

Numerical recipes for Econometrics
Part 1(J6 版)

Masato Shimura
JCD02773@nifty.ne.jp

2008 年 8 月 28 日

目次

第 I 部	第 1 部 回帰分析編	7
第 1 章	3 本のスクリプト-線形回帰入門	9
1.1	データとグラフ	9
1.2	<i>Matrix divide</i>	13
1.3	単回帰モデル	13
1.4	多項式モデル	17
1.5	自己回帰モデル AR	19
1.6	関数一覧	22
第 2 章	データの拵え	25
2.1	平均と比較	25
2.2	ラスパイレス指数	30
2.3	データの標準化、対数化	32
2.4	共分散行列と相関行列	36
2.5	関数一覧	40
第 3 章	古典的回帰モデル	43
3.1	最少自乗法の正規方程式	43
3.2	最少自乗法の性質	44
3.3	重回帰モデル	44
3.4	t 検定	50
3.5	AIC によるモデルの選択	53
3.6	トレンド項とダミー変数	55
3.7	AIC による変数選択の方法	61
3.8	不均一分散・系列相関	63
3.9	関数一覧	69
第 4 章	種々の回帰分析	71
4.1	多項式モデル	71

4.2	乗法モデル	78
4.3	ロジスティック回帰	83
4.4	Newton 法による非線形回帰	85
4.5	動学モデル	92
4.6	Group Logit	101
4.7	ロレンツ曲線とジニ係数	112
4.8	関数一覧	115
第 5 章	スムージングと季節調整	117
5.1	移動平均・Moving Average	117
5.2	指数平滑化	124
5.3	季節調整・Seasonal Adjustment	128
5.4	関数一覧	135
索引		138

Numerical Recipes for Econometrics

2007

Masato Shimura

jcd02773@nifty.com

http://homepage3.nifty.com/asagaya_avenue

Copy right(C) 2007 All right reserved

GNU public Licence is applied to all scripts written by M.Shimura

夏休みになると少年少女の囲碁、将棋の全国タイトル戦が開催される。TV で見た少年少女の何人かは 10 代でプロとして活躍を始める。囲碁で対局者が目指すのは「神の一手」で、サイバースペースに長い修練と苦吟が重なって生まれる。多くのタイトル戦では創造から生まれた妙手の一手が勝負を隔てる。

自然科学の分野では、素晴らしい業績を「神のメモ帳を盗み見る」所業だと例える。計量経済学の教科書の索引欄に人名索引が付いていることがあり、400 名近い業績が一冊の教科書に凝縮されている。^{*1} 古今の研究者は「神の一手」を難解な数式で書き記しており、サイバースペースでない身にはその理解も至難の業である。

後にノーベル賞を受賞したレオンティエフの産業連関表やクラインの計量モデルは、初期の巨大な真空管コンピュータが唸りを上げて計算した。計量経済学の進歩はコンピュータの発達と時を同じくしており、初期の多くの計量経済学者はプログラムを自作した。PC の級数的発展はかつて数十億円もしたコンピュータの何万倍もの機能を研究室や書斎で使うことができ、アメリカを中心に開発された科学技術計算や数理統計、計量経済の定番ツールが活用されている。

最近の多くの教科書が数式を解説した後、パッケージで計算した結果のみを載せるようになってきて、アルゴリズムのブラックボックス化が進行している。ツールは既に研究された結果であり、いくら格闘しても理論の発見にはたどり着けない。一度は、数式と本格的に格闘してみることも良いのではないだろうか。

幸いにもレオンチエフの計算をハーバード大学でサポートし、真空管計算機に線形数学を教え込んだ K.E.Iverson が IBM に移って開発した線形数学に強い言語を高級関数電卓の感覚で使用して線形数学や微積分を計算できる環境が無償で提供されている。

本書は、この人はこんなことを考えていたのだ、これでノーベル記念賞に輝いたのだと数式の森をそぞろ歩きした過程を記してたものである。教科書の例題は、アルゴリズムを理解し、結果を確認できるように出典を記載し、成る可くそのまま収録した。また、公表されている熟練のスク립トは読者の学習に有益であると考え著者の名を付してそのまま紹介し、改悪は避けた。

^{*1} 最近は古今の学者の足跡を世界中の研究者が手分けして探すホームページも有るようだ。



図1 K.E.Iverson(1920-2004)

Keneth E. Iverson (1920-2004)

カナダのアルバータ州に生まれ、第2次大戦にフライトエンジニアとして従軍した後、クイーンズ大学からハーバード大学院へ進み、ハーバード大学で数学とコンピューターによる数学解析を教えた。1962年には *A Programming Language* というタイトルの本を著した。その一つがベクトルやマトリクスをスカラと同じように取り扱う一般多元配列環を装備し、線形数学などを自由に取り扱う言語の開発である。

このアイデアを大型機に搭載するため、IBMに移った。数学者である Iverson は数学をベースに言語のコアの部分の設計を一人で行ない、およそ10年の机上でのアイデアをアルゴリズムに練り上げる作業の後、パワーが追いついた大型機(名機 IBM 360)に1967年に APL が搭載された。大凡初期の LSI の時代である。

晩年はトロントに戻り、APL の高度な関数記述機能を更に洗練し、APL 記号をキーボード記号に移し替えた J 言語を *Roger Hui* の協力を得て創った。

計算機科学の父の一人と称されるアラン・チューリングを讃える ACM のチューリング賞を1979年に受賞。

第 I 部

第 1 部 回帰分析編

第 1 章

3 本のスクリプト-線形回帰入門

単回帰、多項式回帰、自己回帰 (AR) の 3 のアルゴリズムの紹介である。何れも正規方程式は複雑な数式から導出されるが、正規方程式を経ないで直接データから、 $\frac{Y}{X}$ となるような X を作成し、マトリクス X の逆行列を用いて行列除算を行うと、簡潔な一行のスクリプトで書き表すことができる。

1.1 データとグラフ

本書で扱うデータはマトリクスを次のように $X_1 X_2 X_3 \cdots X_n, .Y$ の様に縦に並べる。(y が右端の行になる)

データは EXCEL からのインポートや CSV ^{*1} ファイル、やデータベースから出入れする。EXCEL や CSV ファイルの取扱いは APPENDIX に記載した。

年度	G D P	民間消費	民間投資	政府支出
94	473	272	65	131
95	484	278	67	138
96	502	286	74	142
97	504	283	79	137
98	500	285	75	134
99	503	285	75	137
0	515	287	81	141
1	509	289	78	137
2	514	291	76	139
3	524	293	82	137
4	534	296	86	137

*2

(出典) 伴・中村・跡田「エコノメトリクス」第 2 版 有斐閣 2006

^{*1} Camma Separared Value カンマで一個ずつ区切ったファイル。中身は数値でなくテキストである。

^{*2} データのサンプルは numeric_recipe_data.ijs をロードしたのち手動で読み込むようににしてあるがフォルダの構成が異なるとエラーになるのでフォルダの位置を修正して書き直す。

次の例のように CSV ファイルを読み込む。読み込んだ状態では文字であるので数値化する。

*3

```
require 'files csv plot'
] DN10=. ".@> readcsv '/data/excel/stat_j/csv/ban_1.csv'
```

(フォルダ/data/excel/stat_j/csv に格納されている)

1.1.1 グラフ

散布図やグラフは簡単に描くことができ、EPS、EMF などの形式での出力も簡単に出来る。

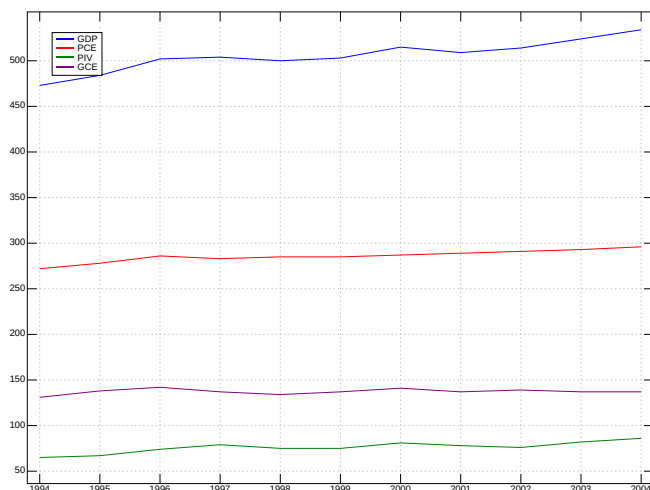


図 1.1 GDP CP IP

```
plot | : 1 2 3 puc DN10
pd 'eps /temp/ban_0.eps' NB. output plot files to /temp folder
```

(/temp に出力される)

J 言語ノート

配列の取出し 配列を簡単に取り出し並び替えるユーティリティを作成した。^{*4}

^{*3} CSV Comma Separated Value カンマで区切った (テキスト) データ。EXCEL などから CSV 形式で出力する。

^{*4} 何番目という順序数で、行や列を数えるとき、0, 1, 2, 3 と数える 0 オリジン派と 1, 2, 3.. と数える 1 オリジン派がいる。本稿は 0 オリジンで統一しているが、1 オリジン派の為の関数も幾らか用意した。

function	name	Usage:
<i>puc</i>	pickup column	0 1 puc DN10

<i>rec</i>	remove column	2 3 rec DN10
<i>puc1</i>	pickup column(1 origin)	1 2 puc1 DN10
<i>rec1</i>	remove column (1 origin)	3 4 rec1 DN10

数値化 `".format"`・数値化 文字化は"`"`: `".@>`はイディオム

ボックスを開く `>open Box` を開く `<`は Box 化

`require` 指定したシステムファイルをロードする

`readcsv` csv ファイルを読み込む 冒頭で `require 'files csv'` と指定する

`plot` グラフを描く

`pd` plot driver グラフの汎用コマンド

`eps` 画像形式の種類 ポストスクリプト形式

1.2 Matrix divide

重相関回帰、多項式回帰、自己相関 (AR) モデルを計算するスクリプトは次の 3 本の一行のスクリプトで表すことができる。

何れの式も中程に % という記号が入っている。matrix - Divide という K.E.Iverson が創った強力な配列除算機能である。

*5

単回帰 重回帰 Ordinaly Least Square OLS	<code>regx=: 3 : '({:"1 y) %. 1,.}:"1 y '</code>
多項式 Polynomial	<code>poly1=:4 : 'y %. (>:i. # y)^/i. >: x '</code>
自己回帰 Auto Regression AR	<code>ar0=:4 : '(x }.tmp) %. .("1)}: >x <\ tmp=:y -(+/%#)y '</code>

1.3 単回帰モデル

1.3.1 OLS 単回帰モデル

(単回帰)

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

(重回帰)

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

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る。

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

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

これらは、更に次のように簡約される。

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

*5 APL ではドミノ 母 であらわす。

線型数学で表現すれば単回帰も重回帰も同じである。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
 $\frac{y}{1, .X}$ と X の左に 1 を付加する。

■Working Example 果実の直径と水分含有率

DN11
X ある果実の直径 *cm*
Y 水分含有率 (%) のデータ。
(データの出典 金谷 「これならわかる応用数学教室」)

サンプルデータ KX の逆行列を手計算で行おうとすれば大変である。教科書には正方行列の数学ノートに記載した方法は紹介されているが、このような縦長の行列はあまり扱っていない。ここでは、素直に *K.E.Iverson* の *matrix – divide (%)* に従おう。

経過と説明

データは X,.Y の形に並べられている。	X に定数項の素の 1 を付加する	$\frac{y}{X} \rightarrow \frac{Y}{1, .X}$ 両項の % . はマトリクスの除算を行う
DN11 X Y ----- 5.6 30 5.8 26 6 33 6.2 31 6.4 33 6.4 35 6.4 37 6.6 36 6.8 33	1, .0 puc DN11 1 5.6 1 5.8 1 6 1 6.2 1 6.4 1 6.4 1 6.4 1 6.6 1 6.8	y % . 1, .X (1 puc DN11) % . 1, . 0 puc DN11 _5.01128 6.03383 reg0 DN11 _5.01128 6.03383 $y = -5.01128 + 6.03383x$

*6

■Script reg0

```

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

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

```

■定数項を用いない回帰 定数項を用いない場合は X の左に 1 のベクトルを付加しない。
 この場合は原点を通る。

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

```
({"1 DN11) %. {"1 DN11
```

5.23401

$$y = 5.23401x$$

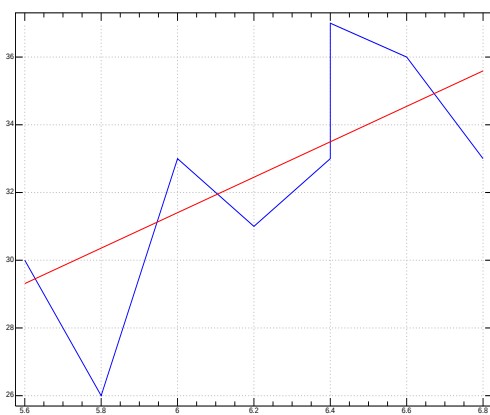


図 1.2 果実の直径と水分含有率

■テスト 相関係数とモデルのテストは *reg_exam_ad* を *reg0* の後ろに連結して行う。

*7

*6 データが X のみで Y は単に順序数でたりる場合には、*reg_t* を用いるが、*reg0* は DN11 のような揃ったデータに用いる *regx* と自動で振り分けを行う。

*7 この様に、動詞 (関数) を引数とする関数を J では副詞と呼ぶ

```

reg0 reg_exam_ad DN11
+-----+-----+
| f=      | _5.01128 6.03383 |
+-----+-----+
| corr=: | 47.8237          | 相関係数
+-----+-----+
| AIC:   | 18.8684          | AIC 情報量規準
+-----+-----+
| DW=    | 2.17534          | ダービン/ワトソンの値
+-----+-----+
| t=:     | _0.33633 2.53299 | t 値
+-----+-----+

```

■**ラインフィットと散布図** 散布図とラインフィットは *linefit_reg0* で作成できる。

回帰係数を求めたなら原データと推計値をグラフに書いてみよう。

```

linefit_reg0 DN11
pd 'eps /temp/reg_0.eps' NB. Save

```

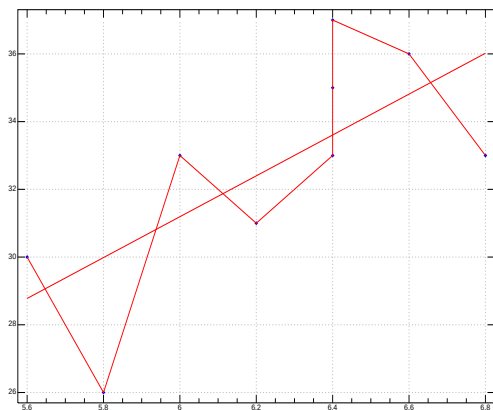


図 1.3 果実の直径と水分含有率

J 言語ノート

副詞 関数（数式を定義した動詞）を引数に取り専ら数式を計算する作用素 `1 : 0` で定義する

数列の生成 `i.` 数列を生成する `i.3 → 0 1 2`（3個） `i: 3 → _3 _2 _1 0 1 2 3`

1を足す `>: 1`を足す. `<: 1`を引く

行列除算 `%. matrix divide` 行列除算を行う 単項では逆行列を計算する

比較演算 $=$ は左右の引数が等しいか否かを判定する。代入ではない

Σ $+/$ は Σ $+/$ 1 2 3 $\rightarrow$ 6

1.4 多項式モデル

1.4.1 多項式モデル

変数 y が t に関する k 次の多項式によって

$$y = C_{00} + C_{01}t + \cdots + c_{0k}t^k + \epsilon$$

のように表されている場合に、パラメータの最尤推定は次の式を解くことによって得られる。

$$\begin{bmatrix} S_0 & S_1 & \cdots & S_k \\ S_1 & S_2 & \cdots & S_{k+1} \\ S_2 & S_3 & \cdots & S_{k+2} \\ \vdots & \vdots & \ddots & \vdots \\ S_K & S_{K+1} & \cdots & S_{2k} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \cdots \\ c_K \end{bmatrix} = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ \cdots \\ T_k \end{bmatrix}$$

多項式はフレキシブルで、次数を上手く選ぶとよくデータを追従する。逆に低次多項式はトレンドを概観でき、全体像を見るのに適している。^{*8}

■Example ある年の2月の東京の最低気温

$$y = 4.89554 + 1.13288x - 0.141456x^2 + 0.00438882x^3$$

3 poly0 DN12

4.89554 1.13288 -0.141456 0.00438882

4 linefit_poly0 DN12

■多項式の構成 データが長いと紙面を取るので、説明のため、10個の乱数データを用意した。

3 poly0 10? . 20

-11.2 17.6678 -3.75583 0.227855

$$f(x) = -11.2 + 17.6678x - 3.75583x^2 + 0.227855x^3$$

$X, a = 10 ? . 20$

X の 0 1 2 3 乗

^{*8} 次数により、両端が上下するという特徴があらわれる。

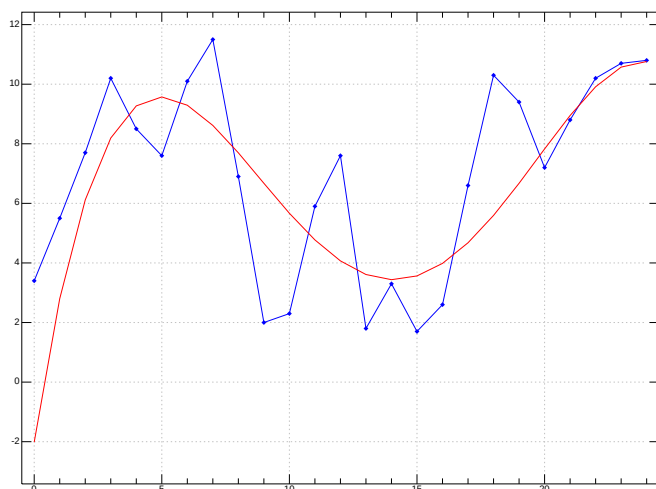


図 1.4 東京の冬の最低気温 (青) と推計値 (赤)4 次

20 までの 10 個の乱数
 X は単に順序数でオリジンを 1 とした。

```
(>:i. # a)
```

```
,.a
```

```
1 6
2 3
3 19
4 15
5 10
6 14
7 0
8 7
9 12
10 17
```

