

UNIVERSITÉ DE LIÈGE



ELÉMENTS DE STATISTIQUES

---

## Projet : rapport

---

3<sup>ÈME</sup> BACHELIER EN INGÉNIEUR CIVIL

*Auteurs :*  
Antoine LOUIS  
Tom CRASSET

*Professeur :*  
L. WEHENKEL

Année académique 2017-2018

# 1 Analyse descriptive

## 1.a Histogrammes

Les histogrammes de la consommation de bière et d'alcool fort dans le monde sur une période d'un an sont données à la Figure 1. Nous constatons que la consommation de canettes de bière par personne est assez faible : entre 0 et 25 canettes pour à peu près un tiers des pays repris dans l'enquête. Les deux autres tiers se répartissent principalement dans des consommations variant de 25 à 200 canettes de bière.

En ce qui concerne la consommation d'alcool fort, les résultats sont plus frappants : plus de la moitié des pays concernés par l'enquête possède une consommation d'alcool fort moyenne par habitant comprise entre 0 et 25 shots par an.

D'une manière générale, la consommation mondiale de bière et d'alcool fort semble assez faible.

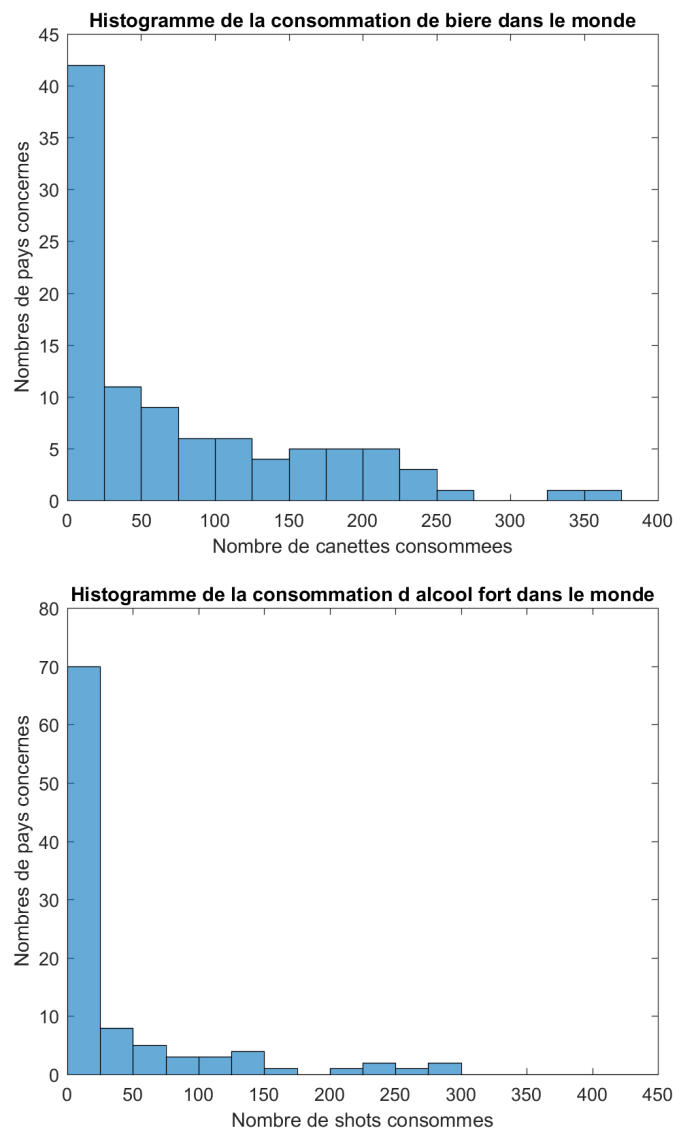


FIGURE 1 – Consommation de bière et d'alcool fort dans le monde

## 1.b Moyenne, médiane, mode et écart-type

Les moyennes, médianes, modes et écart-types de la consommation de bière et d'alcool fort sont calculées à l'aide des fonctions incluses dans Matlab, soient respectivement **mean**, **median**, **mode** et **std**). Les résultats sont donnés dans la Table 1, en nombre de canettes pour la bière et en nombre de shots pour l'alcool fort.

	Moyenne	Médiane	Mode	Écart-type
Bière	98.6	59	0	96.5928
Alcool fort	77.17	41.5	0	91.8158

TABLE 1 – Moyenne, médiane, mode et écart-type de la consommation de bière et d'alcool fort

On constate que plus de la moitié des pays ont une consommation annuelle de bières par habitant inférieure ou égale à 59 canettes, ce qui équivaut approximativement à une canette par semaine. Comparé à la Belgique, qui possède une consommation de 295 canettes par an par habitant, soit environ 5 canettes et demi par semaine, cela paraît relativement faible. On note également que la consommation de bière la plus rencontrée pour chaque pays est une consommation nulle de canettes.

En terme d'alcool fort, on remarque que plus de la moitié des pays ont une consommation d'alcool fort inférieure ou égale à 41 shots et demi par habitant avec une majorité où les habitants n'en consomment aucun. La Belgique possède une consommation d'alcool fort relativement proche de la moyenne globale, soit 84 shots par habitant par an. Toutefois, ce résultat est quand même deux fois supérieur aux résultats de plus de la moitié des pays pris en compte.

Nous pouvons donc dire que la Belgique a une consommation annuelle de bière et d'alcool fort par habitant bien supérieure à la moyenne mondiale.

## 1.c Consommation "normale"

Les résultats normaux sont ceux compris dans l'intervalle  $[\mu_i - \sigma_i, \mu_i + \sigma_i]$  où  $\mu_i$  est une moyenne et  $\sigma_i$  l'écart-type correspondant. On trouve alors pour la consommation :

- De bière : 71% des pays dans l'intervalle  $[2.0072, 195.1928]$ .
- D'alcool fort : 82 % des pays dans l'intervalle  $[-14.6458, 168.9858]$ .

On en déduit que la Belgique n'a pas une consommation de bière "normale" car cette dernière (295 canettes par an par habitant) n'est pas comprise dans l'intervalle de la loi normale. Toutefois, elle possède une consommation d'alcool fort "normale", 84 étant bien compris dans l'intervalle "normal" de la consommation d'alcool fort.

## 1.d Boîtes à moustaches

Les boîtes à moustache de la consommation de bière et d'alcool fort sont représentés à la Figure 2. Les quartiles demandés sont repris dans la Table 2.

