# Université de Liège



# Eléments de statistiques

Projet: rapport

 $3^{\rm \`eme}$  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.

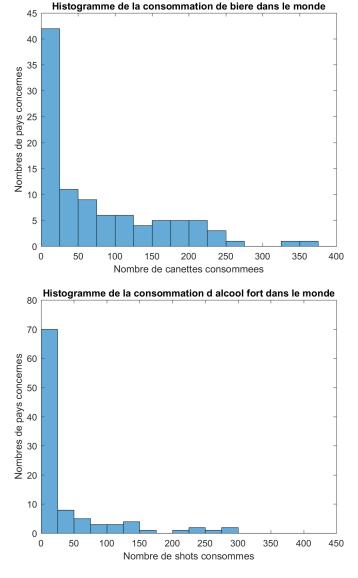


FIGURE 1 – Consommation de bière et d'alcool fort dans le monde

### 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"

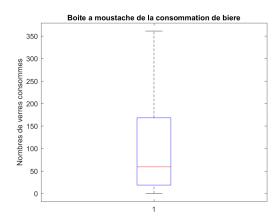
Les résultats normaux sont ceux compris dans l'intervalle  $[\mu_i - \sigma_i, \ \mu_i + \sigma_i]$  où  $\mu_i$  est une moyenne et  $\sigma_i$  l'écart-type correspondant. On trouve alors pour la consommation :

- 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 Boites à moustaches

Les boites 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.



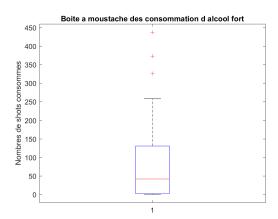


FIGURE 2 – Boites à moustaches relatives à la consommation de bière et d'alcool fort dans le monde

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

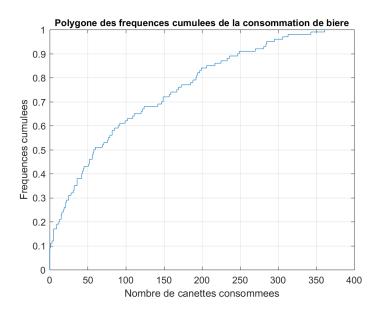


FIGURE 3 – Polygone des fréquences cumulées la consommation de bière

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

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

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.

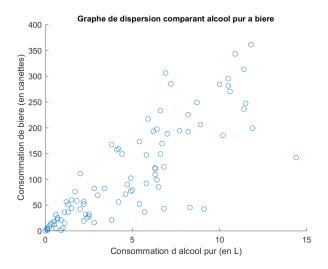


Figure 4 – Graphe de dispersion comparant la consommation d'alcool pur à la consommation de bière

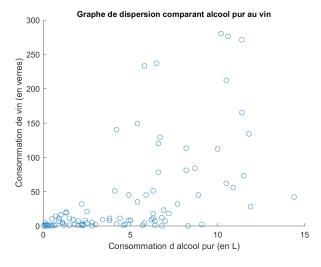


Figure 5 – Graphe de dispersion comparant la consommation d'alcool pur à la consommation de vin

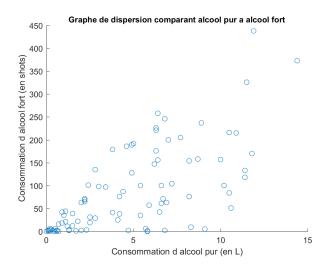


Figure 6 – Graphe de dispersion comparant la consommation d'alcool pur à la consommation d'alcool fort

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

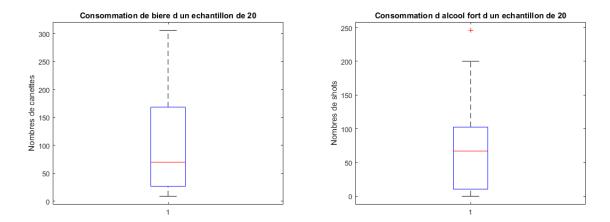
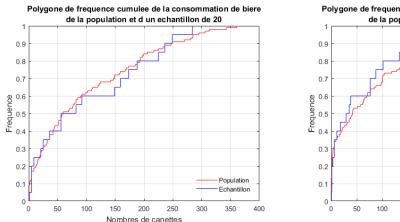


FIGURE 7 – Boites à moustaches de la consommation de bière et d'alcool fort de l'échantillon

On peut voir sur la boite à moustache relative à la consommation de bière que le 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 boite 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.



