

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.

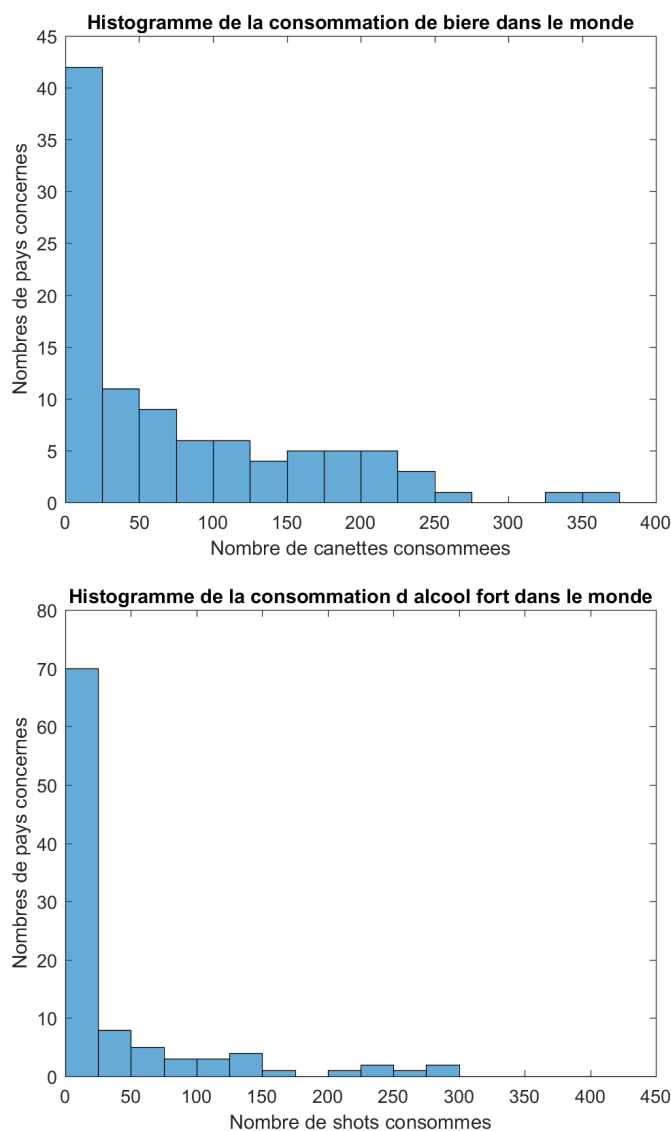


FIGURE 1 – Consommation de bière et d'alcool fort dans le monde

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 donnés dans la Table 1. Ces résultats sont donnés 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 8. Les quartiles demandés sont repris dans la Table 2.

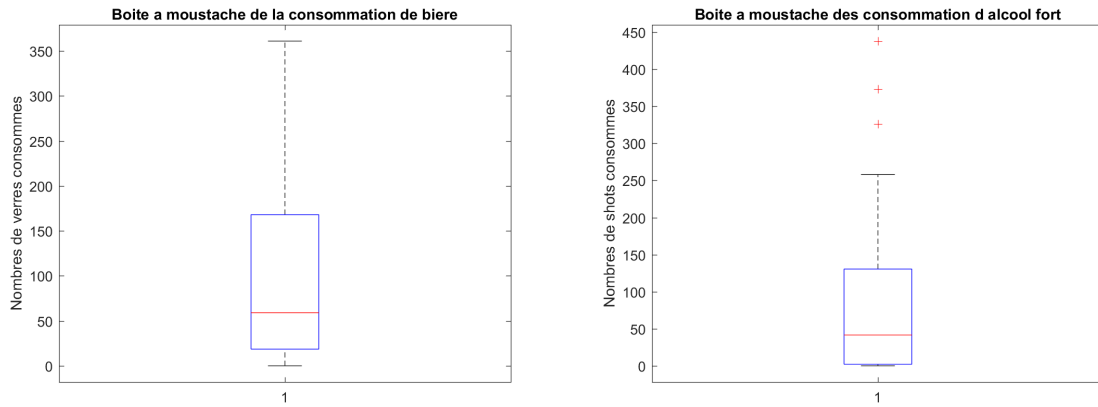


FIGURE 2 – Boîtes à moustaches relatives à la consommation de bière et d’alcool fort dans le monde

	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.

