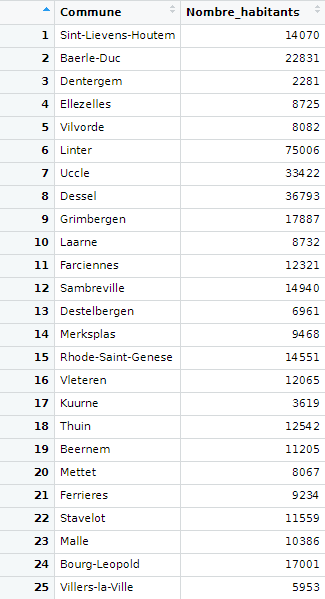
**Synthèse des estimations sur la Belgique en 2004**

**Introduction**

Dans cette étude, nous allons nous intéresser à la population belge ainsi que les revenus moyens de ces dites communes. L’objectif de ce travail est de permettre d’appréhender l’incertitude et la précision de l’estimation d’une grandeur mesurable dans la population belge à l’aide d’un intervalle de confiance réalisé à partir d’un processus d’échantillonnage. Ce processus d’échantillonnage se fera par l’intermédiaire d’un sondage aléatoire simple à probabilités égales. A noter que les résultats peuvent différer en raison des résultats qui se basent sur un échantillon aléatoire de 50 communes parmi 589.

**P**our commencer, nous avons créé une variable Ncom qui va servir à stocker le nombre de communes. Nous avons trouvé qu’il existait 589 communes dans le jeu de données « U ». Ensuite, pour calculer le nombre exact d’habitants en Belgique, nous avons créé une nouvelle variable Nhab et avons obtenu 10 417 122 habitants. Après avoir tiré un échantillon de 50 communes, le nombre d’habitants moyen de ce tirage est d’environ 13 425 habitants.

Par la suite, il serait judicieux de faire une table regroupant les 50 communes sélectionnées et leurs nombres d’habitants. La voici ci-dessous :

**L**e nombre de communes s’arrête évidemment à 50. Par la suite, calculons un estimateur (Nest) du nombre d’habitants (Nhab) à partir de nos 50 communes donc de notre échantillon.

Pour ce faire, il faudrait faire une moyenne du nombre d’habitants de nos 50 communes et le multiplier par le nombre total des communes donc de la population. Ce qui donnerai :

Nest = 13425,14 x 589 = 7 907 407

Donc nous estimons à 7 907 407 le nombre d’habitants en Belgique en 2004.

On remarque que la différence entre le nombre d’habitants exact et estimé n’est pas très éloignée. On relève une différence d’environ 2 509 714 habitants.

On pourrait expliquer cette différence notamment par le mode de tirage qui se fait aléatoirement parmi 589 communes et sans remise.

Ensuite, penchons-nous sur les intervalles de confiances. On retient un intervalle de confiance de -79615 et 106465 habitants. On voit très clairement que le nombre d’habitants exact (Nhab) n’est pas dans cet intervalle.

Après avoir traiter la 1ère partie portant sur l’estimation du nombre d’habitants, penchons-nous sur la 2ème partie qui est dédiée à l’estimation d’une proportion.

**E**n premier lieu, nous allons nous intéresser au salaire moyen par commune. Et pour cela, nous allons créer un tableau qui regroupe l’ensemble des communes de notre jeu de données. Par la suite, nous allons crée une variable Y qui prend comme valeur 1 si le salaire moyen d’une commune est supérieur ou égal à 23500€. Cette variable prend 0 sinon.

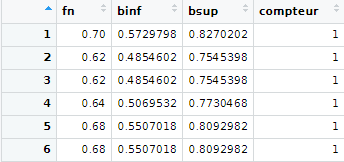
Par la suite, nous avons calculer la proportion p qui vaut 66% (0,66).

**E**n second lieu, nous avons pris un échantillon encore une fois de 50 communes et nous les avons regroupés avec leurs variables Y correspondantes. Nous avons donc un tableau qui regroupe 50 communes avec un indicateur qui indique 1 si le salaire moyen de la commune est supérieur ou égal à 23500€ et 0 sinon.

**E**n troisième lieu, nous allons construire un intervalle de confiance avec une marge d’erreur de 5%. Cet intervalle de confiance est de 0.42 à 0.69. On remarque directement que la valeur p (0.66) est comprise dans cet intervalle. Cependant, cet intervalle est assez large et donc pas très précis.

**E**n quatrième lieu, nous avons considérer 100 échantillons de 50 communes. Avec cette fois-ci la fréquence de ces échantillons que nous allons appeler ‘fn’.

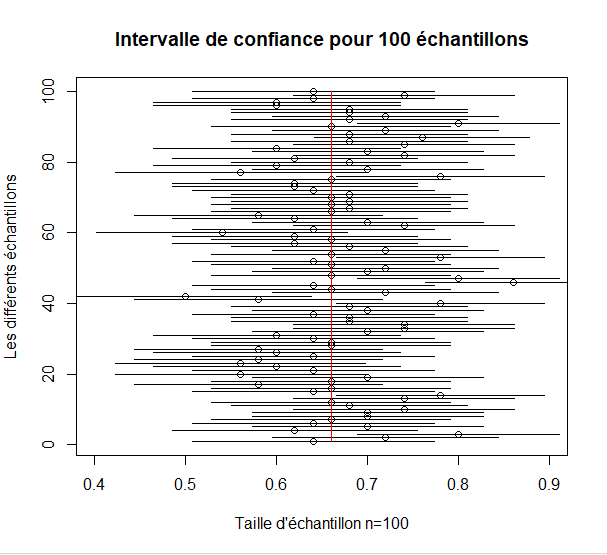
Après avoir calculé les 100 valeurs de ‘fn’ pour l’ensemble des échantillons, nous avons également calculé l’ensemble des intervalles des 100 échantillons. Et tout cela a été regroupé dans un tableau. Le voici ci-dessous :

Le tableau contient bien les 100 échantillons. (100 lignes)

La variable ‘compteur’ a été ajoutée afin compter à chaque fois que p (0.66) est compris dans un intervalle. Si oui, le compteur prend 1 sinon 0.

A la fin, nous avons donc 91 intervalles qui contiennent la valeur de p et 9 intervalles ou p est exclue.

**E**n cinquième et dernier lieu, nous avons choisi de représenter ces 100 intervalles avec un graphique. Voici le graphique ci-dessous :



La ligne rouge représente la valeur p (0.66). On peut voir attentivement sur le graphique les 9 intervalles où p est exclue. Un intervalle qui contient p traverse la ligne rouge.

**P**our conclure, nous avons pu voir dans cette synthèse plusieurs indicateurs que ce soit sur la population de la Belgique en 2004 ou bien du salaire moyen de chaque commune, nous avons pu représenter plusieurs indicateurs à l’aide du logiciel r qui est fourni avec cette synthèse. La plupart des résultats se trouvent dans ce script que nous vous invitons à consulter pour plus d’informations. Le code étant commenté entièrement, vous vous y retrouverez facilement.

**-- FIN --**