

Statistique: Projet 1

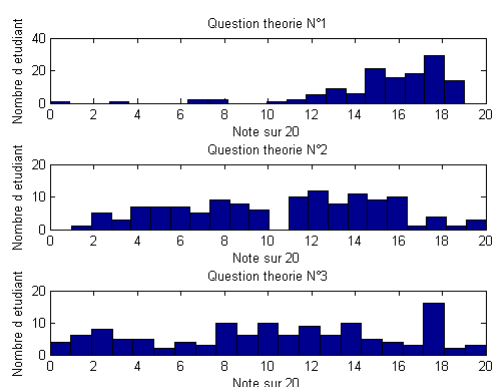
Sauvage Mehdi

November 3, 2016

1 Analyse descriptive

(a)

Vu l'allure des histogrammes, nous pouvons déduire que la question 1 était globalement mieux réussie que la question 2 et la question 3. La question 2 et la question 3 ont, quant à elles, été plus équitablement réparties entre les différentes cotes possibles.



(b)

Dans le tableau, ci-dessous, sont repris les moyennes, médianes, modes et écarts-types des questions d'exercice:

...	Question 1	Question 2	Question 3
Moyennes	11.9685	14.6614	6.5669
Médianes	12	16	5
Modes	20	19	0
écarts-Types	5.8580	4.5639	6.6781

Nous constatons que la question 2 a été bien mieux réalisée que les 2 autres. A la 1er question, les étudiants aillant eu 20 représentent la plus grande proportion alors que la moyenne n'est qu'à 11.96. Pour la question 2, la moyenne et la médiane sont plus élevées que pour la question 1 mais la plus grand partie des étudiants ont eu 19.

Nous pouvons voir que la question 3 est la seule qui a en moyenne pas été réussi. En effet, une majorité des étudiants n'ont pas répondu à la question ou

ont eu zero. En effet, la moyenne ne se situe qu'à 6.57.

Pour qu'un résultat soit estimé "Normal" au sens de la lois normal, il doit être tel que

$$\text{Donnée} \in [\text{Moyenne} - \text{Ecart-type} ; \text{Moyenne} + \text{Ecart-type}]$$

Dés lors, pour qu'une donnée soit normal elle doit être comprise entre:

...	Question 1	Question 2	Question 3
Interval Normal	[6,12 ; 17.81]	[10,10 ; 19,22]	[-0.11 ; 13.25]
Nbr Normal	72	103	99
Proportion	56,69%	81.11%	77.95%

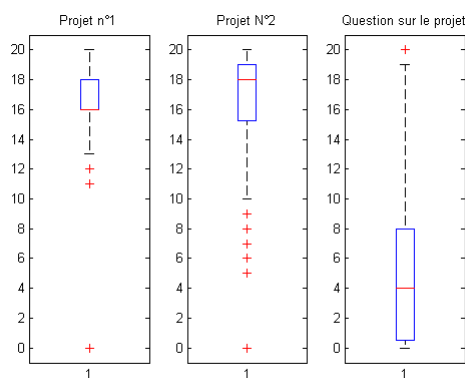
(c)

Comme nous pouvons observer dans les graphiques ci-dessous, il existe des données aberrantes pour les 2 projets et pour la question relative au projet. En effet, pour le projet 1 et 2 , certain étudiant on eu des cotes en dessous de 13 et en dessous de 10 alors que d'autres étudiants on eu 20 à la question du projet lors de l'examen. Toutes ces données constitue des aberrations statistiques.

Notons que les valeurs aberrantes sont défini comme appartenant à:

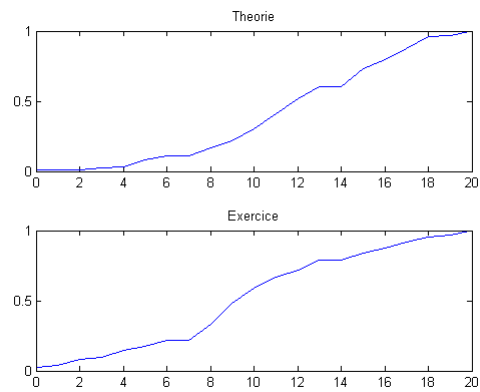
$$[q1 - 1.5 \times (q3 - q1); q3 + 1.5 \times (q3 - q1)]$$

q1 and q3 sont les 25eme and 75eme quartiles des données.



