

UNIVERSITÉ DE LIÈGE



ELÉMENTS DE STATISTIQUES

Projet : rapport

3^{ÈME} 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.

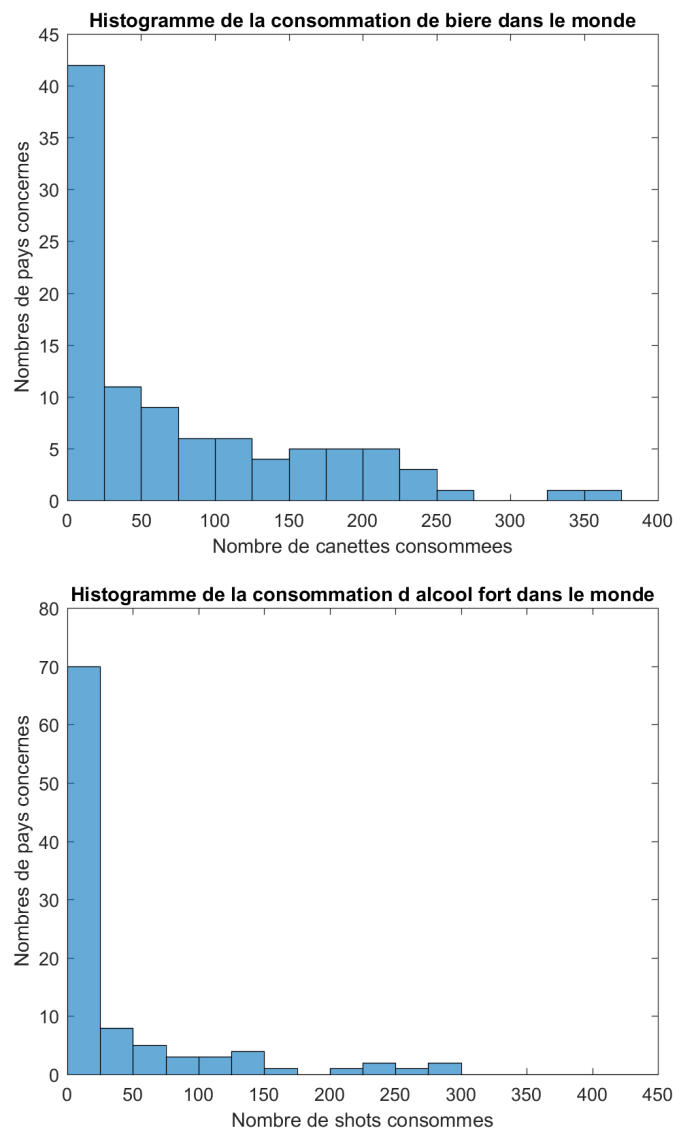


FIGURE 1 – Consommation de bière et d'alcool fort dans le monde

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 μ_i est une moyenne et σ_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.

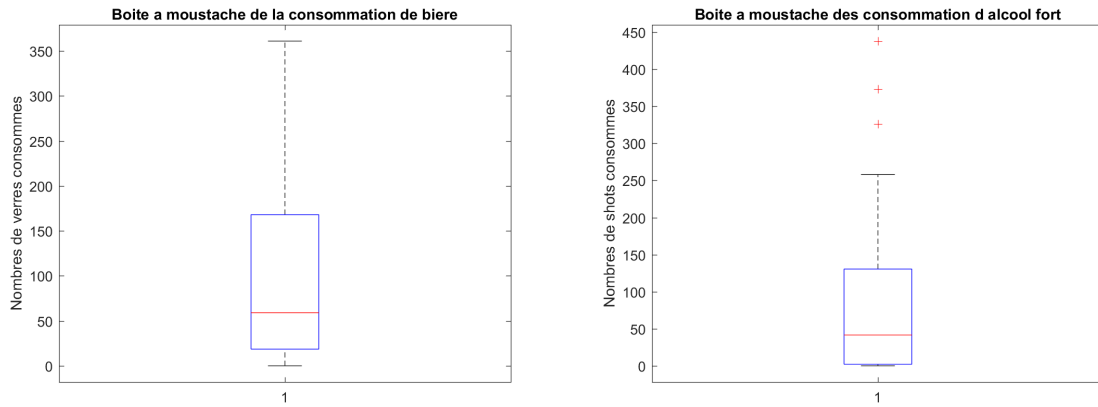


FIGURE 2 – Boites à moustaches relatives à la consommation de bière et d'alcool fort dans le monde

	1 ^{er} quartile	2 ^{ème} quartile	3 ^{ème} 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.

