

## 3 Stratification et probabilités inégales

### 3.1 Sondage stratifié selon les revenus

D'après O. Sautory.

Une entreprise emploie 7 500 salariés et souhaite connaître la proportion  $P$  d'entre eux qui possèdent au moins une voiture. Pour chaque individu de la base de sondage, on dispose de son revenu. On décide de stratifier la population selon 3 strates de revenus : revenus faibles (strate 1), revenus moyens (strate 2), revenus élevés (strate 3). Sauf indication contraire, on utilise dans l'exercice les notations du cours. On note également  $p_h$  l'estimateur de la proportion d'individus possédant au moins un véhicule dans la strate  $h$ .

Les résultats de l'enquête menée avec un SAS au sein de chaque strate sont donnés en table 2.

	Strate 1	Strate 2	Strate 3
$N_h$	3 500	2 000	2 000
$n_h$	500	300	200
$p_h$	0,13	0,45	0,50

TABLE 2 – Table de données pour l'exercice 3.1

#### 3.1.1

Quelle est la valeur de l'estimateur d'Horvitz-Thompson pour ces données d'enquête ?

#### 3.1.2

Proposer un intervalle de confiance associé à l'estimateur d'Horvitz-Thompson.

#### 3.1.3

Que pensez-vous de la stratification choisie ?

#### 3.1.4

Que pensez-vous de l'allocation utilisée ?

### 3.2 Choix d'allocation dans une enquête auprès des entreprises

On cherche à mener une enquête sur l'investissement des entreprises (nature, destination, etc.) d'un secteur donné par un SAS stratifié par tranches d'effectif de 300 unités parmi 1 060.

Grâce aux données fiscales, on dispose d'informations sur la moyenne ( $\bar{y}_h$ ) et la dispersion ( $S_h^2$ ) du montant total d'investissement pour chaque tranche d'effectif (en milliers d'euros). Ces informations sont indiquées en table 3.

Taille	$N_h$	$\bar{y}_h$	$S_h^2$
0-9	500	10	2
10-19	300	50	15
20-49	150	200	50
50-499	100	500	100
500 et plus	10	1 000	2 500

TABLE 3 – Table pour l'exercice 3.2

### 3.2.1

Déterminer les allocations proportionnelle et de Neyman en utilisant le montant total d'investissement comme variable auxiliaire.

### 3.2.2

Pour chaque cas, calculer la variance de l'estimateur du montant total d'investissement construit à partir de l'enquête.

## 3.3 Mesure de surfaces cultivées 2

On souhaite mesurer par sondage de taille fixe 2 la surface cultivée et le montant des subventions obtenues pour une population de fermes de taille 6, dont les données figurent en table 4.

On souhaite comparer l'efficacité de deux plans de sondage : d'une part un sondage aléatoire simple, d'autre part un sondage à probabilités inégales et proportionnelles à la taille des fermes  $X$ .

Ferme ( $k$ )	Taille ( $X_k$ )	Production ( $Y_k$ )	Subventions ( $Z_k$ )
A	50	40	680
B	550	660	20
C	150	120	200
D	250	280	100

TABLE 4 – Données pour l'exercice 3.3

Calculer la variance des estimateurs de la moyenne de  $X$ ,  $Y$  et  $Z$  pour les deux plans de sondage et commenter.

### 3.4 Le retour du directeur de cirque

Un directeur de cirque possède 100 éléphants, et veut estimer le poids total de son troupeau car il veut traverser un fleuve en bateau. Le directeur avait déjà fait peser tous les éléphants de son troupeau l'année dernière et avait obtenu les résultats de la table 5 :

	Effectifs $N_h$	Moyennes $\bar{y}_h$	Variances $S_{yh}^2$
Mâles	60	6	4
Femelles	40	4	2,25

TABLE 5 – Données pour l'exercice 3.4

#### 3.4.1

Calculer la variance dans la population du caractère « poids de l'éléphant » pour l'année précédente.

#### 3.4.2

Si, l'année précédente, le directeur avait procédé à un SAS de 10 éléphants, quelle aurait été la variance de l'estimateur du poids total du troupeau ?

#### 3.4.3

Si le directeur avait procédé à un tirage stratifié à allocation proportionnelle de 10 éléphants, quelle aurait été la variance de l'estimateur du poids total du troupeau ?

#### 3.4.4

Si le directeur avait procédé à un tirage stratifié à allocation optimale de 10 éléphants, quels auraient été les effectifs de strates, et quelle aurait été la variance de l'estimateur du poids total du troupeau ?

### 3.5

Quelle semble être la meilleure stratégie d'échantillonnage pour cette année ?