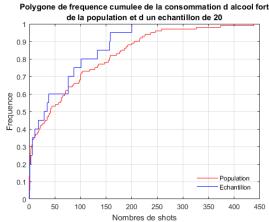


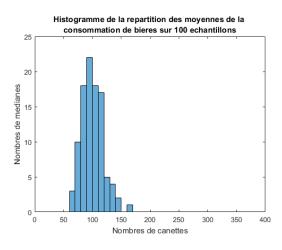
FIGURE 8 – Polygone de fréquences cumulées de la population et de l'échantillon pour la bière et l'alcool fort

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.



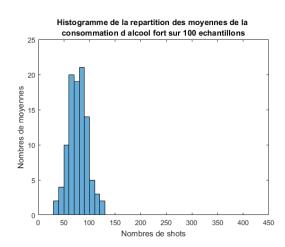


FIGURE 9 – Histogrammes des moyennes de consommation de bières et d'alcool fort de 100 échantillons

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.

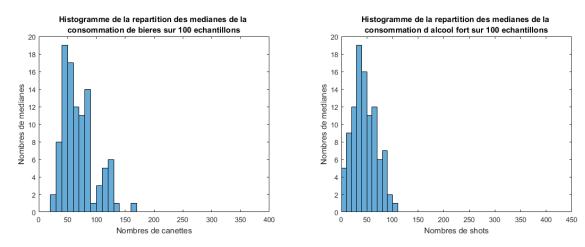
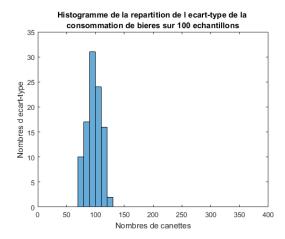


FIGURE 10 – Histogrammes des médianes de consommation de bières et d'alcool fort de 100 échantillons

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.



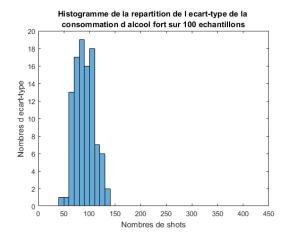


FIGURE 11 – Histogrammes des médianes de consommation de bières et d'alcool fort de 100 échantillons

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.

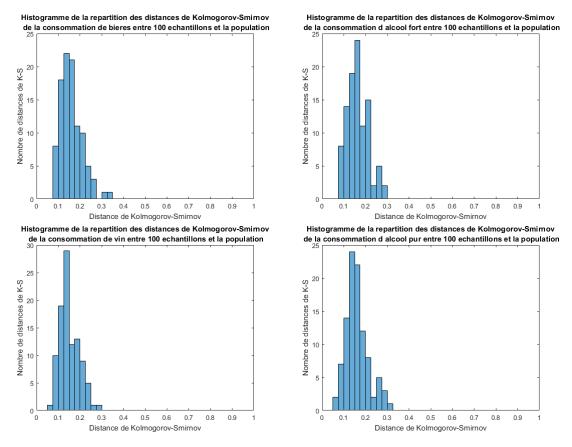


FIGURE 12 – Histogrammes des distances de Kolmogorov-Smirnov entre la population et les échantillons

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 :

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

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

### 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 :

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

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

### 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 :

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

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

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

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 3 4 7 10 5

## 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 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

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

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 somment pas considérés comme un des pays ou 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

Les 5 instituts restant rejettent l'hypothèse  $H_0$  autant de fois /  $\{3,4,7,10,5\}$  On remarque que, chaque instituts individuel rejette à peu près autant de fois l'hypothèse  $H_0$ .

Par contre, le nombre total des rejets, en faisant la somme de chaque institut, 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 grand buveurs de bières si au moins un instituts rejetaient l'hypothèse. Les instituts ont donc 5 fois plus de chance de rejeter l'hypothèse. On va suggérer quelques modifications au prochain point qui pourrait 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  $\alpha$ . Une autre astuce aurait été de laisser l'État belge tirer autant d'échantillons que l'ensemble des autres instituts réuni ou alors faire en sorte que chaque instituts aie le même échantillon.