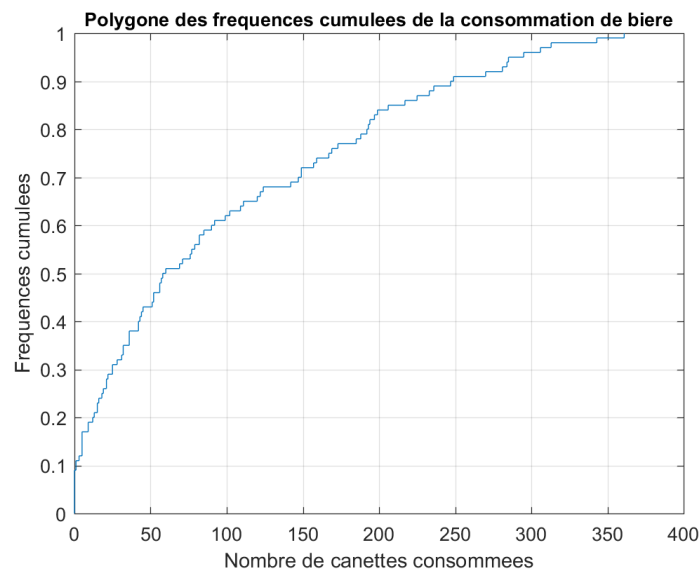


FIGURE 3 – Polygone des fréquences cumulées la consommation de bière

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.

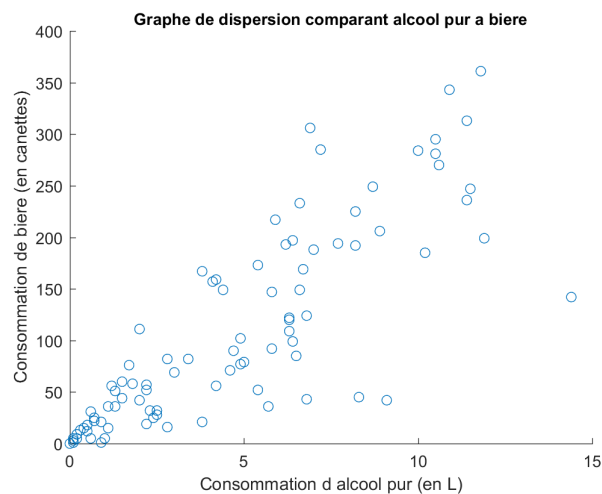


FIGURE 4 – Graphe de dispersion comparant la consommation d'alcool pur à la consommation de bière

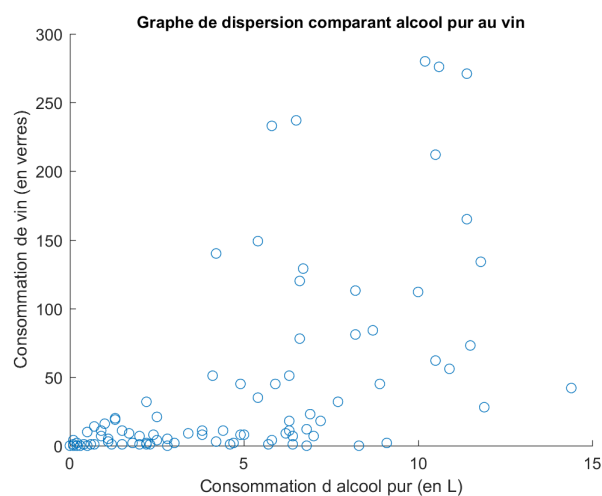


FIGURE 5 – Graphe de dispersion comparant la consommation d'alcool pur à la consommation de vin

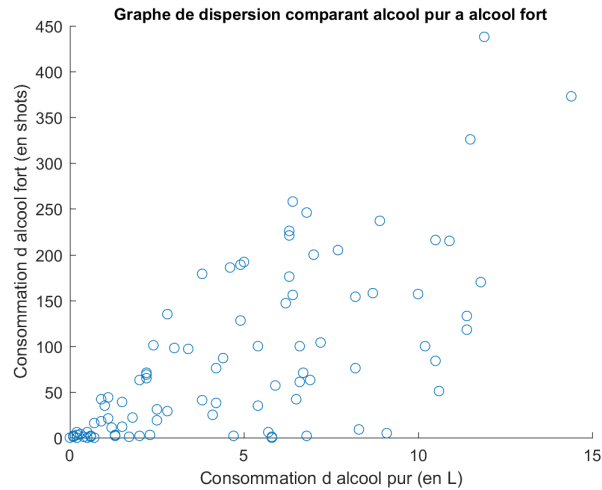


FIGURE 6 – Graphe de dispersion comparant la consommation d’alcool pur à la consommation d’alcool fort

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 1 relatif à toute la population, 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.