...	Projet 1	Projet 2	Question projet
quartile 25	16	15.25	0.5
Médianes	16	18	4
quartile 75	18	19	8
limite aberrance sup	20	20	19.25
limite aberrance inf	13	10	0

(d)

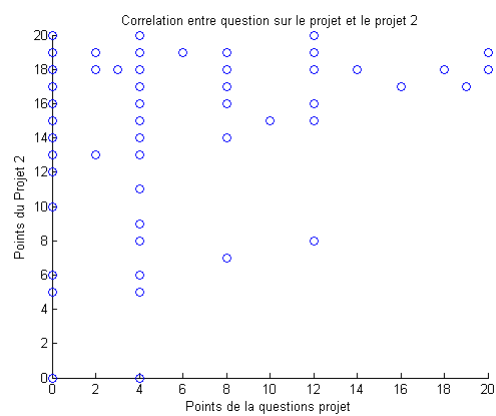


...	Theorie	Exercices
proportion de [10;14]	0.3858	0.3071

(e)

Vu l'allure du graphe, nous ne pouvons pas tirer de relation évidente entre les deux données. Cette conclusion est appuyé par le coefficient de corrélation assez éloigné de 1.

le coefficient de corrélation est 0.2360.



2 Génération d'échantillons i.i.d.

(a)

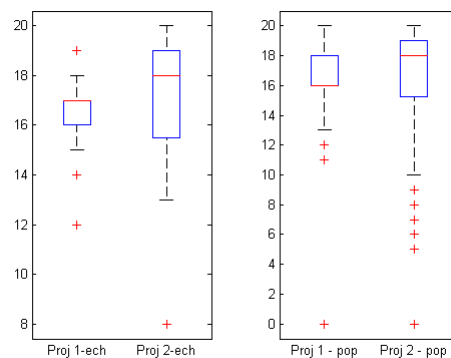
i

Nous pouvons constater que les moyennes, médianes et écarts-types calculées pour les échantillons ne sont pas très éloignées des mêmes valeurs pour la population entière. Néanmoins, ces valeurs pourraient s'en éloigner si l'échantillon considéré était moins représentatif de la population global.

...	Exercice 1	Exercice 2	Exercice 3
Moyennes de l'échantillon	11.15	15	6.75
Médianes de l'échantillon	10.5	16.5	5
Ecart-types de l'échantillon	4.86	4.71	6.92

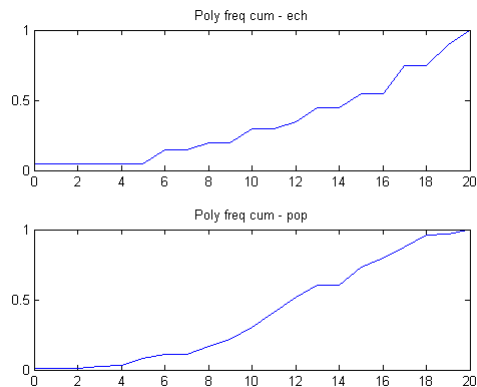
ii

Ci-dessous ce trouve les boites à moustaches relative aux projets. Nous pouvons constater que même si les allures générales sont conservées, il y a des différences importantes entre les boites relative à l'échantillon et à la population global.



iii

Voici, ci-dessous, les polygones des fréquences cumulées relatif aux questions de théorie.



Comme constaté dans les questions précédentes, les allures générales sont conservées mais les détails des graphiques ne sont pas les mêmes.

La distance de Kolmogorov Smirnov a été calculée comme cela:

$$\max\{abs\{ \text{vec_freq_cum}(\text{population}) - \text{vec_freq_cum}(\text{echantillon})\}\} = 0.2453$$

(b)

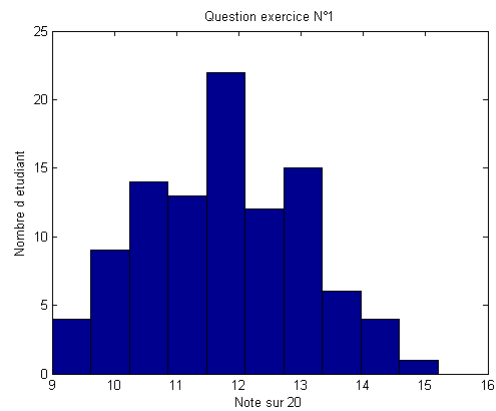
Prenons 100 échantillons de 20 étudiants pour faire une moyenne des moyennes, médianes et écart-types des échantillons de la question 1 d'exercice.

