2 Sondage aléatoire simple

2.1 Mesure de surfaces cultivées

D'après O. Sautory.

On veut estimer la surface moyenne cultivée dans les fermes d'un canton rural. Sur les $N = 2010$ fermes que comprend ce canton, on en tire $n = 100$ par sondage aléatoire simple sans remise. On mesure pour chaque ferme k de l'échantillon la surface cultivée y_k.

On obtient : $\sum_{k \in s} y_k = 2907\ ha$ et $\sum_{k \in s} y_k^2 = 154593\ ha^2$.

2.1.1

Quelle est la valeur de l'estimateur de Horvitz-Thompson pour la moyenne de la surface cultivée $\bar{Y}$?

2.1.2

Proposer un intervalle de confiance pour cet estimateur.

2.2 Nombre d'ecclésiastiques

On veut estimer le nombre d'ecclésiastiques dans la population française. Pour cela, on choisit d'échantillonner par sondage aléatoire simple n individus. Si la véritable proportion (inconnue) d'ecclésiastiques est de 0.1 %, combien faut-il tirer de personnes pour obtenir un coefficient de variation CV de 5% ?

2.3 Mesure de prix à la pompe

D'après Ardilly, 1992.

On veut estimer l'évolution du prix moyen du litre entre mai et juin via la différence des prix moyens, par sondage aléatoire simple (taille d'échantillon $n < 10$). Les prix pour toutes les stations sont donnés en table 1.

2.3.1

Qualitativement, quelle est la meilleure stratégie d'échantillonnage à adopter ? Pourquoi ?

Station	mai	juin
1	5,82	5,89
2	5,33	5,34
3	5,76	5,92
4	5,98	6,05
5	6,2	6,2
6	5,89	6
7	5,68	5,79
8	5,55	5,63
9	5,69	5,78
10	5,81	5,84

TABLE 1 – Table de données pour l'exercice 2.3

2.3.2

On veut comparer deux méthodes d'échantillonnage :

1. On échantillonne n stations en mai ($n < 10$), puis n stations en juin, de manière totalement indépendante
2. On échantillonne n stations en mai, et on interroge de nouveau ces mêmes stations en juin

Que vaut $\sqrt{\dfrac{\mathbb{V}_1(\bar{p}_{juin} - \bar{p}_{mai})}{\mathbb{V}_2(\bar{p}_{juin} - \bar{p}_{mai})}}$, le rapport des écarts-types des estimateurs obtenus via les deux méthodes ?