## 0.13 ジュニア版　　線形回帰とその形状

前章で使った多変数の線形回帰の係数を求める **b=inv(X'X)*X'y** という式を用いて、実際にいろいろな形と意味をもった線形回帰を見ていきます。とりあえず、ここでは式のあてはまりや係数が妥当であるかどうかなどの統計学上の細かい分析は考えないこととします。

**時間がＸ軸のケース**

今、ｙに相当する側にデータが列として設定されているものとしよう。ここで、現在の時間や年月を０として、以降１ずつ時間が増していくものとしよう。例えば、１年後、２年後...というふうにである。このｙのデータを時間に対して直線のトレンドを推定してみよう。この際、前章の最後に行なったように１ばかりからなる列をＸ側にマージすることによって定数項を含んだ２変数を、ｙに対して線形回帰して係数を推定してみます。

プログラム

```
new;
cls;
y={3,5,6,8,9,12,13,14,16,18};
c={1,1,1,1,1,1,1,1,1,1};
t={0,1,2,3,4,5,6,7,8,9};
X=c~t;
b=inv(X'X)*X'y;
yhat=X*b;
print "b=" b;

library pgraph;
graphset;
_plctrl={0,-1};
xlabel("t");
xy(t,yhat~y);
```

プログラム各行の説明

１行目　メモリを初期化する。
２行目　画面をクリアする。
３行目　変数ｙに 10×１のデータを代入する（GAUSS ではカンマまでが１行)。
４行目　変数ｃに 10×１の１ばかりのデータを代入する。
５行目　変数ｔに 10×１のデータを代入する。ここではｔは０から始まっている。
６行目　変数Ｘにｃとｔを水平方向に～でマージしたものを代入する。
７行目　線形回帰の式 inv(X'X)*X'y を計算して変数ｂに代入する。

８行目　Ｘ*b を計算してそれをｙの推定値として変数 yhat に代入する。
９行目　計算された結果ｂの内容を b=というメッセージとともに画面表示する。
11 行目　グラフを描く機能の入った pgraph というライブラリを呼び出す
12 行目　ライブラリ pgraph の各種設定のグローバル変数を初期化する
13 行目　グラフのプロットを結ぶ線を設定するグローバル変数_plctrl の中に、第１番目のＸの値にはデフォルトの０（線で結ぶ)、第２番目の値には-1 で１つごとにプロットするが線で結ばないという意味の数字を入れる（マイナスは線で結ばない)。
14 行目　Ｘ軸のラベルにｔというメッセージを入れる。
15 行目　Ｘ軸方向の値に変数ｔの列を、ｙ軸方向には yhat およびｙの２列を設定してｘｙグラフで描く。共通の値ｔに対して、yhat とｙの２つのグラフが自動的に色分けされて描かれる。

画面表示

```
b=
        3.0363636
        1.6363636
```

すなわち、ここでは定数項が最初の係数、時間に対する項が２番目の係数だから、

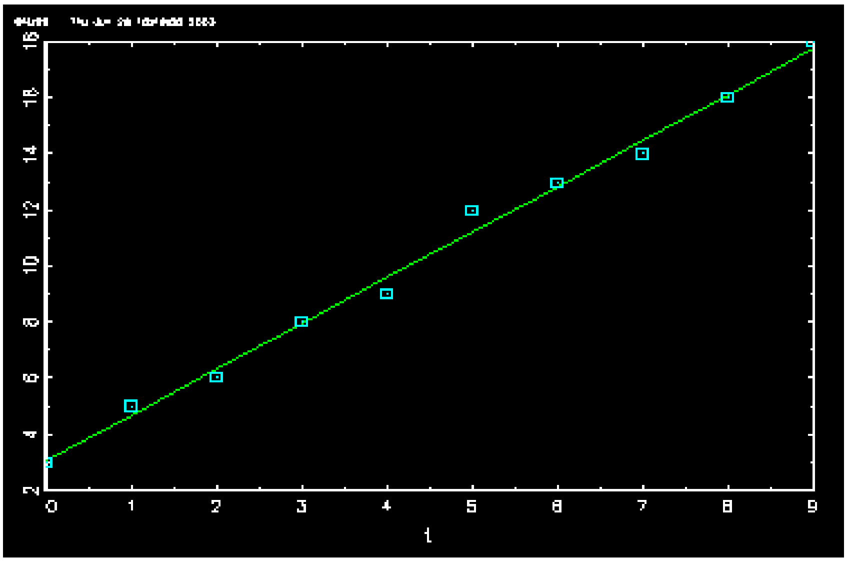
$$\underbrace{y=3.0363636}_{\text{ｙ切片}}+\underbrace{1.6363636\,t}_{\text{傾き}\times t} \qquad t=0,1,2,3,\ldots,9$$

