

1 Bases de la théorie des sondages

1.1 Estimateur pondéré

Dans les questions suivantes, on considère un sondage de 15 individus parmi une population de 300. Le sondage vise à mesurer la fréquentation mensuelle des cinémas par les individus de la population. Les résultats sont donnés dans la table suivante :

| Individu | Frequentation cinéma | Poids sondage |
|----------|----------------------|---------------|
| 1 | 1 | 24 |
| 2 | 2 | 20 |
| 3 | 4 | 18 |
| 4 | 15 | 18 |
| 5 | 2 | 20 |
| 6 | 3 | 14 |
| 7 | 1 | 22 |
| 8 | 0 | 20 |
| 9 | 0 | 17 |
| 10 | 3 | 20 |
| 11 | 6 | 22 |
| 12 | 0 | 21 |
| 13 | 0 | 22 |
| 14 | 1 | 22 |
| 15 | 1 | 20 |

1.1.1

On rappelle que l'estimateur plugin est l'estimateur qui ne prend pas en compte les poids de sondage (et qu'il est, en toute généralité, biaisé!). Quel est l'estimateur plugin de la fréquentation moyenne ?

1.1.2

Quel est l'estimateur pondéré (avec les poids de sondages donnés dans la table) de la fréquentation moyenne ?

1.2 Calcul de probabilités d'inclusion

On se donne une population $\mathcal{U} = \{1, 2, 3\}$ et un plan de sondage p défini par :

$$p(\{1\}) = 0 \quad p(\{1, 2\}) = \frac{1}{2} \quad p(\{1, 2, 3\}) = 0$$

$$p(\{2\}) = 0 \quad p(\{1, 3\}) = \frac{1}{3}$$

$$p(\{3\}) = 0 \quad p(\{2, 3\}) = \frac{1}{6}$$

1.2.1

Calculer les probabilités d'inclusion simples et doubles de toutes les unités de la population, ainsi que les Δ_{kl}

1.2.2

Quelle propriété intéressante possède le plan de sondage p ? Que peut-on dire de $\sum_{k \in \mathcal{U}} \pi_k$ dans ce cas? Le vérifier.

1.3 Introduction à l'estimateur d'Horvitz-Thompson

On se donne la population et le plan de sondage de l'exercice 1.2. On suppose que le revenu des individus de la population est :

$$Y_1 = 1000$$

$$Y_2 = 2000$$

$$Y_3 = 3000$$

1.3.1

Donner la loi de l'estimateur plugin. Calculer le biais et la variance de cet estimateur.

L'estimateur d'Horvitz-Thompson d'un total Y est un estimateur pondéré qui se définit par :

$$\hat{Y}_{HT} = \sum_{k \in s} \frac{y_k}{\pi_k}$$

1.3.2

Donner la loi de l'estimateur d'Horvitz-Thompson. Calculer le biais et la variance de cet estimateur.

1.4 Enquête patrimoine

Un sondeur réalise une enquête ayant pour but de mesurer le patrimoine moyen des ménages d'Île-de-France. Les individus sont tirés parmi la liste des titulaires d'un pass Navigo (carte d'abonnement aux transports en commun d'Île-de-France) pour l'année 2016. Le plan de sondage donne une probabilité de sélection plus forte pour les individus habitant des communes aux revenus médians les plus élevés : Paris 16, Neuilly-sur-Seine, Paris 7, Versailles. L'estimateur utilisé est celui d'Horvitz-Thompson. 70% des individus échantillonnés répondent au questionnaire, mais l'estimateur utilisé ne prend pas en compte la non-réponse.

Que penser des affirmations suivantes ?

1. La non-réponse n'est pas prise en compte dans l'estimation, qui est potentiellement biaisée.
2. La base de sondage présente un défaut de couverture, l'estimation est potentiellement biaisée
3. Donner une probabilité de réponse plus forte pour certains individus conduit à un échantillon non représentatif de la population
4. L'estimateur d'Horvitz-Thompson est biaisé, il vaudrait mieux utiliser l'estimateur plugin (moyenne des patrimoines sur l'échantillon)