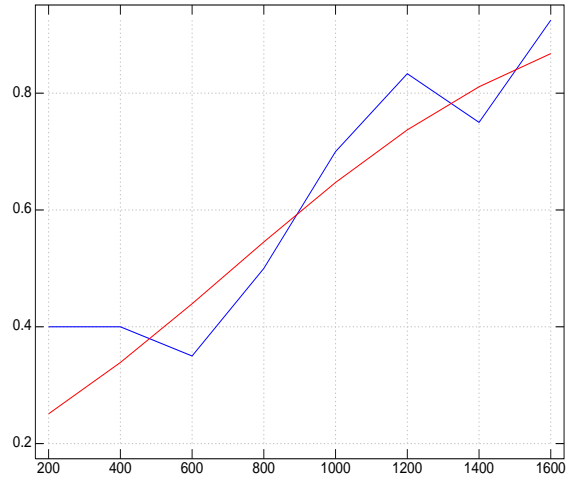


図 21 所得と自動車保有率の logit 回帰

2. $logit = \ln \hat{P} / (1 - \hat{P})$ NB. ロジットの計算

```
L=. ^ . P %-.P
_0.405465 _0.405465 _0.619039 0 0.847298 1.60944 1.09861 2.51231
```

3. $w = N\hat{P}(1 - \hat{P})$

```
W=. ((1{ "1 N0)* P)*-.P
12 24 45.5 75 42 16.6667 15 2.775
```

4. $w_i^{\frac{1}{2}}$ NB. 漸近的有効推定量 = 分散の平方根の逆数

5. $L^* = w_i^{\frac{1}{2}} * logit$

```
LH=. (%: W) * L
_1.40457 _1.98637 _4.17565 0 5.49112 6.5705 4.25491 4.18508
```

6. $X^* = w_i^{\frac{1}{2}} X_i$

logit_1 N0

X	Ni	ni	Pi=n/N	1-Pi	Pi/(1-Pi)	Li	wi= Ni(1-Pi)sqrd	wi	Li*	Xi*
200	50	20	0.4	0.6	0.66666	_0.405465	12	3.4641	_1.40457	692.82
400	100	40	0.4	0.6	0.66666	_0.405465	24	4.89898	_1.98637	1959.59
600	200	70	0.35	0.65	0.53846	_0.619039	45.5	6.74537	_4.17565	4047.22

```
| 800 300 150      0.5      0.5      1      0      75 8.66025      0 6928.2|
|1000 200 140      0.7      0.3  2.3333  0.847298      42 6.48074  5.49112 6480.74|
|1200 120 100 0.833333 0.166667      5  1.60944 16.6667 4.08248  6.5705 4898.98|
|1400  80  60      0.75      0.25      3  1.09861      15 3.87298  4.25491 5422.18|
|1600  40  37      0.925      0.075 12.333  2.51231  2.775 1.66583  4.18508 2665.33|
+-----+
```

8.1.2 reg_logit

```
LDAT=: 9 8 10 {"1 TMP
```

L^* , $W^{\frac{1}{2}}$, X^*

```
_1.40457 3.4641 692.82
_1.98637 4.89898 1959.59
_4.17565 6.74537 4047.22
      0 8.66025 6928.2
5.49112 6.48074 6480.74
6.5705 4.08248 4898.98
4.25491 3.87298 5422.18
4.18508 1.66583 2665.33
```

```
{"1 LDAT %. } {"1 LDAT
```

```
_1.51827 0.0021247
```

```
reg_logit DN50
```

```
_1.51827 0.0021247
```

所得一単位 (1 万円) の増加は、対数オッズを平均して、0.00212 増加させる。

8.1.3 非均一分散その2 Gjarati

3	\hat{P}_i	$\frac{n_i}{N_i}$	持ち家比率 (2)/(3)
4	$1 - \hat{P}_i$		1 - (3)