Notons que les questions i , ii et iii ont pour résultat des graphiques aillant l'allure d'une gaussienne. En effet, ces données semble obéir à des lois Normales.

Pour ces trois mêmes questions, nous pouvons constaté que les moyennes des médianes, moyennes et écart-types prisent sur 100 échantillons sont beaucoup plus proche que lorsque que l'on ne considère qu'un seul échantillon. Cela semble logique puisque la population est mieux représenté par 100 sélections de 20 étudiants que une seule.

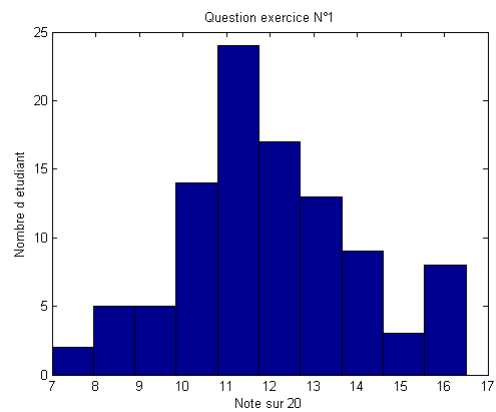
i

La moyenne des moyennes des échantillons est 12.04. Pour la population totale, la moyenne est de 11.96.



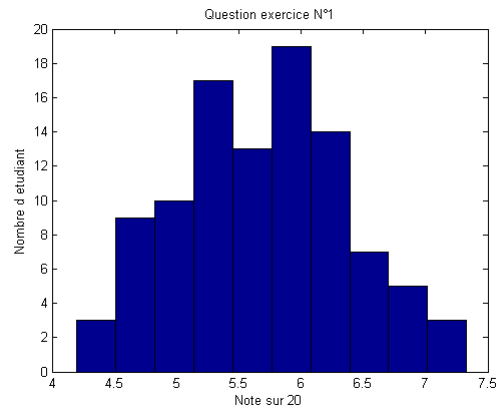
ii

La moyenne des médianes des échantillons est 12,33 points. Pour la population totale, la médiane est de 12 points.



iii

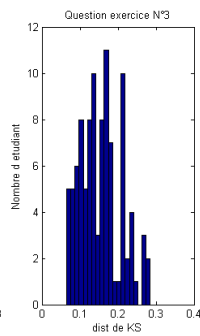
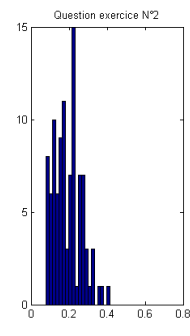
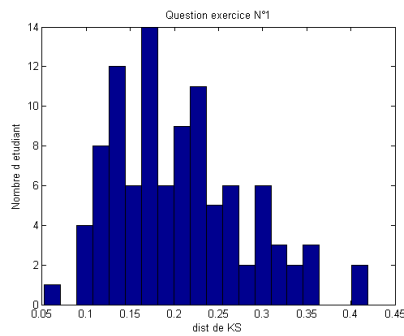
La moyenne des écart-types des échantillons est 5.74 points. Pour la population totale, l'écart-type est de 5.85 points.



iv et v

Les trois histogramme ont la même allure et ressemble vaguement au graphique d'une lois de poisson. Ils croient rapidement jusqu'à un maximum au environ de 0.18 et puis diminue plus doucement jusque qu'a arriver a 0.

La lois de poisson est construite en se basant sur l'idée que la distance de Kolmogorov Smirnov la plus fréquente ce situe vers 0.18 mais il est plus naturelle de trouver des distance plus petite car les 100 echantillons représente relative-ment bien la population. Au dela de 0.18, il y a une brusque diminution du nombre d'étudiant car la l'échantillon devrait , en général, donner une distance de Kolmogorov Smirnov plus petite.