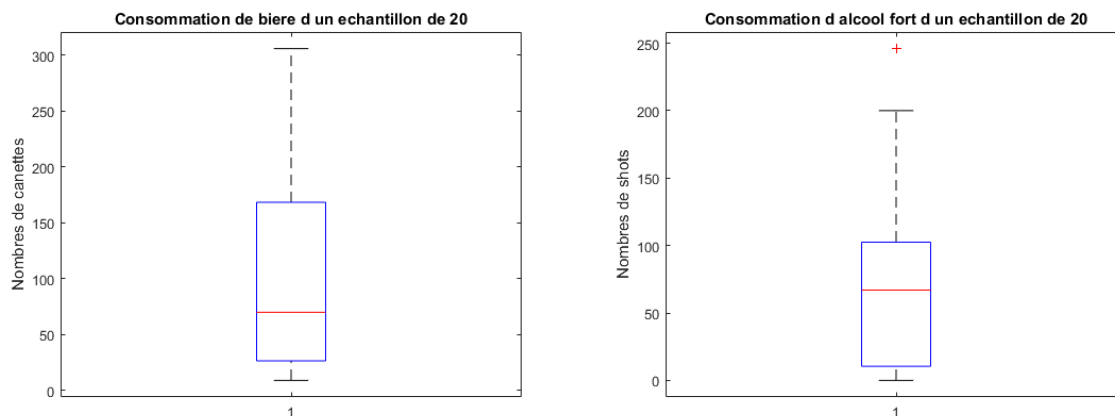


FIGURE 7 – Boîtes à moustaches de la consommation de bière et d'alcool fort de l'échantillon

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.

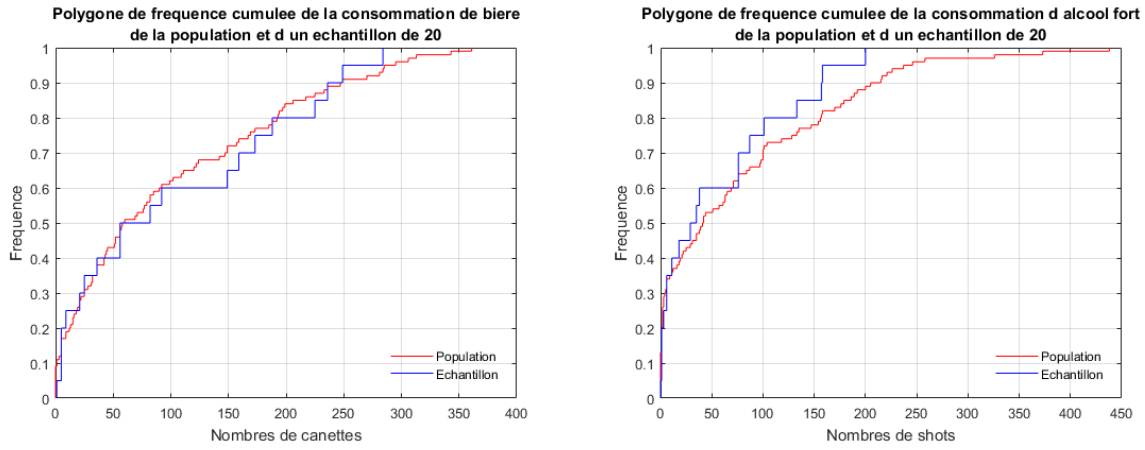


FIGURE 8 – Polygone de fréquences cumulées de la population et de l'échantillon pour la bière et l'alcool fort

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.

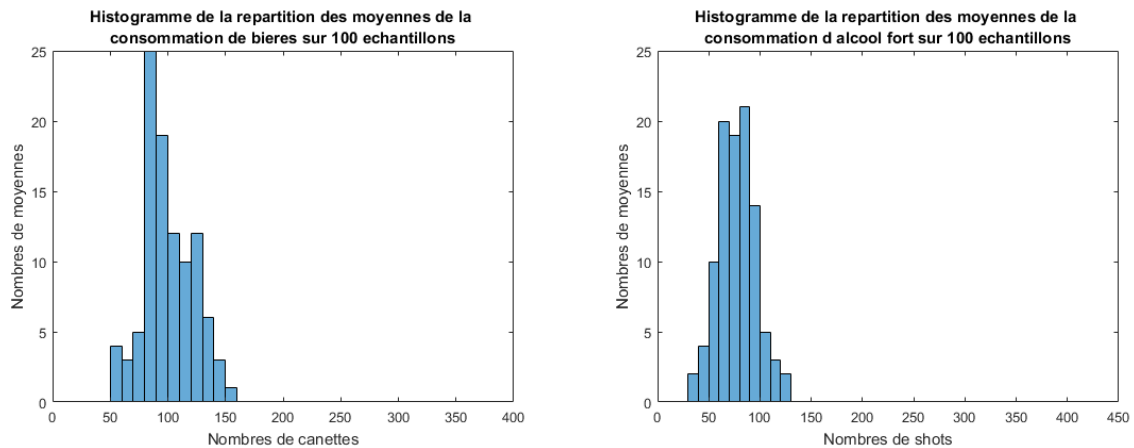


FIGURE 9 – Histogrammes des moyennes de consommation de bières et d'alcool fort de 100 échantillons

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.