5	$\frac{\hat{P}_i}{1-\hat{P}_i}$		(3)/(4)
6	\hat{L}_i	$\ln \frac{\hat{P}_i}{1-\hat{P}_i}$	$\ln(5)$
7	w_i	$N_i \hat{P}_i (1 - \hat{P}_i)$	(1)*(3)*(4)
8	$\sqrt{w_i}$		$\sqrt{(7)}$
9	\hat{L}_i^*	$\hat{L}_i \sqrt{w_i}$	Lstar (6)*(8)
10	\hat{X}_i^*	$X_i \sqrt{w_i}$	Xstar (0)*(8)
11	EL^*		$ELstar = \sum \beta_1 \sqrt{w_i} + \beta_2 X_i^*$
12	logit		$ELstar / \sqrt{w_i}$
13	probability	\hat{P}	$\hat{P}_i = \frac{e^{logit}}{1 + e^{logit}}$
14	Change in probability		$\hat{\beta}(1 - \hat{P})\hat{P}$

この表の10列までは、先の中村・浅野と同一の方法によっている。

Working Example 所得と持ち家

所得 (1000US\$) N_i Sample n_i 持ち家の人の数

DN51		
In	N_i	n_i
6	40	8
8	50	12
10	60	18
13	80	28
15	100	45
20	70	36
25	65	39
30	50	33
35	40	30
40	25	20

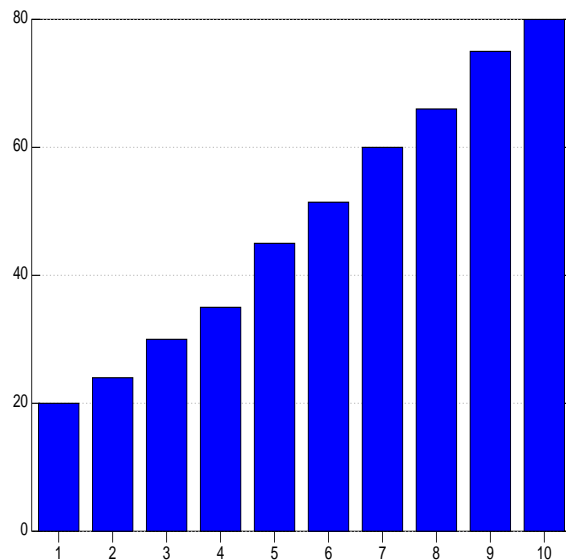


図 22 所得と持ち家

8.1.4 ロジットの基本式

経過と説明 .

$$1. \hat{L}_i = \ln \frac{\hat{P}_i}{1-\hat{P}_i} = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

不均一分散の補正のため、加重最少自乗法を用いる。

$$L_i^* = \hat{\beta}_1 \sqrt{w_i} + \hat{\beta}_2 X_i^* + v_i$$

L^* を $w^{\frac{1}{2}}, X^*$ で回帰する。常数項はつけない。

logit_1 DN51

0	1	2	3	4	5	6	7	8
X	N_i	n_i	\hat{P}_i	$1 - \hat{P}_i$	$\frac{\hat{P}_i}{1 - \hat{P}_i}$	\hat{L}_i	w_i	$\sqrt{w_i}$
			n_i/N_i			$\ln(\frac{\hat{P}_i}{1 - \hat{P}_i})$	$N_i \hat{P}_i (1 - \hat{P}_i)$	
6	40	8	0.2	0.8	0.25	-1.38629	6.4	2.52982
8	50	12	0.24	0.76	0.315789	-1.15268	9.12	3.01993
10	60	18	0.3	0.7	0.428571	-0.8473	12.6	3.54965
13	80	28	0.35	0.65	0.538462	-0.61904	18.2	4.26615
15	100	45	0.45	0.55	0.818182	-0.20067	24.75	4.97494
20	70	36	0.514286	0.485714	1.05882	0.057158	17.4857	4.18159
25	65	39	0.6	0.4	1.5	0.405465	15.6	3.94968
30	50	33	0.66	0.34	1.94118	0.663294	11.22	3.34963
35	40	30	0.75	0.25	3	1.09861	7.5	2.73861
40	25	20	0.8	0.2	4	1.38629	4	2
9	10	11	12	13	14			
\hat{L}_i^*	\hat{X}_i^*	EL^*	logit	probability	change in probability			
$\hat{L}_i \sqrt{w_i}$	$X_i \sqrt{w_i}$			\hat{P}				
-3.50708	15.1789	-2.8365	-1.12123	0.245784	0.014583			
-3.48102	24.1595	-2.91088	-0.96389	0.2761	0.015723			
-3.00761	35.4965	-2.86298	-0.80655	0.308626	0.016786			
-2.64091	55.4599	-2.43403	-0.57055	0.361111	0.01815			
-0.99832	74.6241	-2.05569	-0.41321	0.398143	0.018851			
0.239013	83.6318	-0.08307	-0.01987	0.495034	0.019665			
1.60146	98.7421	1.47511	0.373476	0.592299	0.018997			
2.22179	100.489	2.56856	0.766819	0.682832	0.017037			
3.00867	95.8514	3.17723	1.16016	0.761362	0.014293			
2.77259	80	3.10701	1.5535	0.825419	0.011336			