データの個数を求め、1 から順にナンバーを付け、これを 3 次の場合なら 0 1 2 3 乗する。0 乗が (1 の) 定数項になる。

```
(>:i. # a)^/ i. >: 3
```

```
1 1 1 1
1 2 4 8
1 3 9 27
1 4 16 64
1 5 25 125
1 6 36 216
1 7 49 343
1 8 64 512
1 9 81 729
1 10 100 1000
```

■**多項式の次数の選択** 多項式の次数の選択は AIC を用い、AIC の値を最少にする次数 k を選択する。

AIC については第 3 章を参照

1.4.2 J 言語ノート

べき乗 $\wedge$

乱数 $?, ?$ はシードを固定して同じ乱数を生成する

データの個数 $\#n$ となる。 $\$$ はマトリクスの形を表す

1.5 自己回帰モデル AR

1.5.1 AR モデル

自己回帰モデル (AR :Auto Regressive Model) は離散型の時系列解析の代表的な方法であって、時系列の現時点における値が、自己の過去の時点の線形結合で表される。

M 次までの AR モデルの最少二乗解をすべて求めるには、次のようにすればよい。この行列を逆行列を用いればダイレクトに最少二乗解が得られる。

$$\begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1k} \\ S_{12} & S_{22} & \cdots & S_{2k} \\ S_{13} & S_{23} & \cdots & S_{3k} \\ \vdots & \vdots & \ddots & \vdots \\ S_{1K} & S_{2K} & \cdots & S_{kK} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \cdots \\ b_K \end{bmatrix} = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ \cdots \\ T_k \end{bmatrix}$$

自己回帰のアルゴリズムには *Yull – Walker* 法、*Burk* 法、*Householder* 法があるが、ここでは、データの逆行列から一度に計算してしまうシンプルな方法を紹介する

*9

前出の東京のある年の最低気温のデータ (DN12) のデータで 3 次の自己回帰の仕組みを再現してみよう。

*9 3 方式のアルゴリズムによる回帰係数には若干差があり、*Burk* 法はピークに強く、*Yull – Walker* 法は谷に強い。*householder* 法は中立的であると言われている。

1. $x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_{t-n}$ という型のデータの組み合わせ (X) を作る。
2. 次数相当分のデータ (y) が組み合わせの結果減少するので、先頭から落として個数を合わせる。
3. $\frac{y}{X}$ を計算する。 X の逆行列が求まるので、簡単なマトリクスの除算で係数が求まる。

データ。 NR,y	3 次とする y, x _{t-0} , x _{t-1} , x _{t-2} , x _{t-3}																																																																							
1 3.4 2 5.5 3 7.7 4 10.2 5 8.5 6 7.6 7 10.1 8 11.5 9 6.9 10 2 25 10.7 26 10.8	(3}.DN12),.}: . "1 >3<\DN12 <table><tr><td>Y</td><td>x_{t-1}</td><td>x_{t-2}</td><td>x_{t-3}</td><td></td></tr><tr><td>3.4</td><td>0</td><td>0</td><td>0</td><td>cut</td></tr><tr><td>5.5</td><td>3.4</td><td>0</td><td>0</td><td>cut</td></tr><tr><td>7.7</td><td>5.5</td><td>3.4</td><td>0</td><td>cut</td></tr><tr><td>10.2</td><td>7.7</td><td>5.5</td><td>3.4</td><td></td></tr><tr><td>8.5</td><td>10.2</td><td>7.7</td><td>5.5</td><td></td></tr><tr><td>7.6</td><td>8.5</td><td>10.2</td><td>7.7</td><td></td></tr><tr><td>10.1</td><td>7.6</td><td>8.5</td><td>10.2</td><td></td></tr><tr><td>11.5</td><td>10.1</td><td>7.6</td><td>8.5</td><td></td></tr><tr><td>6.9</td><td>11.5</td><td>10.1</td><td>7.6</td><td></td></tr><tr><td>2</td><td>6.9</td><td>11.5</td><td>10.1</td><td></td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td></td></tr><tr><td>10.7</td><td>10.2</td><td>8.8</td><td>7.2</td><td></td></tr><tr><td>10.8</td><td>10.7</td><td>10.2</td><td>8.8</td><td>cut</td></tr></table>	Y	x _{t-1}	x _{t-2}	x _{t-3}		3.4	0	0	0	cut	5.5	3.4	0	0	cut	7.7	5.5	3.4	0	cut	10.2	7.7	5.5	3.4		8.5	10.2	7.7	5.5		7.6	8.5	10.2	7.7		10.1	7.6	8.5	10.2		11.5	10.1	7.6	8.5		6.9	11.5	10.1	7.6		2	6.9	11.5	10.1		...	...	...	...		10.7	10.2	8.8	7.2		10.8	10.7	10.2	8.8	cut	inifix(\) の強力な機能で X を作成する。一つずつづらした組み合わせを Box(<) と infix(\) で作り Open(>) で開く。 x _{t-1} , x _{t-2} , x _{t-3} の並びになるように縦に回転 Rotate(."1) させる。 Y の時点と個数を合わせるため時点の古い上の方から指定次数分 (3 次) を落とす。 最後の 1 行は回帰には用いないが、(虎の子として) 予測に用いる。 . "1>3<\ >:i.5 3 2 1 4 3 2 5 4 3 この形になれば重相関と同じようにマトリクスの除算 (%) で解が得られる。
Y	x _{t-1}	x _{t-2}	x _{t-3}																																																																					
3.4	0	0	0	cut																																																																				
5.5	3.4	0	0	cut																																																																				
7.7	5.5	3.4	0	cut																																																																				
10.2	7.7	5.5	3.4																																																																					
8.5	10.2	7.7	5.5																																																																					
7.6	8.5	10.2	7.7																																																																					
10.1	7.6	8.5	10.2																																																																					
11.5	10.1	7.6	8.5																																																																					
6.9	11.5	10.1	7.6																																																																					
2	6.9	11.5	10.1																																																																					
...	...	...	...																																																																					
10.7	10.2	8.8	7.2																																																																					
10.8	10.7	10.2	8.8	cut																																																																				

AIC による次数選択

$$Mll = -\frac{n-k}{n}(\log Q_k/(n-k))$$

$$AIC = -2 \times MLL + 2 \times k = (n-k) \times (\log Q_k/(n-k)) + 2 \times k$$

AR の次数選択は AIC の値を最少にする次数 k を見つければよい。

4 exam_ar0 DN12

```
+-----+
|mean=6.904|corr=0.546834|AIC=41.4362|
+-----+
```

3 exam_ar0 DN12

```
+-----+
|mean=6.904|corr=0.536786|AIC=41.5126|
+-----+
```

5 exam_ar0 DN12

```
+-----+
|mean=6.904|corr=0.56474|AIC=41.7886|
+-----+
```

4 ar0 DN12

0.973496 -0.548562 0.232687 0.0641936

4 次を採用する。

$$0.973496t_1 - 0.548562t_2 + 0.232687t_3 + 0.0641936t_4$$

4 linefit_ar DN12

pd 'eps \temp\ar_0.eps'

■Coffee Brake /除算の歴史 現在の人類がアフリカを出て大凡 4 万年を経て、突然エジプトの地にハイテク古代文明が出現した。エッフェル塔に抜かれるまで最高を誇った巨大ピラミッドや神殿、ナイル河の測量、メソポタミアの灌漑設備、フェニキア人の大型船、

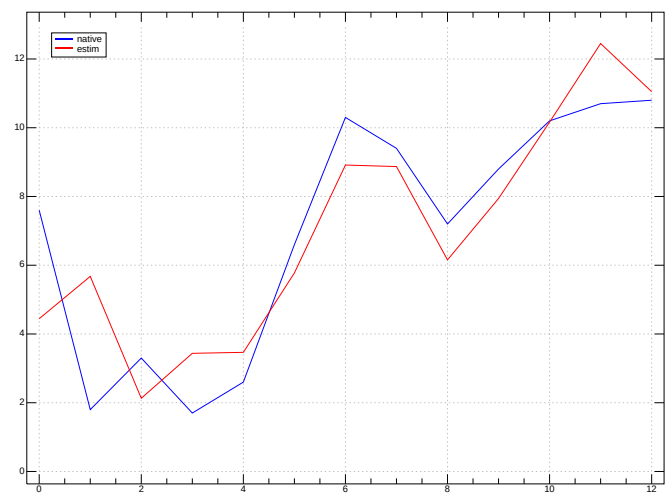


図 1.5 東京の冬の最低気温

ローマのアーチ橋.. 石や日干し煉瓦と人力で作り上げる完成された技術があった。

BC4000 年には 30 日 ×12 月 +5 祝日のエジプト暦が作成され、1582 年のグレゴリオ暦までの間精度で優位に立った。ナイル河の測量では水位 0 point は知られていた。

BC1600 年頃の「リンド・パピルス」と言われる数学書が発見されている。

BC2000 年頃のメソポタミアでは、60 進法で計算が行われ、棒で書かれ乾た赤褐色の粘土板には乗法と特殊な逆数を用いた乗法が書き残されている。

暦、交易、課税、戦争 … 古代から実用の要請があった。

1.6 関数一覧

単回帰・重回帰	<i>reg0</i> <i>reg_exam_ad</i> <i>estim_reg0</i> <i>linefit_reg0</i>	<i>reg0 reg_exam_ad n</i> <i>estim_reg0 n</i> <i>linefit_reg0 n</i>
多項式回帰	<i>poly0</i> <i>poly_exam</i> <i>linefit_poly0</i>	<i>4 poly_exam n</i> <i>4 linefit_poly0 n</i>
自己回帰	<i>ar0</i> <i>exam_ar0</i>	<i>m ar0 n</i> <i>m exam_ar0 n</i>

第2章

データの拵え

データの前処理や比較、加工などに用いる小物ツールを集めた。各種の平均や成長率の求め方、分散共分散行列、相関行列やラスパイレス指数、グラフ表示に必携の標準化や対数化等々である。

2.1

平均と比較

2.1.1 平均

平均には日常用いる算術平均の他にも様々な方法がある。

```
]a=: >: i.10
1 2 3 4 5 6 7 8 9 10
```

項目	数式	Script	Example
算術平均	$\frac{1}{n} \sum_{i=1}^n X_i$	 am=: +/ % # am2=: # %~ +/	 am a 5.5
幾何平均	$\sqrt[n]{\prod_{i=1}^n X_i}$	 gm=: # %: */	 gm a 4.52873 データに 0 があると */ が 0 になり、マイナスがあ ると答えが複素数になるの で使えない。

調和平均	$\frac{n}{\sum_{i=1}^n \frac{1}{X_i}}$	hm=: am &.(%"_)	逆数の平均値の逆数 比率 の平均に用いる hm a 3.41417
common mean 一般化平 均	$^m\sqrt{\frac{x_1^m + \cdots + x_n^m}{n}}$	cm=: [: {. (am,gm)^:_	算術平均、幾何平均の極限 cm a 5.00257 更に一般化できる $f^{-1}\left(\frac{1}{n} \sum_{i=1}^n f x_i\right)$

2.1.2 各種の平均の応用

四半期データの前期比を年率換算

4 半期データの年率換算	$\left(\left(\frac{y_t}{y_{t-1}}\right)^4 - 1\right) \times 100$
--------------	--

四半期伸び率の年率換算

qtr_grow 1 1.02 NB. 1 1.02
8.24322

4 半期で 1.0 から 1.02 への伸びは年率換算 8.2%

Script
qtr_grow=: 3 : '100 * <: ^&4 %/|. y'
NB. e.g. u 1 1.02
NB. exchange rate of growth from quarter to year

■経過と説明

新旧入れ替え 新旧を入れ替えて (rotate |.) から割る

%/|. 1 1.02
1.02

4 乗する 4

1 を引く (<: は 1 を引く)

```
<: ^4 %/|. 1 1.02
0.0824322
```

100 倍する (* times)

```
100 * <: ^4 %/|. 1 1.02
8.24322
```

年平均成長率を求める。

5 年、10 年などの単位で、平均成長率を求める。

単位は % になっている。

年平均成長率	$n \sqrt{\frac{X_{t+n}}{X_t}} - 1 \times 100$
--------	---

■Script

```
grow_ave=: 4 : '100 * <: (x %: %/|. y)'
```

■経過と説明

1. 年次を逆にして比を求める。

```
%/|. 504827 539160      NB. GDP 1995/2000
1.06801
```

2. 5 乗根を求め

```
5 %: %/|. 504827 539160      NB. GDP 1995/2000
1.01325
```

3. 100 倍する

```
100* 5 %: %/|. 504827 539160      NB. GDP 1995/2000
101.325
```

■Working Example

```
5 grow_ave 504827 539160      NB. GDP 1995/2000
1.32463      (1.32%)
```

10 grow_ave 469567 539160 NB. GDP 1990/2000
1.39161 (1.39%)

幾何平均-成長率の平均

成長率のデータしか手許にない場合に、成長率の平均が求められる。

DN20
1.25 1.4 1.07143 1.06667 1.125

gm DN20
1.17608

■Working Example トヨタ自動車の年平均成長率

year 01 02 03 04 05
% 4.2 12.5 6.3 7.3 13.4

DN21
1.042 1.125 1.063 1.073 1.134

gm DN21
1.08681 NB. 8.67% per year

調和平均-平均速度を求める

360 km の区間の最初の 120 km を時速30 km で、次の120 km を時速 40 km で、最後の120 km を60 km で走った場 合の平均速度を求める	hm 30 40 60 40 3 % +/- 1r30 1r40 1r60 40	$\frac{\frac{3}{\frac{1}{3+4+5}}}{120} = \frac{360}{9}$
--	---	---

■J 言語ノート

i.		0 からの順序数を指定個数打ち出す
>:		1 を加える
*	×	かけ算
%	÷	割り算
*:	<i>square</i>	2 乗
%:	<i>square – root</i>	平方根両項は n 乗根
{.	<i>From</i>	最初のスカラ 行 列を取り出す
–	<i>infinity</i>	無限大
~	<i>child</i>	計算順序を逆にする
"1	<i>Rank</i>	マトリクスの演算方向 ランク 1 は行 (ベクトル) 方向
^:	<i>power</i>	一種のループ
&.	<i>under</i>	接続詞
y		右引数
.	<i>Rotate</i>	ベクトルやマトリクスを指定方向に回転させる
/	<i>Insert</i>	計算記号 (演算子) を数列の間に挿入する
<:	<i>Decriment</i>	1 を引く
&	接続詞	数字と演算子を繋いで動詞 (関数) を作る
r	分数	$1r3$ は $\frac{1}{3}$

2.2 ラスパイレス指数

指数は次により求められる。

$$2000 \text{ 年を } 100 \text{ とした指数} = \frac{y_t}{y_{00}}$$

幾つかの品目データを一つの指数にする方法に、ラスパイレス、パーシェ、フィッシャーの3の方法がありウェイトの置き方で区別される。ラスパイレス指数は自己のデータのウェイトで計算できるので便利である。

<i>Laspiress</i>	$\frac{\sum p_t q_0}{\sum p_0 q_0}$
<i>Parshe</i>	$\frac{\sum p_t q_t}{\sum p_0 q_t}$
<i>Fischer</i>	$\sqrt{\frac{\sum p_t q_0}{\sum p_0 q_0} \times \frac{\sum p_t q_t}{\sum p_0 q_t}}$

DN22

A 財 B 財

価格数量 価格数量

```
+-----+-----+
| 75 50|60 70| NB. 基準時
| 85 48|55 72| NB. 第1期
| 95 46|52 75| NB. 第2期
|105 44|48 80| NB. 第3期
+-----+-----+
```

データの入力は次のように行う。

```
DN22=(75 85 95 105,. 50 48 46 44) ;60 55 52 48,. 70 72 75 80
```

(データの出典 日本銀行調査統計局編 計量経済分析の基礎と応用 p 21)

```
lsp_chain0 DN22
```

```
1 1.01887 1.05535 1.08302 NB. 基準 第1期 第2期 第3期
```

```
par_chain0 DN22
```

```
1 1.01515 1.04025 1.04444
```

```
    fis_chain0 DN22
```

```
1 1.01701 1.04777 1.06356
```

■Working Example 生鮮野菜と1所帯あたり購入額

```
DN23=:13 170 15 155,52 49 55 41,13 89 14 85,51 57 53 53,41 55 35 57,:45 33 48 31
```

```
DN23=: ('';1 0 1 0 )<.;1 DN23
```

```
DN23
```

```
2000      2005
```

価格数量価格数量

```
+-----+-----+
```

```
|13 170|15 155| NB. Cabbages
```

```
|52 49|55 41| NB. Spinachs
```

```
|13 89|14 85| NB. Napa
```

```
|51 57|53 53| NB. Leek
```

```
|41 55|35 57| NB. Lettuce
```

```
|45 33|48 31| NB. Broccoli
```

```
+-----+-----+
```

(出典) 白砂第2版表2－14

```
lsp DN23
```

```
Las:      Par:      Fis:
```

```
103.654 103.209 103.431
```

2005年の野菜のラスパイレス、パーシェ、フィッシャー指数はそれぞれ上のとおり。

2.2.1 Script

```
lsp=: 3 : 0
```

```
NB. Calc Laspi Parshe Fischer
```

```
'P0 Q0 P1 Q1'=:{;("2),. |: L:0 y
```

```
las=.(+/ P1 * Q0)% +/ P0 * Q0
```

```
par=.(+/P1 * Q1)% +/ P0 * Q1
```

```
fis=.%:(las * par)
```

```
' Las:      Par:      Fis:',: 9j3 ":100 *(las, par ,fis )
```

```
)
```

■Laspeyres Étinne Laspeyres(1834-1913)

Lass-pey-ress はオックス語でオックス語を使うベルリンに移住したユグノーの家系の生まれ

経済と統計の教授としてベルン、リガ、カールスルーエを経て最後の 26 年はギーセンに落ち着いた。有名な指数は 1871 年に発表した。(Wikipedia より)

n { take 指定したスカラや行、列の取り出し

■J 言語ノート 'Las' character 文字列を入力 (シングルクオート)

12j5 ": format 文字化して出力する (12 桁内少数 5 桁)

