**ELEMENTS DE STATISTIQUES**

**Rapport de la partie 1 du projet**



Pazienza Laurie

S123514

Année académique 2014-2015

**Analyse descriptive**

**Question 1**

**(a)**

Les trois histogrammes ci-dessous représentent les résultats des trois questions de théorie.

La *Figure 1* nous montre que la question 1 a été assez réussie. En effet, le résultat le plus abondant est 17/20, de plus il y a très peu d’échecs à cette question : un zéro et moins de cinq 9/20. Pour cette question, la majorité des étudiants se situent entre 12/20 et 17/20.

La question 2, illustrée à la *Figure 2*, présente cependant beaucoup plus de résultats sous 10/20, plus de 30 étudiants sont dans ce cas, ce qui est nettement supérieur au nombre d’échecs à la question 1. A nouveau, très peu d’étudiants ont obtenu une cote supérieure à 17/20. La majorité des étudiants se situe entre 10/20 et 16/20. Cette question est donc moins bien réussie que la première.

La *Figure 3* nous expose le grand nombre d’échecs à la question 3. En effet, approximativement une cinquantaine d’étudiants ont obtenu un résultat inférieur à 10/20. Il y a très peu de cotes supérieures à 14/20, la majorité étant entre 11/20 et 14/20. Cette question est donc la moins réussie des trois.

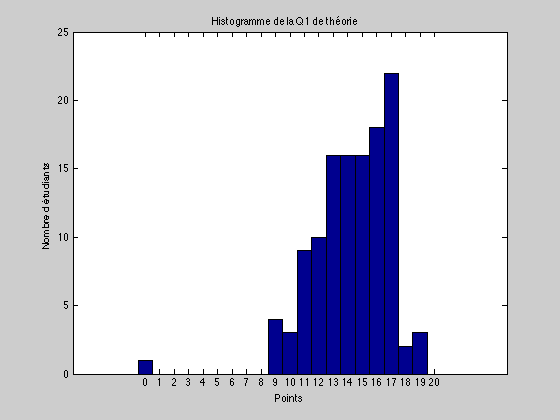


Figure 1 : Histogramme des résultats de la question 1 de théorie

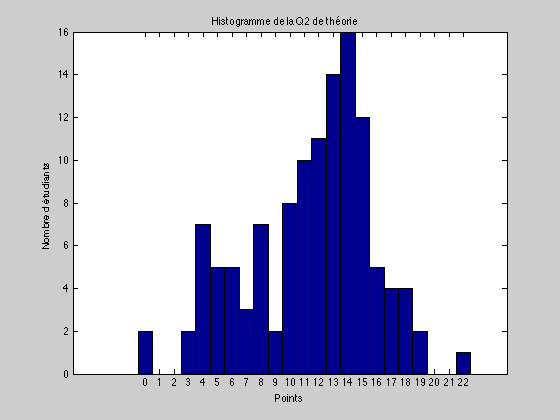


Figure 2 : Histogramme des résultats de la question 2 de théorie

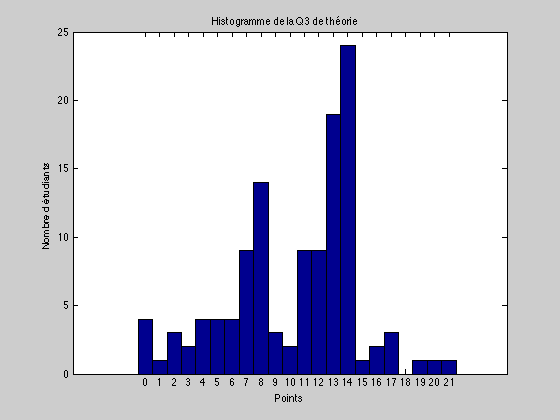


Figure 3 : Histogramme des résultats de la question 3 de théorie

**(b)**

Le *Tableau 1* montre que l’exercice 2 est le plus réussi avec une moyenne de 16.8/20 et une cote dominante de 20/20. L’exercice 3 est le moins réussi, il présente une moyenne beaucoup plus faible de 7.3/20 et un résultat dominant de 0/20. L’exercice 1 se situe entre les deux précédents avec une moyenne de 10.8/20 et une cote dominante de 12/20. On remarque que les médianes sont très proches des moyennes, ceci signifie que la répartition des résultats des étudiants est assez symétrique. Les trois écart-types ont des valeurs semblables, il y a donc une concentration assez similaire autour de la moyenne pour chaque exercice.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Moyennes | Médianes | Modes | Ecart-types |
| Exercice 1 | 10.8167 | 11 | 12 | 5.6717 |
| Exercice 2 | 16.8083 | 18 | 20 | 3.8025 |
| Exercice 3 | 7.7333 | 7.5 | 0 | 5.3038 |