Code

```
1 %Cette fonction genere les 2 histogrammes des r sultats
   des questions de
2 %t h orie
3
4 function [] = Q1a (points)
5
6     theoQ1 = points(: , 4);
7     theoQ2 = points(: , 5);
8     theoQ3 = points(: , 6);
9
10    %Affichage des graphiques
11    figure('name','1a - Questions T h ories ');
12    subplot(3,1,1)
13    hist( theoQ1,21);
14    title('Question theorie N 1 ');
15    xlabel('Note sur 20 ');
16    ylabel('Nombre d etudiant ');
17
18    subplot(3,1,2)
19    hist( theoQ2,21);
20    title('Question theorie N 2 ');
21    xlabel('Note sur 20 ');
22    ylabel('Nombre d etudiant ');
23
24    subplot(3,1,3)
25    hist( theoQ3,21);
26    title('Question theorie N 3 ');
27    xlabel('Note sur 20 ');
28    ylabel('Nombre d etudiant ');
29
30 end
```

```

1  function [moy_ExQ1 ,moy_ExQ2 ,moy_ExQ3 ,mode_ExQ1 ,
      mode_ExQ2, mode_ExQ3 ,...
2  med_ExQ1, med_ExQ2, med_ExQ3 , ET_ExQ1 ,ET_ExQ2 , ET_ExQ3
      ]= Q1b(points)
3
4      %moyennes des points des Exercices de l'exam
5      moy_ExQ1 = mean(points(: , 7));
6      moy_ExQ2 = mean(points(: , 8));
7      moy_ExQ3 = mean(points(: , 9));
8
9      %mediannes des points des exercices de l'exam
10     med_ExQ1 = median(points(: , 7));
11     med_ExQ2 = median(points(: , 8));
12     med_ExQ3 = median(points(: , 9));
13
14     %modes des points des exercices de l'exam
15     mode_ExQ1 = mode(points(: , 7));
16     mode_ExQ2 = mode(points(: , 8));
17     mode_ExQ3 = mode(points(: , 9));
18
19     %L' arts types des points des exercices de l'exam
20     ET_ExQ1 = std(points(: , 7));
21     ET_ExQ2 = std(points(: , 8));
22     ET_ExQ3 = std(points(: , 9));
23
24 end

```

```

1  function [quar25-Qproj, quar50-Qproj, quar75-Qproj] = Q1c
    (points)
2
3      %Boite a moustache pour question projet
4      figure('name','1c - Questions sur le projet')
5      subplot(1,3,1)
6      boxplot(points(:,1));
7      title('Projet n1');
8
9      subplot(1,3,2)
10     boxplot(points(:,2));
11     title('Projet N2');
12
13     subplot(1,3,3)
14     boxplot(points(:,3));
15     title('Question sur le projet');
16
17     %Calcul des quartiles
18     quar25_proj1 = quantile(points(:,1),0.25);
19     quar50_proj1 = quantile(points(:,1),0.5);
20     quar75_proj1 = quantile(points(:,1),0.75);
21
22     quar25_proj2 = quantile(points(:,2),0.25);
23     quar50_proj2 = quantile(points(:,2),0.5);
24     quar75_proj2 = quantile(points(:,2),0.75);
25
26     quar25-Qproj = quantile(points(:,3),0.25);
27     quar50-Qproj = quantile(points(:,3),0.5);
28     quar75-Qproj = quantile(points(:,3),0.75);
29
30 end

```

```

1  function [moy_parEtu_theo , moy_parEtu_ex ,
      Prop_10_14_Theo ...
2  , Prop_10_14_Ex , cumsum_freq_Qtheo] = Q1d(points)
3
4      %Definition de moy_parEtu_theo et moy_parEtu_ex
5      nbrEtu = size(points(:,1));
6      moy_parEtu_theo = zeros(nbrEtu);
7      moy_parEtu_ex = zeros(nbrEtu);
8
9      for i = 1 : nbrEtu
10         moy_parEtu_theo(i) = round((points(i , 4) +
11             points(i , 5) +...
12             points(i , 6))/3);
13
14         moy_parEtu_ex(i) = round( (points(i , 7) + points
15             (i , 8) + ...
16             points(i , 9))/3);
17     end
18
19     freq_parEtu_QTheo = hist(moy_parEtu_theo,21);
20     freq_parEtu_QEx = hist(moy_parEtu_ex,21);
21
22     Prop_10_14_Theo = sum(freq_parEtu_QTheo(11 : 15))/
23         nbrEtu(1);
24     Prop_10_14_Ex = sum(freq_parEtu_QEx(11 : 15))/
25         nbrEtu(1);
26
27     %Affichage du polygone des fr quences
28     figure('name', '1d - Polygone des frequences');
29
30     subplot(2,1,1);
31     cumsum_freq_Qtheo = cumsum (freq_parEtu_QTheo/127);
32     plot( 0:20 , (cumsum (freq_parEtu_QTheo/127) ));
33     title('Theorie')
34
35     subplot(2,1,2);
36     plot(0:20 , cumsum(freq_parEtu_QEx/127) );
37     title('Exercice')
38
39 end

