**TP2 : Intervalles de confiance**

Nous observons des données du fichiers poids.csv qui représente la loi empirique X qui est les poids à la naissance des nouveau-nés d’une maternité observés en une année. Les données sont supposé suivres une loi normale. Ainsi, on pourrait dire que X → (m,).

Nous calculons tout d’abord la moyenne empirique et la variance empirique du poids à la naissance pour l’ensemble des nouveau-nés sur l’ensemble de la population. On peut alors retrouver ces indicateurs dans le tableau ci dessous :

|  |  |
| --- | --- |
| **Moyenne** | 3125.60 |
| **Variance ()** | 191 413.1 |

Il faut rappeler que R donne la variance non biaisé (soit ), il faut alors biaisé cet indicateur pour obtenir la variance biaisé ( ). On fait alors : .

Maintenant, nous formons 3 échantillons de la population de taille 30, 60 et 80. Et nous calculons la moyenne empirique et la variance empirique de ces échantillons. On retrouve ces valeurs dans le tableau ci dessous :

|  |  |  |
| --- | --- | --- |
| **Taille échantillon** | **Moyenne empirique** | **Variance empirique** |
| n = 30 | 3098.7 | 154330.4 |
| n = 60 | 3090.383 | 192081.8 |
| n = 80 | 3113.975 | 187112.1 |

On observe ici, que plus l’échantillon est grand, et plus la moyenne empirique d’un échantillon “ressemble” à la moyenne empirique de la population. Cependant, cela n’est pas exact car la population est de taille : 9863.

On cherche, maintenant l’intervalle de confiance de niveau de risque pour la moyenne en supposant que la variance est connue (on prend ).

La formule de l’intervalle de confiance est :

avec ou u quantile de la loi normale N( 0 ; 1 ).

Pour l’intervalle de confiance de niveau de risque , et on fixe pour la moyenne concernant la population, on trouve : [3116.943 ; 3134.266].

De même, on calcul cet intervalle de confiance pour les différents échantillons :

|  |  |
| --- | --- |
| **Tailles des échantillons** | **Intervalles de confiance à 95%** |
| n = 30 | [ 2941.645 ; 3255.755 ] |
| n = 60 | [ 2979.328 ; 3201.438] |
| n = 80 | [ 3017.799 ; 3210.151] |

Maintenant, on doit calculer l’intervalle de confiance de niveau de risque pour la moyenne en supposant que la variance est inconnue.

On doit utiliser les quantiles de Student lorsque la variance est inconnue, avec un quantile t de degré de liberté n-1. Ainsi la formule de l’intervalle de confiance est alors :

où est la quantile de la loi de Student avec n-1 de degré de liberté.

Et où : c’est l’écart type et est la variance non-biaisé.

Pour la population cet intervalle de confiance est de :

[ 3116.969 ; 3134.24 ]

|  |  |
| --- | --- |
| **Taille de l’échantillon** | **Intervalles de confiance** |
| n = 30 | [ 2949.5 ; 3247.9 ] |
| n = 60 | [ 2976.21 ; 3204.556 ] |
| n = 80 | [ 3017.105 ; 3210.845 ] |

Sous R on utilise la fonction t.test, celle ci donne les résultats suivant :

|  |  |
| --- | --- |
| **Taille de l’échantillon** | **Intervalle de confiance** |
| n = 30 | [ 2949.5 ; 3247.9 ] |
| n = 60 | [ 2976.21 ; 3204.556 ] |
| n = 80 | [ 3017.105 ; 3210.845 ] |
| Population totale | [ 3116.969 ; 3134.24 ] |

On remarque que la fonction t.test calcule exactement les mêmes intervalles de confiance. On remarque également que ces résultats sont légèrements différents que ceux obtenue avec un connue. C’est parce que nous utilisons des indicateurs biaisé, et R utilise des indicateurs non-biaisé. On peut également ajouter que l’approximation par la loi de Student est juste car on a exactement les mêmes valeurs.

Afin d’offrir les premiers soins aux nouveau-nés, la maternité s’approvisionne en lait premier âge auprès d’un fournisseur local. Lors de la livraison, elle constate que seuls 177 cartons de lait sur une commande de 197 respectent les normes d’hygiène de la maternité. En effet, on remarque l’utilisation de proportion dans l’intitulé ci dessus.

On cherche à déterminer un intervalle de confiance à 95% pour la proportion de lait défectueux : on peut même remarquer que cette proportion X → Bin(20/197).

En utilisant les formules du cours on trouve les résultats suivant :

IC = [ 0.0593483 ; 0.1436974 ]

Également, en utilisant la fonction binom.test de R, on trouve :

IC = [ 0.06312328 ; 0.15243447 ]

Ces deux intervalles sont différents, car nous utilisons une méthode approximative. On remplace p par son estimateur dans les calculs de l’intervalle de confiance. En effet, c’est à cause de cela qu’il y a une différence sur ces deux intervalles. R fait le calcul en résolvant toutes les inégalités, il n’utilise pas de méthodes approximative, ainsi l’intervalle de R est plus précis.