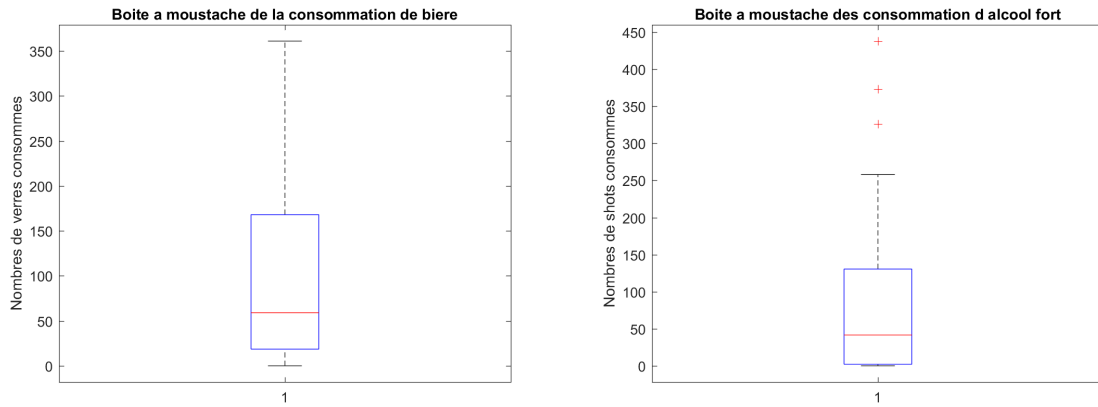


FIGURE 2 – Boites à moustaches relatives à la consommation de bière et d’alcool fort dans le monde

	1 <sup>er</sup> quartile	2 <sup>ème</sup> quartile	3 <sup>ème</sup> quartile
Bière	18.5	59	168
Alcool fort	2	41.5	130.5

TABLE 2 – Quartiles de la consommation de bière et d’alcool fort

Nous remarquons que, dans le cas de l’alcool fort, il y a trois données aberrantes représentées par des croix rouges.

## 1.e Polygone des fréquences cumulées

Le polygone des fréquences cumulées de la consommation de bière est donné à la Figure 3 ci-dessous.

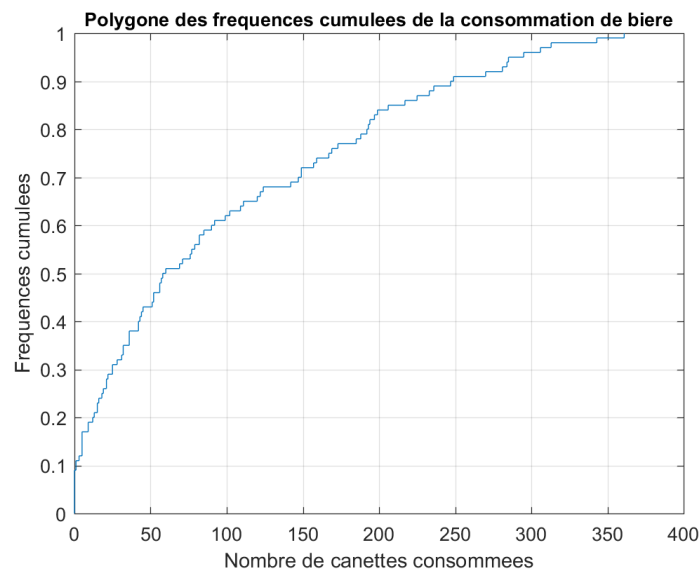


FIGURE 3 – Polygone des fréquences cumulées la consommation de bière

La proportion des pays ayant une consommation comprises dans un certain intervalle  $[a, b]$  est obtenue de telle façon que :

$$F(a \leq x_i \leq b) = F(x_i = b) - F(x_i = a)$$

Ainsi, en observant le graphique, on déduit qu'environ 12% des pays se situent dans l'intervalle de consommation  $[200, 295]$ .

## 1.f Graphes de dispersion

Les graphes de dispersion comparant la consommation d'alcool pur à celle de la bière, du vin et d'alcool fort sont donnés respectivement aux Figures 4, 5 et 6.

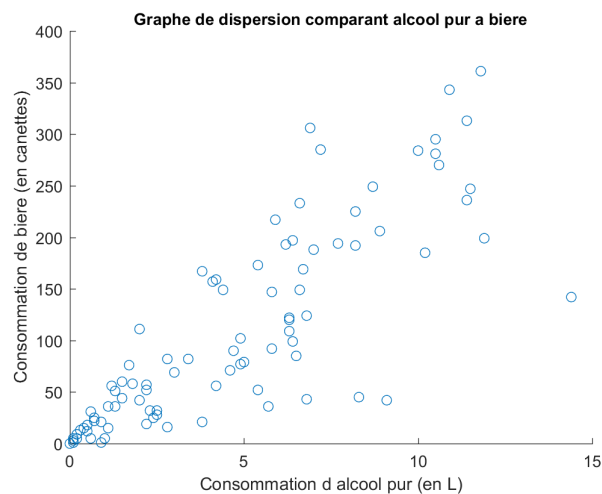


FIGURE 4 – Graphe de dispersion comparant la consommation d'alcool pur à la consommation de bière

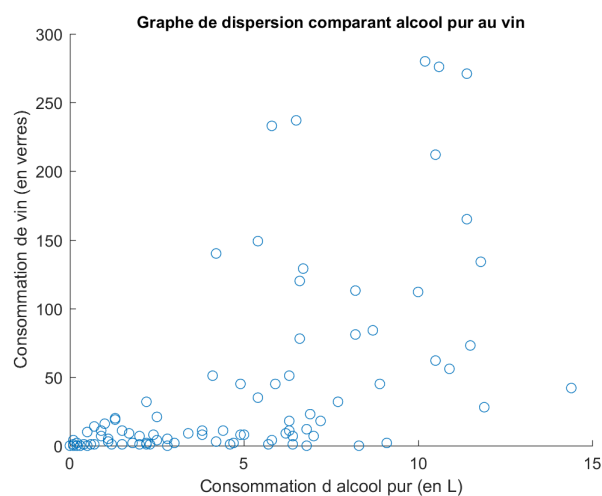


FIGURE 5 – Graphe de dispersion comparant la consommation d'alcool pur à la consommation de vin