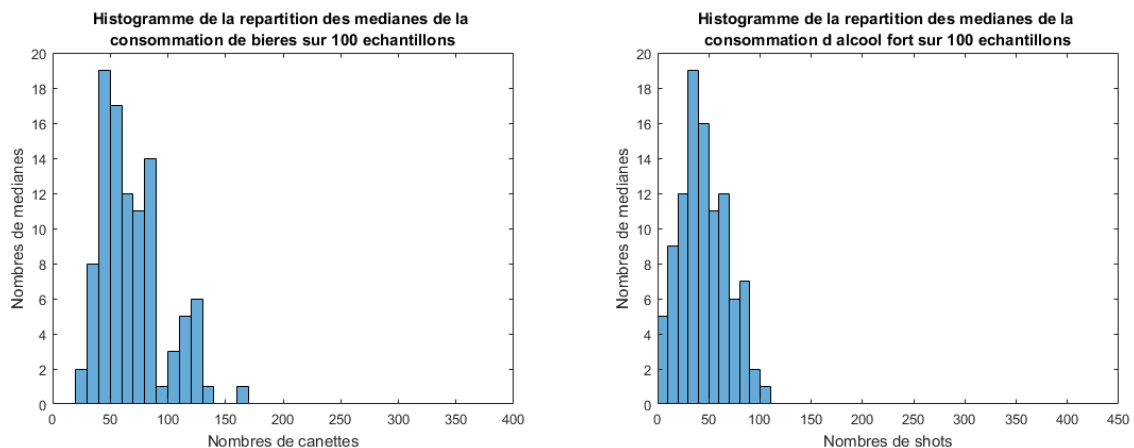


FIGURE 10 – Histogrammes des médianes de consommation de bières et d'alcool fort de 100 échantillons

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.

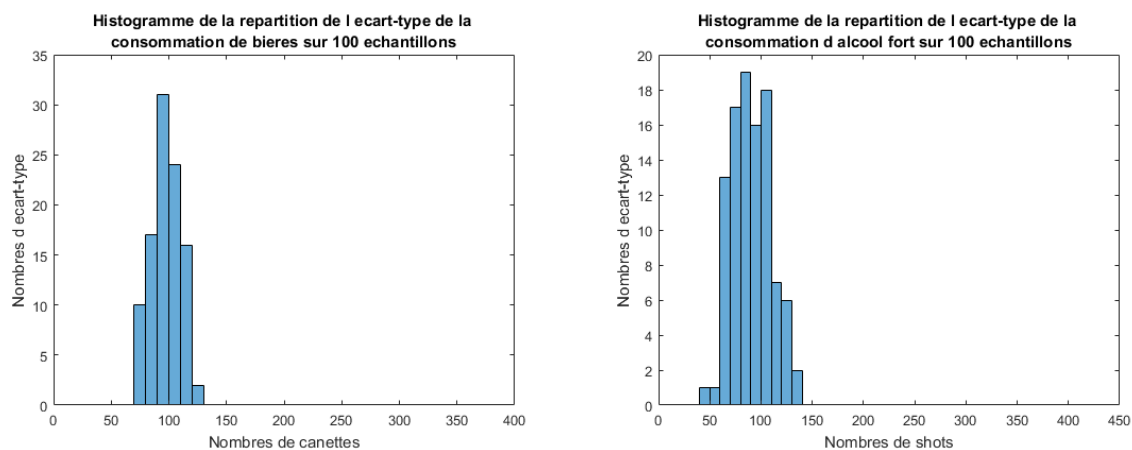


FIGURE 11 – Histogrammes des médianes de consommation de bières et d’alcool fort de 100 échantillons

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.