Tableau 1 : Moyennes, médianes, modes et écart-types des 3 exercices

Les résultats normaux sont ceux compris dans l’intervalle : ***[Moyenne – écart-type ; moyenne + écart-type]*.** Le *Tableau 2* ci-dessous reprend les bornes inférieures et supérieures des différents intervalles.

|  |  |  |
| --- | --- | --- |
|  | Bornes inférieures | Bornes supérieures |
| Exercice 1 | 5.1450 | 16.4883 |
| Exercice 2 | 13.0059 | 20.6108 |
| Exercice 3 | 2.4296 | 13.0371 |

Tableau 2 : Bornes inférieures et supérieures des intervalles de résultats normaux

Pour ces différents intervalles, nous obtenons 65% étudiants ayant réalisé un résultat normal pour l’exercice 1, 83.3% pour le second et 65.83% pour le troisième.

**(c)**

Les trois boites à moustaches sont disponibles aux trois figures suivantes. Les quartiles sont repris dans le *Tableau 3* suivant et sont les deux barres bleues horizontales ainsi que la barre rouge sur le graphe. Pour le projet 1, on peut remarquer que le troisième quartile et la médiane (le deuxième quartile) sont confondus, l’abondance 18/20 en est la cause. La médiane est représentée par la droite rouge horizontale. Il y a quelques données aberrantes qui sont représentées par des croix rouges.

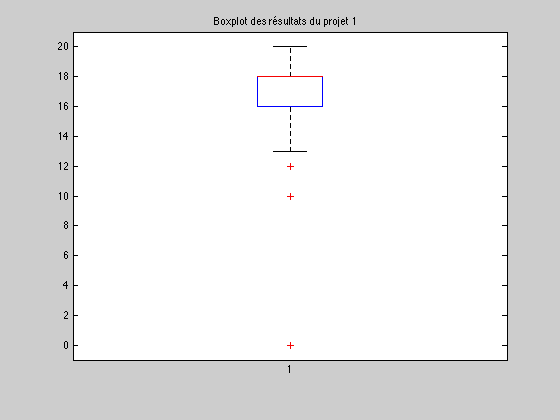


Figure 4 : Boxplot des résultats du projet 1

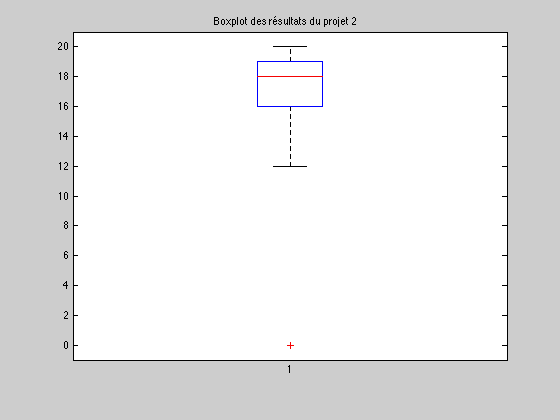


Figure 5 : Boxplot des résultats du projet 2

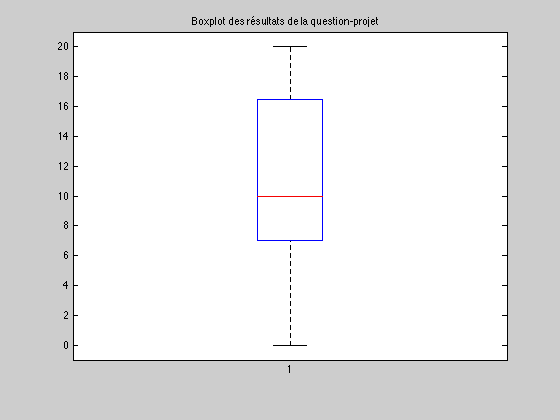


Figure 6 : Boxplot des résultats de la question sur le projet 2

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1er quartile | 2eme quartile | 3eme quartile |
| Projet 1 | 16 | 18 | 18 |
| Projet 2 | 16 | 18 | 19 |
| Q projet 2 | 7 | 10 | 16.5 |

Tableau 3 : Quartiles des deux projets et de la question sur le projet 2

**(d)**

Les figures ci-dessous représentent les polygones des fréquences cumulées pour la théorie et pour les exercices. A l’aide de ces graphes, on remarque qu’environ 43% des étudiants ont obtenu une moyenne des questions de théorie entre 12/20 et 15/20 et qu’environ 30% ont obtenus une moyenne d’exercices dans ce même intervalle.

