
TD 4 : Plans stratifiés

Exercice 1. Dans une population $\mathcal{U} = \{1, 2, 3, 4, 5\}$, on considère le plan de sondage suivant :

$$p(\{1, 2, 4\}) = p(\{1, 2, 5\}) = p(\{1, 4, 5\}) = p(\{2, 3, 4\}) = p(\{2, 3, 5\}) = p(\{3, 4, 5\}) = \frac{1}{6}.$$

Calculez les probabilités d'inclusion d'ordre un et deux ainsi que les Δ_{kl} . Montrez qu'il s'agit d'un plan stratifié.



Exercice 2. On considère une population \mathcal{U} de taille N partitionnée en H strates notées $\mathcal{U}_1, \dots, \mathcal{U}_H$, de tailles respectives N_1, \dots, N_H . On note également $\mu_{y,1}, \dots, \mu_{y,H}$ les moyennes de chaque strate.

Pour chaque strate, on sélectionne un échantillon selon un plan aléatoire simple sans remise de taille n_h , $h = 1, \dots, H$. Les tirages sont indépendants d'une strate à l'autre. Un jeune statisticien propose d'estimer μ_Y par

$$\hat{\mu}_Y = \frac{1}{n} \sum_{k \in S} y_k, \quad n = \sum_{h=1}^H n_h.$$

1. Calculez $\mathbb{E}(\hat{\mu}_Y)$ et en déduire le biais de $\hat{\mu}_Y$.
2. Calculez $\text{Var}(\hat{\mu}_Y)$.
3. Calculez le *ratio du biais*, i.e., le rapport entre le biais et l'écart-type de $\hat{\mu}_Y$.
4. Pourquoi ne faut-il pas utiliser cet estimateur.



Exercice 3. Sur les 7500 employés de l'INSEE, on souhaite connaître la proportion P d'entre eux qui possèdent au moins un véhicule.

Pour chaque individu de la base de sondage, on dispose de la valeur de son revenu. On décide alors de constituer trois strates dans la population : individus de revenu faible (strate 1), de revenu moyen (strate 2), de revenu élevé (strate 3).

Table 1 – Sondage stratifié sur les 7500 employés de l'INSEE.

	$h = 1$	$h = 2$	$h = 3$
N_h	3500	2000	2000
n_h	500	300	200
p_h	0.13	0.45	0.50

On note

N_h :	la taille de la strate h
n_h :	la taille de l'échantillon dans la strate h
p_h :	l'estimateur de la proportion d'individus possédant au moins un véhicule dans la strate h

Les détails du sondage sont donnés au sein du Tableau 1.

1. Quel estimateur, noté \hat{P} , de P proposez vous ? Que peut-on dire de son biais ?
2. Calculez la précision de \hat{P} et donnez un intervalle de confiance à 95% pour P .
3. Estimez-vous que le critère de stratification est adéquat ? Justifiez votre réponse.



Exercice 4. Un directeur de cirque possède 100 éléphants classés en 2 catégories : « mâles et femelles ». Le directeur veut estimer le poids total de son troupeau car il veut traverser un fleuve en bateau. Cependant, l'année précédente, ce même directeur de cirque avait fait peser tous les éléphants de son troupeau et avait obtenu les résultats présentés dans le Tableau 2.

Table 2 – Poids moyens (tonnes) et dispersions selon les strates pour les 100 éléphants du cirque l'année précédente.

	Effectifs N_h	Moyennes $\mu_{y,h}$	Dispersion $S_{y,h}^2$
Mâles	60	6	4
Femelles	40	4	2.25

1. Calculez la dispersion dans la population de la variable « poids de l'éléphant » pour l'année précédente.
2. Le directeur suppose désormais que les dispersions de poids n'évoluent pas sensiblement d'une année sur l'autre. Si le directeur procède à un tirage aléatoire simple sans remise de 10 éléphants, quelle est la variance de l'estimateur du poids total du troupeau ?
3. Si le directeur procède à un tirage stratifié avec allocation proportionnelle de 10 éléphants, quelle est la variance de l'estimateur du poids total du troupeau ?
4. Si le directeur procède à un tirage stratifié optimal de 10 éléphants, quels sont les effectifs de l'échantillon dans chacune des 2 strates et quelle est la variance de l'estimateur total ?