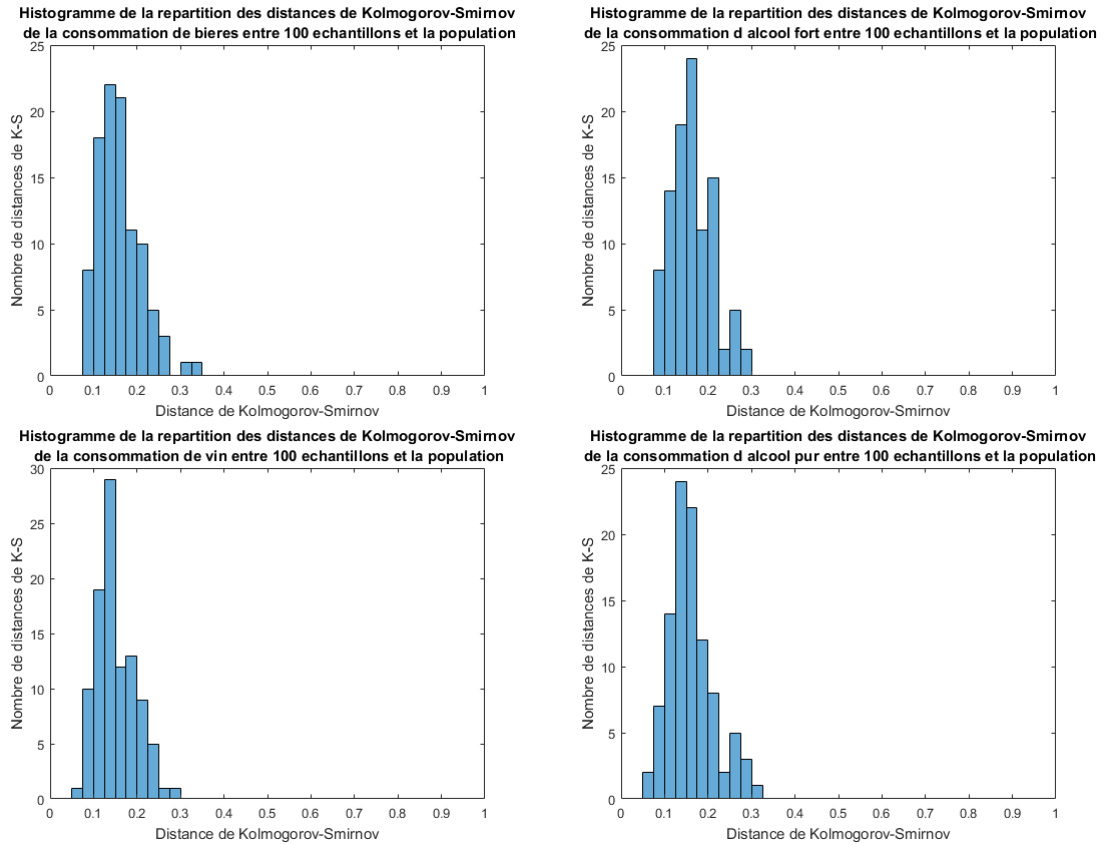


FIGURE 12 – Histogrammes des distances de Kolmogorov-Smirnov entre la population et les échantillons

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 polygones de fréquences cumulées, 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 : $m_x = 39.1155$ 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 :

$$b_{m_x} = m_x - M_x = 1.3455 \text{ verres}$$

Quant à la variance, elle vaut : $v_{m_x} = 260.7961 \text{ verres}^2$

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 : $median_x = 8.32$ verres. Ainsi, le biais de l'estimateur $median_x$ de la consommation moyenne de vin de la population est donné par :

$$b_{median_x} = median_x - M_x = -29.45 \text{ verres}$$

Quant à la variance, elle vaut : $v_{median_x} = 34.8726 \text{ verres}^2$

3.c Échantillons i.i.d de taille 50

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

	Biais	Variance
m_x	0.4208	87.2747
$median_x$	-29.665	9.7065

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 :

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

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

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

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 suivante :

"La proportion de pays consommant plus de bière que la Belgique est de $x\%$."

contre 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 4% , c'est-à-dire que 4% des pays sont de plus gros buveurs de bières que la Belgique, en excluant la Belgique évidemment. 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

$$f \geq p + u_{1-\alpha} \sqrt{\frac{p(1-p)}{n}}$$

où $u_{1-\alpha} = 1.645$, $n = 20$ et $p = 0.04$.

Après avoir fait tourner la fonction `Q4.m`, nous nous rendons compte que l'État belge rejette 4 fois l'hypothèse H_0 . C'est à dire que dans 4% des cas, nous les Belges ne sommes pas considérés comme un des pays où la bière coule à flots. En comparant à $\alpha = 0.05$, on voit que ces valeurs sont très similaires, ce qui traduit bien qu'on voulait un intervalle de confiance de cet ordre là. L'erreur que nous avons commise est le fait que nous avons approché une loi binomiale par une loi gaussienne.

4.b Autres instituts

L'OMS a dû intervenir 23 fois, c'est-à-dire que sur 100 échantillons, il y a eu 23 échantillons où au moins un institut rejetait l'hypothèse H_0 . On remarque que le nombre d'interventions par l'OMS est bien plus grand que le nombre de rejets de l'Etat belge. On a choisi de considérer que les Belges ne font pas partie des plus grands buveurs de bières si au moins un institut rejetait l'hypothèse. Les instituts ont donc 5 fois plus de chance de rejeter l'hypothèse. On va suggérer quelques modifications au prochain point qui pourraient rééquilibrer l'avantage.

4.c Éviter l'avantage des instituts indépendants

Afin d'éviter que l'État belge ne soit pénalisé par rapport aux autres instituts, on aurait pu diminuer la valeur de α . 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 biere
5 histogram(data(:,2),0:25:400);
6 xlabel('Nombre de canettes consommées');
7 ylabel('Nombres de pays concernés');
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 concernés');
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 a 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 a moustache de la 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 frequences cumulees
5 cdfplot(data(:,1));
6 xlabel('Nombre de canettes consommes');
7 ylabel('Frequences cumulees');
8 title('\fontsize{10}Polygone des frequences cumulees de la
    consommation de biere');
9
10 % Verification 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 a la
    consommation de biere
5 scatter(data(:,4), data(:,1));
6 xlabel('Consommation d alcool pur (en L)');
7 ylabel('Consommation de biere (en canettes)');
8 title('\fontsize{10}Graphe de dispersion comparant alcool pur a
    biere');
9
10 % Scatterplot comparant la consommation d'alcool pur a 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 a 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
4 % On prend un echantillon de 20 pays au hasard
5 index_echantillon = randsample(100,20,true);
6
7 echantillon = zeros(20,4);
8 for i = 1:20
9     echantillon(i,:) = data(index_echantillon(i),:);
10 end
11
12 beer_data = echantillon(:,1);
13 spirit_data = echantillon(:,2);
14
15 % Calcul de la mediane pour la consommation de biere et d alcool
    fort
16 median_beer = median(beer_data);
17 median_spirit = median(spirit_data);
18
19 % Calcul de la moyenne pour la consommation de biere et d alcool

```