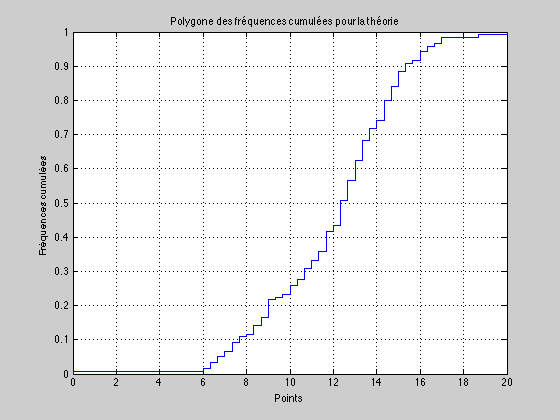
****

Figure 7 : Polygones des fréquences cumulées pour la théorie

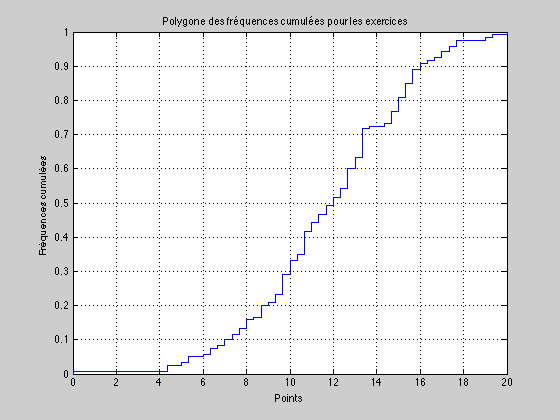


Figure 8 : Polygones des fréquences cumulées pour les exercices

**(e)**

La *Figure 9* illustre graphe de dispersion comparant les résultats du projet 2 aux résultats de la question sur le projet 2. Le coefficient de corrélation vaut 0.1407 ce qui est assez faible. Ceci signifie que les résultats du projet 2 expliquent peu les résultats à la question sur ce projet, ces deux variables sont donc peu liées.

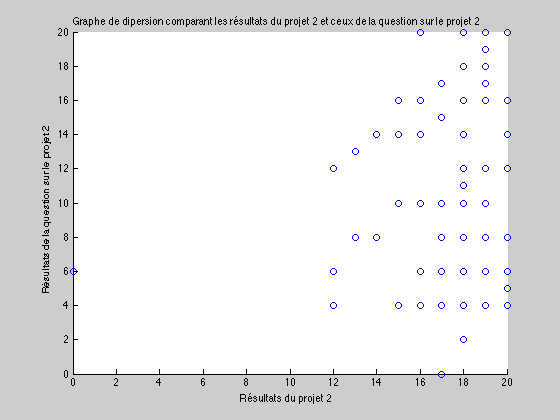
****

Figure 9 : graphe de dispersion comparant les résultats du projet 2 aux résultats de la question sur le projet 2

**Génération d’échantillons i.i.d.**

**Question 2**

**(a)**

**(i)**

Les moyennes, médianes et écart-types de l’échantillon pour les trois exercices sont disponibles dans le *Tableau 4* ci-dessous. On remarque, en comparant au *Tableau 1,* que les moyennes de l’échantillon diffèrent de 0.4 à 1.1 points des moyennes de la population. Les médianes de l’exercice 2 sont identiques, les autres diffèrent de 0.5 à 1 point. Les écart-types sont également différents de 0.3 à 1.2 points par rapport à ceux décrivant la population. Cet échantillon ne représente donc pas tout à fait la population. On peut donc conclure qu’un seul échantillon de 20 étudiants ne suffit pas à représenter la population entière.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Moyennes | Médianes | Ecart-types |
| Exercice 1 | 11.2 | 10 | 6.0228 |
| Exercice 2 | 17.9 | 18 | 2.5935 |
| Exercice 3 | 8.1 | 7 | 5.7574 |

Tableau 4 : Moyennes, médianes et écart-types des 3 exercices pour l’échantillon

**(ii)**

Les différentes boîtes à moustaches sont représentées aux figures suivantes. Celles-ci sont très différentes de celles représentant la population. Ceci nous permet de conclure, à nouveau, que cet échantillon n’est pas représentatif de la population.

