

M2 Statistique et Econométrie

Travaux pratiques du cours Techniques de sondage

ENNONCÉ DU TP 3 AVEC R ET SAS : ESTIMATION DU NOMBRE DE LOGEMENTS VACANTS POUR LES PLANS SYSTÉMATIQUE AVEC ORDRE ALÉATOIRE OU TRI SELON LA VARIABLE LOG

Nous considérons l'extrait du recensement du 1999 présenté lors du premier tp. L'objectif est d'estimer le nombre total de logements vacants or le total de la variable LOGVAC selon un plan systématique quand l'ordre des individus est aléatoire et quand elle est triée selon une variable fortement corélée à LOGVAC.

– **Le plan systématique avec ordre aléatoire**

1. Réaliser une permutation aléatoire des individus de la population REC99HTEGNE. Sélectionner dans cette nouvelle population un échantillon de taille $n = 50$ selon un plan systématique.
2. Calculer une estimation de t_{LOGVAC} dans cet échantillon ainsi que sa variance.
3. Calculer l'estimation de t_{LOGVAC} pour 60 simulations. Déterminer le coefficient de variation.

– **Le plan systématique avec tri selon la variable LOG**

1. Considérer la population REC99HTEGNE et réaliser un tri selon la variable LOG. Sélectionner dans cette nouvelle population un échantillon de taille $n = 50$ selon un plan systématique.
2. Calculer une estimation de t_{LOGVAC} dans cet échantillon ainsi que sa variance.
3. Calculer l'estimation de t_{LOGVAC} pour 60 simulations. Déterminer le coefficient de variation.

– **Comparer les deux types de plans.**

1. **Avec R** : La fonction R du package "sampling" qui sélectionne un échantillon systématique s'appelle `UPsystematic` dont l'argument est `pik`, le vecteur de taille N de probabilités de sélection. On doit avoir la somme des composantes de `pik` égale à n , la taille choisie.
2. **Avec SAS** : la proc `surveyselect` a l'option `method=sys` pour sélectionner un échantillon systématique.

Pour calculer la variance avec R ou SAS, il faut savoir que la variance estimée pour un plan systématique est approximée par la variance d'un plan aléatoire simple sans remise.