Scilabによる統計処理

１．基本統計量・・・合計，最大，最小，平均，メディアン，分散，標準偏差は次の関数で求めることができる。

a=[1,3,6,8,7]

sum(a)	合計
mean(a)	平均
median(a)	メディアン（中央値）
stdev(a)	標準偏差　（平均からのズレの二乗の平均の平方根）
max(a)	最大
min(a)	最小
norm(a)	ノルム（二乗和）

b=[3,5,-1,6,9]

２のベクトル（配列）との相関係数ｒを求めるには

ｒ＝（a*bの平均－aの平均*bの平均）/aの標準偏差 /bの標準偏差

であり a*b の平均はａ とｂとの積和を次元数（length(a)）で割ったものであるから

a*b'/length(a)

で計算できる。ただしｂ'はｂの転置である。

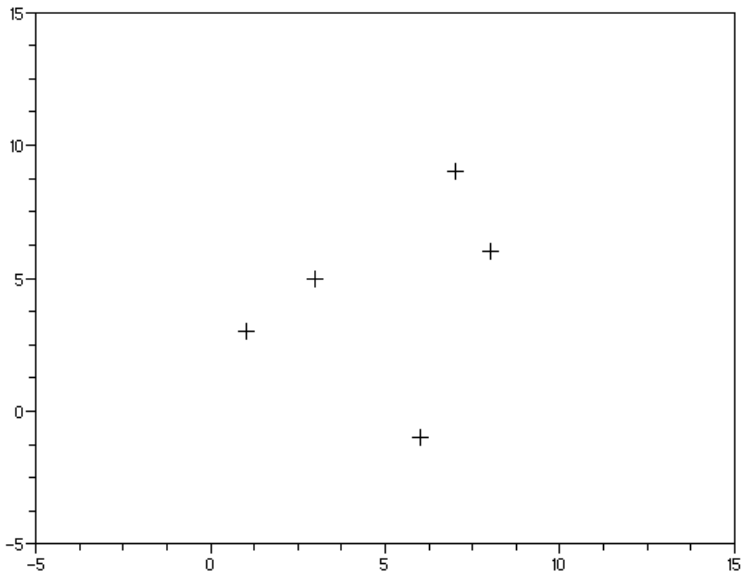
```
r=(a*b'/length(a)-mean(a)*mean(b))/stdev(a)/stdev(b);
```

散布図・・・点列(a,b)を描くことができる。

```
plot2d(a,b,style=-1,rect=[-5,-5,15,15]);
```

styleの値が負のとき点のみを，正のときは折れ線を描き，その数値は色を表す。

rect＝[x1,y1,x2,y2] は描画範囲が左下の点 (x1,y1) 右上の点(x2,y2)であることを表す。



2. データの与え方・・・テキストファイルの読み込み

数値・文字を含んだテキストファイルを読み込んで配列に収めるには **read** 文を使用する。

読み込むべきテキストファイル d.txt に 2 行 3 列の行列が書かれているとする。

1 ,2 ,3
10,20,30

これを読み込ませるにはファイル名はフルパスで書く

a=**read**('D:¥castor¥Saku¥num¥scilab¥ST1¥d.txt',2,3)

（行数が不明なときは明記せずに－１で可）。

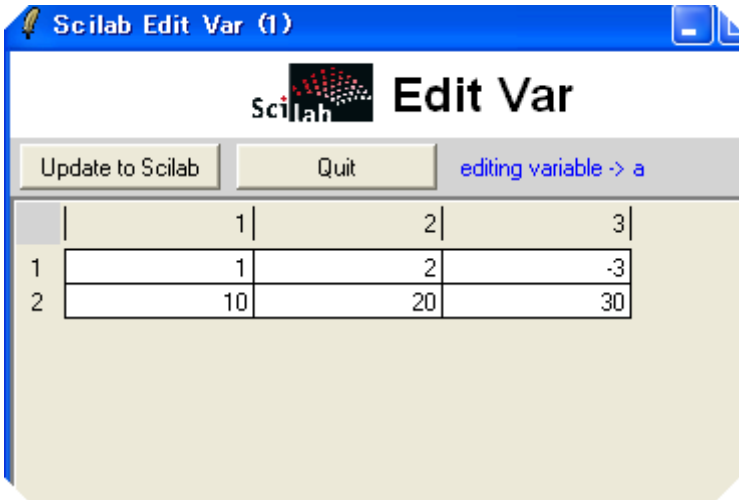
または**[File][Change Directory]よりカレントディレクトリに変更**しておいてから

