STID 1 - Programmation Statistique
TP4
Probabilités ; simulations

Antoine ROLLAND et Anthony SARDELLITTI

Mars 2023

Contents

1 Simulation d'une population

La taille moyenne des français est de 171cm avec un écart-type de 9 centimètres.

1. produire les tailles d'une population simulée de 10.000.000 de français répartis suivant une loi normale de moyenne 171 et d'écart-type 9. Stocker ces tailles dans un vecteur "population"
2. calculer la moyenne et l'écart-type de la population. Retrouvez-vous les valeurs attendues?
3. établir l'histogramme de la taille. Retrouvez-vous la forme bien connues?
4. Combien de personnes ont une taille supérieure à 190cm? Combien devriez-vous en trouver théoriquement?
5. Combien de personnes ont une taille inférieure à 144cm? Combien devriez-vous en trouver théoriquement?

2 Simulation d'échantillons

On va essayer d'estimer la taille moyenne de la population à partir d'un échantillon.

1. Tirez un échantillon de taille 100 dans la population initiale, à l'aide de la fonction sample. Quelle est la taille moyenne dans l'échantillon? Quelle est l'écart-type dans l'échantillon? Ces deux valeurs sont-elles proches de celles de la population?
2. à partir de l'écart-type estimé, calculez la largeur du demi-intervalle de confiance, puis les bornes inférieures et supérieures de l'intervalle de confiance (toujours à 95%)
3. A l'aide de la fonction replicate, tirez 1000 échantillons de taille 100. Stockez dans un dataframe la moyenne et l'écart-type de chaque échantillon (la fonction apply peut être utile).
4. tracez l'histogramme des moyennes des échantillons. Retrouve-t-on une forme connue?
5. Calculez la moyenne des moyennes des échantillons, ainsi que l'écart-type des moyennes des échantillons. Normalement la moyenne des moyennes doit être (à peu près) égale à la moyenne de la population : on dit que la moyenne est un estimateur non biaisé. De même l'écart-type des moyennes des échantillons doit être (à peu près) égal à 0,9 c'est-à-dire $\sigma/\sqrt{n}$.
6. Combien d'échantillons ont une moyenne supérieure à 172,8cm? Quelle est le nombre théorique?
7. Pour chaque échantillon, calculez la largeur du demi-intervalle de confiance en utilisant l'estimation de l'écart-type calculée pour chaque échantillon, puis calculez les bornes inférieures et supérieures des intervalles de confiances (variables à rajouter dans votre dataframe).
8. Tracez les intervalles de confiance sur un graphe en utilisant la fonction plotCI du package gplots. Pour combien d'échantillons la vraie moyenne de la population (171cm) est-elle en dehors de l'intervalle de confiance?

3 Effet taille de l'échantillon

1. créer une fonction "moyenne_echantillon" qui prend en entrée le vecteur V variable d'une population et une taille n d'échantillon et qui donne en sortie la moyenne d'un échantillon aléatoire de taille n tiré dans la population moyenne_echantillon <- function(V, n)
2. à l'aide de la fonction replicate, tirer 1000 échantillons pour chacune des tailles d'échantillons suivantes : 20, 30, 50, 100, 500, 1000 toujours à partir de la population initiale de 10.000.000 d'individus.
3. représentez les histogrammes des moyennes pour chaque taille d'échantillon en gardant les mêmes échelles des axes des abcisses et ordonnées. Que constate-t-on (spoiler alert : cela illustre le *théorème central limite*)?
4. reprendre les 3 questions précédentes avec une nouvelle population de 10.000.000 individus tirés à partir d'une loi uniforme sur [0,1]