2.3 データの標準化、対数化

2.3.1 標準化

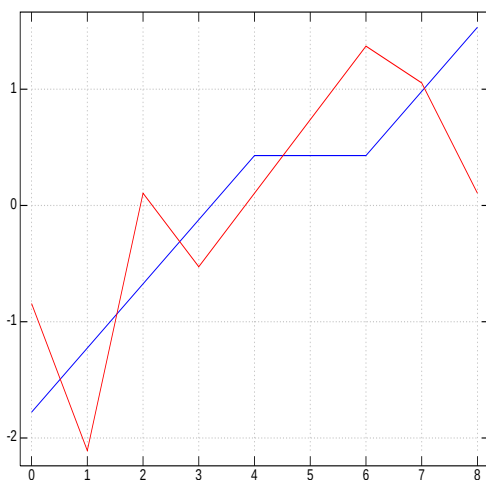
標準化とは、データを平均 0、分散 1 に変換することであり、単位が異なるデータの回帰や図示に用いる。

標準化	$\frac{x - \bar{x}}{\sqrt{\frac{1}{n} \sum (x - \bar{x})^2}}$	データを平均 0、分散 1 に変換する
-----	---	---------------------

x ある果実の直径 cm y 水分含有率 (%) のデータ。	DN24	stand DN24
	5.6 30	_1.7781 _0.843274
	5.8 26	_1.22628 _2.10819
	6 33	_0.674453 0.105409
	6.2 31	_0.122628 _0.527046
	6.4 33	0.429198 0.105409
	6.4 35	0.429198 0.737865
	6.4 37	0.429198 1.37032
	6.6 36	0.981023 1.05409
	6.8 33	1.53285 0.105409

単位が異なるものはうまく一枚のグラフには表せないことが多いが標準化すると単位が無くなり、同一のグラフに表すことができる。

(データの出典 金谷 「これならわかる応用数学教室」)



```
'key fruits water'plot |: stand DN24
pd 'eps \temp\kanaya_02.eps'
```

図 2.1 果実の直径 (直線) と水分含有率

Script

```
stand=: dev % "1 sd

dev=: -"1(+/% #)
sd=. %:@var
var=: # %~ ([:+/[[: *: dev)
```

■J 言語ノート

<code>*</code>	<code>square</code>	2 乗 (^2 と同じ)
<code>%</code>	<code>square_root</code>	2 乗根 (√)
<code>-"1(+/%#)</code>	$x - \bar{x}$	残差
<code>[:</code>	<code>Cap</code>	関数を右から記述順に実行するように指定

2.3.2 平均を引く

平均を引く $x - \bar{x}$	-(+/%#)	時系列解析では前もってデータの平均を計算し、データから平均を引いた数を用いることがよく行われる。(最後に戻しておく)
------------------------	---------	--

2.3.3 対数化

重回帰では対数線形を用いることにより、不均一分散が取り除かれる、係数が弾性値を表し解釈しやすい、単位が取捨されるなどの好ましい性質があるので、頻繁に用いられる。底に e を用いる自然対数を用いる。 $\log$ は $\ln$ と同記述する。J には単項の $\wedge.$ で自然対数が、両項で、任意の数を底にとる対数のプリミティブが用意されている。

$\wedge.100$
 $\wedge 4.60517$

 $\wedge 4.60517$
100

自然対数 $e^y = x$ $\log x = y$	1x1 2.71828 J の 1x1 に e の数値が組み込まれてある。	$\wedge. DN24$ 1.72277 3.4012 1.75786 3.2581 1.79176 3.49651 1.82455 3.43399 1.8563 3.49651 1.8563 3.55535 1.8563 3.61092 1.88707 3.58352 1.91692 3.49651
-----------------------------------	---	--

$\wedge$. *naturel log* 自然対数・常用対数は $10^{\wedge}$. 100 is 2
■J 言語ノート e^n $\wedge$ 4.60517 is 100
 1×1 e e の数値を組込んだ関数 2×1 1×2

2.3.4 ノルム化

$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$ はベクトルの大きさ (長さ) を表す。

長さの 2 乗の合計が 1 となるように基準化したユークリッド (Euclid) ノルムがよく用いられる。

ノルム化	norm=: [: %: [: +/ *:
------	-----------------------

解説

The length(or norm)of $\mathbf{v}$ is the nonnegative scalar $\|\mathbf{v}\|$

$$\|x\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$$

$$\|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v}$$

$\|x - y\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$ は線分 PQ の長さをあらわす。

Example

$$\mathbf{v} = 1 \ -2 \ 2 \ 0 \quad \|\mathbf{v}\| = \sqrt{(1^2 - 2^2 + 2^2 + 0)} = \sqrt{9} = 3 \quad \frac{1}{\|\mathbf{v}\|} \mathbf{v}$$

```
euc_norm 1 _2 2 0
```

```
0.333333 _0.666667 0.666667 0
```

Script

```
norm=: [: %: [: +/ *:
```

```
euc_norm=: 3 : ' y. % norm y.'
```

■Coffee Brake ネイピアと対数

対数とは底を等しくしたときの指数を取り出す関数である。スコットランド人でエディンバラの近くに生まれた *John Napier* (1550 – 1617) が「驚くべき対数法則の記述」を著したのは 1614 年であった。神学と星占術の研究者で数学者ではなかった。勃興してきた天文学の計算の工夫として、対数や e を考案し、天文学者を計算奴隷の仕事から解放した。またネイピアの計算棒といわれる計算器具も開発した。

2.4 共分散行列と相関行列

一度に分散共分散行列と相関行列を求めると、データ構造の見通しが良くなる。

2.4.1 分散と共分散

分散とはデータの平均値からの乖離を2乗したものの平均値である。
標準偏差とは、分散の正の平方根である。

残差平方 和（全変 動）	$\bar{x} = \frac{1}{n} \sum_{i=1}^n X_i$ $\sum (x - \bar{x})^2$	<pre>ss=: [: +/ (*:@dev) mean=: +/ % # NB. am dev=: - mean</pre>
分散 Variance	$\frac{\sum (x - \bar{x})^2}{n}$	<pre>var=: ss%#</pre>
標準偏差 Standard deviation	$\sqrt{\frac{\sum (x - \bar{x})^2}{n}}$	<pre>sd=: %:&(ss%#)</pre>
変動係数	$\frac{sd}{\bar{x}}$	<pre>vr=: sd % mean</pre>
共分散	変数の相関の度合 いを示す $\frac{1}{N} \sum (x - \bar{x})(y - \bar{y})$	<pre>cov=: # %~ ([: +/ [: */"1 dev) dev=: -"1(+ % #) cov=:# %~ ([: +/ [:*/ "1 (-"1(+ % #))) (一行で記述)</pre>

2.4.2 分散共分散行列と相関行列

多変数の場合、分散共分散行列や相関行列を一度に求めた方がデータの見通しがよい。

$$V(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^n V(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j)$$

$$V = \begin{pmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{pmatrix}$$

x	y
x1	y1
x2	y2
x3	y3
x4	y4

$$C = \frac{tXX}{N} = \frac{1}{N} \begin{pmatrix} x_1 - \bar{x} & x_2 - \bar{x} & \cdots & x_n - \bar{x} \\ y_1 - \bar{y} & y_2 - \bar{y} & \cdots & y_n - \bar{y} \end{pmatrix} \begin{pmatrix} x_1 - \bar{x} & y_1 - \bar{y} \\ x_2 - \bar{x} & y_2 - \bar{y} \\ x_3 - \bar{x} & y_3 - \bar{y} \\ x_n - \bar{x} & y_n - \bar{y} \end{pmatrix}$$

$$\frac{1}{N} \begin{pmatrix} (x_n - \bar{x})(x_n - \bar{x}) & (x_n - \bar{x})(y_n - \bar{y}) \\ (y_n - \bar{y})(x_n - \bar{x}) & (y_n - \bar{y})(y_n - \bar{y}) \end{pmatrix} = \begin{pmatrix} x \text{ の分散} & \text{共分散} \\ \text{共分散} & y \text{ の分散} \end{pmatrix}$$

Worked Example

データ あるクラスの生徒 身長 体重 胸囲 座高	列の平均を引く
DN25 NB. (1)	dev2 DN25 NB. (2)
145 30 60 70	_10 _10 _10 _10
145 35 70 75	_10 _5 0 _5
150 35 65 80	_5 _5 _5 0
150 40 70 70	_5 0 0 _10
155 40 75 75	0 0 5 _5
155 45 70 80	0 5 0 0
160 40 80 85	5 0 10 5
160 50 70 90	5 10 0 10
165 40 65 90	10 0 _5 10
165 45 75 85	10 5 5 5

内積演算 (: dev2 DN25) +/ . * (dev2 DN25) NB. (3) 500 275 175 425 275 300 150 250 175 150 300 100 425 250 100 500	n で割る。n = 10 ((: dev2 DN25) +/ . * (dev2 DN25)) % # DN25 NB. (4) 50 27.5 17.5 42.5 27.5 30 15 25 17.5 15 30 10 42.5 25 10 50
variable DN25 50 27.5 17.5 42.5 27.5 30 15 25 17.5 15 30 10 42.5 25 10 50	

Script

```
variable=:# %~ |:@dev2 +/ .* dev2
```

■J 言語ノート +/ . * 内積
|: Transpose 転置・行と列の交換

2.4.3 相関係数と相関行列

先にデータを標準化しておくと同じ手順で相関行列が求められる。

$$R = \begin{pmatrix} 1 & r_{xy} & r_{xz} \\ r_{yx} & 1 & r_{yz} \\ r_{zx} & r_{zy} & 1 \end{pmatrix}$$

$$\rho_{xy} = \frac{(\mathbf{x}, \mathbf{y})}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|}$$

$$X, Y \text{ の相関係数 } \rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)} \cdot \sqrt{V(Y)}}$$

Working Example

```
cortable DN25
```

1 0.710047 0.451848 0.85
0.710047 1 0.5 0.645497
0.451848 0.5 1 0.258199

```
0.85 0.645497 0.258199      1
```

Script

```
cortable=: 3 : 0
ss=. [: +/ [: *: dev2
sd=. %:&(ss%#)
stand=. dev2@[ %"1 sd@]
cortable=. #@] %~ ([:@stand@] +/ . * stand@])
cortable y
)
```

■J 言語ノート @ *Atop* 動詞と動詞を繋いで一つの動詞とする
 -"1(+/%#) -"1 is hook データから縦の平均を縦方向に引く

2.4.4 偏相関行列

相関行列は、各変数間の相関係数の行列表示であり、データの特徴が一目で見られる。

偏相関行列は、相関行列の逆行列とその対角行列から求められる。

対角行列の外積をとり、その平方根で相関行列の逆行列 $-r_{ij}$ を割ればよい。

$$\text{相関行列 } R = [r_{ij}] \text{ の逆行列 } R^{-1} \text{ その要素 } r^{ij} r_{ij.o} = \frac{-r^{ij}}{\sqrt{r^{ii} r^{jj}}}$$

例えば X_1 と X_2 間の関係を残りの変数の影響を除去して計算したい場合には、偏相関を求める。この偏相関を、相関行列から一度に求めてしまう。

相関行列 $R = [r_{ij}]$

R の逆行列 R^{-1} その要素 r^{ij}

$$r_{ij.o} = \frac{-r^{ij}}{\sqrt{r^{ii} r^{jj}}}$$

相関行列	偏相関行列
年間収入 5 分位の月当たりの所得と費用。 DN26 単位円 位 所得 教育 教養娯楽 1 198005 2854 17662 2 254020 5880 24208 3 290237 11995 30538 4 348839 19605 36568 5 452356 23490 48113	
cortable DN26 1 0.967571 0.997407 0.967571 1 0.976533 0.997407 0.976533 1	pcor_table cortable DN26 1 _0.414878 0.965817 _0.414878 1 0.631025 0.965817 0.631025 1 教育と教養娯楽の偏相関は 0.63

2.5 関数一覧

標準化	stand	stand n
統計関数	平均 <i>mean</i> 残差 <i>dev</i> 分散 <i>var</i> 標準偏差 <i>sd</i> 共分散 <i>cov</i>	

分散共分散行列 相関行列 偏相関行列	<i>vartable</i> <i>cortable</i> <i>pcor_table</i>	<i>vartable</i> n <i>cortable</i> n <i>pcor_table</i> <i>cortable</i> n
算術平均 幾何平均 調和平均	<i>am</i> <i>am2</i> <i>gm</i> <i>hm</i>	<i>am</i> i.10 <i>gm</i> i.10 <i>hm</i> i.10
四半期伸び率の年 率変換 年平均成長率	<i>qtr_grow</i> <i>grow_ave</i>	<i>qtr_grow</i> 1 1.02 5 <i>grow_ave</i> 504827 539160
ラスパイレス指数 ノルム	<i>lsp</i> <i>lsp_chain0</i> <i>par_chain0</i> <i>fis_chain0</i> <i>norm</i> <i>norm</i> n	<i>lsp</i> n

2.5.1 Reference

山澤成康 実践 計量経済学入門 日本評論社 2004

日本銀行調査統計局編 [計量経済分析の基礎と応用] 東洋経済新報社 1985

金谷健一 これなら分かる応用数学教室 共立出版 2000

第3章

古典的回帰モデル

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。

■**最少自乗法は誰が** ルジャンドル (Adrean Legendre 1752-1833) 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こったが、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。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。

3.1 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)

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

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

3.2 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。

3.3 重回帰モデル

線形数学では単相関と重相関は同じ方程式で表すことができ、配列計算言語では、単相関と同じプログラムで計算できる。

3.3.1 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。

Model	Form	$Slope = \frac{dY}{dX}$	$Elasticity = \frac{dY}{dX} \cdot \frac{X}{Y}$ 弾性
Linear	$Y = B_1 + B_2X$	B_2	$B_2 \left(\frac{X}{Y} \right)$
Log-linear 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
Log-linear 半対数線形	$\ln Y = B_1 + B_2 X$ ($Y = e^{B_1 + B_2 X}$)	$B_2(Y)$	$B_2(X)$
Linear-log 対数線形	$Y = B_1 + B_2 \ln X$	$B_2 \left(\frac{1}{X} \right)$	$B_2 \left(\frac{1}{Y} \right)$
Reciprocal 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 で回帰することとする。

```
regx=:3 : '{:"1 y } %. 1, .}:"1 y '
```

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 ,.`^{*1}

3.3.2 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'
```

DN0 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
X0, .X1, .Y NB. X, Y
```

^{*1} `stich` とは詩の行のことで K.E.Iverson の命名

(一行では)

```
DN27=: (*/"1) 2 3{"1 DM0),.(*/"1) 4 5{"1 DM0),.1{"1 DM0
```

X0	民間資本ストック×稼働率	LAB	対数変換 (^.)
X1	就業者×労働時間	CAP	
Y	製造業 GDP		
DN27			
X0	X1	Y	
5.49609	11.9897	11.1407	
5.50523	11.9977	11.1846	
5.49767	12.02	11.2269	
5.52197	12.0811	11.2594	
5.55843	12.1935	11.3303	
...			

```
reg0 ^. DN27 NB. ln
1.27786 0.7074 0.25785
```

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

```
reg0 reg_exam_ad DDN27
```

```
+-----+-----+
|f=      |1.27786 0.7074 0.25785 |
+-----+-----+
|corr=:|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 3.1 製造業の時系列データ

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。

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。

3.3.3 最小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)

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

正規方程式を作成して、連立方程式を解いて見よう。^{*2}正規方程式への当てはめ (3×3) は Script を作成した。)

連立方程式はクラメル法を用いて解いた。cr=:%.}:"1

右の列が連立方程式の解で、完全に解ければ対角行列は1になる。

```
test_neq ^. DN27 NB. 正規方程式
23 287.256 126.409 265.19
287.256 3589.44 1578.54 3313.27
```

^{*2} 配列計算言語で OLS に正規方程式を用いることは希である。

```
126.409 1578.54 694.887 1457.37
```

```
cr test_neq ^. DN27 NB. cramer method
```

```
0.999999 1.59432e_5 _1.36559e_5 1.27787
3.28967e_8 1 3.81732e_7 0.7074
1.38614e_7 _1.89109e_6 1 0.257849
```

最右の列が回帰係数 左は単位行列 (ゴミ残り)

*3

$$\ln y = 1.27787 + 0.7074 \ln CAP + 0.257849 \ln LAB$$

```
'key y y^' plot (1980+i.23); |: reg0 reg_estim_ad a
pd 'eps temp/mori_02.eps'
```

Script

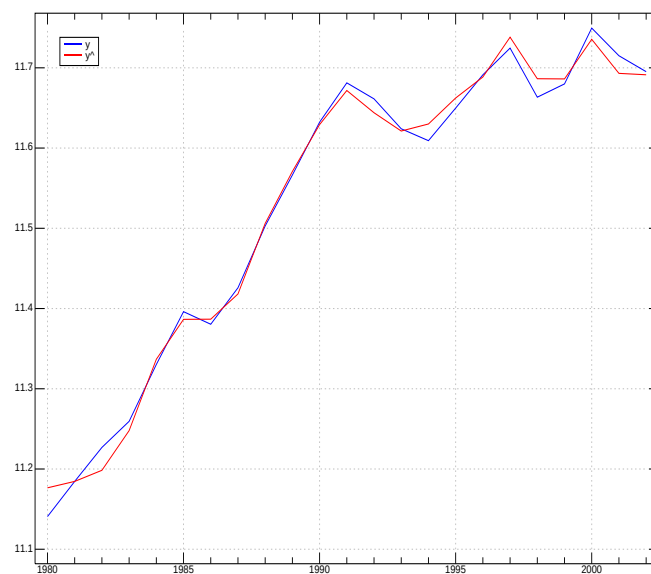


図 3.2 製造業の GDP の実測値と推計値

*3 本件の場合は 0 に近いゴミが入っており、対角行列も 1 以外の数になり、行列に多少歪みが見られるが、行列式の値は 0 ではないのでランクは確保されている。

行列式の値 (-/ . *) を求めて行列のランクを確認する

```
-/ . * }:"1 test_neq DN27
4.50787
```



```

test_neq=: 3 : 0
TMPX=: }:"1 y
TMPY=: {: "1 y
TMP0=:(# TMPY),+ / y NB. n sigba
TMP1=: + / ^&2 TMPX NB. x^2
TMP2=: + / * / "1 TMPX NB. x1x2
TMP3=: + / TMPX * TMPY NB. xnY
L1=. TMP0
L2=. (1{TMP0), ({. TMP1), TMP2, ({.TMP3)
L3=. (2{TMP0), TMP2, ({:TMP1), {:TMP3
L1,L2,:L3
)

```

3.3.4 Worked Example 指数を用いた重相関の例

Working Example 発電量と経済指数 (1969-1994)

重相関回帰は多入力 1 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.  define  3 4 7 2 puc DS0

DAT=: DN28 %("1)  {. DN28  NB. Example data

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|