(第10項 \hat{X}_i^* まで)

L^* , $W^{\frac{1}{2}}$, X^*

LDAT=. 7 8 10 {"1">{: logit_1 DN51

6.4 2.52982 15.1789

9.12 3.01993 24.1595
 12.6 3.54965 35.4965
 18.2 4.26615 55.4599
 24.75 4.97494 74.6241
 17.4857 4.18159 83.6318
 15.6 3.94968 98.7421
 11.22 3.34963 100.489
 7.5 2.73861 95.8514
 4 2 80

2. $ELstar = \beta_1 \sqrt{w_i} + \beta_2 \hat{X}_i^* = \hat{L}_i^* = -1.59324 \sqrt{w_i} + 0.0786686 X_i^*$
 ユニット (所得 1000 ドル) の増加により、8.18% 持ち家が増える。

$1 \times 1 \hat{f}$
 0.203266 1.08185

単純な最少二乗法は次のようになる。図では顕著な差異はない。

$$\hat{L}_i = -1.65867 + 0.0791661 X_i$$

(6 {"1 TMP) %. 1, . (0 {"1 TMP)
 -1.65867 0.0791661

$ELstar$ は上の回帰係数に、各々 $\sqrt{w_i}$ と $\hat{X}_i^* = X_i \sqrt{w_i}$ をかけたものを足し合わせたものである。

$$\frac{\hat{P}_i}{1 - \hat{P}_i} = e^{-1.59324 \sqrt{w_i} + 0.0786686 X_i^*}$$

$$= e^{-1.59324 \sqrt{w_i}} \cdot e^{0.0786686 X_i^*}$$

ELSTAR=: +/ "1 f* ("1) 1 2 {"1 LDAT
 _2.8365 _2.91088 _2.86298 _2.43403 _2.05569 _0.0830732 1.47511 2.56856
 3.17723 3.10701

3. $logit$ は $ELstar$ を $\sqrt{w_i}$ で割ったものである。

(0 2 {"1 LDAT), . +/ "1 f* ("1) 1 2 {"1 LDAT

logit =. EL% 8 {"1 TMP

L^ X^* ELstar

```

_3.50708 15.1789 _2.8365
_3.48102 24.1595 _2.91088
_3.00761 35.4965 _2.86298
_2.64091 55.4599 _2.43403
_0.998324 74.6241 _2.05569
0.239013 83.6318 _0.0830732
1.60146 98.7421 1.47511
2.22179 100.489 2.56856
3.00867 95.8514 3.17723
2.77259 80 3.10701

```

```
LP=(tmp=(+/"1 f* ("1) 1 2{"1 LDAT ))% 1{"1 LDAT
```

```

tmp,.LP
_2.8365 _1.12123
_2.91088 _0.963889
_2.86298 _0.806552
_2.43403 _0.570546
_2.05569 _0.413209
_0.0830732 _0.0198664
1.47511 0.373476
2.56856 0.766819
3.17723 1.16016
3.10701 1.5535

```

4. logit の確率 (13) = $\frac{\hat{P}_i}{1 - \hat{P}_i} = e^{(logit)}$

$$\hat{P}_i = \frac{e^{(logit)}}{1 + e^{(logit)}}$$

これは

$(1x1^{\wedge}LP)\%1+1x1^{\wedge}LP$ で求められる。

```
tmp,.LP,.LP2=(1x1^LP)%1+1x1^LP
```

```

_2.8365 _1.12123 0.245784
_2.91088 _0.963889 0.2761