```

```

1  function [coefCorr_proj2-Qproj] = Q1e(points)
2
3      pts_proj2 = points(:, 2);
4      pts_QProj = points(:, 3);
5
6      %Affichage des graphique
7      figure('name' , ' 1e - Question sur le projet et
          projet 2')
8      title('Projet 2 - Q Projet')
9      scatter( pts_QProj, pts_proj2)
10     title('Correlation entre question sur le projet et le
          projet 2')
11     xlabel('Points de la questions projet')
12     ylabel('Points du Projet 2')
13
14     %Calcul du coef de correlation
15     mat_coefCorr_proj2-Qproj = corrcoef(pts_proj2 ,
          pts_QProj);
16     coefCorr_proj2-Qproj = mat_coefCorr_proj2-Qproj(2,1);
17
18 end

```

```

1 function [moy_ech_Ex1, moy_ech_Ex2, moy_ech_Ex3, ...
2 med_ech_Ex1, med_ech_Ex2, med_ech_Ex3 ...
3 ET_ech_Ex1, ET_ech_Ex2, ET_ech_Ex3 ] = Q2ai(ech)
4
5     %—Les Moyennes
6     moy_ech_Ex1 = mean(ech(:,7))
7     moy_ech_Ex2 = mean(ech(:,8))
8     moy_ech_Ex3 = mean(ech(:,9))
9
10    %—Les Medians
11    med_ech_Ex1 = median(ech(:,7))
12    med_ech_Ex2 = median(ech(:,8))
13    med_ech_Ex3 = median(ech(:,9))
14
15    %—Les Ecart Types
16    ET_ech_Ex1 = std(ech(:,7))
17    ET_ech_Ex2 = std(ech(:,8))
18    ET_ech_Ex3 = std(ech(:,9))
19
20 end

1 function [ ] = Q2aai(ech, points)
2
3     %Boite a moustache pour question projet
4     figure('name','2aai - Questions sur les projet')
5
6     subplot(1,2,1)
7     boxplot([ech(:,1),ech(:,2)], 'label',{'Proj 1-ech','
8             Proj 2-ech'});
9
10    subplot(1,2,2)
11    boxplot([points(:,1),points(:,2)], 'label',{'Proj 1 -
12            pop','Proj 2 - pop'});

```

```

1  function [KS_dist] = Q2aiii(ech , cumsum_freq_Qtheo)
2
3      nbrEch = size(ech(:,1));
4      moy_parEtu_ech_theo = zeros(nbrEch);
5
6      %Calcul de la moyenne aux questions de thorie pour
          chaque udiant
7      for i = 1 : nbrEch
8          moy_parEtu_ech_theo(i) = round((ech(i , 4) + ech(
          i , 5) +...
9          ech(i , 6))/3);
10     end
11
12     freq_parEtu_ech_QTheo = hist(moy_parEtu_ech_theo,21);
13     cumsum_freq_ech_theo = cumsum (freq_parEtu_ech_QTheo
          /20);
14
15     %Affichage du polygone des fr quences
16     figure('name', '2aiii - Polygone des frequences
          cumules ');
17     subplot(2,1,1);
18     plot( 0:20 , cumsum_freq_ech_theo);
19     title('Poly freq cum - ech')
20
21     subplot(2,1,2);
22     plot(0:20 , cumsum_freq_Qtheo );
23     title('Poly freq cum - pop')
24
25     %—distance de KS
26     KS_dist = max(abs(cumsum_freq_ech_theo -
          cumsum_freq_Qtheo));
27
28     end