```

y 発電量
 x_1 人口
 x_2 鉱工業生産指数
 x_3 原油輸入量

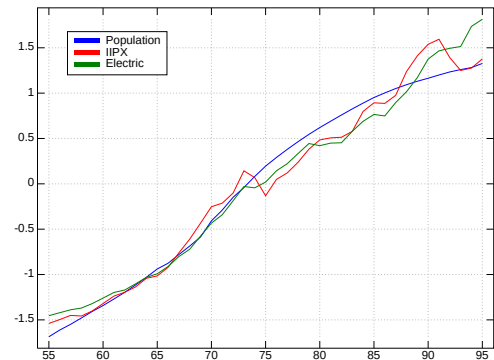


図 3.3 上から発電量 鉱工業生産指数 原油輸入量 人口

```

+-----+-----+
|corr=: |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)

3.4 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。



図 3.4 W.S. Gosset (1876-1937)

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。

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

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。

*4

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。

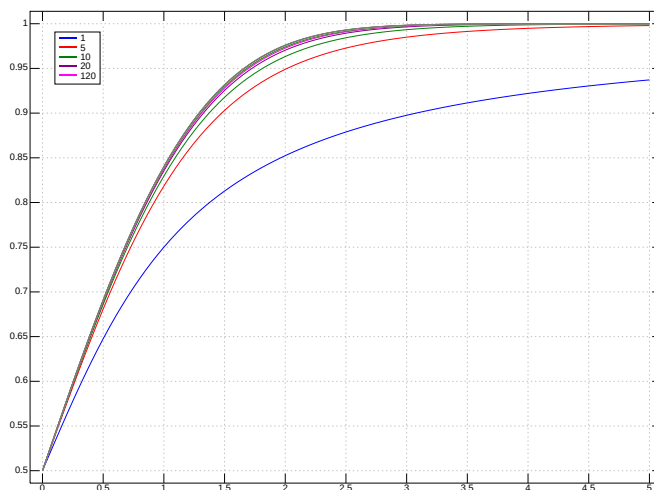


図 3.5 tcdf X-axis is degree of freedom Y-axis is %

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。

```
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
```

*4 サンプル数から説明変数+1 を引いた数

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。

3.4.1 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
```

3.5 AIC によるモデルの選択



図 3.6 H.AKAIKE

赤池弘次 (1927-) 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。

3.5.1 AIC

AIC(Akaike Information Criteria)(赤池) 情報量基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 を選べばよい。

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

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

3.5.2 情報量規準 AIC のアルゴリズム

$$\text{残差平方和 } Q = (y - Xb)'(y - Xb) = y'y - (y'X)b$$

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。

3.6 トренд項とダミー変数

3.6.1 トренд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rendの除去 (1)

GDP と M2CD

```
DMN0=. ".@> readcsv DIRCSV,'minotani_1.csv'
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
```

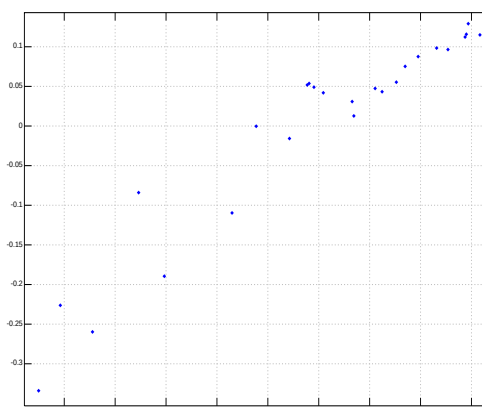


図 3.7 M, 横軸は V

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 t 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
```

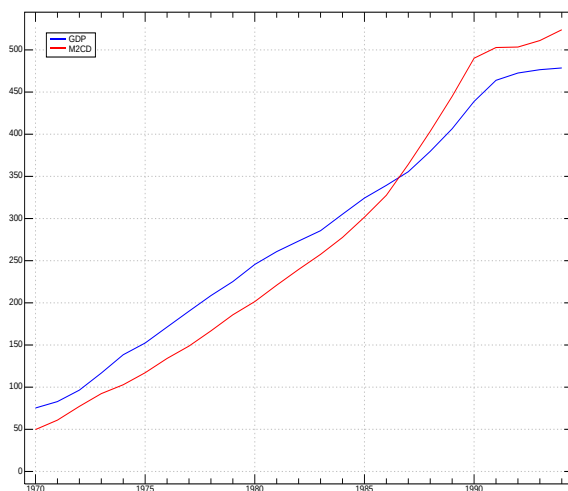


図 3.8 GDP,M2CD

*5

G

■WorkingExample 為替レート、CPI、賃金率

```
DN30=:DA0=. ".@> readcsv DIRCSV,'asano.csv'
```

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</pre> >:i.# a NB.トレンド項を作成(>:は1オリジン) $CPI = 26.4252 + 0.0381768TXR + 3.5465Y_r$
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</pre> $y = 51.4814 - 5487.67(1/EXR) + 4.42389Y_r$
対数とトレンド項	<pre>reg0 (^ . 1{"1 DN30),. (>:i. # DN30) ,. ^ . 2{"1 DN30 1.50601 0.377514 0.0639988</pre> $y = \ln 1.50601 + \ln 0.377514 + 0.0639988Y_r$

*5トレンドのオリジンは0とした

トレンド項は対数にしない。

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|
+-----+-----+
|corr=: |95.478                      |
+-----+-----+
|AIC:   |76.8481                      |
+-----+-----+
|DW=    |0.364896                      |
+-----+-----+
|t=:    |13.2416 _3.71547 11.9488|
+-----+-----+
```

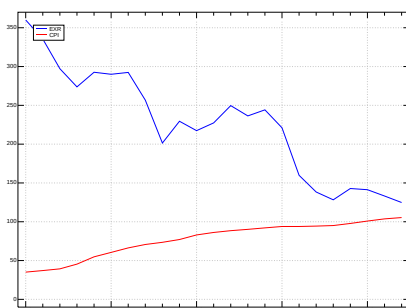


図 3.9 EXC CPI

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。

3.6.2 ダミー変数

ダミー変数は

1. パラメーターの構造変化（バブル, リセセッション）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。

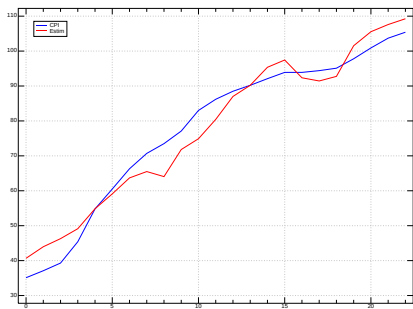


図 3.10 using trend

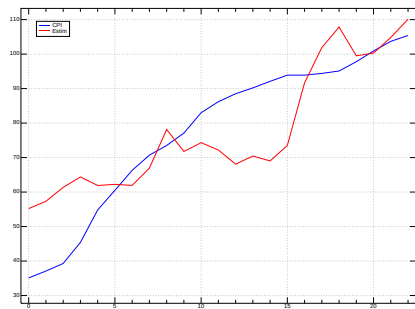


図 3.11 nothing trend

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。

3.6.3 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 0 Spring 1 Summer 2 Autumn 3 Winter
月のダミーを入れる (複数指定可能)	<i>mk_dummy12</i>	8 1 2 <i>mk_dummy12 y.</i>

Worked Example

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

```
DN31
36 44 45 106
38 46 47 112
42 49 48 118
42 50 51 118
      Xmas
```

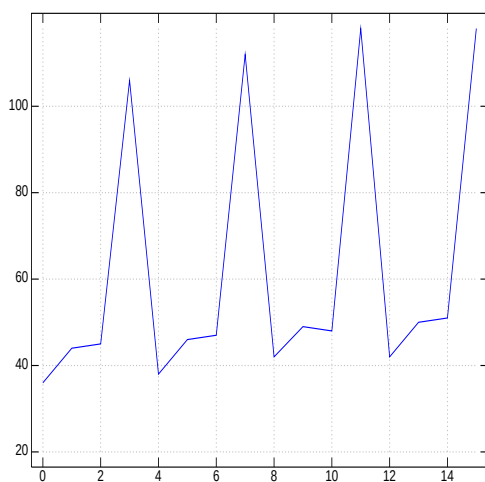


図 3.12 宝石の売り上げ

ダミー変数を作る

```

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

```

■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。

```

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

```

X データにサイズを合わせ,X と結合する.

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 1 2 3 mk_dummy4 DN31
35.6 0.65 72.05 6.95 7.1

```

```
linefit_reg0 1 2 3 mk_dummy4 DN31
```

*6

$$\hat{f} = 35.6 + 0.65Q_0 + 72.05Q_1 + 6.95Q_2 + 7.1Q_3$$

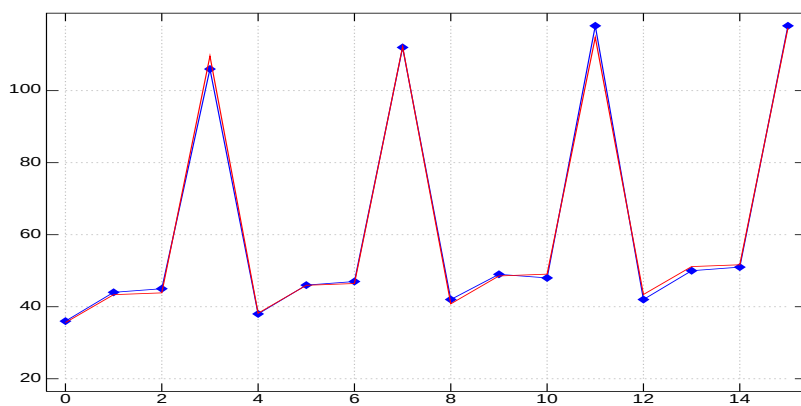


図 3.13 宝石の売り上げと推計値

```
reg0 3 mk_dummy4 DN31
39.69 0.734756 67.1972
```

```
linefit_reg0 3 mk_dummy4 DN31
```

$$\hat{f} = 39.69 + 0.734756Q_3 + 67.1972Q$$

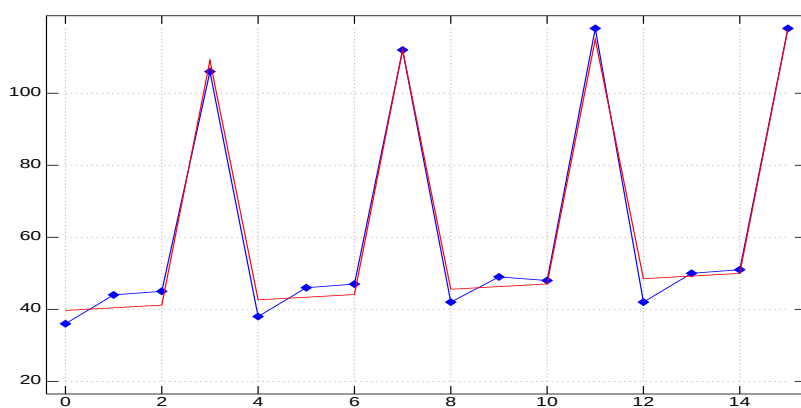


図 3.14 宝石の売り上げと推計値

*6 ダミー項のオリジンを0としている

3.6.4 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。

3.7 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 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 Coinst Lagged) の月次指標から最適指標を選択する方法を考える。

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。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。

竹内・鈴木関数のテスト

```
R=. 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)

*7

ここではデータは Y,X は分離された型となっている。

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 しておく。

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 |209.464|
+-----+
| 1 1 0 1 0 1 1 1 0 1 0 1 |209.542|
+-----+
```

*7 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。

```

|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、 $_mC_n$ の n が大きいと *out_of_memory* を頻発する。予選リーグを行い、L C L G 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。

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|

```

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

3.8 不均一分散・系列相関

3.8.1 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'
```

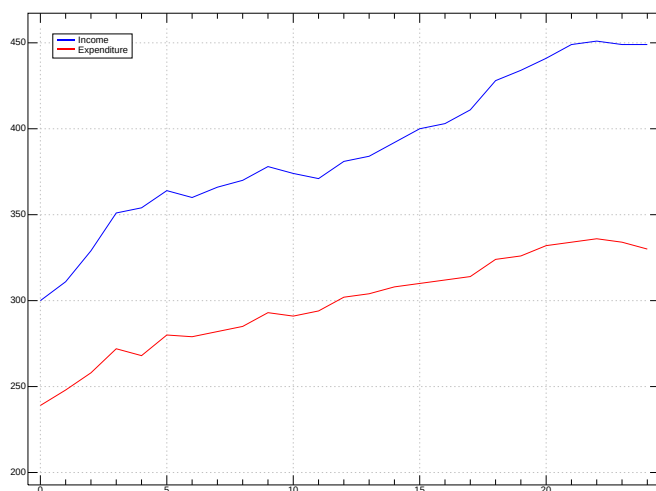


図 3.15 家計の所得と支出（実質）

```

      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|
+-----+-----+
|corr=: |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。

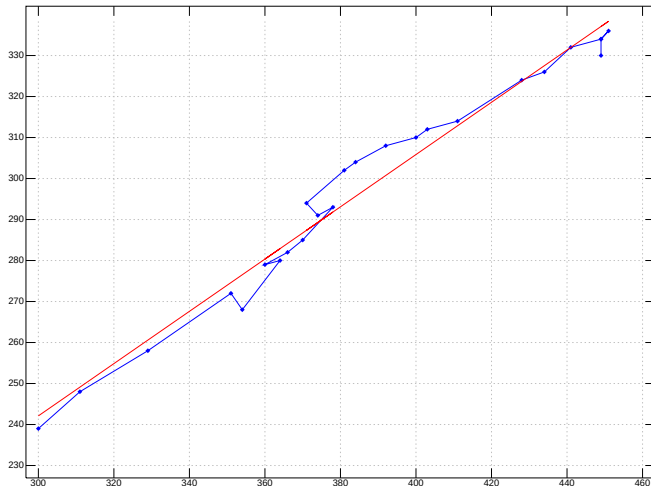


図 3.16 家計の所得 (X) と支出 (実質 Y)

3.8.2 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

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

1 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

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

Y1 消費と X1 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|
+-----+-----+

```

$$\rho = 0.826887$$

$$Y_t = 91.3968 + 0.539686X_t$$

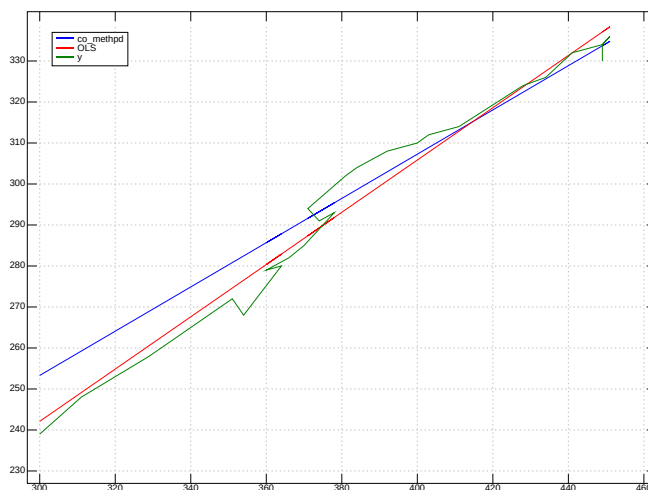


図 3.17 家計の所得 (X) と支出 (Co OLS)

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

```

3.8.3 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 $t = 1$

$$Y_1^* = \sqrt{1 - \hat{\rho}^2} Y_1$$

$$C_1^* = \sqrt{1 - \hat{\rho}^2}$$

$$X_1^* = \sqrt{1 - \hat{\rho}^2} X_1$$

$t = 2, \dots, n$ のとき

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

$$C_t^* = 1 - \hat{\rho}$$

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

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				82	0.149039	65.2935	51.8175
Year	Ct*	Xi*	Yt*	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。

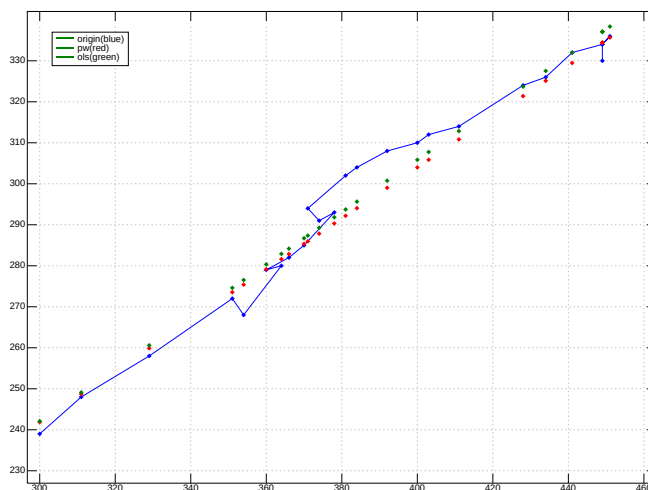


図 3.18 家計の所得と支出 (実質)

3.8.4 不均一分散 (2)・加重最少自乗法 WLS

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 で、割るかまたは、収縮 (deflate) した次のモデルが適用できる。

$$\begin{aligned}\frac{Y_i}{\sigma_i} &= B_1 \frac{1}{\sigma_i} + B_2 \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i} \\ v_i &= \frac{u_i}{\sigma_i} \\ v_i^2 &= \frac{u_i^2}{\sigma_i^2}\end{aligned}$$

$\sqrt{X_i}$ 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.

$$\begin{aligned}\frac{Y_i}{\sqrt{X_i}} &= B_1 \frac{1}{\sqrt{X_i}} + B_2 \frac{X_i}{\sqrt{X_i}} + \frac{u_i}{\sqrt{X_i}} \\ &= B_1 \frac{1}{\sqrt{X_i}} + B_2 \sqrt{X_i} + v_i\end{aligned}$$