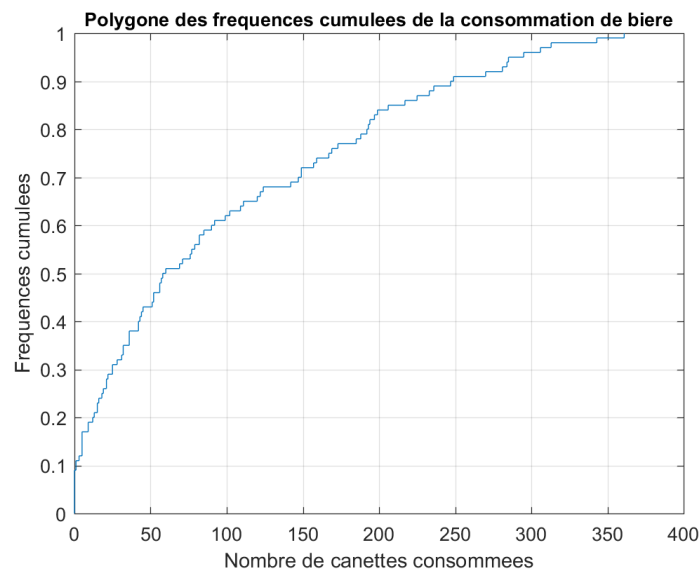


FIGURE 3 – Polygone des fréquences cumulées la consommation de bière

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.

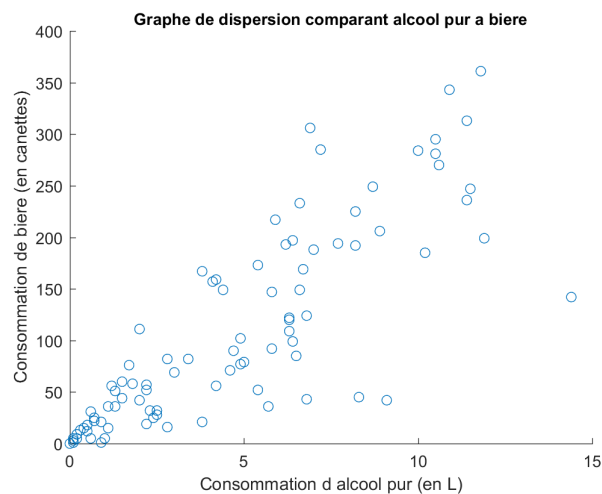


FIGURE 4 – Graphe de dispersion comparant la consommation d'alcool pur à la consommation de bière

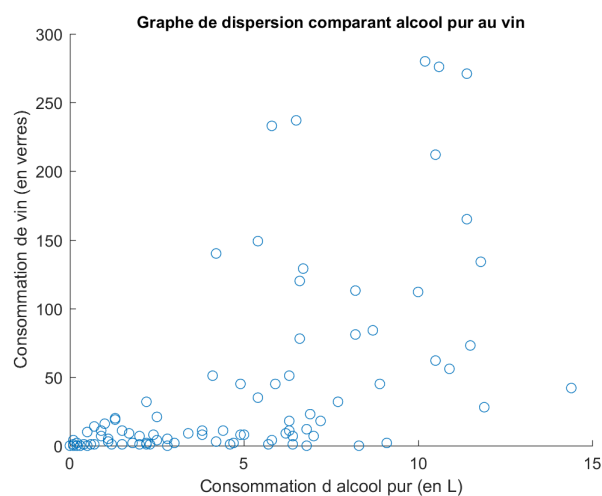


FIGURE 5 – Graphe de dispersion comparant la consommation d'alcool pur à la consommation de vin

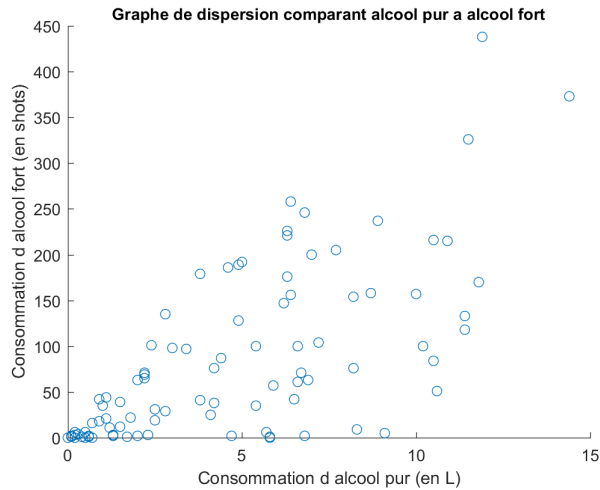


FIGURE 6 – Graphe de dispersion comparant la consommation d’alcool pur à la consommation d’alcool fort