```


2.86298	0.806552	0.308626
2.43403	0.570546	0.361111
2.05569	0.413209	0.398143
0.0830732	0.0198664	0.495034
1.47511	0.373476	0.592299
2.56856	0.766819	0.682832
3.17723	1.16016	0.761362
3.10701	1.5535	0.825419

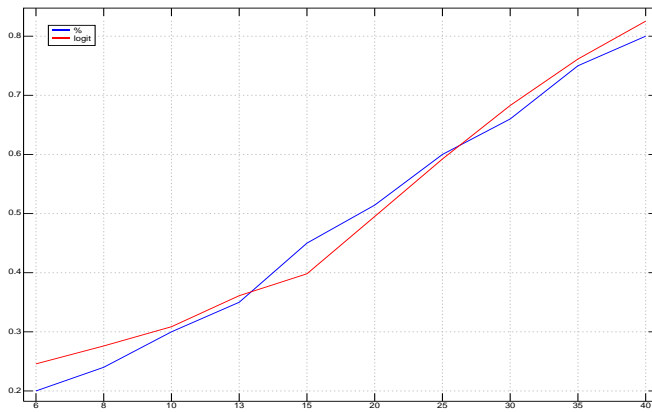


図 23 ni/Ni and Logit

所得階層（左欄 1 0 0 0ドル）とその階層の持ち家の実データ（中欄 比率 $\frac{n_i}{N_i}, \hat{P}_i$ ）logit により推計した確率（右欄 $\hat{P} LP2$ ）を対応させると次のようになる。

(0{"1 TMP) ,. (3{"1 TMP) ,.LP2

6	0.2	0.245784
8	0.24	0.2761
10	0.3	0.308626
13	0.35	0.361111
15	0.45	0.398143
20	0.514286	0.495034
25	0.6	0.592299
30	0.66	0.682832
35	0.75	0.761362
40	0.8	0.825419

```
plot (0{"1 TMP); a=. {"1 TMP04
```

Change in probability 所得階層による持ち家の変化率を求める。

$$\text{Change in probability} = \hat{\beta}(1 - \hat{P})\hat{P}$$

```
estim_logit DN51
X Ni ni n1/Ni(%)logit change in probability
6 40 8 0.2 0.245784 0.0145831
8 50 12 0.24 0.2761 0.0157234
10 60 18 0.3 0.308626 0.016786
13 80 28 0.35 0.361111 0.0181496
15 100 45 0.45 0.398143 0.018851
20 70 36 0.514286 0.495034 0.0196652
25 65 39 0.6 0.592299 0.018997
30 50 33 0.66 0.682832 0.0170374
35 40 30 0.75 0.761362 0.0142933
40 25 20 0.8 0.825419 0.0113363
```

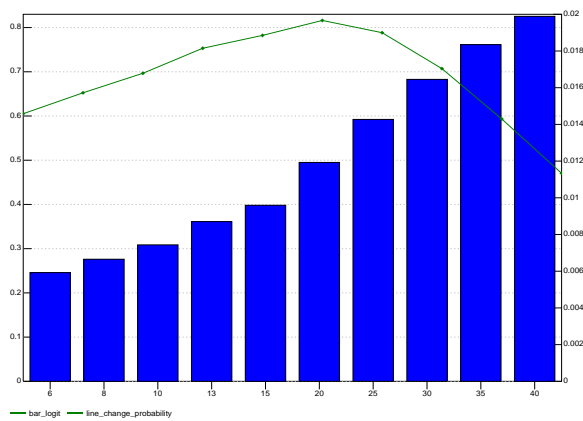


図 24 Logit and change in probability

```
reg_logit DN51
_1.59324 0.0786686
```

```
logit_2 DN51
+-----+
| L*      X*      EL *      Logit  Probability change in P^ |
```

```

+-----+
| _3.50708 15.1789  _2.8365  _1.12123  0.245784  0.0145831  |
| _3.48102 24.1595  _2.91088  _0.963889  0.2761  0.0157234  |
| _3.00761 35.4965  _2.86298  _0.806552  0.308626  0.016786  |
| _2.64091 55.4599  _2.43403  _0.570546  0.361111  0.0181496  |
| _0.998324 74.6241  _2.05569  _0.413209  0.398143  0.018851  |
| 0.239013 83.6318  _0.0830732  _0.0198664  0.495034  0.0196652  |
| 1.60146 98.7421  1.47511  0.373476  0.592299  0.018997  |
| 2.22179 100.489  2.56856  0.766819  0.682832  0.0170374  |
| 3.00867 95.8514  3.17723  1.16016  0.761362  0.0142933  |
| 2.77259 80  3.10701  1.5535  0.825419  0.0113363  |
+-----+