```
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 |
+-----+-----+
| corr=:  | 97.4996           |
+-----+-----+
| AIC:    | 77.3239           |
+-----+-----+
| DW=     | 0.352839          |
+-----+-----+
| t=:     | 5.8238 30.2508    |
+-----+-----+
```

3.9 関数一覧

OLS	<i>reg0</i> <i>reg_exam_ad</i>	<i>reg0 y</i> <i>reg0 reg_exam_ad y</i>
-----	-----------------------------------	--

正規方程式 クラメール法	<i>test_neq</i> <i>cr</i>	<i>test_neq y</i> <i>cr test_neq y</i>
t 検定、p 値	<i>tcd f</i>	<i>m tcd f 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>
reg_select	<i>find_best0</i> <i>find_best</i>	<i>find_best0 y ccompare n;x</i> <i>(m1;m2)find_best y;x</i>
系列相関、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>

第4章

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。

4.1 多項式モデル

4.1.1 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

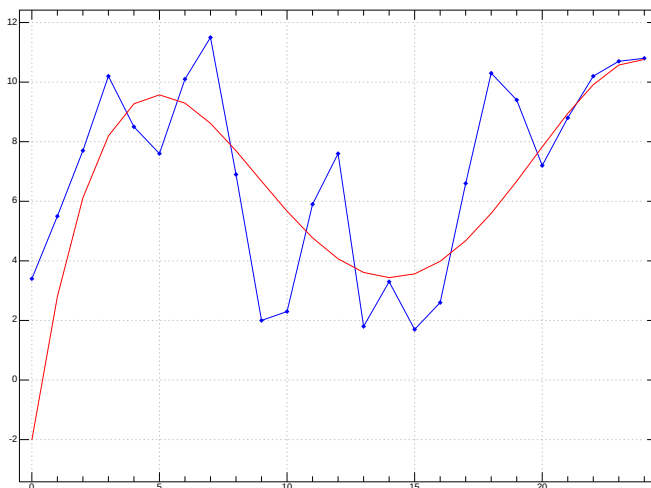


図 4.1 東京の冬の最低気温 (青) と推計値 (赤)4 次

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

$$y = 4.89554 + 1.13288x - 0.141456x^2 + 0.00438882x^3$$

```

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。

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

4.1.2 多項式の解を求める

J の多項式関数 (p .) は非常に強力で、多項式の解や推計値を直ちに求めることができる。

■Worked Example

多項式回帰 4 次多項式で回帰する

```

4 poly0 DN12
_2.01542 5.65633 _0.889141 0.0485201 _0.000848678

```

$$y = 4.89554 + 1.13288x - 0.141456x^2 + 0.00438882x^3$$

複素数の解 (右欄)

```

p. 4 poly0 DN12  NB. 複素数の解
+-----+-----+
|_0.000848678|28.5493 14.1219j4.51938 14.1219j_4.51938 0.378352|
+-----+-----+

```

0.378 が得られる。(図 4.1 参照)

実数解を求める | L:0 p. 4 poly0 DN12 NB. 絶対値で実数解を求める

```

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

```



```
|0.000848678|28.5493 14.8274 14.8274 0.378352|
```

```
+-----+-----+-----+-----+-----+-----+
```

```
plot
```

```
4 linefit_poly0 DN12
```

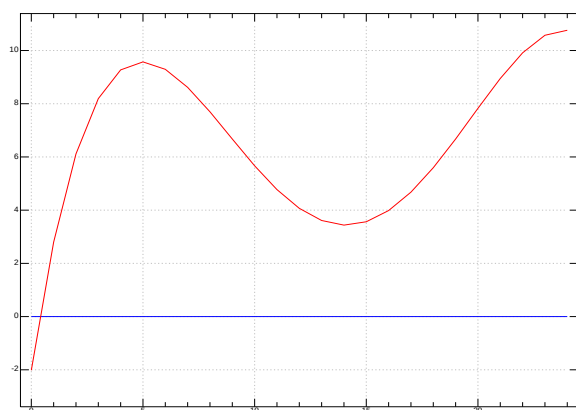


図 4.2 polynomial

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 4.1)

4 *linefit_poly*0 DN12

■J 言語メモ

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

4.1.3 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

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

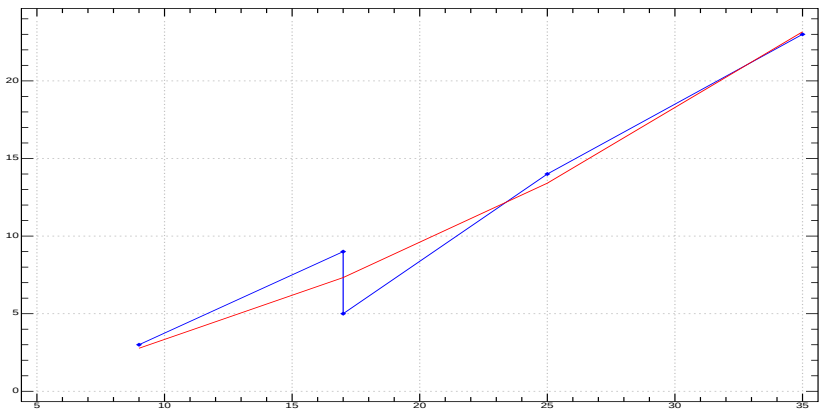


図 4.3 車の速度と停止距離

<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 2 poly0 DN40 y = -0.524753 + 0.259012x + 0.0119353x^2</pre>
--	--

多項式の正規方程式

多項式の正規方程式を用いると

$\begin{bmatrix} \sum v_i^4 & \sum v_i^3 & \sum v_i^2 \\ \sum v_i^3 & \sum v_i^2 & \sum v_i \\ \sum v_i^2 & \sum v_i & \sum 1 \end{bmatrix}, \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum v_i^2 y_i \\ \sum v_i y_i \\ \sum y_i \end{bmatrix}$	<pre>a.=;("1) ,.3<\ +/ V ^/ 4 3 2 1 0 2.06485e6 69055 2509 69055 2509 103 2509 103 5</pre>
--	---

$\begin{bmatrix} \sum v_i^2 y_i \\ \sum v_i y_i \\ \sum y_i \end{bmatrix}$	$b = \frac{Y * V}{\sum V^2} = \frac{1420}{41214} = 0.03445$
	$b \% a$ $0.0119352 \quad 0.259012 \quad -0.524761$ $y = 0.0119352x^2 + 0.259012x - 0.524761$

*1

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。

■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

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

*1 J の多項式関数の統一記述方式で、逆に出る

が多くなると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。

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 数次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。

4.2 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。

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

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

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。

1	2	3	mk_dummy4	DN41	DN41=:1 2 3 mk_dummy4 ^ .DN41
					NB. 対数
0	0	0	0	72	0 0 0 0 4.27667
1	0	0	1	110	1 0 0 1 4.70048
2	0	1	0	117	2 0 1 0 4.76217
3	1	0	0	172	3 1 0 0 5.14749
4	0	0	0	76	4 0 0 0 4.33073
5	0	0	1	112	5 0 0 1 4.7185
6	0	1	0	130	6 0 1 0 4.86753
7	1	0	0	194	7 1 0 0 5.26786
8	0	0	0	78	8 0 0 0 4.35671
9	0	0	1	119	9 0 0 1 4.77912
10	0	1	0	128	10 0 1 0 4.85203
11	1	0	0	201	11 1 0 0 5.3033
12	0	0	0	81	12 0 0 0 4.39445
13	0	0	1	134	13 0 0 1 4.89784
14	0	1	0	141	14 0 1 0 4.94876
15	1	0	0	216	15 1 0 0 5.37528

*2

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。

*2 トренд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。

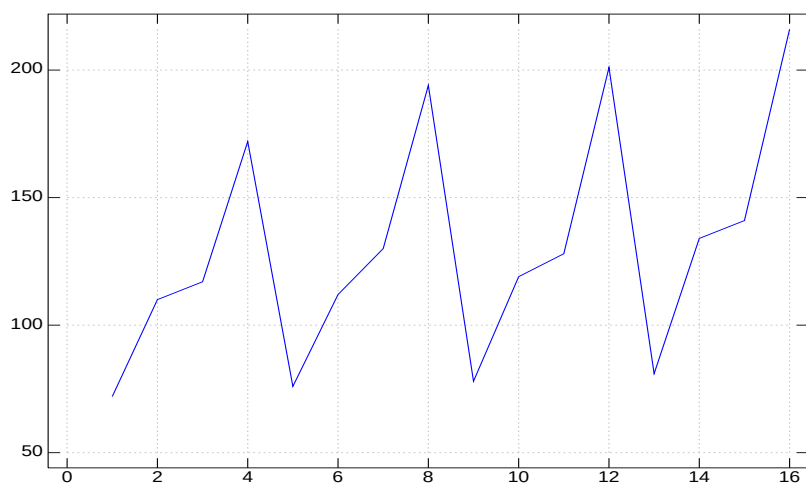


図 4.4 exp_reg

重回帰モデル	<pre>reg0 1 2 3 mk_dummy0 DN41 64.8625 1.98125 113.056 48.2875 40.0188</pre> $\hat{f}_t = 64.8625 + 1.98125t + 113.056Q4_t + 48.2875Q3_t + 40.0188Q2_t$
対数重回帰モデル	対数モデル $\ln(D_t) = a' + b'_1 t + b'_2 Q2_t + b'_3 Q3_t + b'_4 Q4_t + e'_t$ <pre>reg0 1 2 3 mk_dummy0 ^ . DN41 4.25357 0.0143442 0.890812 0.489297 0.420002</pre> $\ln_t = 4.25357 + 0.0143442t + 0.890812Q4_t + 0.489297Q3_t + 0.420002Q2_t$

指数乗法モデル	$Y_t = a_0 \times \alpha_1^{xt1} \times \alpha_2^{xt2} \times \cdots \times \alpha_m^{xtm} \times \epsilon_t, (t = 1, \cdots N)$ $\hat{D}_t = (e^{a'}) (e^{b'_1 t}) (e^{b'_2 Q_{2t}}) (e^{b'_3 Q_{3t}}) (e^{b'_4 Q_{4t}})$ $\hat{D}_t = (e^{4.25357}) (e^{+0.0143442t}) (e^{0.890812 Q_{4t}}) (e^{0.489297 Q_{3t}}) (e^{0.420002 Q_{2t}})$
---------	--

Script

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. $e^n e$ のベキ乗化

```

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

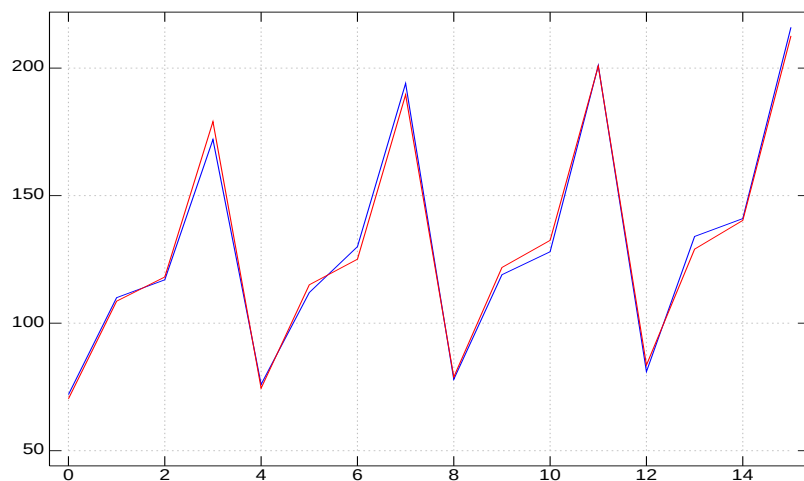


図 4.5 exp.reg

■回帰係数の意味 需要の増加

$$\ln(D_t) = a' + b'_1 t + b'_2 Q2_t + b'_3 Q3_t + b'_4 Q4_t + e'_t$$

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

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

1.445 % increase in demand each quarter 各四半期単位で 1.445% ず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 %)

```
1x1 ^ }. reg0 DN410
1.01445 2.43711 1.63117 1.52196
```

4.2.1 Reference

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

4.3 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

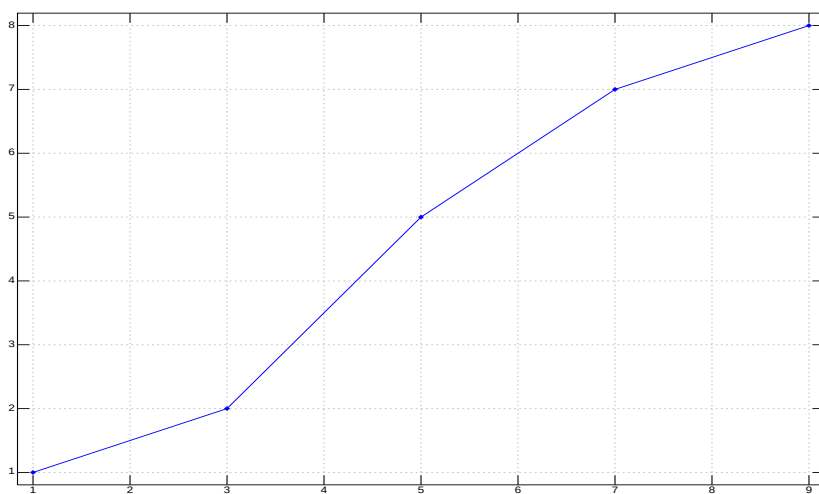


図 4.6 logistic

Script

```

NB. =====
NB. Logistic regression
NB. =====
logis=: 4 : 0
Y=: ^. (snd y.) % x. - snd y.
b0=: (+/ Y*fst y.) - tmp2=: ((+/ Y) * +/ fst y.)% tmp=: > 0{$ L:0 y.
b1=: (+/ (fst y.) ^2) -((+/ fst y. )^2) % tmp
b=: b0 % b1
a0=: (+/ Y)- b* +/ fst y.
a1=: b* tmp
a=: - a0 % a1
a,b
NB. Y reg fst y.
)

```

4.3.1 Reference

石村 貞夫 「グラフ統計のはなし」東京図書 1995

4.4 Newton 法による非線形回帰

4.4.1 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 言語メモ

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.] 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 F is vector-valued
$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$	$x^{(n+1)} = x^{(n)} - \frac{F(x^{(n)})}{F'(x^{(n)})}$

```
ff2 newton
```

```
(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。

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

fn で $x^3 + x - 1$ を定義する。

```
i: 5
```

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

```
fn i:5
```

```
_131 _69 _31 _11 _3 _1 1 9 29 67 129
```

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
```

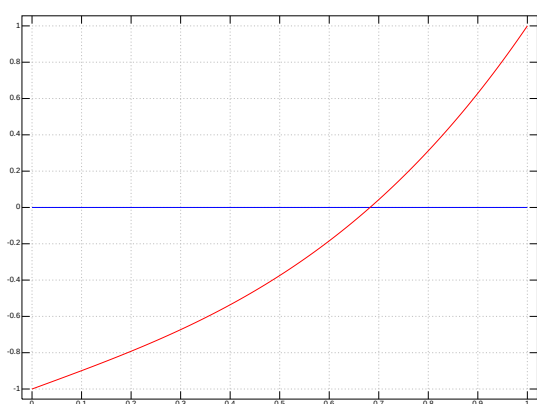
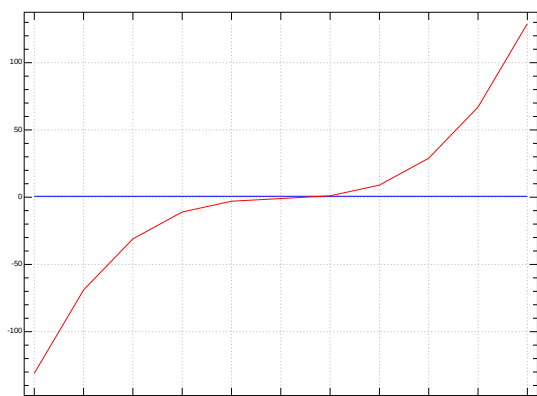
初期値を-5 から 5 までの 11 個とした。i: は i. と異なり、マイナス側にも鏡映の数列を取る。

```
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
```

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
```



