Experiment simulation

Two samples X= $\{X_{i=1}^{n}\}$ and Y= $\{Y_{i=1}^{m}\}$ belong to family of Normal distribution $\boldsymbol{N}(\mu, \sigma^2)$. X has an emplirical characteristics $a_1, s_1^2$ X and Y - $a_2, s_2^2$.

Creating two statistical samples. Generation samples based on characteristics close to those of the general population:

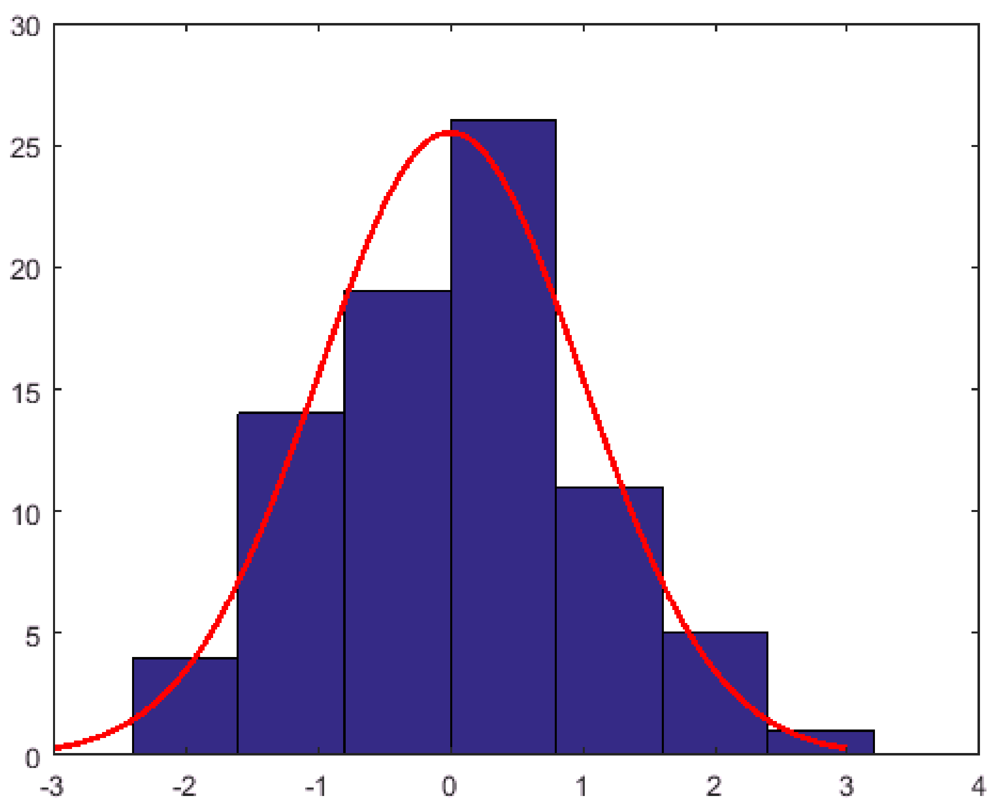
```
n=80; m=100; sigma=2; mu=1;
X=random('Normal', mu, sigma,[n,1]);
Y=random('Normal', 1.05, 1.95,[m,1]);
a1=mean(X); s1=std(X); a2=mean(Y); s2=std(Y);
disp(['a1=',num2str(a1),' s1=',num2str(s1),'  a2=',num2str(a2),'  s2=',num2str(s2)])
```

```
a1=1.3121 s1=1.804  a2=1.0066  s2=2.0066
```

**I) Check the distribution Normality for the single sample (choose the first sample). Evaluate the correspondence of the empirical density function and theoretical one using all the upper center points of each histogram Bin**

a) the first step is to standardize the data:

```
Xs=(X-a1)/s1; Ys=(X-a2)/s2; k=7;
hX=histfit(Xs,k);
```




```
%hX(1) % you can see every field of structure
Xcenter=hX(1).XData
```

```
Xcenter =

   -2.0000   -1.2000   -0.4000    0.4000    1.2000    2.0000    2.8000
```

```
ni=hX(1).YData  % absolute frequencies
```

```
ni =

     4    14    19    26    11     5     1
```

Relative frequency and frequency as in histogram with option 'pdf'

```
Ni=ni'/n; % these relative frequencies
Li=Xcenter(2)-Xcenter(1); %lenght of bin
Npi=Ni/Li; % empirical probability taking
          % into account the length of Bins
```

b) $\rho(X) = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$(Presentation5, page5)

$\rho(X)$ - criterion (measure with known distribution), $p_i$ - theoretical probability to belong $i$ - *th* bin according to Normal Standard Distribution. To determine $p_i$ need to find interval edges and calculate $p_i = \int_{X_i^{left}}^{X_i^{right}} f(\mathrm{x}, \mu, \sigma)\,\mathrm{dx}$

$\mu$=0, $\sigma$ =1 for Normal Standard distribution.