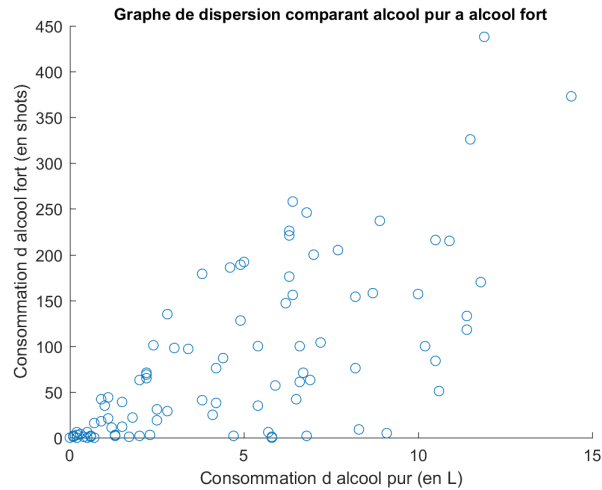


FIGURE 6 – Graphe de dispersion comparant la consommation d’alcool pur à la consommation d’alcool fort

Les coefficients de corrélation obtenus sont donnés à la Table 3 ci-dessous. on note que, d’une façon générale, les coefficients sont relativement élevés, et donc qu’il existe un lien non négligeable entre chacune des paires de variables. C’est principalement le cas lors de la comparaison entre la consommation d’alcool pur avec la consommation de bières. Nous avons un coefficient de corrélation très élevé, ce qui signifie que les deux variables sont fortement liées et donc que la consommation de bière explique bien celle de l’alcool pur.

	Avec bière	Avec vin	Avec alcool fort
Coefficient de corrélation	0.8457	0.5930	0.7107

TABLE 3 – Coefficients de corrélation entre l’alcool pur et les trois types d’alcool

## 2 Calcul de statistiques sur échantillons

### 2.a Échantillon i.i.d unique

(i)

Pour générer un échantillon de 20 pays, nous avons utilisé la fonction **randsample** implémentée dans Matlab. Comme dans la question 1, nous avons utilisé les fonctions incluses dans Matlab (**mean**, **median**, **std**) pour calculer respectivement la moyenne, la médiane et l’écart-type. Les résultats sont affichés dans la Table 4 qui suit.

	Moyenne	Médiane	Écart-type
Bière	93.03	69.50	94.52
Alcool fort	93.85	70.00	104.95

TABLE 4 – Moyenne, médiane et écart-type de la consommation de bière et d’alcool fort d’un échantillon de 20 pays

En comparant au tableau, nous pouvons remarquer que notre échantillon est assez représentatif de la consommation de bière. Par contre, ce n'est pas vraiment le cas pour l'alcool fort. Ceci dit, cela est surtout dû au hasard, au fait qu'on ait choisi dans l'échantillon des pays qui avaient une consommation d'alcool fort assez différente de celles de la population.

(ii)

Les boîtes à moustache de la consommation de bière et d'alcool fort sont représentés à la Figure 7.

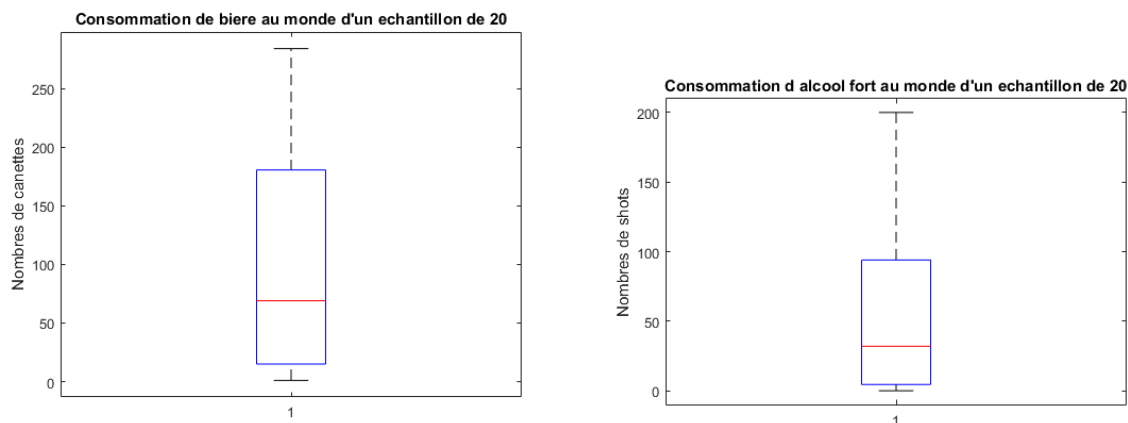


FIGURE 7 – Boîtes à moustaches de la consommation de bière et d'alcool fort de l'échantillon

On peut voir sur la boîte à moustache relative à la consommation de bière que la valeur minimale est 0 et que la valeur max est autour des 290 canettes de bière et pour l'alcool fort ces valeurs sont 0 et 200 shots d'alcool fort. Nous pouvons aussi remarquer, en comparant avec la boîte à moustache de la Figure 2 que les valeurs aberrantes sont moins présentes dans l'échantillon. Cela peut venir du fait que les valeurs aberrantes de la population n'ont pas été tirés dans notre échantillon ou alors du fait que, pour l'échantillon, ces valeurs ne sont pas considérées comme aberrantes.

(iii)

Ci dessous à la Figure 8 sont représenté les polygones des fréquences cumulées de la population et de l'échantillon, une fois pour la bière et une fois pour l'alcool fort.

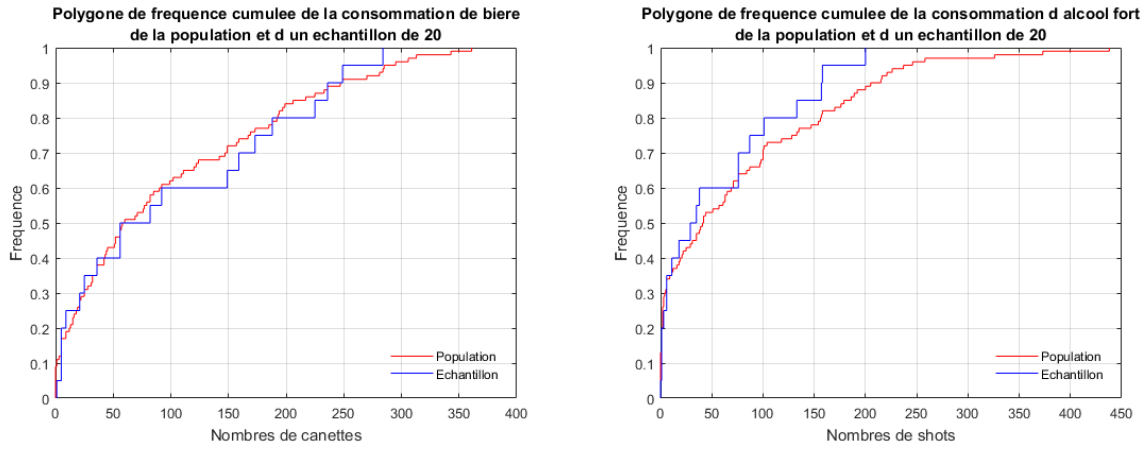


FIGURE 8 – Polygone de fréquences cumulées de la population et de l'échantillon pour la bière et l'alcool fort

On remarque que le polygone des fréquences cumulées de l'échantillon suit bien la tendance de celui de la population, bien que le pas d'escaliers est moindre, ce qui est trivial. De plus, nous avons la distance de Kolmogorov-Smirnov qui, pour la bière vaut 0.1 et pour l'alcool fort vaut 0.13. Ceci confirme la représentation graphique et nous indique que, en effet, l'échantillon fait une bonne approximation de la population.