```

結果の集成 .

logit_1 DN51									
0	1	2	3	4	5	6	7	8	
X	N_i	n_i	\hat{P}_i	$1 - \hat{P}_i$	$\frac{\hat{P}_i}{1 - \hat{P}_i}$	$\hat{L}_i = \ln\left(\frac{\hat{P}_i}{1 - \hat{P}_i}\right)$	$w_i = N_i \hat{P}_i (1 - \hat{P}_i)$	$\sqrt{w_i} = \sqrt{N_i \hat{P}_i (1 - \hat{P}_i)}$	
6	40	8	0.2	0.8	0.25	-1.38629	6.4	2.52982	
8	50	12	0.24	0.76	0.315789	-1.15268	9.12	3.01993	
10	60	18	0.3	0.7	0.428571	-0.8473	12.6	3.54965	
13	80	28	0.35	0.65	0.538462	-0.61904	18.2	4.26615	
15	100	45	0.45	0.55	0.818182	-0.20067	24.75	4.97494	
20	70	36	0.514286	0.485714	1.05882	0.057158	17.4857	4.18159	
25	65	39	0.6	0.4	1.5	0.405465	15.6	3.94968	
30	50	33	0.66	0.34	1.94118	0.663294	11.22	3.34963	
35	40	30	0.75	0.25	3	1.09861	7.5	2.73861	
40	25	20	0.8	0.2	4	1.38629	4	2	

logit_2 DN51					
9	10	11	12	13	14
\hat{L}_i^*	\hat{X}_i^*	EL^*	logit	probability	change in probability
$\hat{L}_i \sqrt{w_i}$	$X_i \sqrt{w_i}$			\hat{P}	
-3.50708	15.1789	-2.8365	-1.12123	0.245784	0.014583
-3.48102	24.1595	-2.91088	-0.96389	0.2761	0.015723
-3.00761	35.4965	-2.86298	-0.80655	0.308626	0.016786

-2.64091	55.4599	-2.43403	-0.57055	0.361111	0.01815
-0.99832	74.6241	-2.05569	-0.41321	0.398143	0.018851
0.239013	83.6318	-0.08307	-0.01987	0.495034	0.019665
1.60146	98.7421	1.47511	0.373476	0.592299	0.018997
2.22179	100.489	2.56856	0.766819	0.682832	0.017037
3.00867	95.8514	3.17723	1.16016	0.761362	0.014293
2.77259	80	3.10701	1.5535	0.825419	0.011336

$$11 = a * \sqrt{w_i} + b * \hat{X}_i^*$$

$$12 = 11 / \sqrt{w_i}$$

$$13 = \frac{\hat{P}}{1 - \hat{P}_i} = e^{(12)} \dots \hat{P}_i = \frac{e^{(12)}}{1 + e^{(12)}}$$

$$14 = \hat{\beta}(1 - \hat{P})\hat{P}$$

8.2 Probit

K.E.Iverson が *D.Knuth* 達 (*G.P.K*) の *Concrete Mathematics Comparison* を著わしたときに超幾何級数 (*H.*) が *J* に組み込まれた。*E.Show* が *H.* を使った統計関数を提供したが、これが最近 *J* の *Mailing List* で更に深められ、*ADD-ON* で *erfinv* 関数も提供されている。この *erfinv* 関数を用いると *probit* が簡単に作れる

J での最初の作業 .

include 最初に require 'plot numeric trig'

ADD ON の読み込み require 'addons/stats/distribs/normal.ijs'

ロケール *normal.ijs* には *locale* が設定されているので *erf_pdistribs_* のように *_pdistribs_* を付けて用いる

erf 関数と *normal.ijs* での関数定義 .