Integral limits:

```
Xedge=Xcenter-Li/2; Xedge=[Xedge,Xedge(end)+Li];
```

To determine pi use normcdf ~ F(x<y), y-integral limits:

```
p=zeros(k,1);
 for i=2:k+1
    p(i-1)=normcdf(Xedge(i),0,1)-normcdf(Xedge(i-1),0,1);
 end
 rhoXs=sum((ni'-n*p).^2./(n*p)) % ni' - column
```

```
rhoXs = 2.1670
```

$\rho(X)$ - criterion statistics is equal to value of calculated rhoX with known distribution $H_{k-1}$ (Presentation5, page6, Pearson Theorem), k=7.

Need to determine critical point for Chi-square distribution of the family $H_6$

Error level $\alpha = 0,05$ (5% ) corresponds to probability greater than 0.95 so critical point is C=$C_{0.95}$. Determine this point

```
C=chi2inv(0.95,6)
```

```
C = 12.5916
```

```
t=0:0.1:12; Py=chi2pdf(t,6);
plot(t,Py,'r-'), hold on, % xline(C)
p=plot(C,0,'b.','markersize',40)
```

```
p =
  Line with properties:

              Color: [0 0 1]
          LineStyle: 'none'
          LineWidth: 0.5000
             Marker: '.'
         MarkerSize: 40
    MarkerFaceColor: 'none'
              XData: 12.5916
              YData: 0
              ZData: [1×0 double]

  Show all properties
```

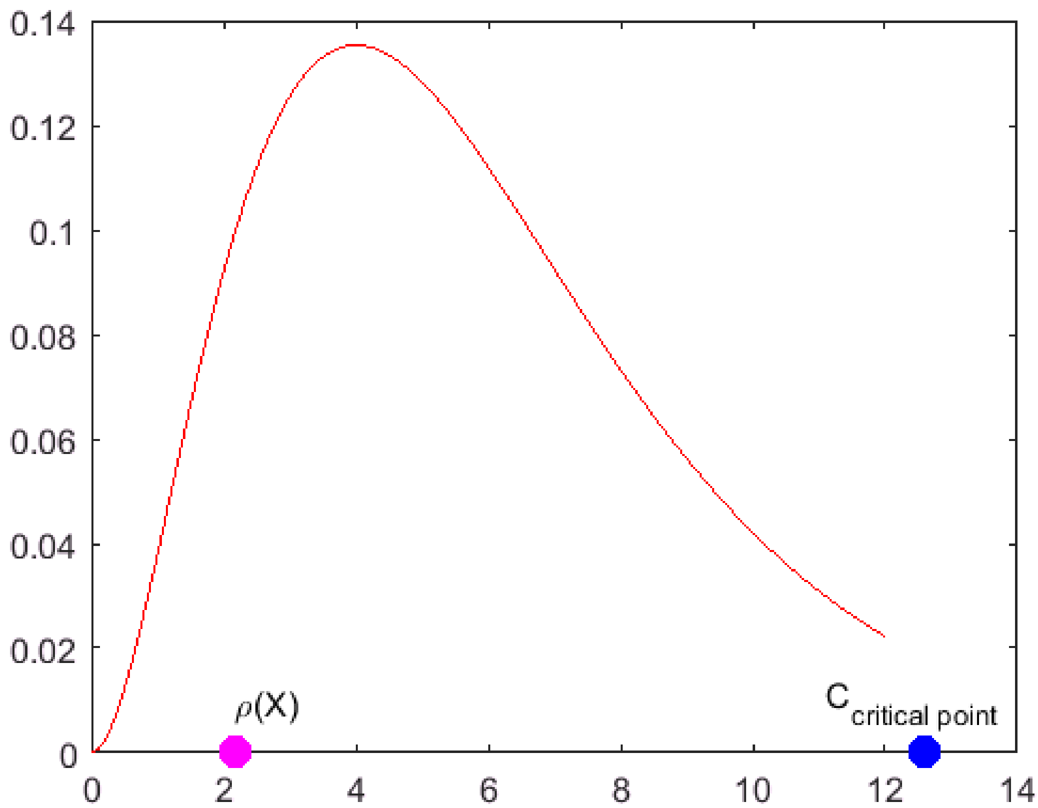
```
text(C-1.5,0.01,'C_{critical point}','fontsize',12)
plot(rhoXs,0,'m.','markersize',40)

        txt=text(rhoXs,0.01,'\rho(X)')
```

```
txt =
  Text (\rho(X)) with properties:

                 String: '\rho(X)'
               FontSize: 10
             FontWeight: 'normal'
               FontName: 'Helvetica'
                  Color: [0 0 0]
    HorizontalAlignment: 'left'
               Position: [2.1670 0.0100 0]
                  Units: 'data'

  Show all properties
```

```
        set(txt,'fontsize',12)
if C>rhoXs
    disp('H_1 - true; sample from N(0,1)')
else
    disp('H_1 reject; sample doesn''t belong N(0,1)')
end
```

```
H_1 - true; sample from N(0,1)
```

```
 set(gca, 'fontsize',12)
```




Does the sample belong to the standard normal distribution? Other way! Use ML functions!