## 2.b Échantillons i.i.d multiples

De nouveau, nous avons généré un échantillon de 20 pays, avec la fonction **rand-sample**. Par contre, nous avons répété cette action 100 fois.

La Figure 9 représente la répartition de la consommation moyenne de bière et d'alcool fort de tout les échantillons.

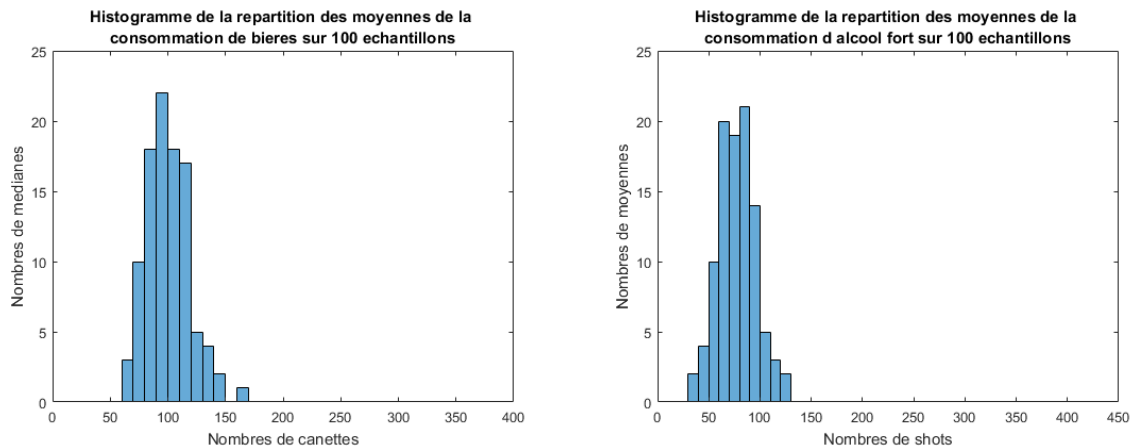


FIGURE 9 – Histogrammes des moyennes de consommation de bières et d'alcool fort de 100 échantillons

Nous voyons ci-dessus qu'un grand nombre d'échantillons ont une moyenne qui est proche de la valeur moyenne de la population, qui vaut 99.6 et 77.17 respectivement



pour la bière et pour l'alcool fort. Nous pouvons confirmer ça en prenant la moyenne des moyennes, cette dernière vaut 99.71 canettes et 78.45 shots. Nous pouvons aussi ajouter que l'allure de cet histogramme fait penser à une loi normale centrée en la moyenne qu'on vient de calculer, ce qui n'est guère surprenant.

(ii)

La Figure 10 représente la répartition de la consommation médiane de bière et d'alcool fort de tout les échantillons.

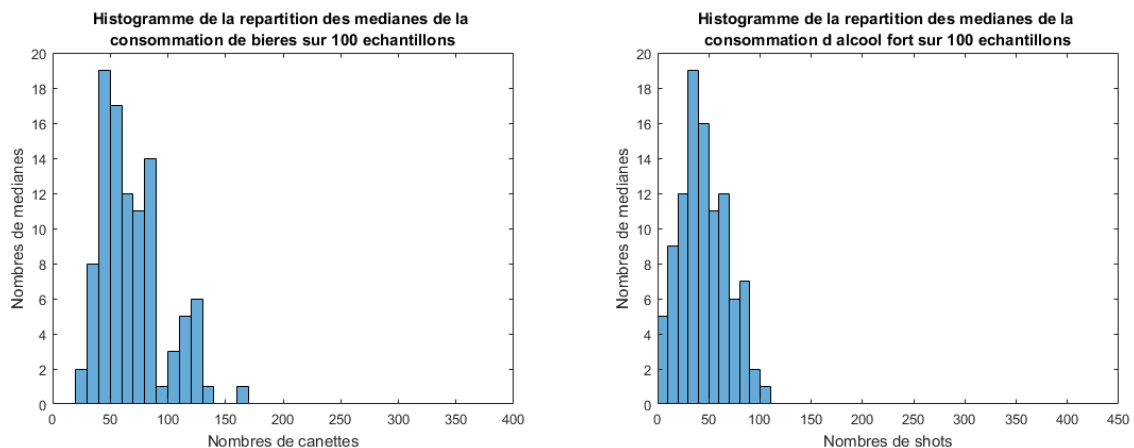


FIGURE 10 – Histogrammes des médianes de consommation de bières et d'alcool fort de 100 échantillons

D'une façon analogue, on prend la moyenne des valeurs médianes obtenues par tout les échantillons, ce qui nous donne 68.51 canettes et 45.76 shots respectivement pour la bière et l'alcool fort. Comparé à la population (59 canettes et 41.5 shots), c'est de nouveau fort similaire. Ici aussi, nous avons une distribution qui ressemble à une distribution normale.

(iii)

La Figure 10 représente la répartition de l'écart-type de la consommation de bière et d'alcool fort de tout les échantillons.

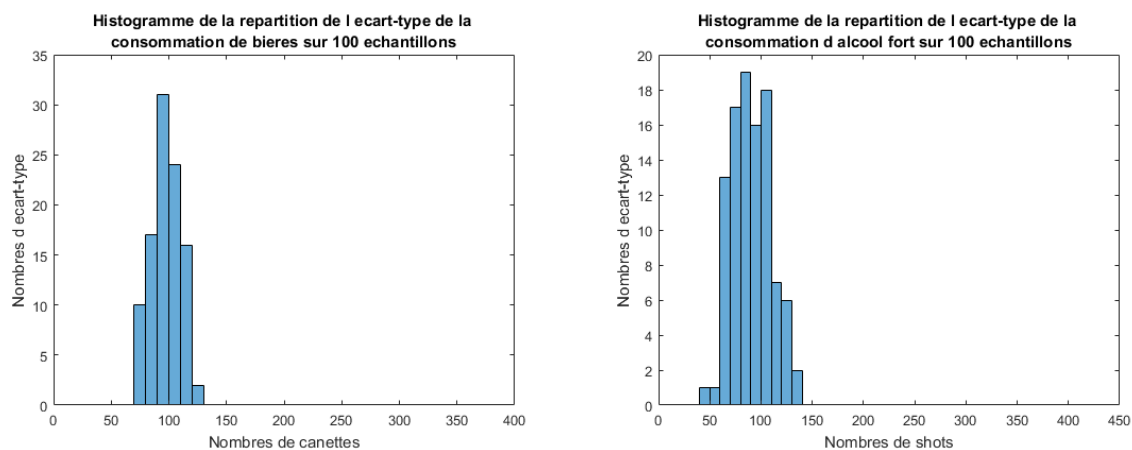


FIGURE 11 – Histogrammes des médianes de consommation de bières et d’alcool fort de 100 échantillons