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, notamment lors de la comparaison entre la consommation d’alcool pur avec la consommation de bières. Cela 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 la moyenne, la médiane et l’écart-type respectivement. 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 de spiritueux 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 fort 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**.

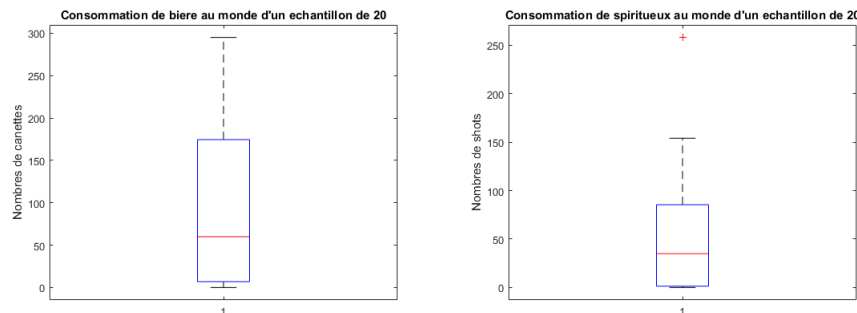


FIGURE 7 – Boîtes à moustaches de la consommation de bière et de spiritueux de l'échantillon

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 350 litres de bière. Nous pouvons aussi remarquer, en comparant avec la boîte à moustache de la **Figure** 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.

Ci dessous à la Figure sont représenté les polygones de fréquence cumulée de la population et de l'échantillon, une fois pour la bière et une fois pour l'alcool fort.

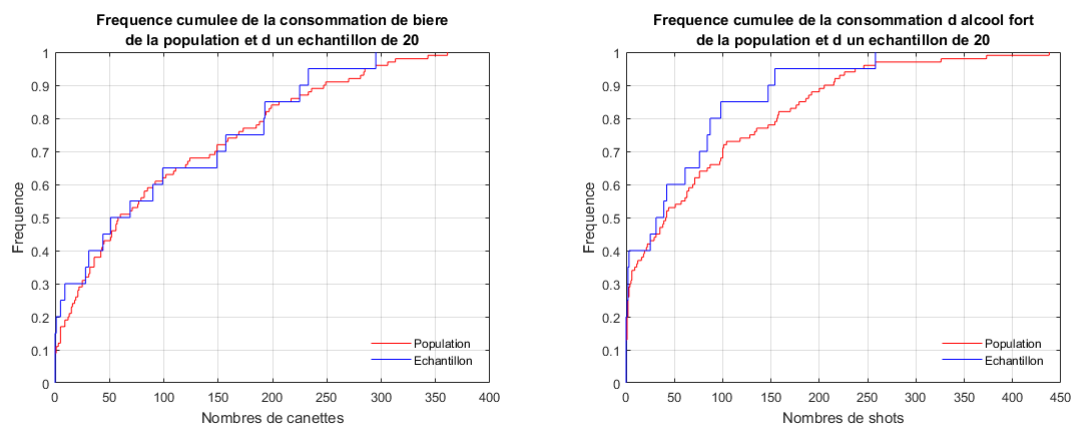


FIGURE 8 – Polygone de fréquences cumulées de la population et de l'échantillon pour la bière et l'alcool fort

On remarque que le polygone de fréquence cumulée de l'échantillon suit bien la tendance de celui de la population, bien que le bas d'escaliers est moindre, ce qui est trivial. De plus, nous avons la distance de Kolmogorov-Smirnov qui, pour la bière vaut et pour l'alcool fort vaut . 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 **Figure9** représente la répartition de la consommation moyenne de bière et d'alcool fort de tout les échantillons.

