

Théorie des sondages  
Série 2 : Sondage stratifié

Exercice 1 L'objectif est d'estimer le poids total de 100 éléphants afin de les embarquer pour traverser le Gange. on dispose des résultats d'une pesée effectuée l'année précédente. Les valeurs sont consignées dans le tableau qui suit (valeurs en tonnes):

	Effectifs $N_h$	Moyennes $\bar{Y}_h$	Dispersions $\sigma_h^2$
Mâles	60	6	4
Femelles	40	4	2,25

1. Calculer, pour l'année précédente, la dispersion dans la population de la variable Y (poids de l'éléphant).

$$\sigma^2 = \frac{\sum_{h=1}^H N_h \sigma_h^2}{N} + \frac{\sum_{h=1}^H N_h (\bar{Y}_h - \bar{Y})^2}{N}$$

2. On procède à un tirage aléatoire simple sans remise de 10 éléphants. En partant de l'hypothèse que la dispersion n'évolue pas sensiblement d'une année à l'autre, quelle serait la variance de l'estimateur du poids total des éléphants?

3. Qu'en est-il dans le cas d'un sondage stratifié à allocations proportionnelles (n vaut toujours 10)?

$$Var(\hat{Y}_{pop}) = \frac{1}{m} (1-f) \sum_{h=1}^H \frac{N_h}{N} \sigma_{hc}^2 = \frac{1}{m} (1-f) \sum_{h=1}^H \frac{N_h}{N} \cdot \frac{N_h}{N} \sigma_h^2$$

4. On opte maintenant pour un sondage stratifié optimal de 10 éléphants. Donner les effectifs de l'échantillon dans chacune des strates et calculer la variance de l'estimateur du total.

Exercice 2 L'ordre des médecins a procédé à la classification des médecins d'une ville en 3 groupes distincts:

- les débutants ou classe 1 (500 médecins)
- les confirmés ou classe 2 (1000 médecins)
- les très expérimentés ou classe 3 (2500 médecins)

Dans chaque classe, on tire, par sondage aléatoire simple sans remise, un échantillon de 200 médecins. On calcule alors le nombre moyen de patients par jour et par médecin dans chacun des échantillons tirés. Les résultats obtenus sont comme suit:

- Classe 1 :  $\hat{\bar{Y}}_1 = 10$   $s_{1c}^2 = 4$

- Classe 2 :  $\hat{Y}_2 = 15$      $s_{2c}^2 = 7$
- Classe 3 :  $\hat{Y}_3 = 20$      $s_{3c}^2 = 10$

1. Estimer le nombre moyen de patients soignés par jour et par médecin.
2. Donner un intervalle de confiance de niveau 95% pour ce paramètre.

**Exercice 3** On souhaite estimer la moyenne d'une variable d'intérêt relative à l'ensemble des entreprises d'un département. Ces entreprises sont classées selon leur chiffre d'affaires et répertoriées en trois classes. Les données issues d'un recensement antérieur sont les suivantes :

Chiffres d'affaires	Nombre d'entreprises
de 0 à 1	1000
de 1 à 10	100
de 10 à 100	10

Nous avons fixé une taille d'échantillon de 111 entreprises. En supposant que la distribution du chiffre d'affaires est une loi uniforme au sein de chaque classe, donner la variance de l'estimateur de la moyenne du chiffre d'affaires pour un plan stratifié avec allocations proportionnelles puis optimales. Commenter. (on utilisera un sondage aléatoire simple dans chaque strate).