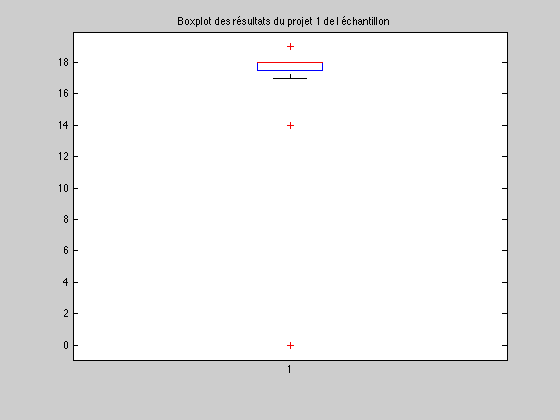
****

Figure 10 : Boxplot des résultats du projet 1 de l’échantillon

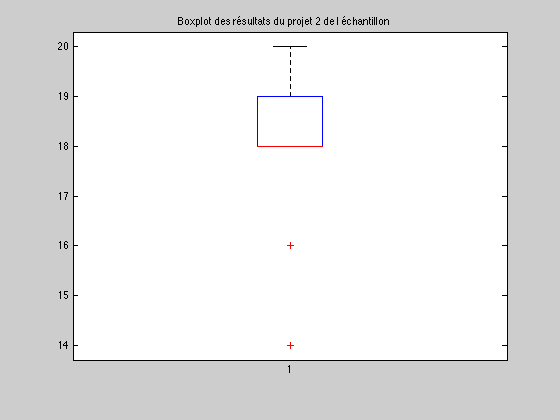
****

Figure 11 : Boxplot des résultats du projet 2 de l’échantillon

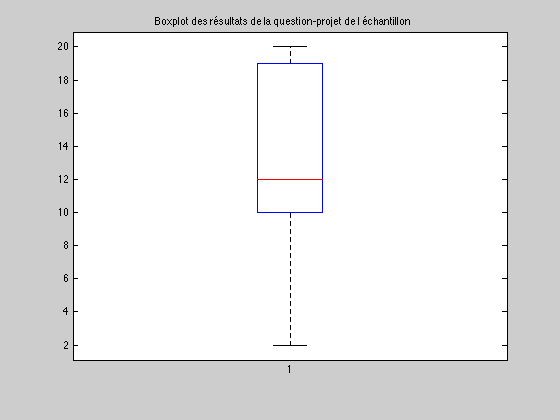
****

Figure 12 : Boxplot des résultats de la question sur le projet 2 de l’échantillon

**(iii)**

La *Figure 13* représente les polygones des fréquences cumulées pour l’échantillon et la population. L’échantillon est tracé en bleu tandis que la population est en rouge. Pour cet échantillon, les deux courbes sont assez similaires mais ne se confondent pas car l’échantillon comprend moins d’étudiants que la population. La distance de Kolmogorov-Smirnov séparant ces deux graphes est de 0.0667, ceci confirme donc la proximité des deux polygones.

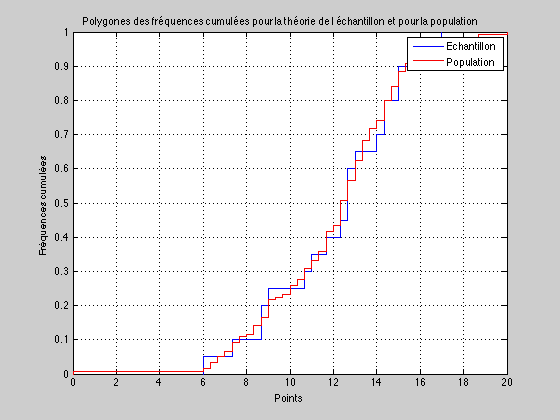
****

Figure 13 : Polygones des fréquences cumulées pour la théorie

**(b)**

**(i)**

L’histogramme des moyennes de l’exercice 1 pour les 100 échantillons est représenté à la *Figure 14*. L’allure de cet histogramme est une loi normale. En effet, ce graphe semble symétrique autour d’une moyenne de 11/20. La moyenne de cette nouvelle variable est de 11.1/20 ce qui est très proche de la moyenne de la population à savoir 10.8/20.

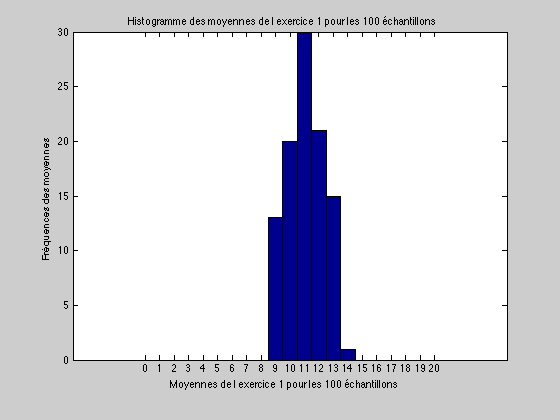


Figure 14 : Histogramme des moyennes de l’exercice 1 pour les 100 échantillons

**(ii)**

La *Figure 15* représente l’histogramme des médianes de l’exercice 1 pour les 100 échantillons. Ce graphe n’a l’allure d’aucune loi connue. En effet, il est parsemé de trous et ne comporte aucune symétrie. La moyenne de cette nouvelle variable est de 10.9/20 ce qui est encore plus proche de la moyenne de la population (10.8/20) que le résultat précédent.