という関係が成り立っていることを示していて、その定数項がｂの第１番目の 3.0363636 であり、時間ｔに対する係数が 1.6363636 である。ｔに０から順番に数字を入れていくとその時間のおおよそのｙの値がわかる。また、その実際のｙの推定値は行列とベクトルの積で表現すると、

yhat＝Ｘ*b

になる。これは、ｂが3.0363636および1.6363636のときのｙ＝3.0363636＋1.6363636ｔを計算した結果に等しい。それをグラフで実線で表し、もともとのｙの値を四角の点で表すと次のようなグラフになる。

グラフ表示



グラフの実線は推定値を結んだもので、直線になっている。この直線になるというのが線形回帰の意味である。時間ｔが０の時（現在）の値はおおよそ３になっている。これが上の計算で求めた定数項の 3.0363636 に相当する。これはグラフでいうとｙ切片でもある。他方、この推定線の傾きが上の計算の 2 番目の係数である 1.6363636 であることは容易に想像することができるだろう。時間ｔが１増えるごとにｙの推定値も 1.6363636 ずつ増えるということである。これは、物理でいうところの速度（スピード）と考えることも可能であろう。このような実線の時間に対するｙのトレンドに対して、四角の点でもともとのデータをプロットしている。実線のトレンドは、このもともとのデータのプロットのほぼ中央を貫くように描かれている。

**Ｘ側が多変数のケース**

上では、Ｘ軸に時間ｔをとって考えたが、これは時間である必要は無論ない。また、定数項以外にＸに相当する側の変数がいくつあっても **b=inv(X'X)*X'y** という式は適用できる。いま、ｙに対して、Ｘ側は定数項を求める１ばかりの列に加えてｘ１というデータとｘ２というデータの２つの列が与えられているものとする。この場合の線形回帰を求める。

プログラム

```
new;
cls;
y={3,5,6,8,9};
c={1,1,1,1,1};
x1={1,2,4,5,6};
x2={11,9,7,5,4};
X=c~x1~x2;
```

```
b=inv(X'X)*X'y;
yhat=X*b;
print "b=" b;
print "yhat=" yhat;
```

画面表示

```
b=
      19.000000
    -0.77777778
     -1.3888889
yhat=
      2.9444444
      4.9444444
      6.1666667
      8.1666667
      8.7777778
```

すなわち、この結果は、

$$\mathrm{y}=19.000000-0.77777778\,x_1-1.3888889x_2$$

ということであって、ｘ1 およびｘ2 がともに０であるときにｙは 19 になるということを表しており（時間ｔに対する回帰の際、時間０に対するｙ切片がおよそ３であったことを思い出そう)、ｘ2 が一定で変わらないときｘ1 を１増やすとｙは 0.77777778 だけ減ることを意味しており、またｘ1 が一定で変わらないときｘ2 を１増やすとｙは 1.3888889 だけ減ることを意味している。この場合は定数項とｘ１とｘ２の２変数をＸ側においてｙに対して線形回帰しているが、３変数以上も同様に計算できる。つまり、x3 や x4 などを列で設定しておいて、

```
X=c~x1~x2~x3~x4;
```

などというふうに水平方向のマージを表す～でつなぎ合わせて列を束ねて行列Ｘにするだけである。定数項がない線形回帰をするにはｃの部分をはずせばよい。なお、それぞれのデータ数がｎ個である場合には、ｃの代わりに１の行列を作る組込み関数 ones を利用して

```
X=ones(n,1)~x1~x2~x3~x4;
```

というふうに定数項つきの線形回帰を設定すれば、わざわざ１ばかりからなる列をつくる必要がなくなる。ｘ軸とｙ軸から表現できる２次元では線形とは傾向線（時間に対する場合はトレンド）であることはわかるが、それ以上の変数の個数に対する多次元では線形とは、それぞれの変数に対する平面の傾向のようなもの言える。また、Ｘ側の変数すべてを０としたときのｙの値が定数項である。