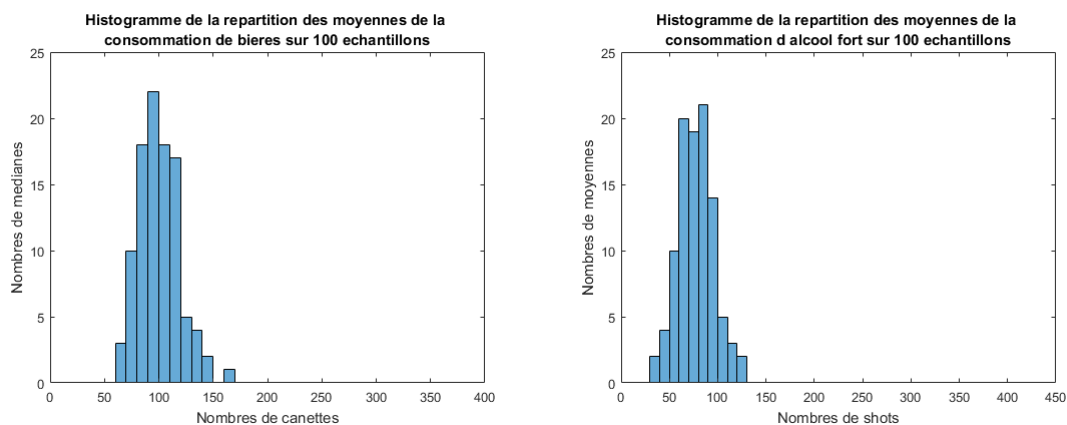


FIGURE 9 – Histogrammes des moyennes de consommation de bières et d'alcool fort de 100 échantillons

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 moyenne, 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 **Figure10** représente la répartition de la consommation médiane de bière et d'alcool fort de tout les échantillons.

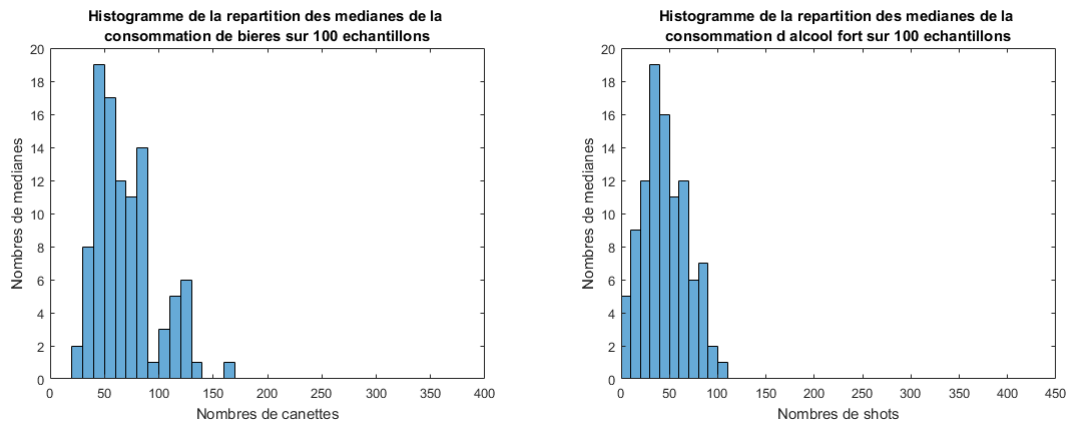


FIGURE 10 – Histogrammes des médianes de consommation de bières et d’alcool fort de 100 échantillons

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

(iii)

La **Figure10**représente la répartition de l’écart-type de la consommation de bière et d’alcool fort de tout les échantillons.