Finalement, on refait la même procédure pour la moyenne de l’écart-type et on obtient comme valeur 97.42 canettes et 91.13 shots respectivement pour la bière et l’alcool fort. Comparé à la population (96.6 canettes et 91.8 shots), c’est de nouveau fort similaire. Ici aussi, nous avons une distribution centrée à la moyenne des écart-type qui ressemble à une distribution normale.

#### (iv - v)

A la Figure 12, nous avons représentés les quatre histogrammes qui représentent la répartition des distance de Kolmogorov-Smirnov entre le polygone des fréquences cumulées des 100 échantillons et celui de la population, pour toutes les boissons ainsi que pour le total d’alcool.

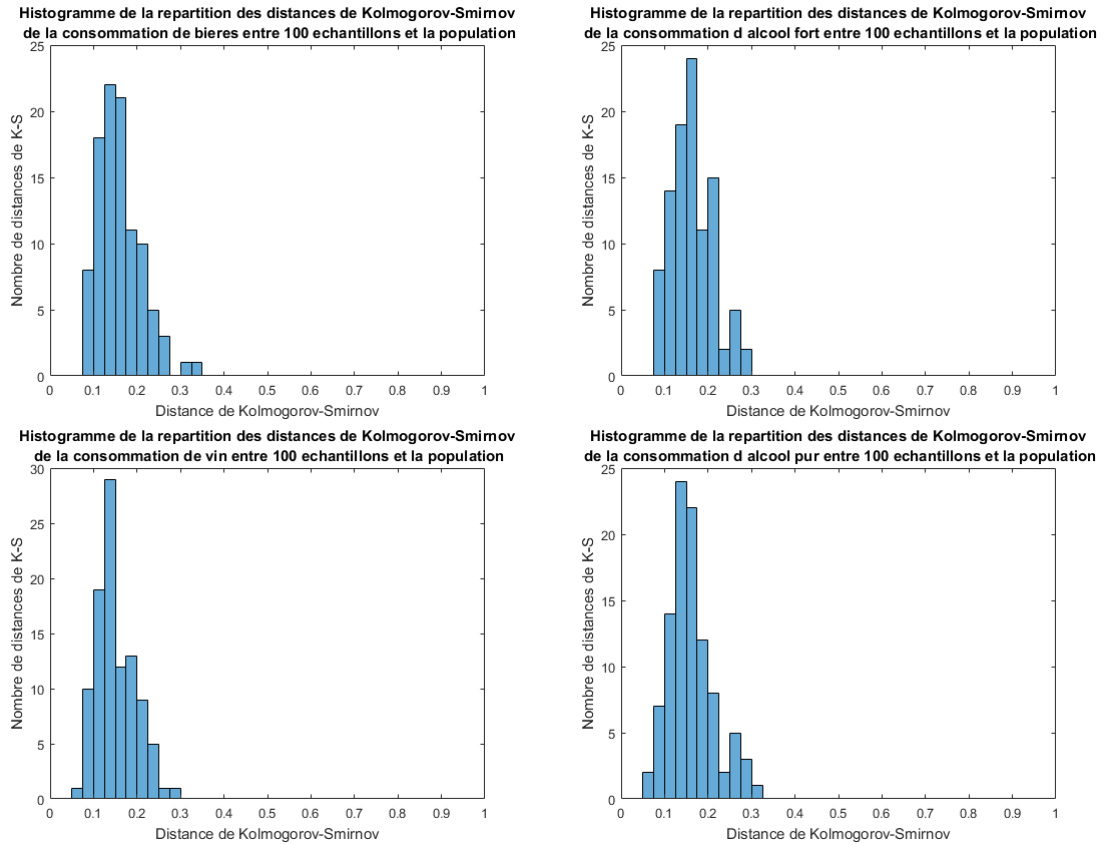


FIGURE 12 – Histogrammes des distances de Kolmogorov-Smirnov entre la population et les échantillons

En observant les quatre graphiques, nous pouvons conclure que ici aussi nous avons à faire à une distribution similaire à une loi normale. De plus, nous pouvons aussi remarquer que les distances de Kolmogorov-Smirnov sont assez faibles, ce qui veut dire que les approximations de la population faites par les échantillon sont assez bonnes, même déjà avec un échantillon de 20 pays. La distance de Kolmogorov-Smirnov est définie par la distance maximale entre deux courbes, ici la population d’une part et l’échantillon de l’autre, ce qui confirme la bonne approximation. Par contre, notons aussi que la distance n’est jamais 0 car la courbe de l’échantillon ne pourrait pas coller parfaitement à la courbe de la population.

### 3 Estimation

#### 3.a Biais et variance de l’estimateur $m_x$

Après avoir tiré 100 échantillons de 20 pays, nous calculons pour chaque échantillon la moyenne de la consommation de vin et nous plaçons chacune de ces moyennes dans une variable nommée `wine_mean_matrix`.

La moyenne de l’ensemble de ces moyennes vaut alors :  $m_x = 36.3745$  verres. Nous calculons à présent la moyenne réelle de la consommation de vin parmi tous les pays.

On obtient :  $M_x = 37.77$  verres. Ainsi, le biais de l'estimateur  $m_x$  de la consommation moyenne de vin de la population se calcule de la manière suivante :

$$b_{m_x} = m_x - M_x = -1.3955 \text{ verres}$$

Quant à la variance, elle vaut :  $v_{m_x} = 135.8917 \text{ verres}^2$

### 3.b Biais et variance de l'estimateur $median_x$

Après avoir tiré 100 échantillons de 20 pays, nous calculons la médiane de la consommation de vin de chacun d'entre eux et plaçons ces résultats dans une variable **wine\_median\_matrix**.