Ｘ側が２次形式のケース

　これまでは、Ｘ変数側に時間ｔや変数ｘをおいて定数項を計算する１の列とマージして全体を行列Ｘとして b=inv(X'X)*X'y という式を用いて、係数を求めました。Ｘが定数項部分も含めて２列であれば、ｂも２つの要素（2 行）からなります。Ｘの設定が３列であればｂも３つの要素からなります。すなわち、Ｘの列数分だけそれに対応する係数がｂに計算されます。このことは、行列Ｘのそれぞれの列を示すものが必ずしも時間ｔ自身やそれぞれのｘ自身である必要はないことは容易に想像できるでしょう。ここでは、定数項の１の列に加えて、０から１ずつ増える時間ｔの列、それにその時間ｔの２乗の列を水平方向にマージして、その全体を行列Ｘとしましょう。それをｙに対して線形回帰した係数ｂ（すなわち、３つの係数からなるはず）を求めます。

プログラム

```
new;
cls;
y={0,9,16,21,23,25,24,20,18,9};
c={1,1,1,1,1,1,1,1,1,1};
t={0,1,2,3,4,5,6,7,8,9};
X=c~t~t^2;
b=inv(X'X)*X'y;
yhat=X*b;
print "b=" b;

library pgraph;
graphset;
_plctrl={0,-1};
xlabel("t");
xy(t,yhat~y);
```

画面表示

```
b=
       0.13636364
        9.7537879
      -0.96590909
```

すなわち、この係数結果は、ｙと時間ｔの関係が

$$y = 0.13636364 + 9.7537879\,t - 0.96590909\,t^2$$

であることを意味しています。最初の係数は定数項、第２番目の係数は時間ｔに対するもの、そしてここで注目すべきは第３番目の係数−0.96590909 というのが時間ｔの２乗した