#### Annexes

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

```
function Q1A()
  data = xlsread('Database.xlsx');
  % Histogramme de la consommation de biere
  histogram (data (:,2), 0:25:400);
  xlabel('Nombre de canettes consommees');
  ylabel ('Nombres de pays concernes');
  title ('Histogramme de la consommation de biere dans le monde');
  % Histogramme de la consommation d'alcool fort
  figure
11
  histogram (data (:,3), 0:25:450);
12
  axis ([0 450 0 80]);
13
  xlabel('Nombre de shots consommes');
  ylabel ('Nombres de pays concernes');
  title ('Histogramme de la consommation d'alcool fort dans le
     monde');
17
  end
18
  Code Matlab répondant à la question 1.(b)
  function [beer, spirit] = Q1B()
  data = xlsread('Database.xlsx');
  beer data = data(:,1);
  spirit_data = data(:,2);
  % Calcul des moyennes
  beer_mean = mean(beer_data);
  spirit_mean = mean(spirit_data);
  % Calcul des medianes
11
  beer_median = median(beer_data);
12
  spirit_median = median(spirit_data);
13
14
  % Calcul des modes
15
  beer mode = mode(beer data);
  spirit_mode = mode(spirit_data);
17
18
  % Calcul des ecart-types
19
  beer std = std (beer data);
20
  spirit_std = std(spirit_data);
21
  beer = [beer_mean, beer_median, beer_mode, beer_std];
```

```
spirit = [spirit_mean, spirit_median, spirit_mode, spirit_std];
25
  end
26
  Code Matlab répondant à la question 1.(c)
  function [beer_rate, spirit_rate] = Q1C()
  data = xlsread('Database.xlsx');
   [beer, spirit] = Q1B();
5
  % Calcul des bornes de la loi normale
  \min_{\text{beer}} = \text{beer}(1) - \text{beer}(4);
  \max_{\text{beer}} = \text{beer}(1) + \text{beer}(4);
  \min_{\text{spirit}} = \operatorname{spirit}(1) - \operatorname{spirit}(4);
10
   \max_{\text{spirit}} = \text{spirit}(1) + \text{spirit}(4);
11
12
  % Calcul du pourcentage de pays compris dans cet intervalle
  nb pays = max(size(data));
14
15
  nb\_norm\_beer = 0;
16
  nb norm spirit = 0;
17
18
   for i=1:nb_{pays}
19
       if data(i,1) >= min_beer \&\& data(i,1) <= max_beer
20
            nb\_norm\_beer = nb\_norm\_beer + 1;
21
       end
22
23
       if data(i,2) >= min_spirit && data(i,2) <= max_spirit
24
            nb_norm_spirit = nb_norm_spirit + 1;
25
       end
  end
28
   beer rate = nb norm beer/nb pays;
29
   spirit_rate = nb_norm_spirit/nb_pays;
30
31
  end
32
  Code Matlab répondant à la question 1.(d)
  function [quartile_beer, quartile_spirit] = Q1D()
  data = xlsread('Database.xlsx');
  % Affichage des boites a moustaches
  boxplot (data (:,1));
  ylabel ('Nombres de verres consommes');
   title ('Boite a moustache de la consommation de biere');
```

```
figure
  boxplot (data(:,2));
  ylabel ('Nombres de shots consommes');
  title ('Boite a moustache de la consommations d'alcool fort');
12
13
  % Calcul des quartiles
