

T. D. n° 2

Sondage stratifié

Exercice 1. Issu du livre « Les techniques de sondage » de Pascal Ardilly.

La variable d'intérêt est le chiffre d'affaires moyen réalisé par un ensemble de 1 060 entreprises. Celles-ci étant de tailles très différentes, nous avons constitué 5 strates en fonction du nombre de salariés dans chaque entreprise.

Nombre de salariés	0 à 9	10 à 19	20 à 49	50 à 499	500 et +
Nombres d'entreprises	500	300	150	100	10

1. À l'intérieur de chaque strate nous réalisons un SAS à PEAR avec les tailles d'échantillon suivantes :

$$n_1 = 130, \quad n_2 = 80, \quad n_3 = 60, \quad n_4 = 25, \quad n_5 = 5.$$

Les résultats sont les suivants :

h	1	2	3	4	5
$\widehat{\mu}_h$	5	12	30	150	600
$s_{h,c}^2$	1,5	4	8	100	2 500

Donner une estimation du chiffre d'affaires moyen avec un intervalle de confiance asymptotique à 95%.

2. En conservant toujours la même taille globale d'échantillon, quels effectifs d'échantillon faut-il considérer dans chaque strate :
- (i) pour une allocation proportionnelle ?
 - (ii) Pour une allocation optimale ? (Faire attention d'avoir toujours $n_h \leq N_h$.)
3. En supposant inchangées les variances dans les strates, calculer les variances de l'estimateur pour le plan avec allocation proportionnelle et celui avec allocation optimale.

Exercice 2. Quel est le plan ?

Dans une population $U = \{1, 2, 3, 4, 5\}$ nous considérons le plan de sondage suivant :

$$\begin{aligned} \mathbb{P}[\{1, 2, 4\}] &= \frac{1}{6}, & \mathbb{P}[\{1, 2, 5\}] &= \frac{1}{6}, & \mathbb{P}[\{1, 4, 5\}] &= \frac{1}{6}, \\ \mathbb{P}[\{2, 3, 4\}] &= \frac{1}{6}, & \mathbb{P}[\{2, 3, 5\}] &= \frac{1}{6}, & \mathbb{P}[\{3, 4, 5\}] &= \frac{1}{6}. \end{aligned}$$

Calculer les probabilités d'inclusion d'ordre un et deux ainsi que les $\Delta_{kl} = \text{Cov}(I_k, I_l)$. Montrer qu'il s'agit d'un plan stratifié.

Exercice 3. D'après le livre « Exercices corrigés de méthodes de sondage » de P. Ardilly et Y. Tillé.

Un directeur de cirque possède 100 éléphants classés en deux 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 précédents dans le tableau ci-dessous (les moyennes sont exprimées en tonnes).

	Effectifs N_h	Moyennes μ_h	Variances corrigées $\sigma_{h,c}^2$
Mâles	60	6	4
Femelles	40	4	2.25

1. Calculer la variance σ^2 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 SAS à PESR de 10 éléphants, quelle est la variance de l'estimateur du poids total ?
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 ?
4. Si le directeur procède à un tirage stratifié optimal de 10 éléphants, quels sont les effectifs de l'échantillon dans chacune des deux strates et quelle est la variance de l'estimateur du poids total ?

Exercice 4. D'après le livre « Exercices de sondage » de A.M. Dussaix et J.M. Grosbras.

Une entreprise veut réaliser une enquête auprès de son personnel, qui comprend 10 000 personnes. Des études ont montré que les variables qu'on cherche à analyser dans l'enquête sont très contrastées selon les catégories du personnel, (ce qui entraîne une stratification selon ces catégories en trois strates). Ces mêmes études ont montré également que les variables sont liées à l'âge des individus. Nous allons donc proposer des plans d'échantillonnage comme si nous voulions étudier l'âge des individus. Nous disposons des renseignements suivants :

Catégorie	Poids dans l'ensemble du personnel	Écart-type des âges
1	20%	18,0
2	30%	12,0
3	50%	3,6
Ensemble	100%	16,0

1. Soit μ_Y l'âge moyen et $\hat{\mu}_Y$ l'estimateur issu d'un SAS à PESR de $n = 100$ individus. Quel est l'écart-type de $\hat{\mu}_Y$?
2. Nous décidons que l'échantillon de 100 individus doit être stratifié selon les trois catégories du personnel. Quelle est la répartition « représentative » ? Quel est l'écart-type de l'estimateur de μ_Y ? Comparer avec le résultat de la question 1.
3. Quelle serait la répartition optimale de l'échantillon ? Quel est l'écart-type de l'estimateur de μ_Y ? Comparer avec le résultat de la question 2.