La moyenne de l'ensemble de ces médianes vaut :  $median_x = 8.455$  verres. Ainsi, le biais de l'estimateur  $median_x$  de la consommation moyenne de vin de la population est donné par :

$$b_{median_x} = median_x - M_x = -29.315 \text{ verres}$$

Quant à la variance, elle vaut :  $v_{median_x} = 26.7005 \text{ verres}^2$

### 3.c Échantillons i.i.d de taille 50

Les résultats obtenus sont donnés à la Table 5 ci-dessous.

	Biais	Variance
$m_x$	-0.8528	51.4263
$median_x$	-29.45	2.6076

TABLE 5 – Biais et variances des estimateurs  $m_x$  et  $median_x$  pour des échantillons i.i.d de 50 pays

On remarque que le biais et la variance diminuent tous les deux pour les deux estimateurs (sauf le biais de l'estimateur  $median_x$  qui reste approximativement le même). Cette diminution est tout à fait normale puisque des échantillons de taille plus importante résulteront en une meilleure traduction de l'ensemble des résultats de l'ensemble de la population.

### 3.d Estimation par intervalle de confiance

L'intervalle de confiance étant fixé à 95%, on s'attend à ce que le nombre d'intervalles de confiance qui contiennent la moyenne réelle, parmi les 100, soit proche de 95. Pour la loi de Student, les intervalles sont donnés par :

$$m_X - t_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq m_X + t_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

De la même manière, les intervalles calculés avec la loi de Gauss sont donnés par :

$$m_X - u_{1-\alpha/2} \frac{S_n}{\sqrt{n-1}} \leq \mu \leq m_X + u_{1-\alpha/2} \frac{S_n}{\sqrt{n-1}}$$

avec  $n = 20$  et  $\alpha = 0.05$ .

Après calculs, on trouve 94 et 92 intervalles contenant la moyenne réelle en utilisant respectivement la loi de Student et la loi de Gauss. Bien évidemment, ces nombres varient en fonction des échantillons tirés mais restent toujours proches de 95, ce qui prouve que l'hypothèse que la variable parente soit gaussienne est justifiée.

On note toutefois que, lorsque les intervalles sont calculés avec la loi de Gauss, le nombre d'intervalles contenant la moyenne réelle est souvent plus éloigné de 95. Ceci s'explique par le fait que l'échantillon est de taille 20 et donc que la loi de Student est plus adaptée au calcul d'intervalle de confiance. En effet, c'est seulement à partir d'une taille d'échantillon de 30 que les coefficients  $u_{1-\alpha/2}$  et  $t_{1-\alpha/2}$  prennent la même valeur et engendrent donc des résultats équivalents.

## 4 Tests d'hypothèse

### 4.a État belge

Nous considérons l'hypothèse  $H_0$  : "**la proportion de pays consommant plus de bière que la Belgique est de  $x\%$** " versus l'hypothèse alternative  $H_1$  : "**la proportion de pays consommant plus de bière que la Belgique est supérieure à  $x\%$** ". Ici,  $x\%$  est la vraie valeur qui résulte du dataset de la population et qui vaut dans notre cas 5%, c'est-à-dire que 5% des pays sont de plus gros buveurs de bières que la Belgique. Nous considérons ici  $\alpha = 0.05$  et nous faisons l'hypothèse qu'on a une variable de Bernoulli, on a donc  $\sigma = \sqrt{\frac{p(1-p)}{n}}$ . On a un test unilatéral à gauche et notre intervalle est donc donné par  $p \geq f - u_{1-\alpha} \sqrt{\frac{f(1-f)}{n}}$  où  $u_{1-\alpha} = 0.82894$ . Les autres données sont  $\alpha = 0.05$ ,  $n = 20$  et  $f = 0.05$ .

Après avoir fait tourner le script *Q4.m*, nous nous rendons compte que l'État belge ne rejette jamais l'hypothèse  $H_0$ . L'erreur que nous avons commise est le fait que nous avons approché une loi binomiale par une loi gaussienne.

### 4.b Autres instituts

De nouveau, dans aucun cas les autres instituts ne rejettent l'hypothèse.

### 4.c Éviter l'avantage des instituts indépendants

Afin d'éviter que l'État belge ne soit pénalisée par rapport aux autres instituts, on aurait pu diminuer la valeur de  $\alpha$ . Une autre astuce aurait été de laisser l'État belge tirer autant d'échantillons que l'ensemble des autres instituts réunis ou alors faire en sorte que chaque institut aie le même échantillon.

# Annexes

## Code Matlab répondant à la question 1.(a)

```
1 function Q1A()
2 data = xlsread('Database.xlsx');
3
4 % Histogramme de la consommation de bire
5 histogram(data(:,2),0:25:400);
6 xlabel('Nombre de canettes consommées');
7 ylabel('Nombres de pays concernes');
8 title('Histogramme de la consommation de biere dans le monde');
9
10 % Histogramme de la consommation d'alcool fort
11 figure
12 histogram(data(:,3),0:25:450);
13 axis([0 450 0 80]);
14 xlabel('Nombre de shots consommés');
15 ylabel('Nombres de pays concernes');
16 title('Histogramme de la consommation d alcool fort dans le
    monde');
17
18 end
```

## Code Matlab répondant à la question 1.(b)

```
1 function [beer, spirit] = Q1B()
2 data = xlsread('Database.xlsx');
3
4 beer_data = data(:,1);
5 spirit_data = data(:,2);
6
7 % Calcul des moyennes
8 beer_mean = mean(beer_data);
9 spirit_mean = mean(spirit_data);
10
11 % Calcul des medianes
12 beer_median = median(beer_data);
13 spirit_median = median(spirit_data);
14
15 % Calcul des modes
16 beer_mode = mode(beer_data);
17 spirit_mode = mode(spirit_data);
18
19 % Calcul des ecart-types
20 beer_std = std(beer_data);
21 spirit_std = std(spirit_data);
22
23 beer = [beer_mean, beer_median, beer_mode, beer_std];
```

```

24 spirit = [spirit_mean , spirit_median , spirit_mode , spirit_std];
25
26 end

```

### Code Matlab répondant à la question 1.(c)

```

1 function [beer_rate , spirit_rate] = Q1C()
2 data = xlsread('Database.xlsx');
3
4 [beer , spirit] = Q1B();
5
6 % Calcul des bornes de la loi normale
7 min_beer = beer(1) - beer(4);
8 max_beer = beer(1) + beer(4);
9
10 min_spirit = spirit(1) - spirit(4);
11 max_spirit = spirit(1) + spirit(4);
12
13 % Calcul du pourcentage de pays compris dans cet intervalle
14 nb_pays = max(size(data));
15
16 nb_norm_beer = 0;
17 nb_norm_spirit = 0;
18
19 for i=1:nb_pays
20     if data(i,1) >= min_beer && data(i,1) <= max_beer
21         nb_norm_beer = nb_norm_beer + 1;
22     end
23
24     if data(i,2) >= min_spirit && data(i,2) <= max_spirit
25         nb_norm_spirit = nb_norm_spirit + 1;
26     end
27 end
28
29 beer_rate = nb_norm_beer/nb_pays;
30 spirit_rate = nb_norm_spirit/nb_pays;
31
32 end