4.4.2 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

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。

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

求める関数	$y = c_0 + c_1 \exp^{-c_2 x}$
-------	-------------------------------

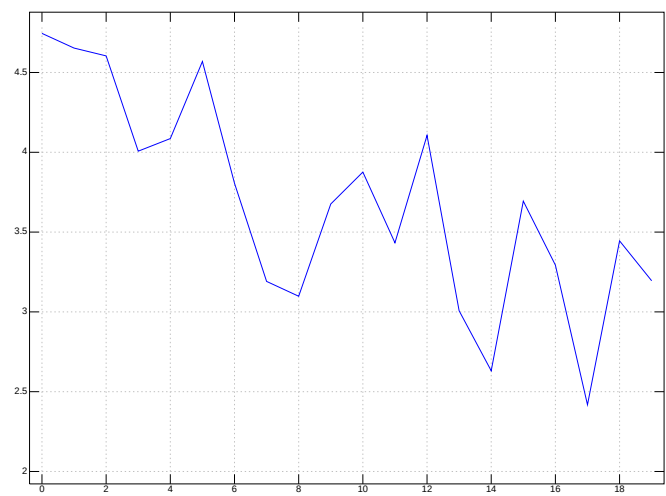


図 4.7 DT2

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>
	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


```
(|. 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 4.8 Newton_fit

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	$y = c_0e^{c_1x}$
Newton 法での定義	$fn2=: 4 : ' (0{y}) * ^x * 1{y}'$

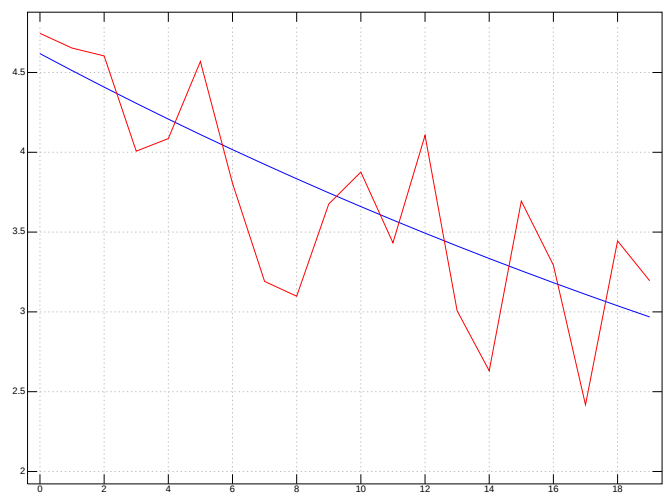


図 4.9 regression with exponential

初期値	C2=:1.3 -1
	DN44 +-----+-----+ 0.1 0.2 0.3 0.4 0.5 1.1112 0.8232 0.6099 0.4518 0.3347 +-----+-----+

(|. DO_0)&(fn2 sub_a2) newton_fit C2=. 1.3 _1
1.49993 _2.9998

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

4.4.3 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'

4.4.4 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

4.5 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。

3.6.1 は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。

4.5.1 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

年電力需要 人口 鉱工業生産指数

	Y	P	IPX
55	442.37	0.8928	6
56	515.81	0.9017	7.3
57	587.98	0.9093	8.7
58	630.76	0.9177	8.5
59	740.81	0.9264	10.2
60	877.46	0.9342	12.7
61	1015.54	0.9429	15.2
62	1076.81	0.9518	16.5
63	1229.44	0.9616	18.3
64	1387.92	0.9718	21.2

(2 3;1) mk_dougaku DN45

Y-1	d-P	d-IPX	Y
442.37	0.0089	1.3	515.81
515.81	0.0076	1.4	587.98
587.98	0.0084	0.2	630.76
630.76	0.0087	1.7	740.81
740.81	0.0078	2.5	877.46
877.46	0.0087	2.5	1015.54
1015.54	0.0089	1.3	1076.81
1076.81	0.0098	1.8	1229.44
1229.44	0.0102	2.9	1387.92

(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

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

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)

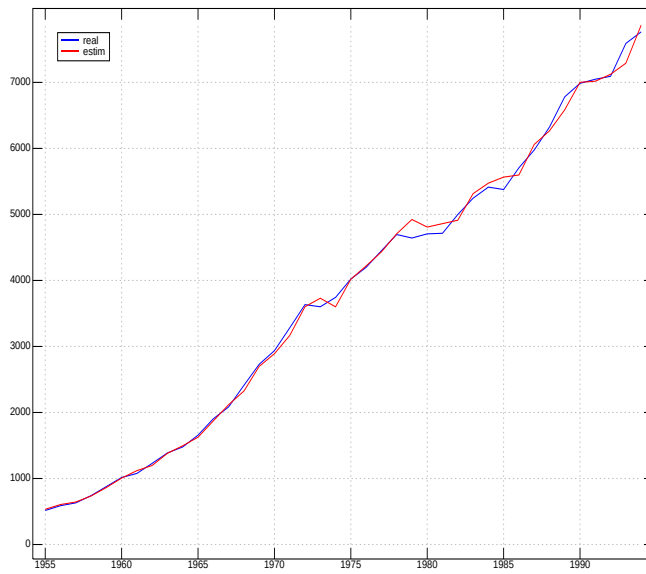


図 4.10 電力需要

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
)
```

4.5.2 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。

$0, A_1, A_2, A_3$ から、ハウタッカ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|
+-----+-----+
```

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。

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

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1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。

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

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

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

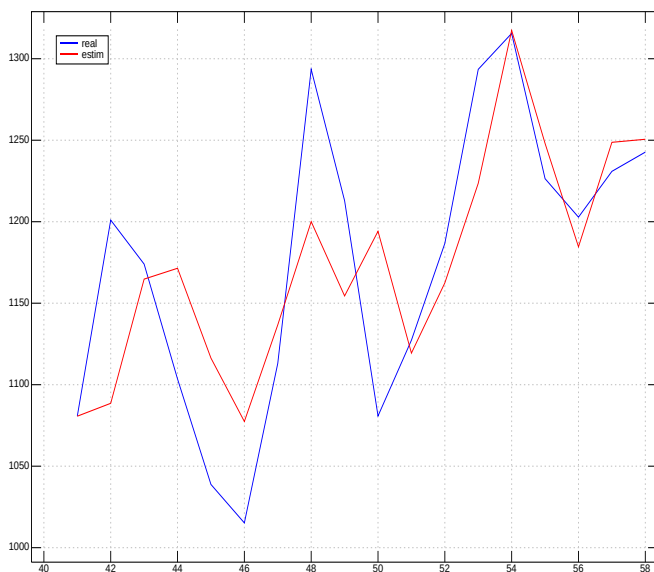


図 4.11 冷蔵庫の需要 HT-Model

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

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

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

(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
```

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

CDS D: 耐久財+半耐久財

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

```
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
)
```

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

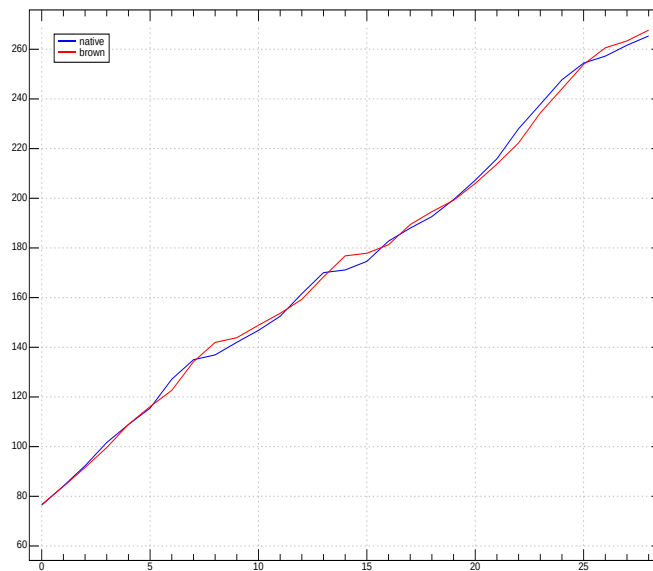


図 4.12 名目家計可分所得と耐久消費財

4.6 Group Logit

Qualitative Response regression Model

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。

4.6.1 不均一分散のある場合-1

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。

従って、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)

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
```

2. $\text{logit} = \ln \hat{P} / (1 - \hat{P})$ 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
```

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

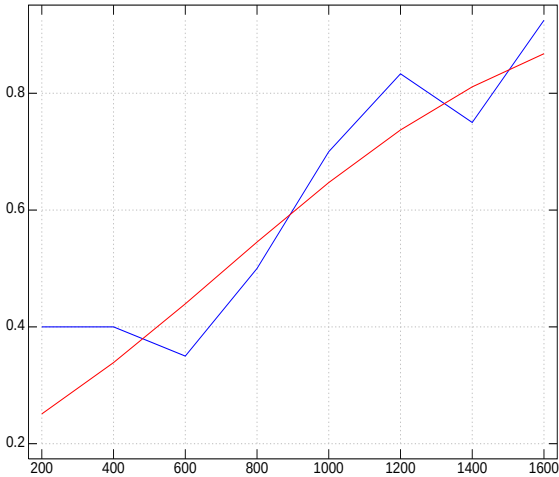


図 4.13 所得と自動車保有率の logit 回帰

```
3. w = N*P-hat*(1 - P-hat)

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

4. w^1/2 NB. 漸近的有效推定量=分散の平方根の逆数
5. L* = w_i^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	692.8	
400	100	40	0.4	0.6	0.66666	_0.405465	24	4.89898	_1.98637	1959.5	
600	200	70	0.35	0.65	0.53846	_0.619039	45.5	6.74537	_4.17565	4047.2	
800	300	150	0.5	0.5	1	0	75	8.66025	0	6928.	
1000	200	140	0.7	0.3	2.3333	0.847298	42	6.48074	5.49112	6480.7	
1200	120	100	0.833333	0.166667	5	1.60944	16.6667	4.08248	6.5705	4898.9	
1400	80	60	0.75	0.25	3	1.09861	15	3.87298	4.25491	5422.1	

```
|1600 40 37 0.925 0.075 12.333 2.51231 2.775 1.66583 4.18508 2665.33|
+-----+
```

```
reg_logit
```

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

```
L*, W1/2, 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。

4.6.2 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	Ni	ni
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

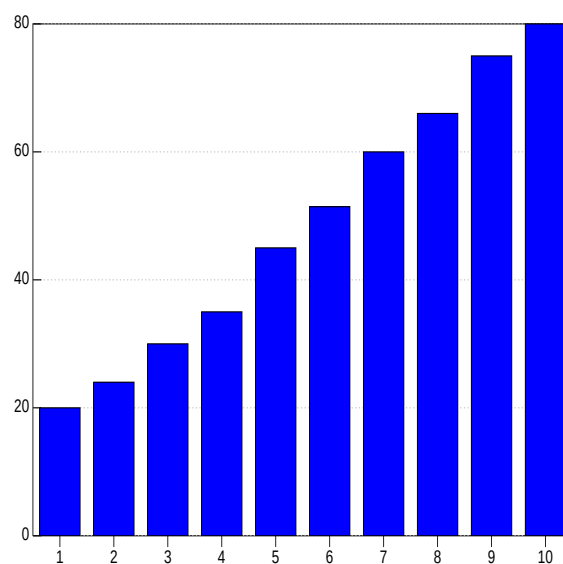


図 4.14 所得と持ち家

4.6.3 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 $w_i^{\frac{1}{2}}, X^*$ 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) $L^*, \quad W^{\frac{1}{2}}, \quad X^*$

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
```

$$2. \text{ELstar} = \beta_1 \sqrt{w_i} + \beta_2 \hat{X}_i^* = \hat{L}_i^* = -1.59324 \sqrt{w_i} + 0.0786686 X_i^*$$

ユニット（所得1000ドル）の増加により、8.18%持ち家が増える。

```
1x1 ^ 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。

$$\begin{aligned} \frac{\hat{p}_i}{1-\hat{p}_i} &= e^{-1.59324 \sqrt{w_i} + 0.0786686 X_i^*} \\ &= e^{-1.59324 \sqrt{w_i}} \cdot e^{0.0786686 X_i^*} \end{aligned}$$

```
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
```

$\hat{L}$	$\hat{X}^*$	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) = $\frac{\hat{P}}{1 - \hat{P}_i} = e^{(logit)}$

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

これは

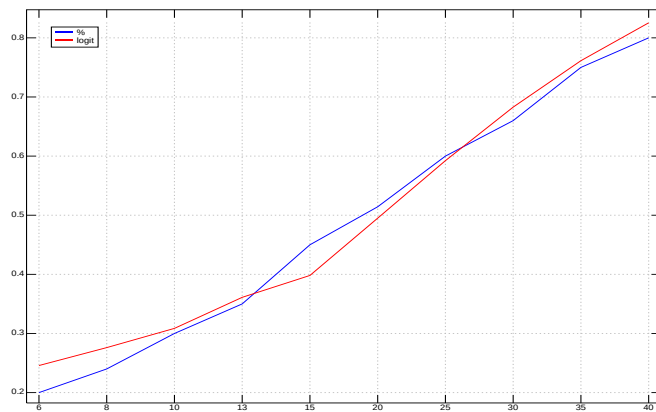
(1x1^LP)%1+1x1 ^ LP で求められる。

```
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 4.15 n_i/N_i and Logit

所得階層（左欄 1000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。

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

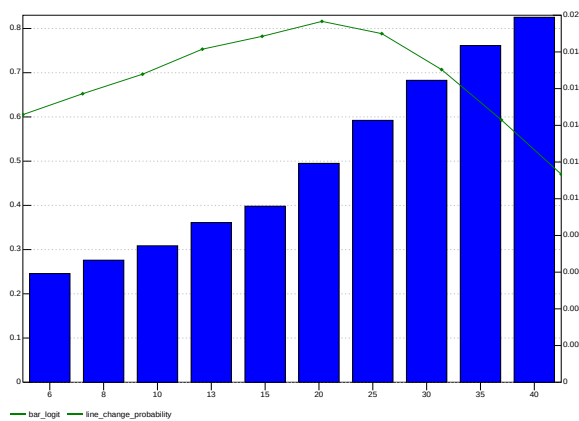


图 4.16 Logit and change in probability

```
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

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

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

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

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

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

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

4.7 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。

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

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

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。

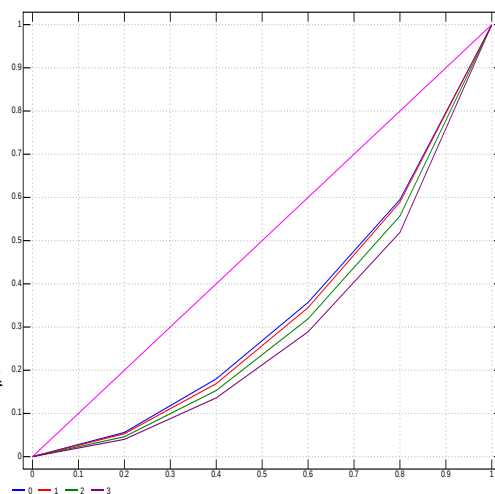


図 4.17 jini

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なる。

```
|: 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|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
```

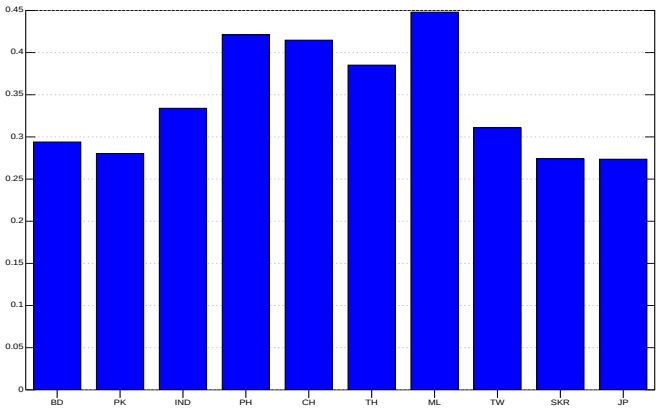


図 4.18 Logistic curve of $\hat{P}$

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>	
ロジット	<i>logit_1</i> <i>logit_2</i> <i>reg_logit</i> <i>estim_logit</i> <i>plot_logit</i> <i>plot_logit2</i>	
ジニ係数	<i>jini_sub</i> <i>calc_jini</i>	

第5章

スムージングと季節調整

Jは移動平均が簡単に作成でき、各種の移動平均が体験できる。月や四半期の時系列データでは季節変動を取り除いた上で用いることが多い。天災や異常気象等の特異な変化を緩和し平準化することも季節調整と言われる。

5.1 移動平均・Moving Average

最初に移動平均を紹介したのは *Poynting*(1884) と言われる。

5.1.1 単純移動平均

Jは、移動平均を次の簡潔なイディオムで表現できる。

```
mav=: +/\ % [
```

これは

```
mav=: (# %~ +/\)\
```

でも表現できる。

経過と説明

Jの infix (\) の機能は強力であり左引数(x)で指定した数で移動平均のためにずらした組み合わせを作る。

乱数生成 ?. はシードを固定

```
3<\ 6?.20
6 3 19 15 10 14
```

3項をずらして 3項を一つずつずらして組上げる

```
3<\ 6?.20
+-----+-----+-----+-----+
|6 3 19|3 19 15|19 15 10|15 10 14|
+-----+-----+-----+-----+
```

平均する ボックス内を平均すると移動平均になる。

```

3(+/%#)\ 6?.20
9.33333 12.3333 14.6667 13

mav      3 mav 6?.20
9.33333 12.3333 14.6667 13

```

?. 同じ乱数を生成

■J 言語メモ +/ Σ
 3 < 1 つずつすらし 3 個毎に Box 化

中心化

移動平均は、次数（左パラメーター x ）が偶数の場合には、原データと整合させるために中心化が行われる。 mav_c は次数の奇数、偶数を自動判別して移動平均を行う。

```

mav=:+/\ % [
mav_c=: 2&mav@mav ' mav @. (2&|@[]) NB.centred MA

```

```

s=.10?.20
a , (2 mav a),: 2 mav_c a

```

```

6 3    19    15 10    14 0    7    12 17
4.5 11    17 12.5 12    7 3.5 9.5 14.5 0
0 7.75 14 14.75 12.25 9.5 5.25 6.5 12 0

```

12 カ月移動平均

$mav12$ は mav の引数を 1 2 に固定して組込み中心化を行っている。

```

mav=: +/\ % [
NB. Season adjustment using 12 month

```

```

mav12=: [: 2&mav 12&mav

```

```

2<\ 12 mav i. 24
+-----+-----+-----+-----+-----+-----+-----+
|5.5 6.5|6.5 7.5|7.5 8.5|8.5 9.5|9.5 10.5|10.5 11.5|11.5 12.5|
+-----+-----+-----+-----+-----+-----+-----+

```

```

mav12 i. 24
6 7 8 9 10 11 12 13 14 15 16 17

```

12 項移動平均は 13 項を取って最初と最後の項に $\frac{1}{2}$ のウェイトを置いた
 $(\frac{1}{2} 1 1 1 1 1 1 1 1 1 1 1 \frac{1}{2})$ 1 3 項加重移動平均と同じである。

5.1.2 加重移動平均

移動平均の項にウェイトを置く方法。

■DATA Stock price of IBM NY and London market
 DN50=: ".@> readcsv DIRCSV,'ibmnyln.csv'

3 項加重移動平均

3 ヶ月の移動平均値を更に移動平均する場合。一層のスムージングがはかれる。

$$\hat{Q} = \frac{\frac{Q_{t-2} + Q_{t-1} + Q_t}{3} \frac{Q_{t-1} + Q_t + Q_{t+1}}{3} + \frac{Q_t + Q_{t+1} + Q_{t+2}}{3}}{3}$$

$$= \frac{1}{9}Q_{t-2} + \frac{2}{9}Q_{t-1} + \frac{3}{9}Q_t + \frac{2}{9}Q_{t+1} + \frac{1}{9}Q_{t+2}$$

$$= \frac{1}{9}[1 \ 2 \ 3 \ 2 \ 1]$$

これは、上記のウェイトによる 5 ヶ月の加重移動平均と同じである。四季の変動の特徴をよく捕らえることができる。

補完

加重 3 月移動平均の欠項は、次のウェイトを用いて両端を補完することができる。

$$Q_{t+1} = \frac{1}{9}Q_{t+3} \rightarrow \frac{1}{9}\left(\frac{Q_t + Q_{t+1} + Q_{t+2}}{3}\right)$$

$$Q_{t+1} = \frac{3 \ 7 \ 10 \ 7}{27}$$

$$Q_{t+2} = \frac{2}{9}Q_{t+3} \rightarrow \frac{2}{9}\left(\frac{Q_t + Q_{t+1} + Q_{t+2}}{3}\right) + \frac{1}{9}\left(\frac{Q_t + Q_{t+1} + Q_{t+2}}{3}\right)$$

$$Q_{t+2} = \frac{2 \ 3 \ 4}{9}$$

cpl_mavw3 は両端の補完を行う。plot_w3 は cpl_mavw3 を組み込んだ。

```
plot |: a,. cpl_mavw3 a=. {. |: DN50
```