a=read('d.txt',－1,3)　とすると

a は 2 行 3 列の行列となり　2 次元 a(1,1)～a(2,3)　または 1 次元 a(1)～a(6) に格納される。

第 1 行は　b1=a(1,:)
第 2 行は　b2=a(2,:)
第 2 列は　c2=a(:,2)
行数 列数は　**length**(c2)　length(b2)
行列のサイズは　**size**(a)　で与えられる。

また GUI による行列値表示ができる　**editvar** a



Sample program 1

```
a=read('d.txt',-1,3)
// a は 2 行 3 列の matrix となり　a(1,1)～a(*,3) に格納される。

//第 1 行
b1=a(1,:)
//第 2 行
b2=a(2,:)
//b1 の統計諸量
m1=mean(b1);
md1=median(b1);
st1=stdev(b1);
mx1=max(b1);
mn1=min(b1);
mprintf("¥n  mean=%f  median=%f  std=%f ¥n max=%f  min=%f¥n",  m1, md1,
st1,mx1,mn1);
mprintf("¥n");

m2=mean(b2);
st2=stdev(b2);
// b1*b2 の平均
m12=b1*b2'/length(b1);
//相関係数
r=(m12-m1*m2)/st1/st2
//プロット
plot2d(b1,b2,style=－1,rect=[-5,-5,15,15]);
```

p88 の例題 1, p90 の例題２, p91 の例題３などをデータとしてプログラムを試してみよう。

3. 頻度分布・・・度数分布表　ヒストグラム

y=[78, 59, 50, 90, 85,90, 89, 78, 82,63,72,100,72,56,62,58,52,62]
と与えられたとき，ｙの度数分布は**nfreq**(y)または**tabul**(y)で求まる。
tabul(y,”d”)では降順に（”d”は省略可），**tabul**(y,”i”)では昇順に並べられる。
下記の例ではyyy=floor(y/10)*10;によって切捨てを行ってから処理した。
またヒストグラムを描くには**histplot**([x1:step:x2],y,style=2)
ただしx1, x2はｘ軸の始点終点であり，stepは間隔（下記例では10）であり,style=の次の数は色コードである。