変数に対するものであることです。上の式は、おおよそのところ

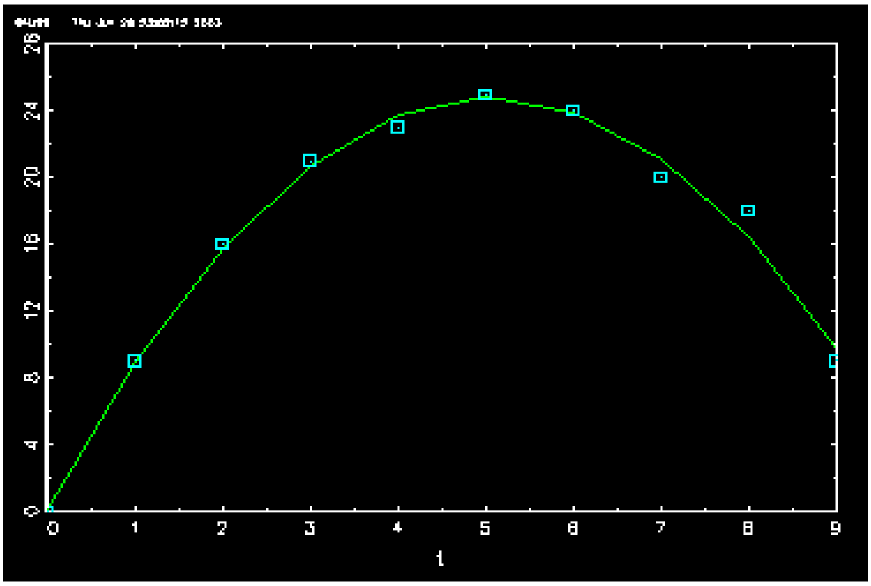
$$y = 0 + 10t - 1t^2$$

ですから、簡単な変形をしてやると

$$y = -t(t - 10)$$

という０と 10 をｘ切片にもつ山のかたちをした２次関数であることがわかります。

グラフ表示




このことは、ｔに０を入れてやるとｙはおおよそ０になり、ｔに 10 を入れてやるとｙは（グラフでは途切れていますが）おおよそ 10 になることを意味します。この場合は、たまたま偶然に定数項はほぼ０になっています。行列Ｘ側の作り方は、定数項、ｔ、それにｔの２乗の順番にしてありますが、通常の代数と同じように次数の大きいものから、ｔの２乗、それからｔ、最後に定数項の順にしても、対応する係数が逆の順番に計算されて出てくるだけです。これらは、まったく同じことです。変数自体がすでに２乗や３乗などになっているかぎりは、１つの変数と見なせますから、全体としても多変数の線形回帰と考えることが可能なのです。上のグラフのような形になる意味というのは、時間ｔが増すにつれてｙの増加のスピードが減少していき、ｔが５あたりでピークに達して、その後は時間ｔとともにｙは減少に転じ、その減少スピードはさらに増していくような形になっています。このことは、$y = -t(t - 10)$のｔ関する微分（すなわちスピード）$dy/dt$ が$-2t + 10$ となって、ｔが５までは正の値をとり、ｔが５のとき０、その後ｔが大きくなるにつれて負の値が増していくことからもわかります。さらに、ｔに関してもう１回微分をしてやると$-2$となって、時間ｔが増すにつれてこのスピードはコンスタントに２ずつ減少していくことがわかります。すなわち、物理でいうところの加速度は$-2$となっています。

X側が単純な原点を通る２次形式のケース

　いま、X側に考える変数を時間ではなくて何かの値であるものとします。また、考える２次形式はほぼ原点を通り（xが０に対してyも０になるということ）xが増えるとともにyもどんどんと大きくなるという２次の形式を考えてみましょう。

プログラム

```
new;
cls;
y={0,2,10,16,30,48,70,100,130,165};
x1={0,1,2,3,4,5,6,7,8,9};
X=x1^2;
b=inv(X'X)*X'y;
yhat=X*b;
print "b=" b;

library pgraph;
graphset;
_plctrl={0,-1};
xlabel("x");
xy(x1,yhat~y);
```

画面表示

```
b=          2.0198917
```

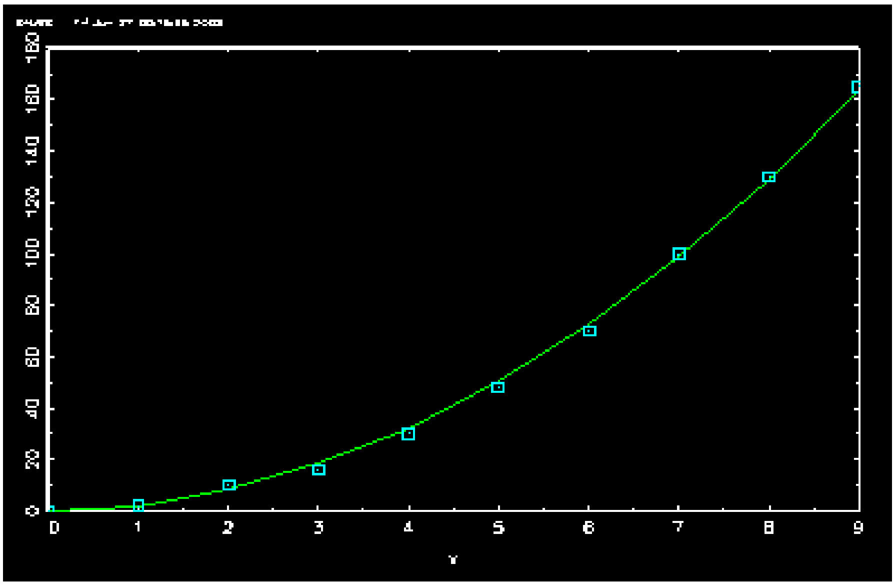
この結果は、X側に x の２乗を定数項なしで単独で用いていることから、１つだけの係数がでてきており、

$$y = 2.0198917x^2$$

という、おおよそ$y = 2x^2$というおわん型のパラボラのxが正の部分の部分を表しています。xが増すにつれて、ますますyの増加も増す形となっています。実際、xに関する微分 $dy/dx = 2x$であって、このyの増加率はxが大きい値ほど大きい値になっています、また、xに関してもう１回微分をすると２でコンスタントに２ずつこの増加率は増していることがわかります。上の例は、原点を通るケースですが、同じパラボラで上か下に移動したものは定数項の１の列をつけて線形回帰すればその平行移動分が計算できます。また、平行移動を考えるにはさらにxそのものを時間tの２次形式のケースと同じように加えてやれば、一般的な２次形式が考えられます。時間tの２次形式のケースでは、山のかたちをしていましたが、これはxの２乗にたいする係数がマイナスの値であったためです。もし、ここでxの２乗に対する係数が正であれば、逆にそれをひっくり返したおわん型の係数が推定できるわけです。どちらも一般的な２次形式ですから、結果として２乗の項に対