14
  quartile\_beer = quantile(data(:,1), [0.25 0.50 0.75]);
  quartile\_spirit = quantile(data(:,2), [0.25 0.50 0.75]);
17
  end
  Code Matlab répondant à la question 1.(e)
  function F = Q1E()
  data = xlsread('Database.xlsx');
  % Affichage du polygone des frequences cumulees
  cdfplot (data (:,1));
  xlabel('Nombre de canettes consommees');
  ylabel('Frequences cumulees');
  title ('\fontsize {10} Polygone des frequences cumulees de la
     consommation de biere');
  % Verification de la proportion de pays ayant une consommation
     comprise entre [200, 295]
  [F, x] = cdfcalc(data(:,1));
  F_a = F(find(x >= 200, 1));
  F_b = F(find(x > 295, 1));
13
14
  F = F_b - F_a;
15
16
  \operatorname{end}
17
  Code Matlab répondant à la question 1.(f)
function [cc_beer, cc_wine, cc_spirit] = Q1F()
  data = xlsread('Database.xlsx');
  % Scatterplot comparant la consommation d'alcool pur a la
     consommation de bire
  scatter (data(:,4), data(:,1));
  xlabel('Consommation d alcool pur (en L)');
  ylabel ('Consommation de biere (en canettes)');
  title ('\fontsize {10} Graphe de dispersion comparant alcool pur a
     biere');
  % Scatterplot comparant la consommation d'alcool pur a la
     consommation de vin
11 figure
```

```
scatter(data(:,4), data(:,3));
  xlabel('Consommation d alcool pur (en L)');
  ylabel ('Consommation de vin (en verres)');
  title ('\fontsize {10} Graphe de dispersion comparant alcool pur au
      vin ');
16
  % Scatterplot comparant la consommation d'alcool pur a la
17
     consommation d alcool fort
  figure
  scatter (data (:, 4), data (:, 2));
19
  xlabel('Consommation d alcool pur (en L)');
  ylabel ('Consommation d alcool fort (en shots)');
21
  title ('\fontsize {10} Graphe de dispersion comparant alcool pur a
     alcool fort');
  % Calcul des coefficients de correlation avec la consommation d'
     alcool pur
  cc\_beer\_matrix = corrcoef(data(:,4), data(:,1));
25
  cc\_wine\_matrix = corrcoef(data(:,4), data(:,3));
26
  cc_spirit_matrix = corrcoef(data(:,4), data(:,2));
27
28
  cc\_beer = cc\_beer\_matrix(2,1);
29
  cc\_wine = cc\_wine\_matrix(2,1);
  cc_spirit = cc_spirit_matrix(2,1);
31
32
  end
33
  Code Matlab répondant à la question 2.(a)
 function [data1Ech, DistKS] = Q2A()
  data = xlsread ('Database.xlsx');
  % On prend un echantillon de 20 pays au hasard
  index_echantillon = randsample(100,20,true);
  echantillon = zeros(20,4);
  for i = 1:20
       echantillon(i,:) = data(index_echantillon(i),:);
  end
10
11
  beer_data = echantillon(:,1);
  spirit_data = echantillon(:,2);
14
  % Calcul de la mediane pour la consommation de biere et d alcool
      fort
  median_beer = median(beer_data);
  median_spirit = median(spirit_data);
  % Calcul de la moyenne pour la consommation de biere et d alcool
```