```

### Code Matlab répondant à la question 1.(d)

```

1 function [quartile_beer , quartile_spirit] = Q1D()
2 data = xlsread('Database.xlsx');
3
4 % Affichage des boites moustaches
5 boxplot(data(:,1));
6 ylabel('Nombres de verres consommés');
7 title('Boite a moustache de la consommation de biere');
8

```

```

9 figure
10 boxplot(data(:,2));
11 ylabel('Nombres de shots consommés');
12 title('Boite à moustache des consommations d'alcool fort');
13
14 % Calcul des quartiles
15 quartile_beer = quantile(data(:,1), [0.25 0.50 0.75]);
16 quartile_spirit = quantile(data(:,2), [0.25 0.50 0.75]);
17
18 end

```

#### Code Matlab répondant à la question 1.(e)

```

1 function F = Q1E()
2 data = xlsread('Database.xlsx');
3
4 % Affichage du polygone des fréquences cumulées
5 cdfplot(data(:,1));
6 xlabel('Nombre de canettes consommées');
7 ylabel('Fréquences cumulées');
8 title('\fontsize{10}Polygone des fréquences cumulées de la
    consommation de bière');
9
10 % Vérification de la proportion de pays ayant une consommation
    comprise entre [200, 295]
11 [F, x] = cdfcalc(data(:,1));
12 F_a = F(find(x >= 200, 1));
13 F_b = F(find(x > 295, 1));
14
15 F = F_b - F_a;
16
17 end

```

#### Code Matlab répondant à la question 1.(f)

```

1 function [cc_beer, cc_wine, cc_spirit] = Q1F()
2 data = xlsread('Database.xlsx');
3
4 % Scatterplot comparant la consommation d'alcool pur et la
    consommation de bière
5 scatter(data(:,4), data(:,1));
6 xlabel('Consommation d'alcool pur (en L)');
7 ylabel('Consommation de bière (en canettes)');
8 title('\fontsize{10}Graphe de dispersion comparant alcool pur et
    bière');
9
10 % Scatterplot comparant la consommation d'alcool pur et la
    consommation de vin
11 figure

```



```

12 scatter(data(:,4), data(:,3));
13 xlabel('Consommation d alcool pur (en L)');
14 ylabel('Consommation de vin (en verres)');
15 title('\fontsize{10}Graphe de dispersion comparant alcool pur au
    vin');
16
17 % Scatterplot comparant la consommation d'alcool pur la
    consommation d alcool fort
18 figure
19 scatter(data(:,4), data(:,2));
20 xlabel('Consommation d alcool pur (en L)');
21 ylabel('Consommation d alcool fort (en shots)');
22 title('\fontsize{10}Graphe de dispersion comparant alcool pur a
    alcool fort');
23
24 % Calcul des coefficients de correlation avec la consommation d'
    alcool pur
25 cc_beer_matrix = corrcoef(data(:,4), data(:,1));
26 cc_wine_matrix = corrcoef(data(:,4), data(:,3));
27 cc_spirit_matrix = corrcoef(data(:,4), data(:,2));
28
29 cc_beer = cc_beer_matrix(2,1);
30 cc_wine = cc_wine_matrix(2,1);
31 cc_spirit = cc_spirit_matrix(2,1);
32
33 end

```

### Code Matlab répondant à la question 2.(a)

```

1 function [data1Ech, DistKS] = Q2a()
2 data = xlsread("Database.xlsx");
3 % On prend 1 echantillon de 20 pays au hasard
4 index_echantillon = randsample(100,20,false);
5
6 echantillon = zeros(20,4);
7 for i = 1:20
8     echantillon(i,1:4) = data(index_echantillon(i),:);
9 end
10
11
12
13 beer_data = echantillon(:,1);
14 spirit_data = echantillon(:,2);
15
16 % Calcul de la mediane pour la consommation de biere et d alcool
    fort
17 median_beer = median(beer_data);
18 median_spirit = median(spirit_data);
19

```

```

20 % Calcul de la moyenne pour la consommation de biere et d alcool
    fort
21
22 mean_beer = mean(beer_data);
23 mean_spirit = mean(spirit_data);
24
25 % Calcul de l ecart-type pour la consommation de biere et d
    alcool fort
26
27 std_beer = std(beer_data);
28 std_spirit = std(spirit_data);
29
30 figure;
31 boxplot(beer_data);
32 title("Consommation de biere au monde d un echantillon de 20");
33 ylabel("Nombres de canettes");
34
35
36 figure;
37 boxplot(spirit_data);
38 title("Consommation d alcool fort au monde d un echantillon de
    20");
39 ylabel("Nombres de shots");
40
41 figure;
42 pop = cdfplot(data(:,1));
43 hold on;
44 set(pop, 'color', 'r');
45 ech = cdfplot(beer_data);
46 set(ech, 'color', 'b');
47 title({"Polygone de frequence cumulee de la consommation de
    biere"; "de la population et d un echantillon de 20"});
48
49 legend("show");
50 legend("Population", "Echantillon");
51 legend("Location", "southeast");
52 legend("boxoff");
53 ylabel("Frequence");
54 xlabel("Nombres de canettes");
55 hold off;
56
57
58
59 figure
60 pop = cdfplot(data(:,2));
61 hold on;
62 set(pop, 'color', 'r');
63 ech = cdfplot(spirit_data);

```

```

64 set(ech, 'color', 'b');
65 legend("show");
66 legend("Population", "Echantillon");
67 legend("Location", "southeast");
68 legend("boxoff");
69 ylabel("Frequence");
70 xlabel("Nombres de shots");
71 title({"Polygone de frequence cumulee de la consommation d
        alcool fort"; "de la population et d un echantillon de 20"});
72
73 [~,~,KS_beer] = kstest2(beer_data,(data(:,1)));
74 [~,~,KS_spirit] = kstest2(spirit_data,(data(:,2)));
75
76 DistKS = [KS_beer, KS_spirit];
77 data1Ech = [median_beer, median_spirit, mean_beer, mean_spirit,
             std_beer, std_spirit];
78 end