する係数がマイナスの場合は山のかたち、プラスの場合はおわん型となるだけです。なお、もし一般的な２次形式に対する係数を求める際に、２乗の項に対する係数がほぼ０となるようであれば、一般的な定数項＋ｘの線形回帰を行なってみることもできるでしょう（とりあえず、ここでは、どちらが統計学上あてはまりがよいのかは問わないこととします）。

グラフ表示




**自然対数とｅとの関係**

通常、計量経済学や経済学でいう log というのは自然対数 ln のことを、ほとんどの場合、意味します。また、初等代数では次のような関係が成り立つことがわかっています。

$$\ln e = 1$$

$$\ln e^{a} = a \ln e = a$$

$$\ln B^{a} = a \ln B$$

$$\ln(BC) = \ln B + \ln C$$

$$\ln(B/C) = \ln B - \ln C$$

$$\ln(B^{a}C^{d}) = a \ln B + d \ln C$$

**セミログの時間に関するケース**

Ｘ側の変数には２次や３次などの変形を加えた変数をそれぞれ１変数とみなして線形回帰ができることがわかったと思います。これは、その変数が変数の自然対数であろうと、またその他の関数形であろうと同じことができます。もちろん、ｙ変数自体を変形した変数を用いて線形回帰することも可能です。ここでは、まずｙ変数だけを自然対数変換したものを使って、それに対して定数項つきの通常の時間ｔに対する線形回帰を行なってみましょう。今、もともとのｙ変数には０やマイナスの数は含まれていないものとします。な

ぜならログは正の数に対して変換ができるからです。

プログラム

```
new;
cls;
y1={6,8,10,16,30,40,70,100,130,200};
t={0,1,2,3,4,5,6,7,8,9};
c={1,1,1,1,1,1,1,1,1,1};
y=ln(y1);
X=c~t;
b=inv(X'X)*X'y;
yhat=X*b;
print "b=" b;

library pgraph;
graphset;
_plctrl={0,-1};
xlabel("t");
ylabel("lny");
xy(t,yhat~y);
```

画面表示

```
b=
       1.6700333
      0.40790302
```

この結果は、左側のｙの変数だけをログ（自然対数）変換したものを、定数項つきで時間ｔに対して線形回帰していますから

$$\ln y = 1.6700333 + 0.40790302\,t$$

ということを表しています。これはどういうことを意味しているのでしょうか？ぱっと見ただけではわかりません。しかしながら、自然対数とｅの関係を知っていれば、

$$y = e^{1.6700333+0.40790302\,t} = e^{1.6700333}\,e^{0.40790302\,t}$$

とこの式を容易に変形し直すことができるでしょう。確かめるには、逆方向に下の式の両辺をログ変換したものが上の式になることが容易にわかると思います。また、ｅの計算ではその部分の足し算は分解するとｅどおしの掛け算になります。すなわち、$e^{a+b} = e^{a}e^{b}$ であることが言えることからわかると思います。ただし、$e^{1.6700333}$ の部分は変数が含まれない何かの数ですから、全体を定数ｃと考えることができます。すなわち、

$$y = e^{1.6700333}\,e^{0.40790302\,t} = c\,e^{0.40790302\,t}$$

ということです。あらためて、この両辺のログ変換をし直してみると

$$\ln y = \ln c e^{0.40790302t} = \ln c + \ln e^{0.40790302t} = 定数 + 0.40790302t$$

となります。ここで $\ln c$ も単なる定数と考えることができますので、この部分はあえて気にせずに、通常の定数項つきの回帰分析をＸ側にｔをとってやった結果であるということがわかります。すなわち、

$$\ln y = 定数 + 0.40790302t$$

というふうに、左側のｙはログ変換した変数を、右側のＸ変数側は定数項つきの１ばかりの列に時間ｔをマージしたものの全体を行列Ｘとして回帰分析したものであるのです。右側の変数ｔにはログ変換がなされていないことに注意してください。ここで少し面倒なのですが、両辺を時間ｔに関して微分してみましょう。左辺は、$\ln A$ のＡに関する微分は 1/A であることがわかっています。また、この場合、ｙ自身による微分ではなくてｔに関する微分なので、関数の関数の微分をルールを用いて、この 1/y に dy/dt というｔに関する微分を掛け合わせたものが左辺になることがわかります。右辺は 0.40790302 そのものになります。つまり、

$$(dy/dt)/y = 0.40790302$$

ということが言えるのです。これはｙの成長率が 0.40790302 ということを意味しています。自然界や人の営みには成長率が一定で指数関数的に増加していくものが数多く見られます。このｙ側だけにログ変換を行なうセミログは、成長率が一定であるということを前提とした「線形回帰」であるということができます。