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

```
ylabel('Frequence');
  xlabel('Nombres de shots');
  title ({ 'Polygone de frequence cumulee de la consommation d
      alcool fort'; 'de la population et d un echantillon de 20'});
68
  [\sim, \sim, KS\_beer] = kstest2(beer\_data, (data(:,1)));
69
  [\sim, \sim, KS\_spirit] = kstest2(spirit\_data, (data(:,2)));
70
71
  DistKS = [KS\_beer, KS\_spirit];
72
  data1Ech = [median_beer, median_spirit, mean_beer, mean_spirit,
      std_beer, std_spirit];
  end
74
  Code Matlab répondant à la question 2.(b)
1 function [meanOfAll, data100Ech, DistKS] = Q2B()
  data = xlsread('Database.xlsx');
  echantillons = zeros(20,4,100);
  median\_beer\_100 = zeros(100,1);
  median\_spirit\_100 = zeros(100,1);
  mean\_beer\_100 = zeros(100,1);
  mean spirit 100 = zeros(100,1);
  std\_beer\_100 = zeros(100,1);
  std\_spirit\_100 = zeros(100,1);
10
  KS beer 100 = zeros(100,1);
11
  KS_{spirit}_{100} = zeros(100,1);
12
  KS_{wine_{100}} = zeros(100,1);
  KS pure 100 = zeros(100,1);
15
16
  % On prend 100 echantillons de 20 pays au hasard
17
  for j = 1:100
18
       index_echantillon = randsample(100,20,true);
19
       for i = 1:20
           echantillons (i,1:4,j) = data (index_echantillon(i),:);
21
       end
22
23
       beer_data = echantillons(:,1,j);
24
       spirit_data = echantillons(:, 2, j);
25
       wine_data = echantillons (:,3,j);
26
       pure\_data = echantillons(:, 4, j);
27
28
29
      % Calcul de la mediane pour la consommation de biere et d
30
          alcool fort pour chaque echantillon
       median_beer_100(j) = median(beer_data);
31
       median_spirit_100(j) = median(spirit_data);
32
```

```
% Calcul de la moyenne pour la consommation de biere et d
34
          alcool fort pour chaque echantillon
       mean_beer_100(j) = mean(beer_data);
35
       mean_spirit_100(j) = mean(spirit_data);
36
37
      % Calcul de l ecart-type pour la consommation de biere et d
38
          alcool fort pour chaque echantillon
       std\_beer\_100(j) = std(beer\_data);
39
       std_spirit_100(j) = std(spirit_data);
40
41
      % Calcul de la distance de Kolmogorov-Smirnov entre la
42
          population et l echantillon pour toutes les boissons
       [\sim,\sim, KS\_beer] = kstest2(beer\_data,(data(:,1)));
43
       KS\_beer\_100(j) = KS\_beer;
44
       [\sim, \sim, KS\_spirit] = kstest2(spirit\_data, (data(:,2)));
45
       KS_spirit_100(j) = KS_spirit;
47
       [\sim,\sim, KS\_wine] = kstest2(wine\_data,(data(:,3)));
48
       KS\_wine\_100(j) = KS\_wine;
49
       [\sim, \sim, KS\_pure] = kstest2(pure\_data, (data(:,4)));
50
       KS_pure_100(j) = KS_pure;
51
  end
53
54
  %Histogrammes
55
  figure
56
  histogram (median_beer_100,0:10:400);
57
  title ({ 'Histogramme de la repartition des medianes de la '; '
      consommation de bieres sur 100 echantillons' });
  ylabel ('Nombres de medianes');
  xlabel ('Nombres de canettes');
60
61
  figure
62
  histogram (median_spirit_100,0:10:450);
63
  title ({ 'Histogramme de la repartition des medianes de la '; '
      consommation d alcool fort sur 100 echantillons');
  ylabel ('Nombres de medianes');
  xlabel('Nombres de shots');
66
67
  figure
68
  histogram (mean_beer_100,0:10:400);
69
  title ({ 'Histogramme de la repartition des moyennes de la '; '
      consommation de bieres sur 100 echantillons'});
  ylabel ('Nombres de medianes');
  xlabel ('Nombres de canettes');
72
  figure
  histogram (mean_spirit_100,0:10:450);
  title ({ 'Histogramme de la repartition des moyennes de la '; '
```

```
consommation d alcool fort sur 100 echantillons');
   ylabel ('Nombres de moyennes');
   xlabel('Nombres de shots');
78
  figure
79
  histogram (std beer 100,0:10:400);
80
   title ({ 'Histogramme de la repartition de l ecart-type de la '; '
      consommation de bieres sur 100 echantillons'});
   ylabel('Nombres d ecart-type');
   xlabel ('Nombres de canettes');
83
84
   figure
85
  histogram (std_spirit_100,0:10:450);
86
   title ({ 'Histogramme de la repartition de l ecart-type de la '; '
      consommation d alcool fort sur 100 echantillons');
   ylabel('Nombres d ecart-type');
   xlabel('Nombres de shots');
89
90
91
  % Calcul de la moyenne de la mediane des echantillons
92
  mean_median_beer = mean(median_beer_100);