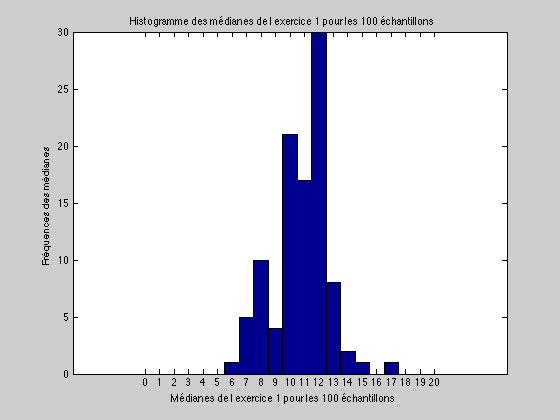


Figure 15 : Histogramme des médianes de l’exercice 1 pour les 100 échantillons

**(iii)**

L’histogramme des écart-types de l’exercice 1 pour les 100 échantillons est disponible à la *Figure 16*. Cet histogramme présente une allure de loi normale, il semble symétrique mais pas parfaitement, cette allure est moins marquée que pour l’histogramme des moyennes mais plus que pour celui des médianes. La moyenne de cette nouvelle variable est de 5.6 alors qu’elle était de 5.7 pour la population, ces deux résultats sont très proches, ceci confirme à nouveau l’allure de loi normale.

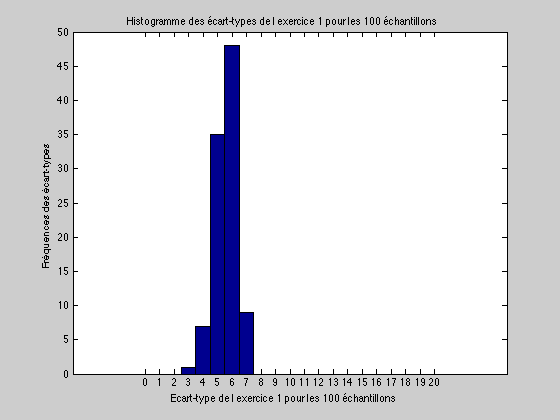


Figure 16 : Histogramme des écart-types de l’exercice 1 pour les 100 échantillons

**(iv)**

La *Figure 17* illustre l’histogramme des distances de Kolmogorov-Smirnov entre les polygones des fréquences cumulées de la population et des 100 échantillons pour l’exercice 1.

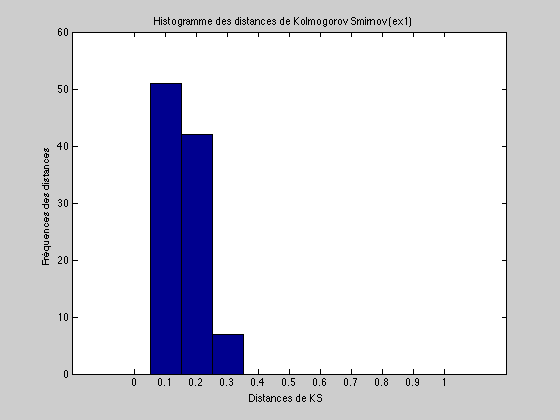


Figure 17 : Histogramme des distances de Kolmogorov Smirnov (exercice1)

**(v)**

Les *Figures 18* et*19* représentent les histogrammes des distances de Kolmogorov-Smirnov entre les polygones des fréquences cumulées de la population et des 100 échantillons pour les exercices 2 et 3. On remarque que les graphes pour les exercices 1 et 3 sont très similaires, on peut donc en conclure qu’ils suivent la même loi de répartition. Le maximum à 0.1 du graphe relatif à l’exercice 2 est nettement plus élevé que pour les deux autres exercices, les échantillons sélectionnés pour cet exercice sont donc représentatifs de la population.

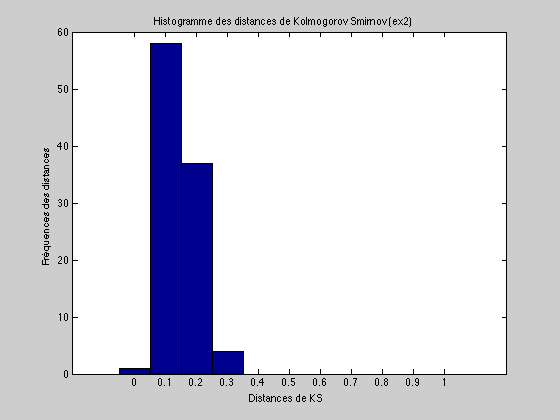


Figure 18 : Histogramme des distances de Kolmogorov Smirnov (exercice2)