5.1.3 Spencer 移動平均

スペンサー 15 ポイント移動平均は U.S. 商務省センサス局法 X-10 で用いられた。
 ウェイトを次のように置いており、隣接に重心をおくので、三角形や低位の数値が破壊され
 れないでよく残っている。

これは、次のウェイトに依っている。

$$\frac{1}{4} \begin{bmatrix} -3 & 3 & 4 & 3 & -3 \end{bmatrix}$$

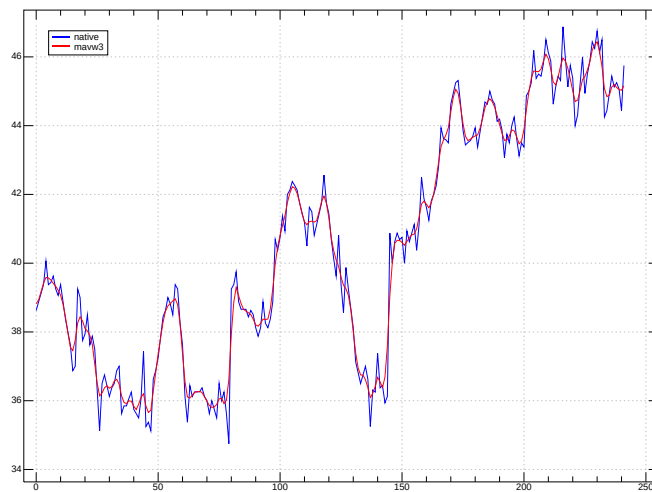


図 5.1 Stock Price of IBM NY Market/WMAV3r

$$\frac{1}{5} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\frac{1}{4} \begin{bmatrix} -3 & 3 & 4 & 3 & -3 \end{bmatrix}$$

$$\frac{1}{4} \begin{bmatrix} -3 & 3 & 4 & 3 & -3 \end{bmatrix}$$

これは、実際には、次のウェイトになる。[?] [?]

$$a_0 = \frac{74}{320}, a_{\pm 1} = \frac{67}{320}, a_{\pm 2} = \frac{46}{320}, a_{\pm 3} = \frac{21}{320},$$

$$a_{\pm 4} = \frac{3}{320}, a_{\pm 5} = \frac{-5}{320}, a_{\pm 6} = \frac{-6}{320}, a_{\pm 7} = \frac{-3}{320}$$

Spencer 移動平均のウェイトは 15 項限定であり、加重 15 項移動平均のパラメーターと一致する。

```
plot_spencer {.: DN50
```

Script

```
spencer2=: 3 : 0
```

```
WEIGHT=: _3 _6 _5 3 21 46 67 74 67 46 21 3 _5 _6 _3 % 320
```

```
y,.; WEIGHT +/. * (L:0) 15<\ ( (7 # {.),] ,7 #{: ) y
```

C.Reiter によるスマートな Tacit のスクリプト

```
NB. =====
```

```
NB. C.Reiter Fractals Visuarization and J (2nd Edition 2000)
```

```
NB. =====
```

```
wts=: (|. ,|. ) 74 67 46 21 3 _5 _6 _3%320
```

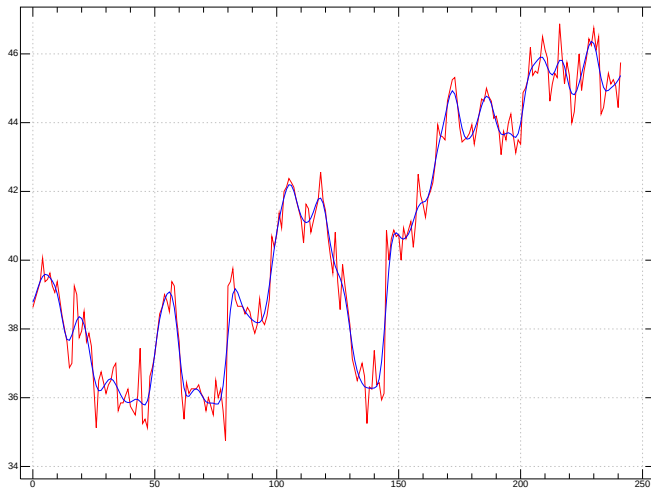



図 5.2 Stock Price of IBM NY Market/Spencer

```
locspen=: (+/ . *)&wts
spencer=: 15"_ locspen\ (7: # {.),] ,7:#{:
```

5.1.4 ヘンダーソン移動平均

Henderson 移動平均は U.S. 商務省センサス局法 X-11,X-12 の採用する季節調整法に採り入れられている手法で、基本的には、全変動を、Trendcycle(TC)、季節変動 (S)、不規則変動 (I) の 3 つの成分に分けるときの、TC を求めるのに使用される重み付き移動平均法である。

$$H^n(B) = \sum_{i=-(n-1)/2}^{(n-2)/2} h_i^{(n)} B^i, \quad i = 0, \pm 1, \pm 2, \dots, \pm(n-1)/2$$

B, 時間の遅れ、 F、時間の前のパラメータ

B is B(back) F is F(Foward)

$h_i^{(n)}$ 重み係数 $m = (n+3)/2$

$$h_i^{(n)} = \frac{315((m-1)^2 - i^2)(m^2 - i^2)((m+1)^2 - i^2)((3m^2 - 16) - 11i^2)}{8m(m^2 - 1)(4m^2 - 1)(4m^2 - 9)(4m^2 - 25)}$$

5 項のウェイト

$$H^{(5)}(B) = -\frac{21}{286}B^2 + \frac{84}{286}B + \frac{160}{286} + \frac{84}{286}F - \frac{21}{286}F^2$$

7項のウェイト

$$H^{(7)}(B) = -\frac{42}{715}B^3 + \frac{42}{715}B^2 + \frac{210}{715}B + \frac{295}{715} + \frac{210}{715}F + \frac{42}{715}F^2 - \frac{42}{715}F^3$$

Spencer や Henderson の移動平均は、原系列を追いつながら滑らかさを実現していることである。Henderson の7項は、原系列を追っていくが、15項では滑らかになる。Spencer の15項は、Henderson の15項と図で表せば、概ね同じような滑らかさを持っており、ウェイトの差は微小である。

```
plot_henderson {.: DN50
15 plot_henderson2 {.:DN50 NB. spencer and henderson
15 plot_comb_all {.:DN50 NB. combination all
```

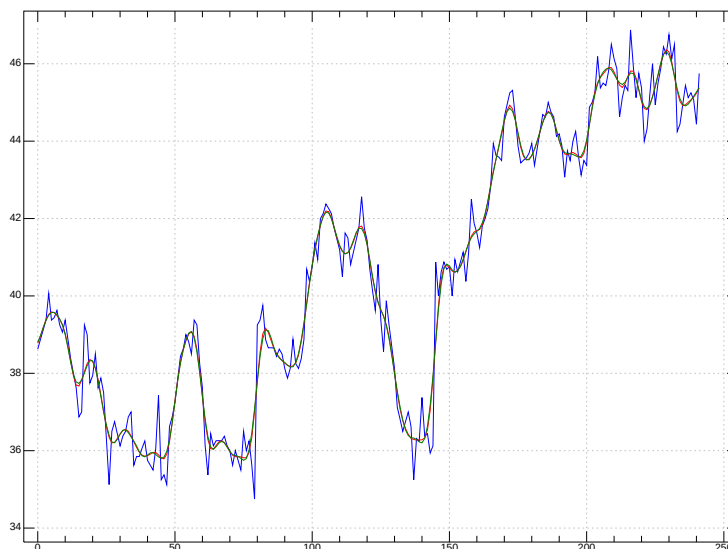


図 5.3 Stock Price of IBM NY Market Spencer,Henderson

IBM Stock Price of NY Market の最初の100ポイントを Spencer と Henderson の15ポイントで plot してみると、ほとんど線が重なる。henderson は項数が任意に取れるところが優れている。

交通量やデパート売り上げのデータでは、8月や12月のピークが大きいですが、henderson は高次になるほど、追従よりも滑らかさを求めていく保守的 (Concerbative) な振舞いがあると言われる。

Script

スクリプトは竹内による。[?] henderson の移動平均の項数は任意（奇数）で、13－15項では Spencer と極めて類似した結果となる。

```

henderson=:4 : 0
NB. Usage: e.g. 15&henderson y //or 7 henderson y
wt=: x hnwgt i:-:<:x
x (+/ . *)&wt\((<.@-:x)#{.y),y,(<.@-:x)#{:y
)
hnwgt=:4 : 0
m=: -: x +3 NB. (n+3)/2
nm=: 315*((m-1)^2)-*:y)*((m^2)-*:y)*((m+1)^2)-*:y)*((3**m)-16)-11**y)
hi=:nm%8*m*((m^2)-1)*((4**m)-1)*((4**m)-9)*((4**m)-25)
)

```

5.1.5 移動平均の PLOT

移動平均を求め原データと同時にグラフにあらわす関数を用意した。

- ・ 任意の次数（例では 3＝四半期）と原系列を plot する

偶数の場合はグラフにあらわすため中心化の処理を行っている。

<i>plot_wmav</i>	weighted mav	plot_wmav n
<i>plot_mav12</i>	12 months mav	plot_mav12 n
<i>plot_mav12cpl</i>	12 months mav	plot_mav12cpl n
<i>plot_w13</i>	13 weighted mav and mav12cpl	plot_w13 n
<i>plot_spencer</i>	spencer	plot_spencer n
<i>plot_henderson</i>	henderson 15 項	plot_henderson n
<i>plot_henderson2</i>	henderson 任意の項 and spencer	m. plot_henderson2 n
<i>plot_comb_all</i>	henderson 任意の項と and spencer ,mav12cpl	m. plot_comb_all n henderson は次数は奇数のみ 31 33 付近は中心化 1 2 月移動 平均と類似 両端は計算方法の 差によるばらつきがある。

5.2 指数平滑化

5.2.1 指数平滑化

時系列データの平滑化で、計算が簡単でよく使い込まれた便利な方法に指数平滑化がある。任意の $0 < w < 1$ のウェイトを与えるとウェイトにより平滑化を行う。最初の方の値の影響がが後まで残る方法であるが、ウェイトを決める方法はないので、グラフを見ながら適当な値を捜す。

$$\begin{aligned}x_1^- &= x_1 \\x_2^- &= wx_1^- + (1-w)x_2 \\x_3^- &= wx_2^- + (1-w)x_3 = w^2x_1 + w(1-w)x_2 + (1-w)x_3\end{aligned}$$

■Working Example DATA is DN12 NB. ある年の東京の2月の平均気温

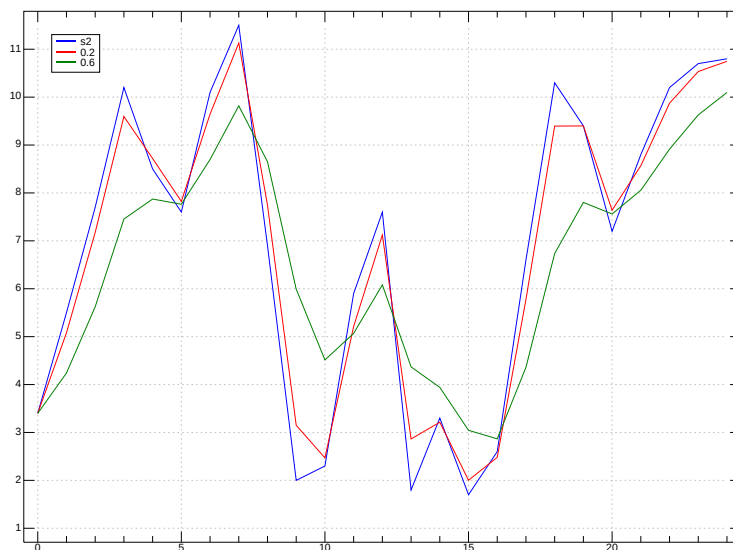


図 5.4 東京の冬の気温と指数平滑化

Script

NB. written by Giichiro Suzuki

```
smexp=:4 : 0
s=. {.y=. y
while. 1<#y do.
  s=.s,(x *{:s)+(1-x )*{:y=.}.y
```

```

DN12,. 0.2 smexp DN12
DN12      estim              7.6 7.12279
-----
3.4      3.4              3.3 3.21291
5.5      5.08            1.7 2.00258
7.7      7.176          2.6 2.48052
10.2     9.5952          6.6  5.7761
8.5      8.71904        10.3 9.39522
7.6      7.82381        9.4  9.39904
10.1     9.64476        7.2  7.63981
11.5     11.129         8.8  8.56796
6.9      7.74579       10.2 9.87359
2 3.14916          10.7 10.5347
2.3      2.46983       10.8 10.7469
5.9      5.21397

end.
)

```

5.2.2 ホルトの2段階指数平滑化

Holt's two-parameter double exponential smoothing model

Holt の2段階指数平滑化で指数平滑化のパラメータとして α に加えて β を用いる。

$$S_t = \alpha Y_t + (1 - \alpha)(S_{t-1} + b_{t-1})$$

$$b_t = \beta(S_t - S_{t-1}) + (1 - \beta)b_{t-1}$$

$$F_{t+m} = S_t + b_t m$$

	Starting base
α = level smoothing constant	Y_1 $(Y_2 - Y_1 + Y_4 - Y_3)/2$
S_t = smoothed at end of period t	
β = trend smoothing constant	
b_t = smoothed trend in period t	
m = forecast horizon	

α, β は一意には定まらない。この例では、 $\alpha = 0.2$ $\beta = 0.6$ が最良であるが、周辺の数字を入れ替えてトライして、残差が最少のものを選択して、モデルとする。

(2) Smoothed Level

(3) Smoothed Trend

$$RSE = \sqrt{\sum \frac{e^2}{n-1}}$$

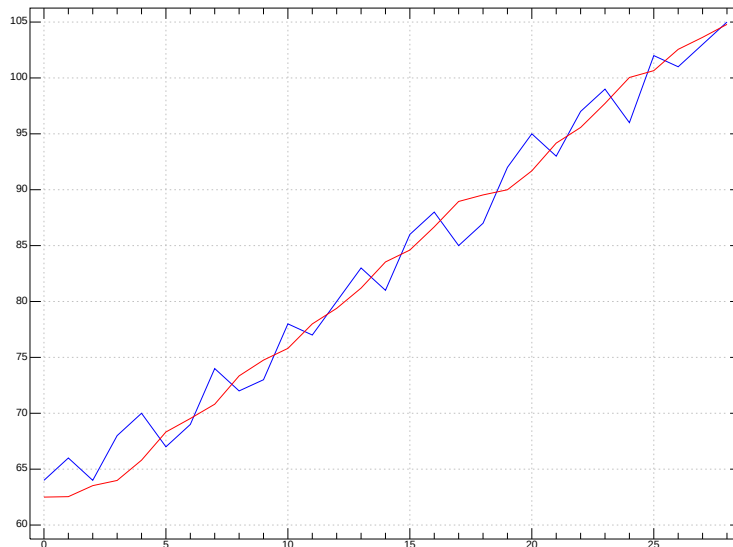


図 5.5 holt's 2 parameter trend model

残差分析の自動計算

残差計算が大変なので、1に近いパラメーターを与えると、0まで0.05ピッチで0までぐるぐる回し、最少値を選択して、ホルツ法を計算するルーチンを作った。結果は降順にソートしてあるので最下行をとればよい。(初期値によって結果が多少異なる)

```
0.2 0.8 holt_loop DN54
```

```
0.2 0.45 2.46177
0.15 0.7 2.41542
0.2 0.5 2.39981
0.15 0.75 2.37576
0.2 0.55 2.35378
0.15 0.8 2.34396
0.2 0.6 2.3198
0.2 0.65 2.29505
0.2 0.7 2.27745
0.2 0.75 2.26552
0.2 0.8 2.25821    NB. minimum residual
```

```
α = 0.2,β = 0.8 が残差が最少

0.2 0.8 holt DN54
data smoothed smoothed fitted
  level      trend
-----
63      63      _0.5      0
64      62.8    _0.26 62.5
66 63.232    0.2936 62.54
64 63.6205 0.369504 63.5256
68 64.792    1.01111 63.99
70 66.6425    1.68261 65.8031
67 68.0601    1.4706 68.3251
69 69.4245    1.38569 69.5307
74 71.4482    1.89606 70.8102
72 73.0754    1.68098 73.3442
73 74.4051    1.39996 74.7564
78 76.244    1.75115 75.8051
77 77.7962    1.59192 77.9952
80 79.5105    1.68983 79.3881
83 81.5602    1.97778 81.2003
81 83.0304    1.5717 83.538
86 84.8817    1.79536 84.6021
88 86.9416    2.00703 86.6771
85 88.1589    1.37525 88.9487
87 89.0273    0.969777 89.5342
92 90.3977    1.29024 89.9971
95 92.3503    1.82017 91.6879
93 93.9364    1.63288 94.1705
97 95.8554    1.8618 95.5693
99 97.9738    2.06704 97.7172
96 99.2327    1.42051 100.041
102 100.923    1.636 100.653
101 102.247    1.38663 102.559
103 103.507    1.28528 103.633
105 104.834    1.31855 104.792
```

```
plot 0 1 3{ |: }. 0.2 0.8 holt DN54
```

指数平滑化	smexp	m smexp n
Holt の 2 段階指数平滑化	holt	m_1 m_2 holt n
	hol_loop	holt_loop n

5.3 季節調整・Seasonal Adjustment

季節調整法は経験の積重ねの中で発展してきたもので、絶対的な理論が存する訳ではない。

Pindyck and Rubinfeld の *Econometric models and Forecast*(1998) に目を通していたら、次のような記述があった。

Seasonal adjustment techniques are basically ad hoc methods of computing seasonal indices.

