

Glossaire

Une **population** désigne un ensemble d'unités auxquelles on s'intéresse. Il peut s'agir par exemple de l'ensemble des élèves de 2^{nde} d'un lycée (dans ce cas, 1 unité = 1 élève), des entreprises du secteur automobile (1 unité = 1 entreprise), ou encore des ménages vivant en France métropolitaine (1 unité = 1 ménage).

Une **base de sondage** désigne une liste des unités de la population. On peut par exemple disposer de la liste de tous les élèves inscrits en 2^{nde} dans un lycée. On sélectionne dans cette base de sondage une partie des unités de la population appelée **échantillon**, qui va servir à représenter la population entière.

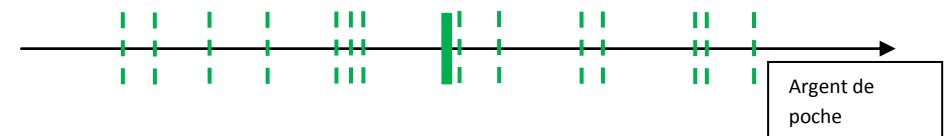
Un échantillon peut être sélectionné de façon **aléatoire**, en donnant à chaque unité une certaine probabilité d'être sélectionnée. On peut en particulier introduire des contraintes sur la composition de l'échantillon, par exemple choisir de sélectionner un échantillon contenant la même proportion d'hommes et de femmes que dans le reste de la population. L'échantillon peut aussi être constitué en sélectionnant des groupes d'éléments, ce qui est parfois plus pratique mais souvent moins précis (on parle alors d'**effet de grappe**). Supposons par exemple que l'on sélectionne 40 élèves pour estimer la note moyenne en mathématiques dans un lycée, en prenant tous les élèves d'une même classe. L'estimation pourra varier fortement selon la sévérité du professeur de mathématiques de la classe sélectionnée.

L'échantillon peut également être sélectionné de façon non aléatoire, mais **empirique**. Par exemple on retiendra tous les élèves d'une classe dont la note moyenne en mathématique est égale à la note moyenne de leur classe (méthode des « unités-type »). On peut par exemple laisser aux enquêteurs le

soin de construire un échantillon contenant la même proportion d'hommes et de femmes que dans le reste de la population (**méthode des quotas**).

Une **estimation** désigne une quantité calculée sur l'échantillon afin d'évaluer une quantité (inconnue) mesurée sur la population entière. Exemple : l'argent de poche moyen des élèves de l'échantillon est utilisé comme estimation de l'argent de poche moyen des élèves de la classe (voir le sujet 4).

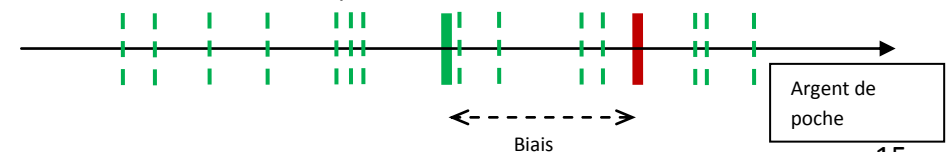
L'**espérance** désigne la valeur moyenne des estimations calculées sur tous les échantillons possibles. Plaçons-nous dans un cas simplifié où l'on pourrait sélectionner et enquêter tous les échantillons possibles. Les estimations associées (traits verticaux verts en pointillés) sont représentées sur une ligne :



L'espérance est symbolisée par le trait vertical épais, en vert. On peut également représenter sur le schéma la quantité à estimer (argent de poche moyen dans l'ensemble de la classe), sous la forme d'un trait vertical épais en rouge. Si l'espérance se confond avec la quantité à estimer, on dit que l'estimation est **sans biais** :



Dans le cas contraire, on dit que l'estimation est **biaisée** :



On parle de **non-réponse** quand une partie (ou la totalité) des questions de l'enquête n'est pas renseignée pour une unité. On parle de **non-réponse totale** quand aucune question n'est renseignée pour une unité (en cas de perte du questionnaire, ou si la personne refuse de répondre par exemple), et de **non-réponse partielle** quand une partie seulement des questions est renseignée pour une unité (par exemple, si le questionnaire est trop long).

La non-réponse partielle pour une unité est généralement corrigée par **imputation**, ce qui consiste à remplacer une valeur manquante par une valeur plausible, à l'aide de l'information dont on dispose.

Dans une enquête par sondage, **l'erreur totale** due à l'estimation comporte plusieurs composantes :

- le biais d'estimation ;
- la dépendance (plus ou moins forte) de l'estimation à l'échantillon tiré, qui renvoie donc à une notion de variabilité de l'estimation ;
- l'effet de la non-réponse ;
- les défauts de la base de sondage ;
- l'erreur de mesure, qui existe dès qu'une unité fournit une réponse qui diffère de la réalité.