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

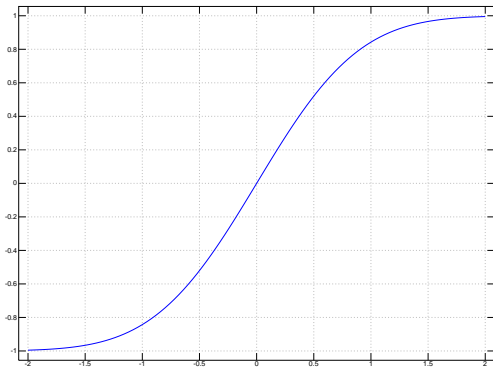
$$erfc(x) = 1 - erf(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt$$

erf の定義 .

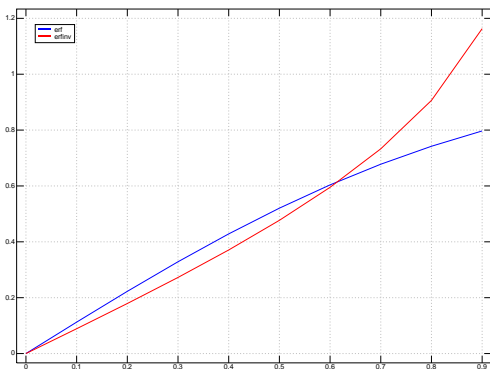
$$\operatorname{erfc}(z) = 1 - \operatorname{erf}(z) = \frac{2z}{\sqrt{\pi}} {}_1F_1\left(\frac{1}{2}; \frac{3}{2}; -z^2\right)$$

$$= \frac{2ze^{-z^2}}{\sqrt{\pi}} {}_1F_1\left(1; \frac{3}{2}; z^2\right)$$

H. Hypergeometrics 関数を用いて積分を計算している。



inverf 幸いに J でサポートされたが、単独での作成は相当難しい。



plot _2 2; 'erf_pdistribs_'
_2 から 2 までを plot

正規分布関数を定義する .

erf を用いた正規分布の定義 (pnormh)

NB. erf v error function

NB. ref Abramovitz and Stegun

7.1.21 (right)

```
erf=:(*&(%:4p_1)%^@:*)*[:1 H. 1.5*:
```

NB. erfc erf の補関数

```
erfc=: >:@-@erf
```

```
a=. steps 0 1 10
```

```
a,. (erf_pdistribs_ a)
```

```
,. erfinv_pdistribs_ a
```

a	erf	erfinv
0	0	0
0.1	0.112463	0.088856
0.2	0.222703	0.179143
0.3	0.328627	0.272463
0.4	0.428392	0.370807
0.5	0.5205	0.476936
0.6	0.603856	0.595116
0.7	0.677801	0.732869
0.8	0.742101	0.906194
0.9	0.796908	1.16309
1	0.842701	-

$$\Phi(x) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right) = \frac{1}{2} \operatorname{erfc} \left(-\frac{x}{\sqrt{2}} \right)$$

NB. pnormh v Standard Normal CDF

NB. ref Abramovitz and Stegun

26.2.29 (solved for P)

pnormh = (-: @:>:@erf@(%&(%:2))) f.

8.2.1 erfinv を用いた probit

$$\operatorname{probit}(p) = \Phi^{-1}(p) = \sqrt{2} \operatorname{erf}^{-1}(2p - 1)$$

erf^{-1} があれば $\operatorname{probit}(x)$ は容易に定義できる。

normal.ijs の *qnorm* を用いた erf^{-1} の定義は直接定義ファイルを見て欲しい。

NB. erfinv v inverse of error function

erfinv =: (0,%:2) qnorm 0.5 + -:

Worked Example (出典 *Gujarati table 15.4.*)

DN51

X_i 所得階層 (1000 ドル)