c) Chi-square goodness-of-fit test: [H,p] = chi2gof(X); X - sample, H=0, hypothesis H1 - true, p - value of probability to obtain criterion statistics

```
[H,P] = chi2gof(X)
```

```
H = 0
P = 0.7735
```

```
if H==0
    disp('Standard Normal Distribution')
else
    disp('H1 is Rejected, chi2gof')
end
```

```
Standard Normal Distribution
```

d) Kolmogorov-Smirnov test: [h,p]=kstest(X); X - sample, h=0 - hypothesis H1 is true, p - value of probability to obtain criterion statistics

```
[h,p]=kstest(Xs)   % 0.025<p<0.975
```

```
h =    0
p = 0.8926
```

```
if h==0
    disp('Standard Normal Distribution')
else
    disp('H1 is Rejected, kstest')
end
```

```
Standard Normal Distribution
```

**II) Compare distribution of two samples. They are standardized**

a) Samples Xs, Ys have different volume n and m, respectively. Comes from normal family population with the same variance. Compare means

```
[hts,pts]=ttest2(Xs,Ys)
```

```
hts = 0
pts = 0.3128
```

```
if hts==0
    disp('Standard Normal Distribution')
else
    disp('H1 is Rejected, ttest2')
end
```

```
Standard Normal Distribution
```

b) The same (a) without standardization

```
[ht,pt]=ttest2(X,Y)
```

```
ht = 0
pt = 0.2901
```

```
if ht==0
    disp('Standard Normal Distribution, ttest2')
else
    disp('H1 is Rejected, ttest2')
end
```

```
Standard Normal Distribution, ttest2
```

**III) Prove equality of empirical and theoretical mean**

a) a test decision for the null hypothesis that the data in the vector X comes from a normal distribution with mean $\mu$ and a standard deviation $\sigma$

Hmz - indicator of chosen hypothesis; pmz - probability of criterion statistics zval ; ci - confidence interval for mean of X

```
[Hmz,pmz,ci,zval]=ztest(X,mu,sigma,'tail','left')
```

```
Hmz = 0
pmz = 0.9186
ci =

      -Inf
    1.6799
```

```
zval = 1.3957
```

```
if Hmz==0
    disp('X∈N(μ=1, σ=2)')
else
    disp('Hmz is Rejected, ztest')
end
```

```
X∈N(μ=1, σ=2)
```

Using statistical Z- criterion and construct $\rho(X)$: (presentation5, p.17)

$$\rho(X) = \frac{\sqrt{n}\,(\overline{X} - \mu)}{\sigma} \in N(0,1)$$

two-side ctrterion, for size criterion equal to $\alpha = 0.05$, critical points $C_1 = F^{-1}(0.025)$ and $C_2 = F^{-1}(0.975)$.

$H_1$ - true if $C_1 < \rho < C_2$

```
rhoz=sqrt(n)*(a1-mu)/sigma
```

```
rhoz = 1.3957
```

```
Prhoz=normcdf(rhoz,0,1)
```

```
Prhoz = 0.9186
```

```
C1=norminv(0.025,0,1);
C2=norminv(0.975,0,1);
```

b) a test decision for the null hypothesis that the data in x comes from a normal distribution with mean equal to given $\mu$ and unknown variance, using the one-sample ttest

```
[Hmt,pmt]=ttest(X,1)
```

```
Hmt = 0
pmt = 0.1258
```

```
if Hmt==0
    disp('X∈N(μ=1, σ-unknown), ttest')
else
    disp('Hmz is Rejected, ttest2')
end
```

```
X∈N(μ=1, σ-unknown), ttest
```

Try realize statistical t-criterion according to presentation.5 p.18 and construct $\rho(X)$ = , having ??? distribution

and determine result, using MatLab functions:

```
%rhot=
%Prhot=
```

```
%C1t=
%C2t=
```

**IV) Equality of two samples variances**

presentation5, p. 8 $H_1$

$$\rho(X,Y) = \frac{S_0^2(X)}{S_0^2(Y)} \in F_{n-1,m-1}$$

```
sigmaX=var(X)*n/(n-1) % explain!
```

```
sigmaX = 3.2955
```

```
sigmaY=var(Y)*n/(n-1) % explain!
```

```
sigmaY = 4.0775
```

```
rhoVars=sigmaX/sigmaY
```

```
rhoVars = 0.8082
```

```
Ccr2=finv(0.95,n-1,m-1)
```

```
Ccr2 = 1.4174
```

```
if(rhoVars<Ccr2)
    disp('H1 - true, variancies of the samples are the same')
else
    disp('H1 is rejected: variancies are different')
end
```

```
H1 - true, variancies of the samples are the same
```

**V) Equality of the sample variance to General Population variance, presentation5, p.8**

**You need create** $\rho(X) \in H_{n-1}$

```
RhoX=sigmaX/sigma^2 % sigma - standard deviation
```

```
RhoX = 0.8239
```

```
CcritI=tinv([0.025 0.975],n-1)
```

```
CcritI =

   -1.9905    1.9905
```

```
%Draw Conclusion!
```