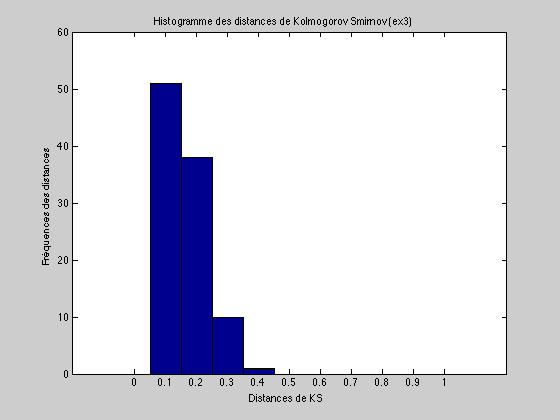


Figure 19 : Histogramme des distances de Kolmogorov Smirnov (exercice3)

**Annexes**

**Q1A**

function Q1A

resultats=xlsread('Proba1ereSession20132014.xls'); %rÈcupÈration des donnÈes Excel

points=0:22; %vecteur abscisse

hist(resultats(:,4),points); %fonction hist trace l'histogramme des donnÈes

set(gca,'xtick',[0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20]); % commande pour afficher toutes les graduations

xlabel('Points');

ylabel('Nombre d Ètudiants');

title('Histogramme de la Q1 de thÈorie');

figure

hist(resultats(:,5),points);

set(gca,'xtick',[0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22]); % commande pour afficher toutes les graduations

xlabel('Points');

ylabel('Nombre d Ètudiants');

title('Histogramme de la Q2 de thÈorie');

figure

hist(resultats(:,6),points);

set(gca,'xtick',[0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21]); % commande pour afficher toutes les graduations

xlabel('Points');

ylabel('Nombre d Ètudiants');

title('Histogramme de la Q3 de thÈorie');

end

**Q1B**

function Q1B

resultats=xlsread('Proba1ereSession20132014.xls');%rÈcupÈration des donnÈes Excel

exercice1=resultats(:,7); %Stockage des rÈsultats de l'exercice 1

exercice2=resultats(:,8); %Stockage des rÈsultats de l'exercice 2

exercice3=resultats(:,9); %Stockage des rÈsultats de l'exercice 3

%Calcul des moyennes :

moyenne\_exercice1=mean(exercice1) %la fonction mean calcule la moyenne d'un vecteur

moyenne\_exercice2=mean(exercice2)

moyenne\_exercice3=mean(exercice3)

%Calcul des mÈdianes :

mediane\_exercice1=median(exercice1) %la fonction median calcule la mÈdiane d'un vecteur

mediane\_exercice2=median(exercice2)

mediane\_exercice3=median(exercice3)

%Calcul des modes :

mode\_exercice1=mode(exercice1) %%la fonction mode calcule la mÈdiane d'un vecteur

mode\_exercice2=mode(exercice2)

mode\_exercice3=mode(exercice3)

%Calcul des Ècarts types :

ecart\_type\_exercice1=std(exercice1) %la fonction std calcule l'Ècart-type d'un vecteur

ecart\_type\_exercice2=std(exercice2)

ecart\_type\_exercice3=std(exercice3)

%RÈsultats normaux :

min\_exercice1 = moyenne\_exercice1-ecart\_type\_exercice1; %bornes de l'intervalle

min\_exercice2 = moyenne\_exercice2-ecart\_type\_exercice2;

min\_exercice3 = moyenne\_exercice3-ecart\_type\_exercice3;

max\_exercice1 = moyenne\_exercice1+ecart\_type\_exercice1;

max\_exercice2 = moyenne\_exercice2+ecart\_type\_exercice2;

max\_exercice3 = moyenne\_exercice3+ecart\_type\_exercice3;

elevesnormaux\_exercice1=0;

elevesnormaux\_exercice2=0;

elevesnormaux\_exercice3=0;

nombre\_etudiants=max(size(resultats));

for i=1:nombre\_etudiants

if exercice1(i)>=min\_exercice1 && exercice1(i)<=max\_exercice1

elevesnormaux\_exercice1=elevesnormaux\_exercice1 + 1;

end

if exercice2(i)>=min\_exercice2 && exercice2(i)<=max\_exercice2

elevesnormaux\_exercice2=elevesnormaux\_exercice2 + 1;

end

if exercice3(i)>=min\_exercice3 && exercice3(i)<=max\_exercice3

elevesnormaux\_exercice3=elevesnormaux\_exercice3 + 1;

end

end

elevesnormaux\_exercice1/1.2

elevesnormaux\_exercice2/1.2

elevesnormaux\_exercice3/1.2

end

**Q1C**

function Q1C

resultats=xlsread('Proba1ereSession20132014.xls'); %rÈcupÈration des donnÈes Excel

boxplot(resultats(:,1)); %crÈe et affiche le boxplot

title('Boxplot des rÈsultats du projet 1');

figure

boxplot(resultats(:,2));

title('Boxplot des rÈsultats du projet 2');

figure

boxplot(resultats(:,3));

title('Boxplot des rÈsultats de la question-projet');