なお、この線形回帰は、

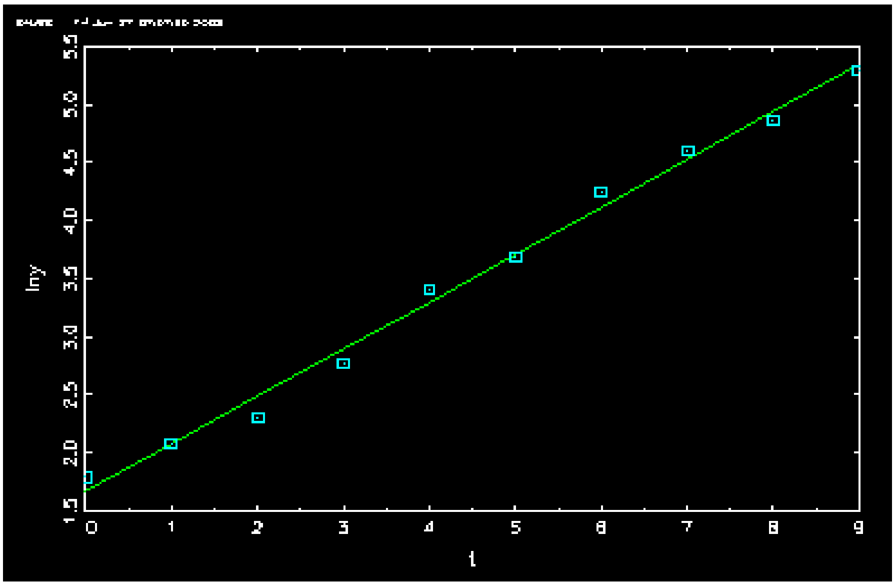
$$y = c e^{gt}$$

の両辺にログをとったものと考えることができます。指数的に増加していくというのは $e^{gt}$ の部分からもわかるでしょう。成長率ｇで増加していっているのです。おもしろいことに

$$y = c e^{rt}$$

という具合に成長率ｇをｒに代えれば、ｃが元本でｒが利子率の連続複利の場合の受取額がｙということになります。上と同じように、$\ln y$ と定数つきのｔについて線形回帰をすればｒは計算できます（ただし、この場合にはｙも yhat もコンピュータ上の計算誤差を除いて正確に一致します）。

グラフ表示




上のグラフでは、実線の傾きは計算結果の 0.40790302 にあたります。すなわち、yは 40%の割合で一定に成長していっているということが言えます。このことはログ変換をした変数yが時間tが１増えるごとに、おおよそ 0.4 増えることの言い換えです。なお、もう１つの定数に対応する係数 1.6700333 はグラフの ln yに対する切片です。なお、実際のyに対してグラフを描いたときの理想上のy切片は、自然対数とｅの関係から、 $e^{1.6700333}$ であることは容易に確かめることができます。

**ログログのケース**

　例えば、一番簡単なケースで、データyとkの列が与えられていて、$y=k^{\alpha}$ というような関係になるαを求めたいとします。これには、これまでと同様に両辺の自然対数をとり式を変形します（ただし両辺ともに正の数である必要があります)。そうすると、

$$\ln y=\alpha \ln k$$

というような関係式になります。y変数側にyのログ変換した値を、X側に定数項をつけないでログ変換したkの値だけを考えて、単純な線形回帰をします。

プログラム

```
new;
cls;
y1={1,3,10,17,24,38,50,63,80};
k={1,2,3,4,5,6,7,8,9};
y=ln(y1);
X=ln(k);
b=inv(X'X)*X'y;
```

```
print "b=" b;
```

画面表示

```
b=          2.0004847
```

上では、ｙがｋ$^{\alpha}$になるようになるようなαを求めていますが、もし指数的に増加するｙ、すなわちｅ$^{y}$がｋ$^{\alpha}$になるようなαを求めたい場合には、ｙ側はｙのままで、Ｘ側が定数項なしのログ変換されたｋがくるセミログになります。その場合のセミログは、成長率を求めるときのセミログの設定とは全く逆の側にログ変換をほどこすものになることに注意してください。

