TD 5 : Plans par grappes et à plusieurs degrés

Exercice 1. L'objectif est d'estimer le revenu moyen des ménages. Dans un arrondissement d'une ville composée de 60 îlots de maisons, on sélectionne 3 îlots à probabilités égales et sans remise. On sait, en outre, que 5000 ménages résident dans cet arrondissement. Le résultat est donné dans le tableau suivant :

Table 1 – *Étude du revenu moyen des ménages par un sondage simple sans remise sur 3 îlots pris parmi 60.*

Numéro de l'îlot	Nombre de ménages dans l'îlot	Revenu total des ménages dans l'îlot
1	120	1500
2	100	2000
3	80	2100

1. Estimez le revenu moyen des ménages de l'arrondissement et les revenus totaux des ménages de l'arrondissement par le π–estimateur.
2. Estimez la variance du π–estimateur de la moyenne. (Il est demandé de retrouver les formules du cours et pas de les appliquer).
3. Estimez le revenu moyen des ménages dans l'arrondissement par le ratio de Hájek. Commentez.

✎ ✎ ✎

Exercice 2. Soit la population $\mathcal{U} = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ et le plan de sondage suivant :

$$
\begin{array}{lll}
p(\{1,2\}) = 1/6, & p(\{1,3\}) = 1/6, & p(\{2,3\}) = 1/6, \\
p(\{4,5\}) = 1/12, & p(\{4,6\}) = 1/12, & p(\{5,6\}) = 1/12, \\
p(\{7,8\}) = 1/12, & p(\{7,9\}) = 1/12, & p(\{8,9\}) = 1/12.
\end{array}
$$

1. Donnez les probabilités d'inclusion d'ordre un.
2. Ce plan est–il simple, stratifié, en grappes, à deux degrés ou aucun de ces plans particuliers ? Justifiez votre réponse.

✎ ✎ ✎

Exercice 3. On veut faire une enquêtes sur les hôtels hors région parisienne (90 départements). On met en place un plan à deux degrés en tirant dans un premier temps 10 départements (simple, sans remise) puis en interrogeant 20% des hôtels de chaque département sur la proportion d'étrangers dans sa clientèle (encore simple, sans remise). Les résultats sont résumés dans le tableau suivant :

1. Montrez qu'un estimateur de la proportion moyenne p sur toute la France (sauf la région parisienne) est donné tout simplement par

$$\hat{p} = \frac{\sum_{i \in S_1} N_i \hat{p}_i}{\sum_{i \in S_1} N_i},$$

où vous prendrez soin de définir les quantités N_i, S_1 et $\hat{p}_i$.

Table 2 – *Résultat du sondage sur la proportion d'étrangers dans les hôtels.*

N° département	18	43	25	45	3	57	23	32	28	68
Nb. hôtels	50	65	45	48	52	58	42	66	40	56
Nb. hôtels sondés	10	13	9	10	10	12	8	13	8	11
Proportion	0.40	0.38	0.22	0.30	0.50	0.25	0.38	0.31	0.25	0.36

2. Selon vous, cet estimateur est–il sans biais ? (pas de calcul demandé)
3. Estimez la proportion d'étranger sur les 90 départements.

Exercice 4. On considère un plan à deux degrés pour lequel le π–estimateur du total t_y s'écrit

$$\hat{t}_{y,\pi} = \sum_{i \in S_1} \frac{\hat{t}_{y,i}}{\pi_{1,i}}, \qquad \hat{t}_{y,i} = \sum_{k \in S_{2,i}} \frac{y_k}{\pi_{k|i}}.$$

1. Montrez que la variance de cet estimateur peut s'écrire sous la forme $V_{\text{up}} + V_{\text{us}}$ où V_{up} est un terme de variance relié aux unités primaires et V_{us} aux unités secondaires.
2. Que devient cette expression lorsque les unités primaires et secondaires sont choisies selon un plan simple sans remise ?

Astuce : On se rappellera (ou pas !) que $Var(Y) = Var\{\mathbb{E}(Y \mid X)\} + \mathbb{E}\{Var(Y \mid X)\}$.