%Calcul des quartiles :

quartiles\_projet1 = quantile(resultats(:,1), [.25 .50 .75]) %la fonction quantile calcule les quartiles pour 25%, 50%, 75%

quartiles\_projet2 = quantile(resultats(:,2), [.25 .50 .75])

quartiles\_qprojet = quantile(resultats(:,3), [.25 .50 .75])

end

**Q1D**

function moyenne\_theorie = Q1D

resultats=xlsread('Proba1ereSession20132014.xls'); %rÈcupÈration des donnÈes Excel

%Calcul des moyennes :

moyenne\_theorie = (resultats(:,4) + resultats(:,5) + resultats(:,6))/3;

moyenne\_exercices = (resultats(:,7) + resultats(:,8) + resultats(:,9))/3;

%Polygones des frÈquences cumulÈes :

cdfplot(moyenne\_theorie);

xlabel('Points');

ylabel('FrÈquences cumulÈes');

title('Polygone des frÈquences cumulÈes pour la thÈorie');

figure

cdfplot(moyenne\_exercices);

xlabel('Points');

ylabel('FrÈquences cumulÈes');

title('Polygone des frÈquences cumulÈes pour les exercices');

end

**Q1E**

function Q1E

resultats=xlsread('Proba1ereSession20132014.xls'); %rÈcupÈration des donnÈes Excel

scatter(resultats(:,2),resultats(:,3)); % Scatter dessine le graphe de dispersion comparant les rÈsultats du projet 2 ‡ ceux de la question sur le projet 2

xlabel('RÈsultats du projet 2');

ylabel('RÈsultats de la question sur le projet 2');

title('Graphe de dipersion comparant les rÈsultats du projet 2 et ceux de la question sur le projet 2');

%Coefficient de corrÈlation

cc=corrcoef(resultats(:,2),resultats(:,3));

coefficient\_correlation=cc(1,2)

end

**Q2A**

function Q2A

resultats=xlsread('Proba1ereSession20132014.xls'); %rÈcupÈration des donnÈes Excel

echantillon = randsample(120,20,true); %CrÈation d'un Èchantillon de 20 Ètudiants

for i=1:20

exercice1(i)=resultats(echantillon(i),7); %rÈcupÈration des rÈsulats des exercices de ces 20 Ètudiants

exercice2(i)=resultats(echantillon(i),8);

exercice3(i)=resultats(echantillon(i),9);

projet1(i)=resultats(echantillon(i),1);

projet2(i)=resultats(echantillon(i),2);

qprojet2(i)=resultats(echantillon(i),3);

theorie1(i)=resultats(echantillon(i),4);

theorie2(i)=resultats(echantillon(i),5);

theorie3(i)=resultats(echantillon(i),6);

end

%Calcul des moyennes :

moyenne\_exercice1=mean(exercice1)

moyenne\_exercice2=mean(exercice2)

moyenne\_exercice3=mean(exercice3)

%Calcul des mÈdianes :

mediane\_exercice1=median(exercice1)

mediane\_exercice2=median(exercice2)

mediane\_exercice3=median(exercice3)

%Calcul des Ècart-types:

ecart\_type\_exercice1=std(exercice1)

ecart\_type\_exercice2=std(exercice2)

ecart\_type\_exercice3=std(exercice3)

%Boxplot

boxplot(projet1); %crÈe et affiche le boxplot

title('Boxplot des rÈsultats du projet 1 de l Èchantillon');

figure

boxplot(projet2);

title('Boxplot des rÈsultats du projet 2 de l Èchantillon');

figure

boxplot(qprojet2);

title('Boxplot des rÈsultats de la question-projet de l Èchantillon');

%Moyenne de chaque Ètudiant de l'Èchantillon pour la thÈorie

moyenne\_theorie=(theorie1(:)+theorie2(:)+theorie3(:))/3;

%Polygone des frÈquences cumulÈes

moyenne\_theoriepopu=Q1D;

cdfplot(moyenne\_theorie);

xlabel('Points');

ylabel('FrÈquences cumulÈes');

title('Polygone des frÈquences cumulÈes pour la thÈorie de l Èchantillon');

figure

cdfplot(moyenne\_theorie);

hold on

h=cdfplot(moyenne\_theoriepopu);

set(h,'color','r')

xlabel('Points');

ylabel('FrÈquences cumulÈes');

title('Polygones des frÈquences cumulÈes pour la thÈorie de l Èchantillon et pour la population');