N_i 階層別構成所帯数

n_i その階層での住宅保有所帯

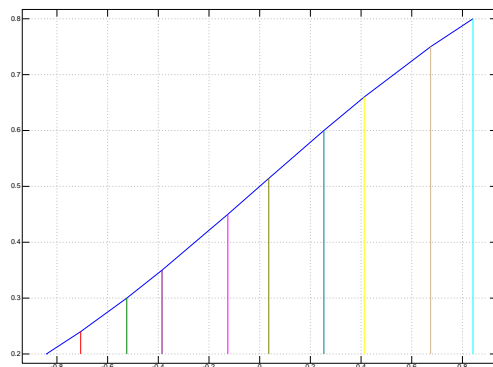
calc_probit_all DN51

Xi	Ni	ni	p	probit	Estim_nid	prob
6	40	8	0.2	_0.841621	_0.72478	0.234294
8	50	12	0.24	_0.706303	_0.627847	0.265052
10	60	18	0.3	_0.524401	_0.530914	0.297739
13	80	28	0.35	_0.38532	_0.385515	0.349928
15	100	45	0.45	_0.125661	_0.288582	0.386451
20	70	36	0.514286	0.0358166	_0.0462501	0.481555
25	65	39	0.6	0.253347	0.196082	0.577727
30	50	33	0.66	0.412463	0.438414	0.669457
35	40	30	0.75	0.67449	0.680746	0.751984
40	25	20	0.8	0.841621	0.923078	0.822017

+-----+-----+

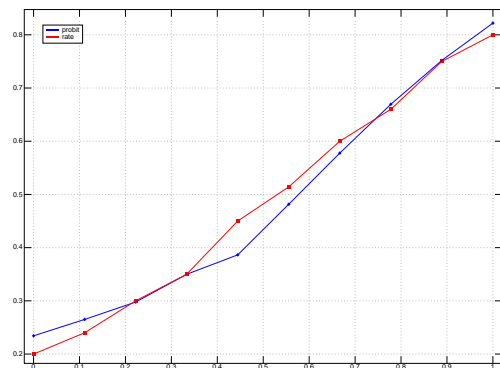
<p>p</p> <p><i>probit</i></p>	<p>$p = \frac{n_i}{N_i}$</p> <p>$probit(p) = \sqrt{2}erf^{-1}(2p - 1)$</p> <p>分布曲線上の p の x 軸上の値を 求める</p>	<pre>probit=(%:@2:) * erfinv_pdistribs_@<:@+:</pre>
<p><i>Estim- n.i.d</i></p>	<p><i>probit</i> の値を X_i で回帰して推 計値を求める。</p>	<pre>find_rate=: (%/)%@2 1 &{} "1 probit=(%:@2:) * erfinv_pdistribs_@<:@+ : NB. fork NB. probit@find_rate DN51 reg_probit=: probit@find_rate %. 1&,.@{"1 NB. OLS estim_probit=: 3 : '(reg_probit y)&p.{"1 y' NB. estimate calc_probit=: (pnormh_pdistribs_)@estim_probit calc_probit_all=: 3 : 0 tmp=.y,.(find_rate y),.(probit@find_rate y),.(estim_pro c=.<'Xi Ni ni p probit Estim_nid prob' c,:<tmp)</pre>
<p><i>proba- bility</i></p>	<p>回帰直線上の推計値を正規分布 の確率変数の値に戻す</p>	<pre>pnormh_pdistribs_@estim_probit DN51</pre> <p><i>normal.ijs</i> の関数</p> <p>p の値が確率 (<i>prob</i>) に変換されており、階層間で の変化も求められる。</p>

(probit@find_rate;find_rate)L:0 DN51



p の割合 (累積密度関数) を x 軸上の値 (逆関数) に変換する

```
'key probit rate;type marker,line'
plot (steps 0 1 9);
|:&(calc_probit,.find_rate) DN51
```



当初の確率（比率）と *probit* の確率

参考文献

浅野・中村「計量経済学」有斐閣 2000

D.N.Gujarati [Basic Economics 4th Edition] McGrawHill 2002

付録 A SNA のデータで

「*Numeric Recipes for Econometrics (0)* 入門」に便利なホームメイドの関数を紹介している

A.1 SNA データの取り出しとグラフ

ESRI の SNA の年次データから CSV データを作成する

- 2 . 経済活動別の国内総生産・要素所得 20s2n_jp.xls を用いた
- カンマをとる
- 空白セルは 0 をいれ、左詰め防止
- 空白行は *cut*
- 文字列は見出し行のみとする
- *csv* で出力 (*gdp_products.csv*)
- *Editor* でファイルを確認。最下行以降にカンマのみの行があれば *cut*

A.2 Jでの読み込み

準備 .

```
require 'files csv plot'
```

csv ファイルの読み込み `a=. readcsv '/data/sna/esri/gdp/2010/gdp_products.csv'`

```
] a1=. |: ;("1) ". L:0 }."1 a
```