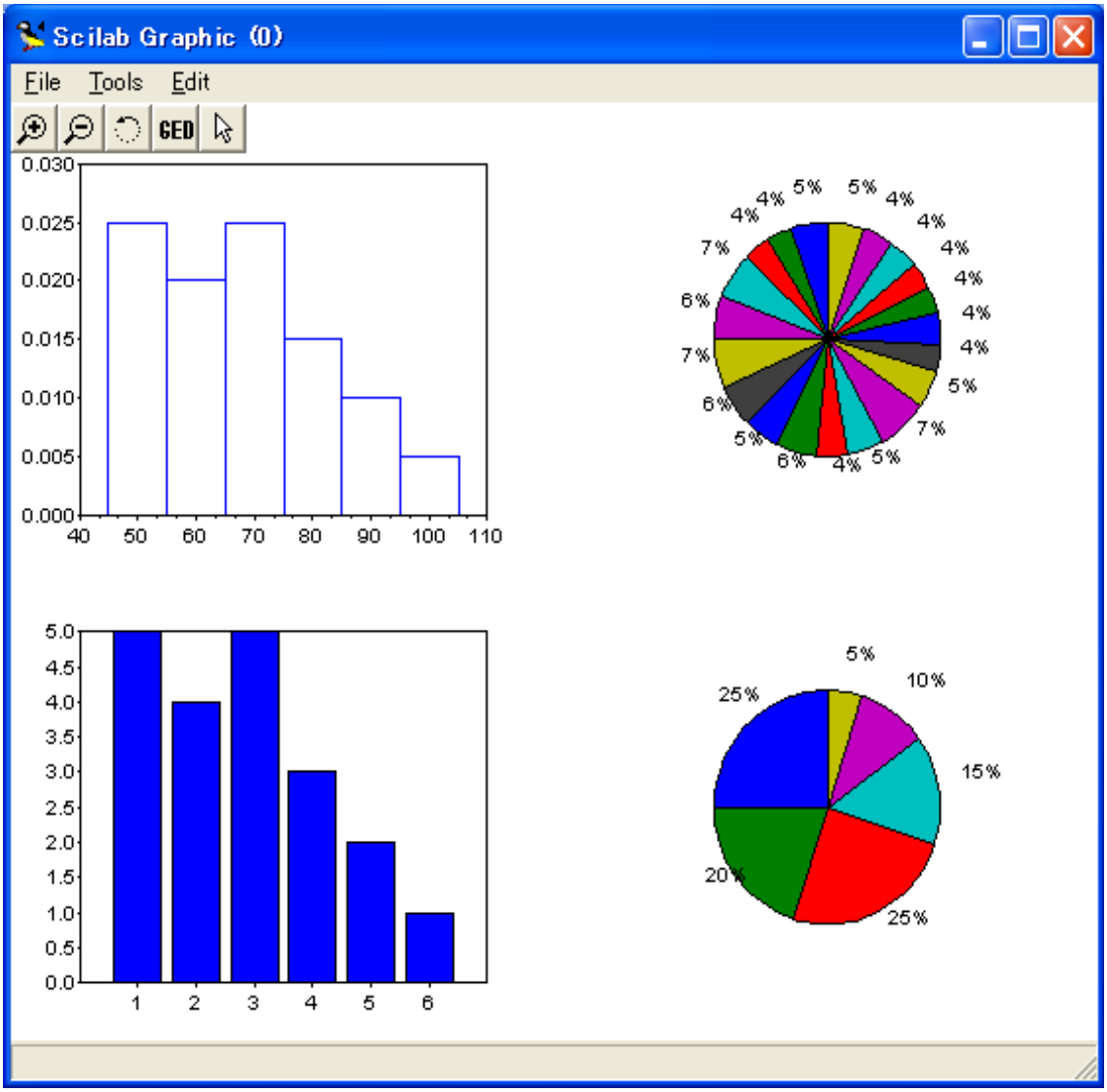
```
   50.      5.
   60.      4.
   70.      5.
   80.      3.
   90.      2.
   100.     1.
```

ここでｍは行列（左では6行2列）であり，第1行はr1=m(1,:)で，第2列はc2=m(:,2)である。ｍの（3,2）成分＝r3の第2成分＝c2の第3成分＝5である。ｍの行数は列ベクトルのサイズでlength(r1)，列数は列ベクトルのサイズでlength(c1)となる。

棒グラフ描画には　**bar**(y)･･･縦向，**barh**(y)･･･横向　円グラフには**pie**(y)を使うが注意を要する。

Sample program

```
y=[78,59,50,90,85,90,89,78,82,63,72,100,72,56,62,58,52,62,66,70];
yyy=floor(y/10)*10;
//度数分布表
m=tabul(yyy,"i")
//第1列,第2列，行数
c1=m(:,1);
c2=m(:,2);
l=length(c1);
//グラフ
subplot(2,2,1)
 t=[c1(1)-5:10:c1(l)+5];
 histplot(t,yyy,style=2)
subplot(2,2,2)
 pie(yyy)
subplot(2,2,3)
 bar(c2)
subplot(2,2,4)
 pie(c2)
```