%Distance de Kolmogorov Smirnov

[~, ~, distance\_ks]=kstest2(moyenne\_theoriepopu, moyenne\_theorie);

distance\_ks

end

**Q2B**

function Q2B

resultats=xlsread('Proba1ereSession20132014.xls'); %rÈcupÈration des donnÈes Excel

matrice\_echantillon=zeros(20,100); %Matrice 200 lignes/100 colonnes pour stocker les Èchantillons

for i=1:100

echantillon = randsample(120,20,true); %CrÈation d'un Èchantillon de 20 Ètudiants

matrice\_echantillon(:,i)=echantillon; %Remplissage de la matrice contenant les 100 Èchantillons, une colonne = un Èchantillon

end

%Calcul des moyennes (et nouvelle variable)

points\_exercice1=zeros(20,100);

points\_exercice2=zeros(20,100);

points\_exercice3=zeros(20,100);

for j=1:100

for k=1:20

points\_exercice1(k,j) = resultats(matrice\_echantillon(k,j),7);

points\_exercice2(k,j) = resultats(matrice\_echantillon(k,j),8);

points\_exercice3(k,j) = resultats(matrice\_echantillon(k,j),9);

end

moyenne\_exercice1(j)=mean(points\_exercice1(:,j));

mediane\_exercice1(j)=median(points\_exercice1(:,j));

ecart\_type\_exercice1(j)=std(points\_exercice1(:,j));

end

%Histogramme de la variable moyenne

moyennes=0:20; %vecteur abscisse

hist(moyenne\_exercice1,moyennes); %fonction hist trace l'histogramme des donnÈes

set(gca,'xtick',[0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20]); % commande pour afficher toutes les graduations

xlabel('Moyennes de l exercice 1 pour les 100 Èchantillons');

ylabel('FrÈquences des moyennes');

title('Histogramme des moyennes de l exercice 1 pour les 100 Èchantillons');

%Moyenne de la moyenne

moyenne\_moyenne=mean(moyenne\_exercice1)

%Histogramme de la variable mediane

medianes=0:20; %vecteur abscisse

figure

hist(mediane\_exercice1,medianes); %fonction hist trace l'histogramme des donnÈes

set(gca,'xtick',[0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20]); % commande pour afficher toutes les graduations

xlabel('MÈdianes de l exercice 1 pour les 100 Èchantillons');

ylabel('FrÈquences des mÈdianes');

title('Histogramme des mÈdianes de l exercice 1 pour les 100 Èchantillons');

%Moyenne de la mÈdiane

moyenne\_mediane=mean(mediane\_exercice1)

%Histogramme de la variable Ècart-type

ecart\_type=0:20; %vecteur abscisse

figure

hist(ecart\_type\_exercice1,ecart\_type); %fonction hist trace l'histogramme des donnÈes

set(gca,'xtick',[0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20]); % commande pour afficher toutes les graduations

xlabel('Ecart-type de l exercice 1 pour les 100 Èchantillons');

ylabel('FrÈquences des Ècart-types');

title('Histogramme des Ècart-types de l exercice 1 pour les 100 Èchantillons');

%Moyenne de l'Ècart-type

moyenne\_ecart\_type=mean(ecart\_type\_exercice1)

%Distance Kolmogorov-Smirnov pour l'exercice 1

for m=1:100

[~, ~, dks]=kstest2(resultats(:,7), points\_exercice1(:,m));

distance\_ks(m)=dks;

end

%Histogramme distance Kolmogorov-Smirnov (ex1)

distances=0:0.1:1; %vecteur abscisse

figure

hist(distance\_ks,distances); %fonction hist trace l'histogramme des donnÈes

set(gca,'xtick',[0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1]); % commande pour afficher toutes les graduations

xlabel('Distances de KS');

ylabel('FrÈquences des distances');

title('Histogramme des distances de Kolmogorov Smirnov (ex1)');

%Distance Kolmogorov-Smirnov pour l'exercice 2

for m=1:100

[~, ~, dks2]=kstest2(resultats(:,8), points\_exercice2(:,m));

distance\_ks2(m)=dks2;

end

%Histogramme distance Kolmogorov-Smirnov (ex2)

distances=0:0.1:1; %vecteur abscisse

figure

hist(distance\_ks2,distances); %fonction hist trace l'histogramme des donnÈes

set(gca,'xtick',[0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1]); % commande pour afficher toutes les graduations

xlabel('Distances de KS');

ylabel('FrÈquences des distances');

title('Histogramme des distances de Kolmogorov Smirnov (ex2)');

%Distance Kolmogorov-Smirnov pour l'exercice 3

for m=1:100

[~, ~, dks3]=kstest2(resultats(:,9), points\_exercice3(:,m));

distance\_ks3(m)=dks3;

end

%Histogramme distance Kolmogorov-Smirnov (ex2)

distances=0:0.1:1; %vecteur abscisse

figure

hist(distance\_ks3,distances); %fonction hist trace l'histogramme des donnÈes

set(gca,'xtick',[0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1]); % commande pour afficher toutes les graduations

xlabel('Distances de KS');

ylabel('FrÈquences des distances');

title('Histogramme des distances de Kolmogorov Smirnov (ex3)');

end