```

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

```

```

65 ylabel('Frequence');
66 xlabel('Nombres de shots');
67 title({'Polygone de frequence cumulee de la consommation d
        alcool fort'; 'de la population et d un echantillon de 20'});
68
69 [~,~,KS_beer] = kstest2(beer_data,(data(:,1)));
70 [~,~,KS_spirit] = kstest2(spirit_data,(data(:,2)));
71
72 DistKS = [KS_beer,KS_spirit];
73 data1Ech = [median_beer,median_spirit,mean_beer,mean_spirit,
             std_beer,std_spirit];
74 end

```

Code Matlab répondant à la question 2.(b)

```

1 function [meanOfAll, data100Ech, DistKS] = Q2B()
2 data = xlsread('Database.xlsx');
3
4 echantillons = zeros(20,4,100);
5 median_beer_100 = zeros(100,1);
6 median_spirit_100 = zeros(100,1);
7 mean_beer_100 = zeros(100,1);
8 mean_spirit_100 = zeros(100,1);
9 std_beer_100 = zeros(100,1);
10 std_spirit_100 = zeros(100,1);
11 KS_beer_100 = zeros(100,1);
12 KS_spirit_100 = zeros(100,1);
13
14 KS_wine_100 = zeros(100,1);
15 KS_pure_100 = zeros(100,1);
16
17 % On prend 100 echantillons de 20 pays au hasard
18 for j = 1:100
19     index_echantillon = randsample(100,20,true);
20     for i = 1:20
21         echantillons(i,1:4,j) = data(index_echantillon(i),:);
22     end
23
24     beer_data = echantillons(:,1,j);
25     spirit_data = echantillons(:,2,j);
26     wine_data = echantillons(:,3,j);
27     pure_data = echantillons(:,4,j);
28
29
30     % Calcul de la mediane pour la consommation de biere et d
        alcool fort pour chaque echantillon
31     median_beer_100(j) = median(beer_data);
32     median_spirit_100(j) = median(spirit_data);
33

```

```

34 % Calcul de la moyenne pour la consommation de biere et d
    alcool fort pour chaque echantillon
35 mean_beer_100(j) = mean(beer_data);
36 mean_spirit_100(j) = mean(spirit_data);
37
38 % Calcul de l ecart-type pour la consommation de biere et d
    alcool fort pour chaque echantillon
39 std_beer_100(j) = std(beer_data);
40 std_spirit_100(j) = std(spirit_data);
41
42 % Calcul de la distance de Kolmogorov-Smirnov entre la
    population et l echantillon pour toutes les boissons
43 [~,~,KS_beer] = kstest2(beer_data,(data(:,1)));
44 KS_beer_100(j) = KS_beer;
45 [~,~,KS_spirit] = kstest2(spirit_data,(data(:,2)));
46 KS_spirit_100(j) = KS_spirit;
47
48 [~,~,KS_wine] = kstest2(wine_data,(data(:,3)));
49 KS_wine_100(j) = KS_wine;
50 [~,~,KS_pure] = kstest2(pure_data,(data(:,4)));
51 KS_pure_100(j) = KS_pure;
52
53 end
54
55 %Histogrammes
56 figure
57 histogram(median_beer_100,0:10:400);
58 title({'Histogramme de la repartition des medianes de la ','
    consommation de bieres sur 100 echantillons'});
59 ylabel('Nombres de medianes');
60 xlabel('Nombres de canettes');
61
62 figure
63 histogram(median_spirit_100,0:10:450);
64 title({'Histogramme de la repartition des medianes de la ','
    consommation d alcool fort sur 100 echantillons'});
65 ylabel('Nombres de medianes');
66 xlabel('Nombres de shots');
67
68 figure
69 histogram(mean_beer_100,0:10:400);
70 title({'Histogramme de la repartition des moyennes de la ','
    consommation de bieres sur 100 echantillons'});
71 ylabel('Nombres de medianes');
72 xlabel('Nombres de canettes');
73 figure
74 histogram(mean_spirit_100,0:10:450);
75 title({'Histogramme de la repartition des moyennes de la ','

```