  mean_median_spirit = mean(median_spirit_100);
  % Calcul de la moyenne de la moyenne des echantillons
96
  mean\_mean\_beer = mean(mean\_beer\_100);
97
  mean_mean_spirit = mean(mean_spirit_100);
98
99
  % Calcul de la moyenne de l ecart type des echantillons
100
  mean\_std\_beer = mean(std\_beer\_100);
101
   mean_std_spirit = mean(std_spirit_100);
102
103
104
  % Histogrammes
105
  figure
106
  histogram (KS_beer_100,0:0.025:1)
107
   title ({ 'Histogramme de la repartition des distances de
      Kolmogorov-Smirnov'; ' de la consommation de bieres entre 100
      echantillons et la population'});
   ylabel ('Nombre de distances de K-S');
109
   xlabel ('Distance de Kolmogorov-Smirnov');
110
111
  figure
112
  histogram (KS_spirit_100, 0:0.025:1)
   title ({ 'Histogramme de la repartition des distances de
      Kolmogorov-Smirnov'; ' de la consommation d alcool fort entre
      100 echantillons et la population');
  ylabel ('Nombre de distances de K-S');
  xlabel ('Distance de Kolmogorov-Smirnov');
```

```
117
   figure
118
   histogram (KS_wine_100, 0:0.025:1)
   title ({ 'Histogramme de la repartition des distances de
120
      Kolmogorov-Smirnov'; ' de la consommation de vin entre 100
      echantillons et la population'});
   ylabel ('Nombre de distances de K-S');
121
   xlabel ('Distance de Kolmogorov-Smirnov');
122
123
   figure
124
   histogram (KS_pure_100, 0:0.025:1)
125
   title ({ 'Histogramme de la repartition des distances de
126
      Kolmogorov-Smirnov'; ' de la consommation d alcool pur entre
      100 echantillons et la population');
   ylabel ('Nombre de distances de K-S');
127
   xlabel ('Distance de Kolmogorov-Smirnov');
129
   meanOfAll = [mean_median_beer, mean_median_spirit, mean_mean_beer,
130
      mean_mean_spirit, mean_std_beer, mean_std_spirit];
   data100Ech = [median_beer_100, median_spirit_100, mean_beer_100,
131
      mean_spirit_100, std_beer_100, std_spirit_100];
   DistKS = [KS_beer_100, KS_spirit_100, KS_wine_100, KS_pure_100];
132
  end
134
   Code Matlab répondant à la question 3.(a)
  function [mean_biais, mean_var] = Q3A()
   data = xlsread('Database.xlsx');
  % Matrice des echantillons
   echantillons = zeros(20,100);
  % Matrice des moyennes de chaque echantillon
   wine_mean_matrix = zeros(100,1);
  % Creation des echantillons
   for j = 1:100
10
       index_echantillon = randsample(100,20,true);
11
       for i = 1:20
12
           echantillons(i,j) = data(index_echantillon(i),3); % On
13
              reprend seulement la consommation de vin
       end
14
       % Calcul des moyennes de chaque echantillon
15
       wine_mean_matrix(j) = mean(echantillons(:,j));
16
   end
17
18
  % Calcul du biais
19
  real wine mean = mean(data(:,3));
   mean_biais = mean(wine_mean_matrix) - real_wine_mean;
```

```
22
  % Calcul de la variance
23
  mean_var = var(wine_mean_matrix,1);
  end
26
  Code Matlab répondant à la question 3.(b)
  function [median_biais, median_var] = Q3B()
  data = xlsread('Database.xlsx');
  % Matrice des echantillons
  echantillons = zeros(20,100);
  % Matrice des medianes de chaque echantillon
  wine_median_matrix = zeros(100,1);
  % Creation des echantillons
  for j = 1:100
10
      index_echantillon = randsample(100,20,true);
      for i = 1:20
12
           echantillons(i,j) = data(index_echantillon(i),3); % On
13
              reprend seulement la consommation de vin
14
      % Calcul des medianes de chaque echantillon
15
      wine_median_matrix(j) = median(echantillons(:,j));
  end
18
  % Calcul du biais
19
  real wine mean = mean(data(:,3));
20
  median_biais = mean(wine_median_matrix) - real_wine_mean;
22
  % Calcul de la variance
  median_var = var(wine_median_matrix,1);
25
  end
26
  Code Matlab répondant à la question 3.(c)
function [mean_biais, mean_var, median_biais, median_var] = Q3C()
  data = xlsread('Database.xlsx');
  % Matrice des echantillons
  echantillons = zeros(50,100);
6 % Matrice des moyennes de chaque echantillon
  wine_mean_matrix = zeros(100,1);
 % Matrice des medianes de chaque echantillon
  wine_median_matrix = zeros(100,1);
11 % Creation des echantillons
```

```
for j = 1:100
12
       index_echantillon = randsample(100,50,true);
       for i = 1:50
           echantillons(i,j) = data(index_echantillon(i),3); % On
15
              reprend seulement la consommation de vin