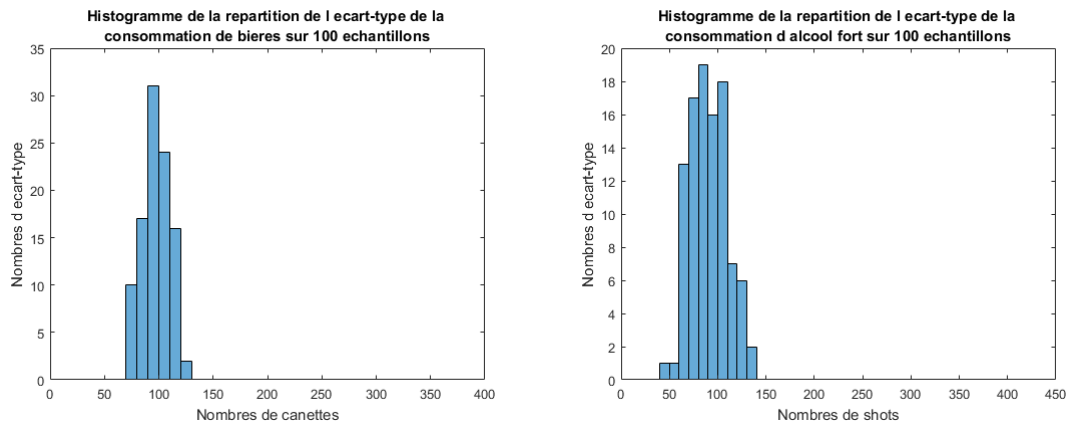


FIGURE 11 – Histogrammes des médianes de consommation de bières et d’alcool fort de 100 échantillons

Finalement, on refait la même procédure pour la moyenne de l’écart-type et on obtient comme valeur **97.42** et **91.13** pour la bière et l’alcool fort. Comparé à la population , 96.6 et 91.8, 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 4 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.

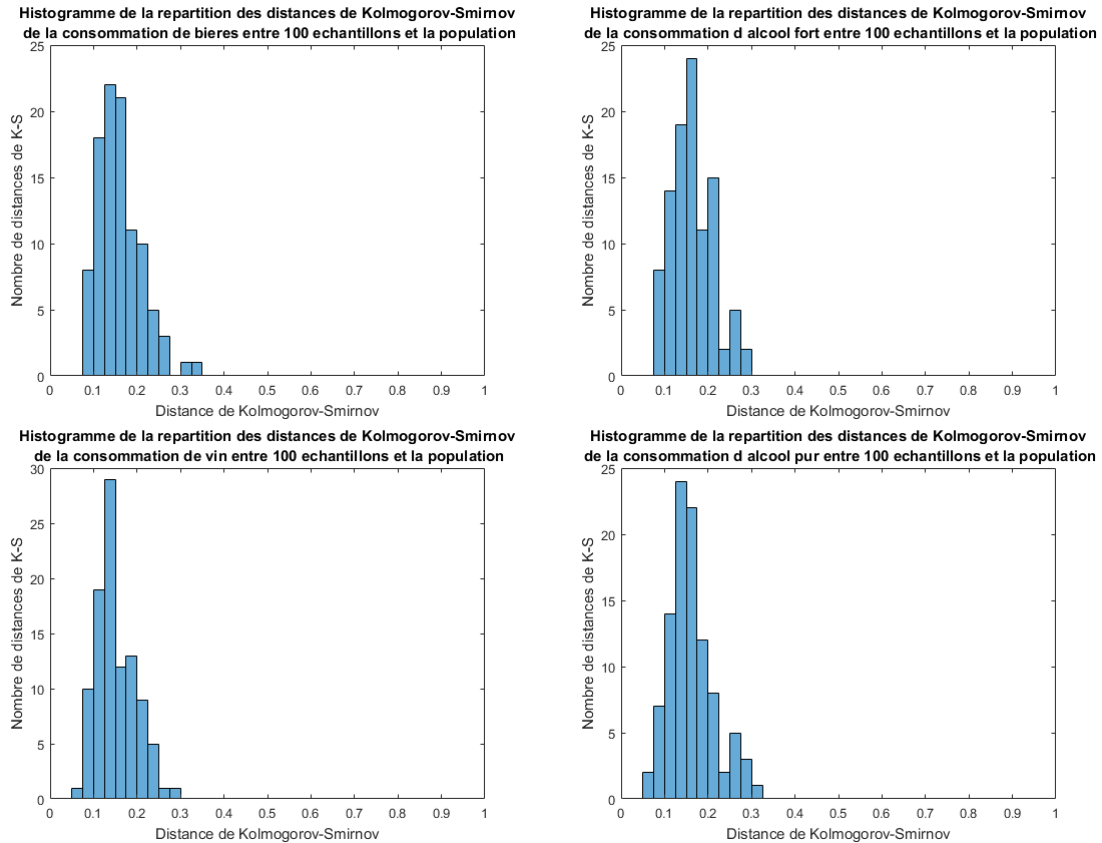


FIGURE 12 – Histogrammes des distances de Kolmogorov-Smirnov entre la population et les échantillons

En observant les 4 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 2 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