```

    consommation d alcool fort sur 100 echantillons'}));
76 ylabel('Nombres de moyennes');
77 xlabel('Nombres de shots');
78
79 figure
80 histogram(std_beer_100,0:10:400);
81 title({'Histogramme de la repartition de l ecart-type de la ','
    consommation de bieres sur 100 echantillons'}));
82 ylabel('Nombres d ecart-type');
83 xlabel('Nombres de canettes');
84
85 figure
86 histogram(std_spirit_100,0:10:450);
87 title({'Histogramme de la repartition de l ecart-type de la ','
    consommation d alcool fort sur 100 echantillons'}));
88 ylabel('Nombres d ecart-type');
89 xlabel('Nombres de shots');
90
91
92 % Calcul de la moyenne de la mediane des echantillons
93 mean_median_beer = mean(median_beer_100);
94 mean_median_spirit = mean(median_spirit_100);
95
96 % Calcul de la moyenne de la moyenne des echantillons
97 mean_mean_beer = mean(mean_beer_100);
98 mean_mean_spirit = mean(mean_spirit_100);
99
100 % Calcul de la moyenne de l ecart type des echantillons
101 mean_std_beer = mean(std_beer_100);
102 mean_std_spirit = mean(std_spirit_100);
103
104
105 % Histogrammes
106 figure
107 histogram(KS_beer_100,0:0.025:1)
108 title({'Histogramme de la repartition des distances de
    Kolmogorov-Smirnov',' de la consommation de bieres entre 100
    echantillons et la population'}));
109 ylabel('Nombre de distances de K-S');
110 xlabel('Distance de Kolmogorov-Smirnov');
111
112 figure
113 histogram(KS_spirit_100, 0:0.025:1)
114 title({'Histogramme de la repartition des distances de
    Kolmogorov-Smirnov',' de la consommation d alcool fort entre
    100 echantillons et la population'}));
115 ylabel('Nombre de distances de K-S');
116 xlabel('Distance de Kolmogorov-Smirnov');

```

```

117
118 figure
119 histogram(KS_wine_100, 0:0.025:1)
120 title({'Histogramme de la repartition des distances de
        Kolmogorov-Smirnov'; ' de la consommation de vin entre 100
        echantillons et la population'});
121 ylabel('Nombre de distances de K-S');
122 xlabel('Distance de Kolmogorov-Smirnov');
123
124 figure
125 histogram(KS_pure_100, 0:0.025:1)
126 title({'Histogramme de la repartition des distances de
        Kolmogorov-Smirnov'; ' de la consommation d alcool pur entre
        100 echantillons et la population'});
127 ylabel('Nombre de distances de K-S');
128 xlabel('Distance de Kolmogorov-Smirnov');
129
130 meanOfAll = [mean_median_beer, mean_median_spirit, mean_mean_beer,
              mean_mean_spirit, mean_std_beer, mean_std_spirit];
131 data100Ech = [median_beer_100, median_spirit_100, mean_beer_100,
               mean_spirit_100, std_beer_100, std_spirit_100];
132 DistKS = [KS_beer_100, KS_spirit_100, KS_wine_100, KS_pure_100];
133
134 end

```

Code Matlab répondant à la question 3.(a)

```

1 function [mean_biais, mean_var] = Q3A()
2 data = xlsread('Database.xlsx');
3
4 % Matrice des echantillons
5 echantillons = zeros(20,100);
6 % Matrice des moyennes de chaque echantillon
7 wine_mean_matrix = zeros(100,1);
8
9 % Creation des echantillons
10 for j = 1:100
11     index_echantillon = randsample(100,20,true);
12     for i = 1:20
13         echantillons(i,j) = data(index_echantillon(i),3); % On
            reprend seulement la consommation de vin
14     end
15     % Calcul des moyennes de chaque echantillon
16     wine_mean_matrix(j) = mean(echantillons(:,j));
17 end
18
19 % Calcul du biais
20 real_wine_mean = mean(data(:,3));
21 mean_biais = mean(wine_mean_matrix) - real_wine_mean;

```

```

22
23 % Calcul de la variance
24 mean_var = var(wine_mean_matrix,1);
25
26 end

```

Code Matlab répondant à la question 3.(b)

```

1 function [median_biais, median_var] = Q3B()
2 data = xlsread('Database.xlsx');
3
4 % Matrice des echantillons
5 echantillons = zeros(20,100);
6 % Matrice des medianes de chaque echantillon
7 wine_median_matrix = zeros(100,1);
8
9 % Creation des echantillons
10 for j = 1:100
11     index_echantillon = randsample(100,20,true);
12     for i = 1:20
13         echantillons(i,j) = data(index_echantillon(i),3); % On
            reprend seulement la consommation de vin
14     end
15     % Calcul des medianes de chaque echantillon
16     wine_median_matrix(j) = median(echantillons(:,j));
17 end
18
19 % Calcul du biais
20 real_wine_mean = mean(data(:,3));
21 median_biais = mean(wine_median_matrix) - real_wine_mean;
22
23 % Calcul de la variance
24 median_var = var(wine_median_matrix,1);
25
26 end

```

Code Matlab répondant à la question 3.(c)

```

1 function [mean_biais, mean_var, median_biais, median_var] = Q3C()
2 data = xlsread('Database.xlsx');
3
4 % Matrice des echantillons
5 echantillons = zeros(50,100);
6 % Matrice des moyennes de chaque echantillon
7 wine_mean_matrix = zeros(100,1);
8 % Matrice des medianes de chaque echantillon
9 wine_median_matrix = zeros(100,1);
10
11 % Creation des echantillons

```