```

```

1  function [moy_moy_ech100] = Q2bi(ech100)
2
3      moy_ech100 = zeros(100,1);
4
5      for i = 1 : 100
6          moy_ech100(i) = mean(ech100(i,:),7));
7      end
8
9      %Affichage du graphique
10     figure('name','2bi- ');
11     hist(moy_ech100);
12     title('Question exercice N1');
13     xlabel('Note sur 20');
14     ylabel('Nombre d etudiant');
15
16     %Calcul de la moyenne des moyennes
17     moy_moy_ech100 = mean(moy_ech100);
18 end

1  function [moy_med_ech100] = Q2bii(ech100)
2
3      med_ech100 = zeros(100,1);
4
5      for i = 1 : 100
6          med_ech100(i) = median(ech100(i,:),7));
7      end
8
9      %Affichage du graphique
10     figure('name','2bii - ');
11     hist(med_ech100);
12     title('Question exercice N1');
13     xlabel('Note sur 20');
14     ylabel('Nombre d etudiant');
15
16     %Calcul de la moyenne des medianes
17     moy_med_ech100 = mean(med_ech100);
18 end

```



```

1  function [moy_ET_ech100] = Q2biii(ech100)
2
3      ET_ech100 = zeros(100,1);
4
5      for i = 1 : 100
6          ET_ech100(i) = std(ech100(i,:),7));
7      end
8
9      %Affichage du graphique
10     figure('name','2biii- ');
11     hist(ET_ech100);
12     title('Question exercice N1');
13     xlabel('Note sur 20');
14     ylabel('Nombre d etudiant');
15
16     %Calcul de la moyenne de arts -types
17     moy_ET_ech100 = mean(ET_ech100);
18 end

```

```

1  function [distKS_echPop] = Q2biv(ech100 , points)
2
3      %Calcul des polygone de fr quences cumules
4      freq_parEtu_QEx1_POP = hist(points(:,7),21);
5      cumsum_freq_QEx1_POP = cumsum (freq_parEtu_QEx1_POP
6      /127);
7      distKS_echPop = zeros(1,100);
8
9      for i = 1 : 100
10         freq_parEtu_QEx1_ech = hist(ech100(i,:),7),21);
11         cumsum_freq_QEx1_ech = cumsum (
12             freq_parEtu_QEx1_ech/20);
13         distKS_echPop(1,i) = max(abs(cumsum_freq_QEx1_POP
14             -cumsum_freq_QEx1_ech));
15     end
16
17     %Affichage des graphiques
18     figure('name','2biv- ');
19     hist(distKS_echPop,20);
20     title('Question exercice N1');
21     xlabel('dist de KS');
22     ylabel('Nombre d etudiant');
23 end

```

```

1 function [] = Q2bv(ech100, points)
2
3 %-----Question N*2
4 %Calcul des polygone de frquences cumules
5 freq-parEtu-QEx2-POP = hist(points(:, 8), 21);
6 cumsum_freq-QEx2-POP = cumsum (freq-parEtu-QEx2-POP
7 /127);
8 distKS_echPop-QEx2 = zeros(1,100);
9
10 for i = 1 : 100
11     freq-parEtu-QEx2_ech = hist(ech100(i, :, 8), 21);
12     cumsum_freq-QEx2_ech = cumsum (
13         freq-parEtu-QEx2_ech/20);
14     distKS_echPop-QEx2(1, i) = max(abs(
15         cumsum_freq-QEx2-POP-cumsum_freq-QEx2_ech));
16
17 end
18
19 %-----Question N*3
20 %Calcul des polygone de frquences cumules
21 freq-parEtu-QEx3-POP = hist(points(:, 9), 21);
22 cumsum_freq-QEx3-POP = cumsum (freq-parEtu-QEx3-POP
23 /127);
24 distKS_echPop-QEx3 = zeros(1,100);
25
26 for i = 1 : 100
27     freq-parEtu-QEx3_ech = hist(ech100(i, :, 9), 21);
28     cumsum_freq-QEx3_ech = cumsum (
29         freq-parEtu-QEx3_ech/20);
30     distKS_echPop-QEx3(1, i) = max(abs(
31         cumsum_freq-QEx3-POP-cumsum_freq-QEx3_ech));
32
33 end
34
35 %Affichage des graphiques
36 figure('name', '2bv- Q EX 2 et QEx 3 ');
37 subplot(1,2,1);
38 hist(distKS_echPop-QEx2, 20);
39 title('Question exercice N 2 ');
40 subplot(1,2,2);
41 hist(distKS_echPop-QEx3, 20);
42 title('Question exercice N 3 ');
43 xlabel('dist de KS');
44 ylabel('Nombre d etudiant ');
45 xlabel('dist de KS');
46 ylabel('Nombre d etudiant ');
47
48 end

```