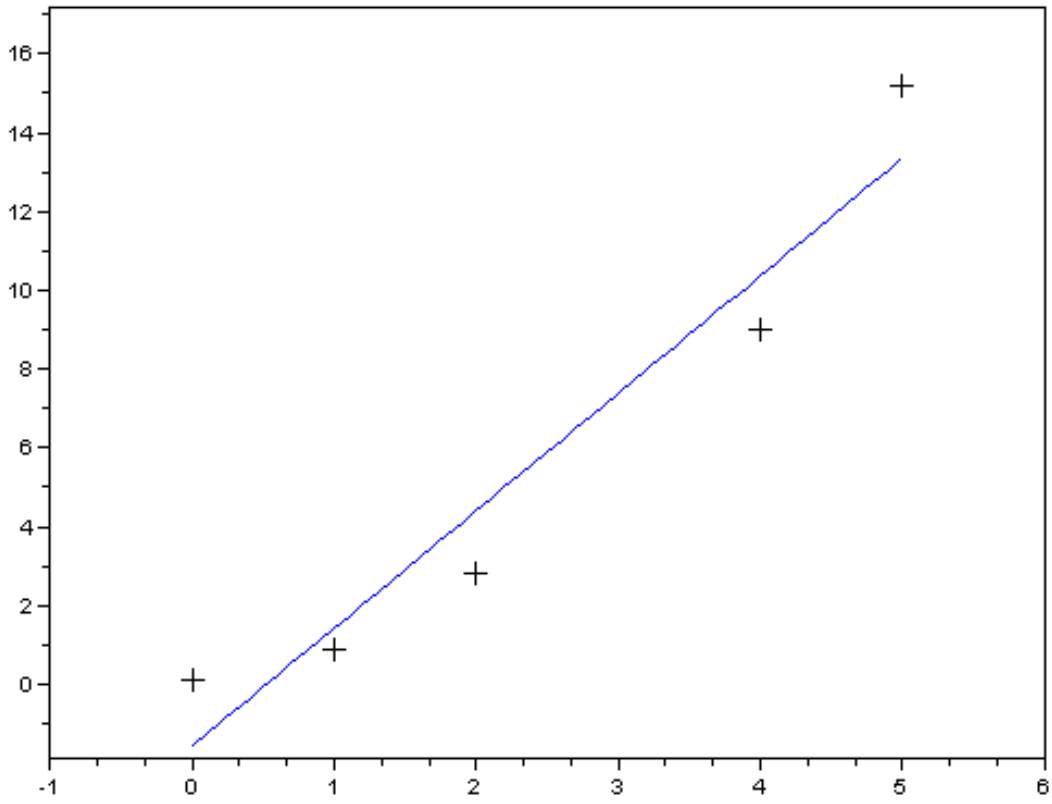
左上　yyy のヒストグラム　　比率で表示　　右上　yyy の円グラフ
左下　c2 の棒グラフ　　個数で表示　　右下　c2 の円グラフ

4. 回帰分析・・・最小二乗法

与えられた２つの変数間の最小二乗法関係式を得る。

x=[0, 1, 2, 4, 5]
y=[0.1, 0.9, 2.8, 9.0,15.2]　が与えられたとき
ｘとｙとの回帰直線を
　yy=p*xx+q
とすればその係数は
[p,q]=reglin(x,y)
または
coefs=regress(x,y)
で求まる。ただしp=coefs(2)　q=coefs(1) であり,x=x0に対する予測値は
p*x0+qである。

plot2d(x,y,style=－1,rect=[〇,〇,〇,〇])　　(x,y)のプロット
plot2d(xx,yy,style=2,rect=[〇,〇,〇,〇])　　回帰直線のプロット