もう一度、ログログに戻ってもう少し複雑なケースに取り組んでみましょう。ここで

$$Q = A L^{\alpha} K^{\beta}$$

というような、例えばＬを労働の時間などの単位、Ｋをかけるお金などの資本の単位であるとすると、２つの指数関数の積からなるような関係式で生産される量Ｑが決まる。

プログラム

```
new;
cls;
let data[8,3]=
6.8385212   3   9.1
2.0000000   2   2.1
4.2294851   5   4.0
4.7287080   4   5.5
3.5676213   6   3.2
6.2357393   7   6.1
7.2376242   8   6.8
1.0000000   1   1.0
;
Q=data[.,1]; L=data[.,2]; K=data[.,3];
y=ln(Q);
X=ones(rows(data),1)~ln(L)~ln(K);
b=inv(X'X)*X'y;
print "b=" b;
```

画面表示

```
b=
     -0.022656379
        0.25064944
        0.75173301
```

上のプログラムではdataという変数にコピー＆ペーストで３列のデータを貼り付けている。御自分で何か他の実際のデータを張り付けてもよいだろう。その場合には、data の配列の縦と横の数を正確に変更する必要がある。１列目は生産量Q、２列目は労働量L、３列目は資本の量Kに対応するデータがあるものとする。ここで、 $Q=L^{\alpha}K^{\beta}$は両辺が正の数であるから、両辺の自然対数をとっても等号関係は変更されない。したがって、

$$
\begin{aligned}
\ln Q &= \ln A L^{\alpha} K^{\beta} \\
&= \ln A + \ln L^{\alpha} + \ln K^{\beta} \\
&= \text{定数} + \alpha \ln L + \beta \ln K
\end{aligned}
$$

と表すことができる。このケースも $\ln L$ および $\ln K$ をそれぞれ１つの変数とみなして定数項つきで、$\ln Q$ に対して線形回帰すると定数項と係数 $\alpha$ と $\beta$ が求まる。この場合、おおよそ

$$\ln Q = 0 + 0.25 \ln L + 0.75 \ln K$$

であると言える。すなわち、もとのログ変換をする前の式で考えれば、$e^0=1$ であることに注意すると $A=1$ であるから、

$$Q = L^{0.25} K^{0.75}$$

であることを意味している。

**弾力性の計算について**

一つ前に取り上げたログログのケースを利用して、今度は定数項をつけて線形回帰することによって、その係数を調べることによって、弾力性という概念を計算してみよう。まずは、代数的にどのようなことが言えるかはさておいて、とりあえず計算をすることにしよう。いま、$\ln y = \text{定数} + \alpha \ln x$ の $\alpha$ を定数項つき計算で求めることにする。

プログラム

```
new;
cls;
y1={3.00,3.17,3.21,3.28,3.30,3.33,3.34,3.11,3.25};
x1={1,3,4,6,7,8,9,2,5};
c={1,1,1,1,1,1,1,1,1};
y=ln(y1);
X=c~ln(x1);
b=inv(X'X)*X'y;
print "b=" b;
```

画面表示

b=

1.0994052

0.049051974

これは、$\ln y = 1.0994052 + 0.049051974 \ln x$ が言えるということであって、さらにこの２番目の係数（ほぼ 0.05）はｘが所得であればｙの所得弾力性の値であり、ｘが価格であればｙの価格弾力性であると言える。この場合の弾力性とは「ｘが１％増加した際のｙの増加のパーセント値」ということである。この場合、ｘが１％増加すればｙは約 0.05%増加するという状態を表している。