```

12 for j = 1:100
13     index_echantillon = randsample(100,50,true);
14     for i = 1:50
15         echantillons(i,j) = data(index_echantillon(i),3); % On
            reprend seulement la consommation de vin
16     end
17     % Calcul des moyennes de chaque echantillon
18     wine_mean_matrix(j) = mean(echantillons(:,j));
19     % Calcul des medianes de chaque echantillon
20     wine_median_matrix(j) = median(echantillons(:,j));
21 end
22
23 % Calcul des biais
24 real_wine_mean = mean(data(:,3));
25 mean_biais = mean(wine_mean_matrix) - real_wine_mean;
26 median_biais = mean(wine_median_matrix) - real_wine_mean;
27
28 % Calcul des variances
29 mean_var = var(wine_mean_matrix,1);
30 median_var = var(wine_median_matrix,1);
31
32 end

```

Code Matlab répondant à la question 3.(d)

```

1 function [counter_student, counter_gauss] = Q3D()
2 data = xlsread('Database.xlsx');
3
4 % Matrice des echantillons
5 echantillons = zeros(20,100);
6 % Matrice des moyennes de chaque echantillon
7 wine_mean_matrix = zeros(100,1);
8
9 % Creation des echantillons
10 for j = 1:100
11     index_echantillon = randsample(100,20,true);
12     for i = 1:20
13         echantillons(i,j) = data(index_echantillon(i),3); % On
            reprend seulement la consommation de vin
14     end
15     % Calcul des moyennes de chaque echantillon
16     wine_mean_matrix(j) = mean(echantillons(:,j));
17 end
18
19 % Donnees
20 n = 20;
21 alpha = 0.05;
22 real_wine_mean = mean(data(:,3));
23

```

```

24 % Initialisation des variables
25 s_tmp = zeros(100,1);
26 s_student = zeros(100,1);
27 s_gauss = zeros(100,1);
28 min_student = zeros(100,1);
29 max_student = zeros(100,1);
30 min_gauss = zeros(100,1);
31 max_gauss = zeros(100,1);
32 counter_student = 0;
33 counter_gauss = 0;
34
35 % i) Intervalle construit avec la loi de Student
36 t = tinv(1-alpha/2, n-1);
37
38 for j=1:100
39     for i=1:20
40         s_tmp(j) = s_tmp(j) + (echantillons(i,j) -
41             wine_mean_matrix(j))^2;
42     end
43     % Calcul des S_n
44     s_student(j) = sqrt(s_tmp(j)/(n-1));
45 end
46
47 for j=1:100
48     % Calcul des bornes des intervalles
49     min_student(j) = wine_mean_matrix(j) - (t*s_student(j)/sqrt(
50         n));
51     max_student(j) = wine_mean_matrix(j) + (t*s_student(j)/sqrt(
52         n));
53
54     if real_wine_mean >= min_student(j) && real_wine_mean <=
55         max_student(j)
56         counter_student = counter_student + 1;
57     end
58 end
59
60 % ii) Intervalle construit avec la loi de Gauss
61 u = norminv(1-alpha/2, 0, 1);
62
63 for j=1:100
64     for i=1:20
65         % Calcul de l'ecart-type
66         s_gauss(j) = sqrt(var(echantillons(:,j),1));
67     end
68 end
69
70 for j=1:100

```

```

68 % Calcul des bornes des intervalles
69 min_gauss(j) = wine_mean_matrix(j) - (u*s_gauss(j)/sqrt(n));
70 max_gauss(j) = wine_mean_matrix(j) + (u*s_gauss(j)/sqrt(n));
71
72 if real_wine_mean >= min_gauss(j) && real_wine_mean <=
    max_gauss(j)
73     counter_gauss = counter_gauss + 1;
74 end
75 end
76
77 end

```

Code Matlab répondant à la question 4.(a) et (b)

```

1 data = xlsread("Database.xlsx");
2 beerBelgium = data(93,1);
3 nbAboveBelgium = zeros(100,6);
4 n = 50;
5 echantillons = zeros(n,1,100,6);
6 interventionOMS = 0;
7 rejetUlg = 0;
8
9 for j = 1:100
10
11     rejetInstituts = 0;
12
13     for k = 1:6
14
15
16         index_echantillon = randsample(100,n,true);
17
18         for i = 1:n
19             echantillons(i,1,j,k) = data(index_echantillon(i),1)
20             ;
21
22             %Pays buvant plus de bieres que la Belgique
23             if(echantillons(i,1,j,k) > beerBelgium)
24                 nbAboveBelgium(j,k) = nbAboveBelgium(j,k) + 1;
25             end
26         end
27
28         % Proportion de pays de l echantillon qui boivent plus
29         % de bieres que BE
30         propNbAboveBelgium = nbAboveBelgium/n;
31
32         %Calcul de la borne
33         proportion = 0.04;
34         ecart_type = sqrt(proportion*(1-proportion)/n);
35         u = 1.645;

```

```

34         borne = proportion + (u*ecart_type);
35
36
37         if propNbAboveBelgium(j,k) > borne
38             if k == 1
39                 rejeUlg = rejeUlg + 1;
40             else
41                 rejeInstituts = rejeInstituts +1;
42             end
43         end
44
45
46     end
47     % Si au moins 1 institut rejette l'hypothse, l'OMS
48     intervient
49     if rejeInstituts >= 1
50         interventionOMS = interventionOMS +1;
51     end
52 end
53
54 rejeUlg
55 interventionOMS

```