- 見出しを切り離し数値のみにする }."1
- 数値化 ".@>ではばらけたので、;("1) ". L:0 を用いた
- *transpose verb/(—:)* 時間軸が上下になるようにデータを転置する
- データの取り出し 2{"1 a1
列の取り出しはランクを用いる
from({.}) は 0 オリジンである

xaxis グラフ用の X 軸の見出し作成

```
xaxis=. {"1 a1
```

A.3 グラフの作成

- J のグラフは次のフォーマット

(i.6);?. 3 6 \$ 18

```

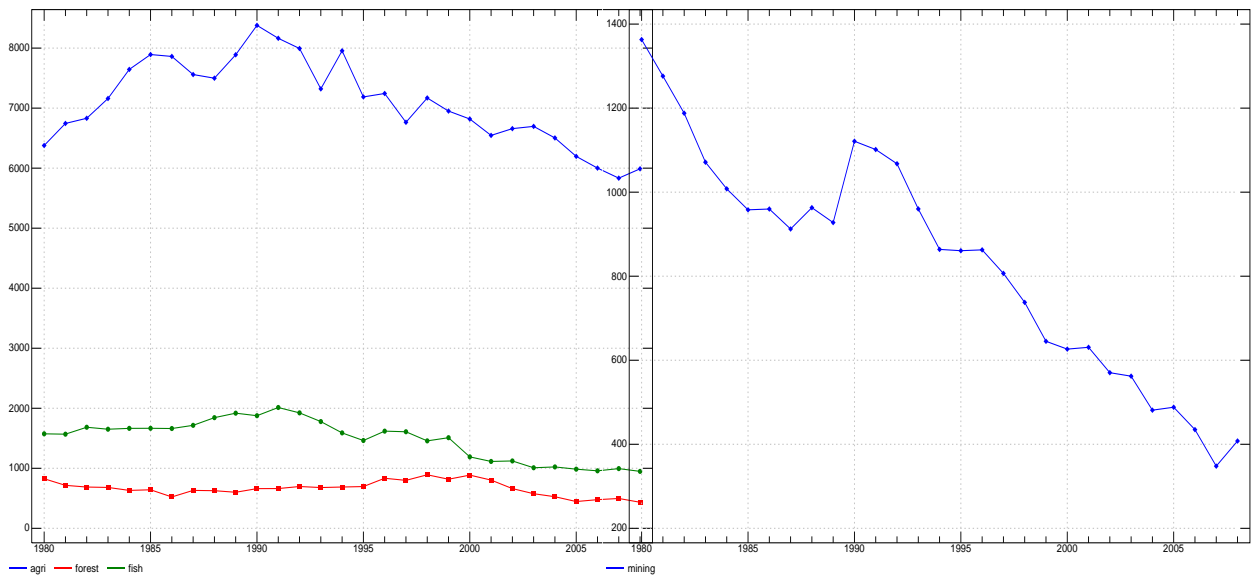
+-----+-----+
|0 1 2 3 4 5|12 17 17 4 2 9|
|           | 2 7 4 2 12 14|
|           | 1 10 7 2 16 6|
+-----+-----+
X 軸          Y の値

```

- *xaxis* は先に抜き出したのを用いる
- データは (再度) 転置 (*verb*(—:)) する

農林水産業

鉱業



Miscellane

J602 is download available (No charge)

<http://www.jsoftware.com>

Scripts are accessible

<http://japla.sakura.ne.jp>

索引

p value, 21

赤池弘次, 22

一般化最少自乗法, 36, 38

AIC, 41, 42, 44–46

erfinx, 84

加重最少自乗法, 39, 74

コクランオーカット法, 34

最少自乗法, 3, 4, 36

指数回帰, 66

指数乗法モデル, 48

ジニ, 69–71, 73

重回帰, 4, 5, 31

定数項, 5

ダービン・ワトソンテスト, 32

対数, 46, 48, 76

多項式, 41, 43, 45, 46, 62, 72

多重共線性, 23, 34

トレンド項, 46

Newton, 63–65

非線形回帰, 64, 65

不均一分散, 74, 77

probit, 84

Logit, 74

ロレンツ, 69