National economic data in the United States usually are seasonal adjusted by the Census II method, which was developed by the Bureau of the Census of the U.S. Department of Commerce.

The Census II method is a rather detailed and complicated procedure (and is amazingly ad hoc),

各種の移動平均を5回以上駆使する X12 は歌舞伎の隈のような厚塗りであり、大型研磨機で鉛筆を削るようなものである。身軽な方法から始めよう。

Pindic & Rubinfeld は次の簡潔な方法を紹介している。

Additive $Y = L + S + C + I$

Multiplicative $Y = L \times S \times C \times I$

L value of the long-term secular trend in series

S value of seasonal component

C (long term) cyclical component

I irregular component

ここで S を取り除くには、先ず単純移動平均により $L \times C$ を推計し

$$\hat{y} = \frac{1}{12}(y_{t+6} + \cdots y_t + y_{t-1} + \cdots y_{t-5})$$

オリジナルデータを $L \times C$ で割ると $S \times I$ が求められる。

$$\frac{L \times S \times C \times I}{L \times C} = S \times I = \frac{y_t}{\hat{y}} = z_i$$

I を求める

$$\hat{z}_1 = \frac{1}{4}(z_1 + z_{13} + z_{25} + z_{37})$$

$$\hat{z}_2 = \frac{1}{4}(z_2 + z_{14} + z_{26} + z_{38})$$

.....

$$\hat{z}_{12} = \frac{1}{4}(z_{12} + z_{24} + z_{36} + z_{48})$$

trend and cyclical(TC) components are estimated using moving averages

5.3.1 四季の調整

Working Example

DeLurgio の BIGQTR のデータと計算メソッドに従って、季節データを季節変動指数を

算出して、 TCe を求め、更に、時間との単回帰により、トレンド T を求め、グラフに表してみよう。

この例では中心化4項移動平均で S を求めている。データは四季単位である。

adj_4season DN41

Year	Quarter	Sales	Y	deseasonal	Trend	Predict
1	1	72				
1	2	110				
1	3	117				
1	4	172				
2	1	76				
2	2	112				
2	3	130				
2	4	194				
3	1	78				
3	2	119				
3	3	128				
3	4	201				
4	1	81				
4	2	134				
4	3	141				
4	4	216				

4季の変動係数を求める

adj_4season_sub DN41

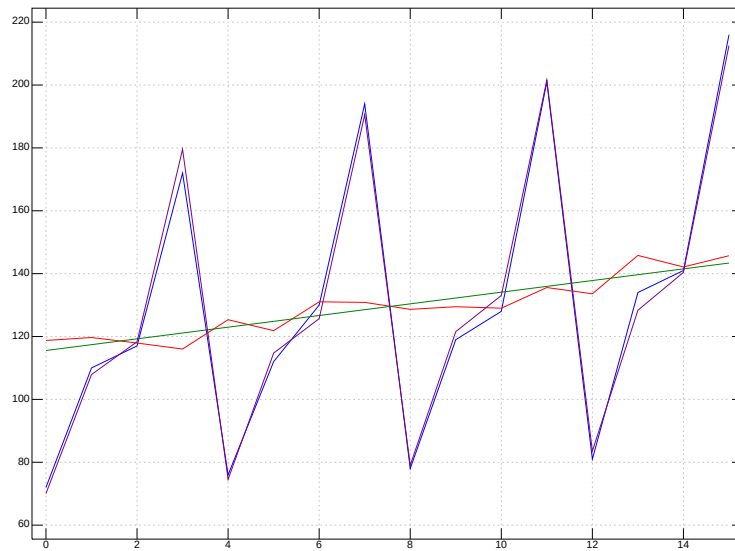
0.606313 0.919069 0.992121 1.4825

原データを変動係数で割ると *deseasonal* が求められる。

Deseasonal を一次の最少自乗法で回帰して *Trend* を求め、これを変動係数で割り戻すと推計値が得られる。

ここではトレンドを1次式で回帰しているが、ここを低次の多項式などに変更すればさらに詳細に分解できる。

plot |: 0 3{"1 adj_4season DN41

図 5.6 季節データと TCe

mav_c	移動平均の標準型。偶数では中心化を行う。	m. mav_c n.
adj_4season adj_4season_sub	4季の変動指数を求め、回帰とBest Fitting	adj_4season n

Script

```

adj_4season_sub=: 3 : 0
NB. find season_index
TMP1=. (2}. _2}. y) ,. 4&mav_c Y0=: (-4|#y)}. y
SEASON0=. 2|. "1 >_4<\ %/"1 TMP1
SEASON1=. +/ SEASON0 % (# TMP1) % 4
SEASON1 * 4 % +/ SEASON1
)

adj_4season=: 3 : 0
NB. using mav_c and reg0
Y0=. (-4|#y)}. y
SEASON=. adj_season_sub Y0
TMP0=. Y0,., (_4<\Y0) % L:0 SEASON NB. const data
f=. reg0 {"1 TMP0 NB. reg /find ternd
TMP1=. TMP0,., (1,.>: i. # TMP0) +/ . * f NB. reg_t use >: i. # y
TMP1,., ;SEASON * (L:0) _4<\ 2{"1 TMP1
)

```

Working Example GDP(1994-2007 Real) の deseasonalize

$LSCI$ から S を除くと LCI が残りこれを deseason とする。ここから 1 次回帰で T を求めると T の周りで CI が変動する。 T を S で割り戻したものが *fitted* である。

右は加重 deseason(red) と 3 項移動平均 (green) との比較。

(1994 1) plot_4season GDP

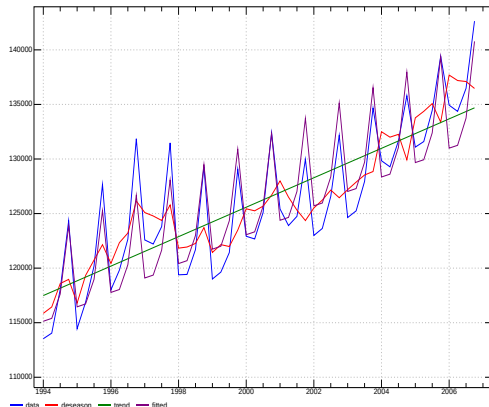


図 5.7 GDP(Japan 1994-2007 real)

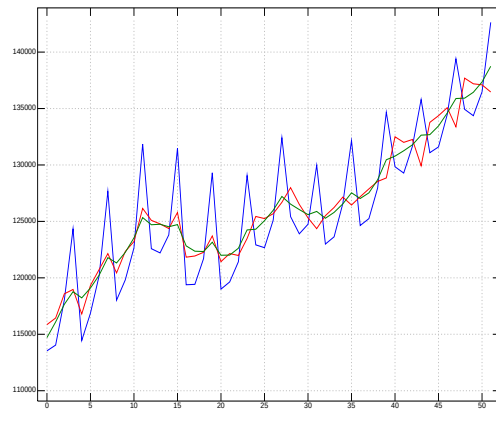


図 5.8 GDP(real)=blue sdeseason=red
wma3=green

5.3.2 12 月の deseason

移動平均法により季節変動指数を求め、季節変動を除去することで、トレンドをもとめる。

季節変動指数

12 月の時系列データが短い時は全期間で季節変動指数を求める。

旧通商産業省の統計マニュアル「指数の作成と利用（1999）」には、直近の 5 年をとり最大値と最小値を除いた 3 年分で、季節変動指数を求める手法が紹介されている。特異値が除かれて便利なので、この方法も紹介する。係数の合計を 12 と微調整して最終の季節変動指数とする。

Working Example Delurgio Seriesf

Delurgio の seriesf のデータ (DN55) とプロセスを再現する。

5j2 " : (T1 % T0)

1.21 1.15 1.02 0.95 0.90 0.81 0.84 0.93 0.95 1.02 1.09 1.14
1.20 1.10 1.04 0.95 0.88 0.82 0.91 0.96 1.02 1.03 1.12 1.14
1.10 1.06 0.96 0.89 0.83 0.91 0.94 0.95 1.03 1.05 1.12 1.19

5j2 " : (+/%#) (T1 % T0) NB. sum each month

1.17 1.10 1.01 0.93 0.87 0.85 0.89 0.95 1.00 1.04 1.11 1.15

```

+ / (+/%#) (T1 % T0) NB. sum whole year
12.0701

```

```

12% + / (+/%#) (T1 % T0) NB. adjust to 12
0.994196

```

```

5j2 ": 6|. tmp* 12% + / (+/%#) (T1 % T0) NB. rotate 6
0.89 0.94 0.99 1.03 1.11 1.15 1.17 1.10 1.00 0.93 0.86 0.84

```

```

5j2 ": saj DN55
0.89 0.94 0.99 1.03 1.11 1.15 1.17 1.10 1.00 0.93 0.86 0.84

```

5.3.3 Trend と OLS による回帰

季節変動指数 (2) が求まると、データから季節変動を除いて取り除き、最少自乗法による単回帰で Trend を求め、季節変動指数を乗じて、Fitted Value を求めている。

```

12{. adj_season DN55
data deseason trend fitted
-----
546 631.847 661.704 571.801
578 615.624 668.9 628.021
660 676.391 676.096 659.712
707 692.655 683.292 697.443
738 670.808 690.488 759.651
781 691.507 697.683 787.975
848 707.979 704.879 844.287
818 732.673 712.075 795.003
729 714.054 719.271 734.325
691 731.976 726.466 685.798
658 747.319 733.662 645.976
604 744.527 740.858 601.024
.....

```

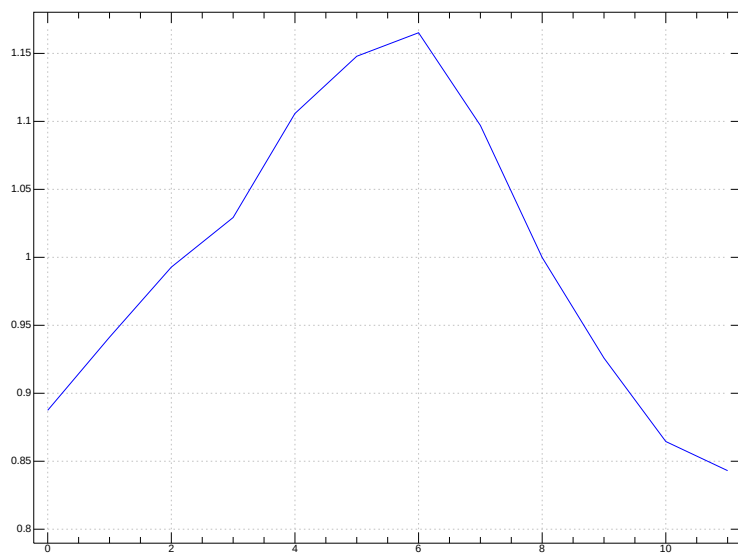
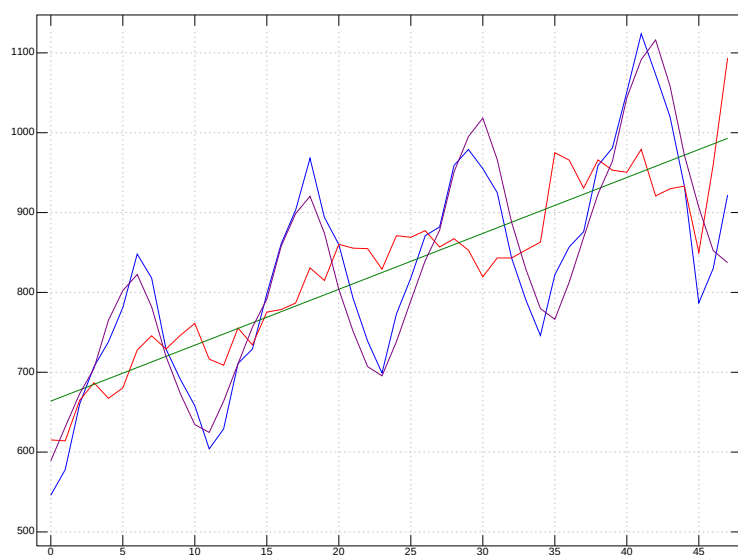


図 5.9 季節変動指数

図 5.10 *Trend, TredCyclee*

Working Example

デパートの売上高（店舗調整なし、1997/2 - 2007/1）の deseason

データの出典 東洋経済統計月報

(1997 2) plot_season ;{."1 DP

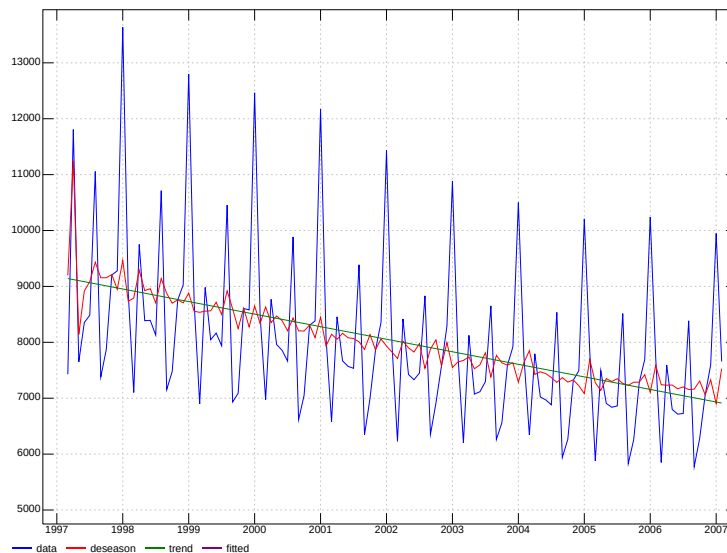


図 5.11 $deseason = red, Trend = green$

saj=:3 : 0

NB. using whole months to calc

NB. last update 31 July 2007

NB. Season Adjustment using Moving Average

NB. Usage: saj y // y is data (ex 5 years=60 months)

T0=: (TMP1=:(<.(>: # TMP0)%12) , 12) \$ TMP0=: mav12 y

T1=: TMP1 \$ 6}. _6}. y

6 |. T3* 12 % +/ T3=:(+/%#) T1%T0

)

adj_season=: 3 : 0

NB. 12 season data //using centered moving average

Y0=. (- (12| # y)) }. y NB. drop out of 4 times from last //if 0 drop nothing

S40=.(6}. _6}. Y0),. mav12 Y0

S41=. 6|. "1 >_12<\ %/ "1 S40 NB. _4 non overlap,2|. is rotate

S42=.+ / S41 % ({. \$ S40) % 12

SEASON=:S42 * 12 %(+ / S42)

NB. Sasonal

```

S44=:Y0,.,>(_12<\Y0)%L:0 SEASON      NB. Deseasonalized: TCe
f=:({:"1 S44)%X=.1,.,>:i.#S44 NB. f with OLS
TREND=:X+/. * f
S44,.,TREND,.,(_12<\TREND)*L:0 SEASON      NB. trend
)

```

参考文献

- Stepaen A. Delurgio Forecasting Principles and Applications McGraw-Hill 1998
 Gareth Janacek Practical Time series Arnold 2001
 R.S.Pingyck D.L.Rubinfeld Economic Models and Economic Forecasts (4th.edition)
 McGraw-Hill 1998
 C.Reiter [Fractal Visualization ans J 2nd ed.]
 鈴木義一郎 J言語による統計分析 森北出版 1996

5.4 関数一覧

<i>mav</i>	Moving Average	m mav n
<i>mav_c</i>	Centred mav auto classified	m mav_c n
<i>mav12</i>	12 month mav centred	mav12 n
<i>spencer</i>	spencer's 15 point moving average	spencer n
<i>henderson</i>	henderson n point moving average	m (次数) hender- son n 次数は奇数を指定

<i>plot_mav12</i>	12 months mav	plot_mav12 n
<i>plot_mav12cpl</i>	12 months mav	plot_mav12cpl n
<i>plot_w13</i>	13 weighted mav and mav12cpl	plot_w13 n
<i>plot_spencer</i>	spencer	plot_spencer n
<i>plot_henderson</i>	henderson 15 項	plot_henderson n
<i>plot_henderson2</i>	henderson 任意の項 and spencer	m. plot_henderson2 n
<i>plot_comb_all</i>	henderson 任意の項と and spencer ,mav12cpl	m. plot_comb_all n henderson は次数は奇数のみ 31 33 付近は中心化 1 2 月移動 平均と類似 両端は計算方法の 差によるばらつきがある。

<i>saj</i>	季節変動指数を求める.whole months	saj n.(List)
<i>saj2</i>	季節変動指数を求める。 last 3 years	saj2 n.(List),usualy using saj2
<i>adj_season</i>	季節変動指数を求め、回帰 と Best Fitting を行う	adj_season n.(List)
<i>adj_season2</i>	季節変動指数を求め、回帰 と Best Fitting を行う.saj2	adj_season2 n.(List),usualy using adj_season2
<i>plot_adj_season</i>		

索引

mav, 135

p value, 53

Iverson, 5, 13, 14

一般化最少自乗法, 67, 68
移動平均, 117–119, 123, 128, 131

AIC, 19, 21, 71, 72, 76, 77
X-10, 119
X-12, 121

加重移動平均, 119
加重最少自乗法, 69, 101

幾何平均, 25, 28, 41
季節変動指数, 128, 131, 132, 137
逆行列, 9, 14, 19, 20, 39

コ克蘭オーカット法, 65

最少自乗法, 43, 44, 67, 129, 132
算術平均, 25, 41
散布図, 10, 16

時系列解析, 19, 34
自己回帰, 9, 13, 19
指数回帰, 89
指数平滑化, 124, 125, 127
自然対数, 34, 35
ジニ, 112–114, 116
重回帰, 13, 14, 22, 34, 43, 61
定数項, 14

スペンサー 1 5 ポイント移動平均, 119

相関行列, 36–39

対角行列, 39
対数, 34, 35, 78, 79, 103
多項式, 9, 13, 17, 19, 22, 71, 72, 74, 76, 77, 85, 116, 129
多重共線性, 65

中心化, 118
調和平均, 26, 41

トレンド項, 78

Newton, 86–88

非線形回帰, 87
標準化, 32, 33, 38, 40

不均一分散, 34, 101, 104
分散共分散行列, 36, 37, 41

偏相関行列, 39–41
Henderson, 122

ラスパイレス, 30, 31

Logit, 101
ロレンツ, 112