      end
16
      % Calcul des moyennes de chaque echantillon
17
      wine_mean_matrix(j) = mean(echantillons(:,j));
      % Calcul des medianes de chaque echantillon
      wine_median_matrix(j) = median(echantillons(:,j));
20
  end
21
22
  % Calcul des biais
23
  real\_wine\_mean = mean(data(:,3));
  mean_biais = mean(wine_mean_matrix) - real_wine_mean;
  median_biais = mean(wine_median_matrix) - real_wine_mean;
27
  % Calcul des variances
  mean var = var (wine mean matrix, 1);
29
  median_var = var(wine_median_matrix,1);
30
31
  end
32
  Code Matlab répondant à la question 3.(d)
  function [counter_student, counter_gauss] = Q3D()
  data = xlsread ('Database.xlsx');
  % Matrice des echantillons
  echantillons = zeros(20,100);
  % Matrice des moyennes de chaque echantillon
  wine_mean_matrix = zeros(100,1);
  % Creation des echantillons
  for j = 1:100
      index_echantillon = randsample(100,20,true);
       for i = 1:20
12
           echantillons(i,j) = data(index_echantillon(i),3); % On
13
              reprend seulement la consommation de vin
      end
14
      % Calcul des moyennes de chaque echantillon
15
      wine_mean_matrix(j) = mean(echantillons(:,j));
  end
17
18
  % Donnees
19
  n = 20;
20
  alpha = 0.05;
21
  real wine mean = mean(data(:,3));
```

```
% Initialisation des variables
  s_{tmp} = zeros(100,1);
  s_student = zeros(100,1);
  s gauss = zeros(100,1);
  \min_{\text{student}} = \text{zeros}(100,1);
28
  \max_{\text{student}} = \text{zeros}(100,1);
29
  \min_{\text{gauss}} = \operatorname{zeros}(100,1);
  \max_{\text{gauss}} = \text{zeros}(100,1);
   counter student = 0;
   counter\_gauss = 0;
33
34
  % i) Intervalle construit avec la loi de Student
35
  t = tinv(1-alpha/2, n-1);
36
37
   for j = 1:100
38
       for i = 1:20
            s_{tmp}(j) = s_{tmp}(j) + (echantillons(i,j) -
40
               wine_mean_matrix(j))^2;
       end
41
       % Calcul des S n
42
       s_student(j) = sqrt(s_tmp(j)/(n-1));
43
  end
44
45
   for j = 1:100
46
       % Calcul des bornes des intervalles
47
       min_student(j) = wine_mean_matrix(j) - (t*s_student(j)/sqrt(
48
       max_student(j) = wine_mean_matrix(j) + (t*s_student(j)/sqrt(
49
          n));
50
       if real wine mean >= min student(j) && real wine mean <=
51
           max student(j)
            counter_student = counter_student + 1;
52
       end
53
  end
54
56
  % ii) Intervalle construit avec la loi de Gauss
57
  u = norminv(1-alpha/2, 0, 1);
58
59
   for j = 1:100
60
       for i = 1:20
61
           % Calcul de l'ecart-type
62
           s_{gauss}(j) = sqrt(var(echantillons(:, j), 1));
63
       end
64
  end
65
66
  for j = 1:100
```

```
% Calcul des bornes des intervalles
68
       \min_{gauss(j)} = \min_{mean_{matrix}(j)} - (u*s_{gauss(j)}/sqrt(n));
69
       \max_{gauss(j)} = \min_{mean_{matrix(j)}} + (u*s_{gauss(j)}/sqrt(n));
70
71
       if real_wine_mean >= min_gauss(j) && real_wine_mean <=
72
          max_gauss(j)
           counter_gauss = counter_gauss + 1;
73
       end
74
  end
75
76
  end
77
  Code Matlab répondant à la question 4.(a) et (b)
  data = xlsread ("Database.xlsx");
  beerBelgium = data(93,1);
  nbAboveBelgium = zeros(100,6);
  n = 50;
  echantillons = zeros(n,1,100,6);
  rejets = zeros(1,6);
  for j = 1:100
       for k = 1:6
10
11
12
           index_echantillon = randsample(100,n,true);
13
           for i = 1:n
15
                echantillons(i,1,j,k) = data(index_echantillon(i),1)
16
17
               %Pays buvant plus de bieres que la Belgique
18
                if (echantillons (i, 1, j, k) > beerBelgium)
19
                    nbAboveBelgium(j,k) = nbAboveBelgium(j,k) + 1;
                end
           end
22
23
           % Proportion de pays de l echantillon qui boivent plus
24
              de bieres que BE
           propNbAboveBelgium = nbAboveBelgium/n;
25
           %Calcul de la borne
27
           proportion = 0.04;
28
           ecart\_type = sqrt(proportion*(1-proportion)/n);
29
           u = 1.645;
30
           borne = proportion + (u*ecart_type);
31
32
```

```
34
              if \ propNbAboveBelgium(j\ ,k)\ >\ borne
35
                    rejets(k) = rejets(k) +1;
36
              end
37
38
39
        \quad \text{end} \quad
40
   end
41
42
43
  rejets
44
  rejetUlg = rejets(1)
```