　p91　問題4, p94　問題7, p95　問題8を上記の方法で解いてみよう

重回帰のときには回帰平面　**zz=p*xx+q*yy+r**　では
[p,q,r]=reglin(x,y,z)　より係数が求まる。

5. 離散的確率分布・・・2項分布，ポアッソン分布，超幾何分布

2項分布　${}_nC_r*p^r*(1-p)^{n-r}$

nとpを与えてpr=**binomial**(p,n)とすれば，各事象が起こる確率は配列pr(1)～pr(n+1)に収められる。全く起こらない確率はpr(1)であり，r回起こる確率はpr(r+1)であることに注意！例えば，サイコロを30回振って1の目（どんな数でも同じ）が10回出る確率はpr=binomial(1/6,30)のpr(11)である。
この配列の和は当然1である。sum(pr)＝1
r1≦i≦r2までの確率は次の関数で求まる。

Sample program 3

```
function [fx]=bioms(r1,r2)
  s=0;
 for i=r1+1:r2+1
  s=s+pr(i);
  fx=s;
 end;
endfunction
```

したがってサイコロを30回振って1の目が3回以下の確率はbioms(0,3)

```
for j=0:n
 expect(j+1)=pr(j+1)*j;
 mprintf( ″  %d  %f %f¥n″ ,j,pr(j+1),expect(j+1))
end;
//期待値
Expected=sum(expect)
```

pr(i)をグラフ表示すると図のようになりnが大ききなると左右対称に近づく

```
j=0:n
plot2d(j,pr,1,rect=[0,0,n,1]);
```

ポアッソン分布　　$e^{-m}*m^r/r!$

未完

超幾何分布　　${}_rC_x\ {}_sC_y/{}_{r+s}C_{x+y}$

未完

6. 連続的確率分布　・・・正規分布, t 分布, χ^2 分布

正規分布

例　試験の点数　身長　測定値

標準正規分布関数 f(x)=exp(-x^2/2)/√(2π)を 0 から z まで積分しそれを z の関数とする。
ただし正規化されていないときは x=(t-mean)/sigma で変換する。t は生の値

```
function [pr]=integ(z)
   pr=integrate('exp(-x^2/2)/sqrt(%pi*2)','x',0,z)
endfunction
```

integ(1.7)により 0～1.7 の積分値,すなわち 0＜z＜1.7 である確率が求まる。
prob1=integ(z_1),prob2=integ(z_2)　とすれば
確率($z<z_1$)　は　0.5+prob1
確率($z>z_1$)　は　0.5-prob1
確率($z<|z_1|$) は 2*prob1
確率($z_1<z<z_2$)　は　prob2-prob1
であることは明らかである。

Scilab には erf という関数が装備されているが,それは被積分関数として exp(-x^2)/√π を使っているので注意を要する。上記の値との関係は integ(m)＝erf(m/sqrt(2))/2

Sample program 4

```
//正規分布
//グラフ
x=-4:0.1:4;
y=exp(-x^2/2)/sqrt(%pi*2);
plot2d(x,y,1,rect=[-4,0,4,1]);

//平均　標準偏差を与える
mean=168.4;sigma=5.5;
//値を入力して標準化
x1=input("x1= ");
m1=(x1-mean)/sigma
x2=input("x2= ");
m2=(x2-mean)/sigma

//積分する関数
function [pr]=integ(z)
 pr=integrate('exp(-x^2/2)/sqrt(%pi*2)','x',0,z)
endfunction

//関数を呼び出すとｚが 0～ｍである確率が求まる
prob1=integ(m1)
prob2=integ(m2)

// 確率(<m)                    prob+0.5
// 確率(-m～m)                 2*prob
// 確率(>m)
ans1= 0.5-prob1;
ans2= 0.5-prob2;
ans=ans2-ans1

//グラフに追加
plot2d4([m1;m1],[0.7;0],5);
plot2d4([m2;m2],[0.7;0],5);
xset("font size",3);
xstring(-1,0.8,"probabitity = "+ string(ans1));
xstring(-1,0.7,"probabitity = "+ string(ans2));
xtitle('Normal Distribution');
```

ｔ分布　　**未完**

χ^2分布　　**未完**