代数的に簡単に説明をすると、

$$\ln y = c + \alpha \ln x$$

を対数をとるときのように d という微分の記号をとることにしよう。これはその瞬間の変分を表す。

$$d \ln y = d(c + \alpha \ln x)$$

ここで、$\ln A$ のＡに関する微分が 1/A になることを思い出そう。すなわち、

$$d \ln y / dy = 1/y \quad \text{および} \quad d \ln x / dx = 1/x$$

というふうに逆数の形になる。いま、その分母のところを右辺に移行すると

$$d \ln y = dy/y \quad \text{および} \quad d \ln x = dx/x$$

と言えることは容易にわかるであろう。これを両辺のｄをとった $d \ln y = d(c + \alpha \ln x)$ にあてはめてみる。なお、定数ｃの変分はコンスタントの変分は０であり、αは変分の外側に何倍という形でくくり出せるものとすると、結局

$$d \ln y = \alpha\, d \ln x$$

$$\Rightarrow \quad dy/y = \alpha \ dx/x$$

$$\Rightarrow \quad \alpha = (dy/y)/(dx/x)$$

すなわち、αは「ｘのパーセント変化分のｙのパーセント変化」であると言えるのである。これを経済学では弾力性という。

もう 1 回、$\ln y = c + \alpha \ln x$ の式に戻ると、このαを線形回帰で推定していることは、とりもなおさず「ｘのパーセント変化分のｙのパーセント変化」の値、すなわち「弾力性」を求めていることにほかならないのである。

ここで、１つ前の $\ln Q = 0.25 \ln L + 0.75 \ln K$ で表された線形回帰の結果を思い出そう。この式に同じ弾力性の概念とその計算法をあてはめてみれば、$\ln L$ の係数 0.25 は「Ｌが１％変化したときのＱのパーセント変化」が 0.25%であることを意味している。同様にして、$\ln K$ の係数 0.75 は「Ｋが１％変化したときのＱのパーセント変化」が 0.75%であることを意味している。なお、この 0.25 と 0.75 の係数の部分の和が１であるときには、これらは労働と資本の「分配率」であることは、経済学の基礎中の基礎である。どうして、そうなるかは、もともとの Q の式をＬで偏微分したものを賃金ｗと置いて、その係数イコールの式

に移項して直すことによりその係数が何を表しているかを御自分で確かめてみよう。また、Ｑの式をＫで偏微分したものをｒとおいて、そうようにその係数が何を表しているかを確かめてみよう。

注意事項

1）「係数に関して」線形である（$\alpha x_1 + \beta x_2 + \ldots$）などとなるケース、または $x_1, x_2, \ldots$ に相当する部分に変換された変数がまとまりとして考えられる場合、ログ変換などで係数に関して線形に変換できる場合、それらはすべて線形回帰で分析できる。

2）ログ変換する際のもともとの変数は正の値である必要がある。０であってはいけない。同様に、平方根変換するケースも、もともとの数は非負の数である必要がある。

3）「係数に関して」どのようにしても線形化できないモデル（例えば、ロジスティック関数や複雑な三角関数のモデル）は線形回帰することができない。その場合には別の方法で係数を求めることになる（本編で取り上げる残差平方和の最小化や Likelihood の最大化による）。

4）ｘ軸に時間ｔをとって、１次や２次もしくはそれ以上のモデルについて線形回帰する場合、ｔ＝0,1,2,..と０から始めるとｙ切片はｔ＝０のときの値を示している。時間ｔを実際の値、例えばｔ＝1950,1951,1952,…と０以外から始めるとｙ切片はｔ＝０のとき、すなわちキリスト誕生の年の値を示す。ただし、両者ともに傾き自体は同一。

**練習問題**

【問１】x={3.2,1.0,3.9,2.1,4.9,6.0,9.1,8.0}; y={4.5,3.0,5.1,3.8,5.4,4.9,7.1,6.7};に対して

$$y = \alpha + \beta\sqrt{x}$$

というモデルが考えられるとする。係数αとβを線形回帰で求めよ。また、そのような係数関係のとき、ｘとｙには言葉で表すとどういう関係が成立しているか説明してみよう。（ヒント：ｘを組込み関数 sqrt で変換したものを１つの変数として、１ばかりからなる列とマージすることによって、その２列全体をＸとする）

【問２】x={0,1,2,3,4,5,6,7,8,9}; y={29,50,54,45,30,15,6,9,30,76};に対して、一般的な３次方程式で表されるモデルが適用できるものとする。その場合の係数を線形回帰で求めよ。その推定値はどのような形状になっているのかグラフで示そう。（ヒント：$y = \alpha + \beta x + \gamma x^2 + \delta x^3$という関係で表される係数αβγδの４つのパラメータを求める。推定値はＸ*bで表される。GAUSSにおいては小文字ｘと大文字Ｘには区別がないのでプログラムの際は、例えばxをx1とするなど置き換えが必要）