```

### Code Matlab répondant à la question 2.(b)

```

1 function [meanOfAll, data100Ech, DistKS] = Q2b()
2
3
4 data = xlsread("Database.xlsx");
5
6 echantillons = zeros(20,4,100);
7 median_beer_100 = zeros(100,1);
8 median_spirit_100 = zeros(100,1);
9 mean_beer_100 = zeros(100,1);
10 mean_spirit_100 = zeros(100,1);
11 std_beer_100 = zeros(100,1);
12 std_spirit_100 = zeros(100,1);
13 KS_beer_100 = zeros(100,1);
14 KS_spirit_100 = zeros(100,1);
15
16 KS_wine_100 = zeros(100,1);
17 KS_pure_100 = zeros(100,1);
18
19 % On prend 100 echantillons de 20 pays au hasard
20 for j = 1:100
21     index_echantillon = randsample(100,20,false);
22     for i = 1:20
23         echantillons(i,1:4,j) = data(index_echantillon(i),:);
24     end
25
26
27     beer_data = echantillons(:,1,j);
28     spirit_data = echantillons(:,2,j);
29     wine_data = echantillons(:,3,j);

```

```

30     pure_data = echantillons(:,4,j);
31
32
33     % Calcul de la mediane pour la consommation de biere et d
        alcool fort
34     %pour chaque echantillon
35     median_beer_100(j) = median(beer_data);
36     median_spirit_100(j) = median(spirit_data);
37
38     % Calcul de la moyenne pour la consommation de biere et d
        alcool fort
39     %pour chaque echantillon
40
41     mean_beer_100(j) = mean(beer_data);
42     mean_spirit_100(j) = mean(spirit_data);
43
44     % Calcul de l ecart-type pour la consommation de biere et d
        alcool fort
45     %pour chaque echantillon
46
47     std_beer_100(j) = std(beer_data);
48     std_spirit_100(j) = std(spirit_data);
49
50
51
52     % Calcul de la distance de Kolmogorov-Smirnov entre la
        population et
53     % l echantillon pour toutes les boissons
54     [~,~,KS_beer] = kstest2(beer_data,(data(:,1)));
55     KS_beer_100(j) = KS_beer;
56     [~,~,KS_spirit] = kstest2(spirit_data,(data(:,2)));
57     KS_spirit_100(j) = KS_spirit;
58
59     [~,~,KS_wine] = kstest2(wine_data,(data(:,3)));
60     KS_wine_100(j) = KS_wine;
61     [~,~,KS_pure] = kstest2(pure_data,(data(:,4)));
62     KS_pure_100(j) = KS_pure;
63
64 end
65
66 %Histogrammes
67 figure
68 histogram(median_beer_100,0:10:400);
69 title({"Histogramme de la repartition des medianes de la ","
        consommation de bieres sur 100 echantillons"});
70 ylabel("Nombres de medianes");
71 xlabel("Nombres de canettes");
72

```

```

73 figure
74 histogram(median_spirit_100,0:10:450);
75 title({"Histogramme de la repartition des medianes de la ";
    consommation d alcool fort sur 100 echantillons"});
76 ylabel("Nombres de medianes");
77 xlabel("Nombres de shots");
78
79 figure
80 histogram(mean_beer_100,0:10:400);
81 title({"Histogramme de la repartition des moyennes de la ";
    consommation de bieres sur 100 echantillons"});
82 ylabel("Nombres de medianes");
83 xlabel("Nombres de canettes");
84 figure
85 histogram(mean_spirit_100,0:10:450);
86 title({"Histogramme de la repartition des moyennes de la ";
    consommation d alcool fort sur 100 echantillons"});
87 ylabel("Nombres de moyennes");
88 xlabel("Nombres de shots");
89
90 figure
91 histogram(std_beer_100,0:10:400);
92 title({"Histogramme de la repartition de l ecart-type de la ";
    consommation de bieres sur 100 echantillons"});
93 ylabel("Nombres d ecart-type");
94 xlabel("Nombres de canettes");
95
96 figure
97 histogram(std_spirit_100,0:10:450);
98 title({"Histogramme de la repartition de l ecart-type de la ";
    consommation d alcool fort sur 100 echantillons"});
99 ylabel("Nombres d ecart-type");
100 xlabel("Nombres de shots");
101
102
103
104 % Calcul de la moyenne de la mediane des echantillons
105 mean_median_beer = mean(median_beer_100);
106 mean_median_spirit = mean(median_spirit_100);
107
108 % Calcul de la moyenne de la moyenne des echantillons
109
110 mean_mean_beer = mean(mean_beer_100);
111 mean_mean_spirit = mean(mean_spirit_100);
112
113 % Calcul de la moyenne de l ecart type des echantillons
114
115 mean_std_beer = mean(std_beer_100);

```

```

116 mean_std_spirit = mean(std_spirit_100);
117
118
119
120
121 figure
122 histogram(KS_beer_100,0:0.025:1)
123 title({"Histogramme de la repartition des distances de
        Kolmogorov-Smirnov";" de la consommation de bieres entre 100
        echantillons et la population"});
124 ylabel("Nombre de distances de K-S");
125 xlabel("Distance de Kolmogorov-Smirnov");
126
127 figure
128 histogram(KS_spirit_100, 0:0.025:1)
129 title({"Histogramme de la repartition des distances de
        Kolmogorov-Smirnov";" de la consommation d alcool fort entre
        100 echantillons et la population"});
130 ylabel("Nombre de distances de K-S");
131 xlabel("Distance de Kolmogorov-Smirnov");
132
133 figure
134 histogram(KS_wine_100, 0:0.025:1)
135 title({"Histogramme de la repartition des distances de
        Kolmogorov-Smirnov";" de la consommation de vin entre 100
        echantillons et la population"});
136 ylabel("Nombre de distances de K-S");
137 xlabel("Distance de Kolmogorov-Smirnov");
138
139 figure
140 histogram(KS_pure_100, 0:0.025:1)
141 title({"Histogramme de la repartition des distances de
        Kolmogorov-Smirnov";" de la consommation d alcool pur entre
        100 echantillons et la population"});
142 ylabel("Nombre de distances de K-S");
143 xlabel("Distance de Kolmogorov-Smirnov");
144
145 meanOfAll = [mean_median_beer,mean_median_spirit,mean_mean_beer,
        mean_mean_spirit,mean_std_beer,mean_std_spirit]
146 data100Ech = [median_beer_100,median_spirit_100,mean_beer_100,
        mean_spirit_100,std_beer_100,std_spirit_100];
147 DistKS = [KS_beer_100,KS_spirit_100,KS_wine_100,KS_pure_100];
148
149
150 end

```