

Un échantillon correspond à un sous-ensemble de la population totale, choisi pour la représenter lorsqu'il est impossible, trop long ou trop coûteux d'observer chaque individu. Cela permet par exemple d'estimer un phénomène sans devoir mesurer toute la population. On distingue deux grandes familles de méthodes d'échantillonnage. On peut tout d'abord noter les méthodes probabilistes qui regroupent l'aléatoire simple, le tirage systématique, l'échantillonnage en grappe ou encore le stratifié. Les méthodes non probabilistes reposent au contraire sur des choix guidés par des critères pratiques, comme les quotas, la convenance ou la méthode dite "boule de neige". Le choix des méthodes dépend de plusieurs éléments, à savoir l'objectif de l'étude, la structure de la population et les contraintes matérielles ou organisationnelles.

Un estimateur se définit comme étant une statistique utilisée pour approcher un paramètre inconnu de la population. Il est calculé à partir d'un échantillon, avec l'idée que cette valeur soit la meilleure approximation possible de ce que l'on obtiendrait si l'on observait toute la population. L'estimation correspond, elle, à la valeur numérique prise par l'estimateur une fois calculé sur un échantillon. Autrement dit, c'est la valeur réellement observée dans les données.

L'intervalle de fluctuation est lié à la variabilité d'une statistique lorsqu'on répète différents échantillonnages. Il constitue un outil important en statistique, car il permet l'aide à la décision avec un risque d'erreur contrôlé. L'intervalle de confiance, lui, vise à encadrer la valeur du paramètre qui est utilisée pour estimer un inconnu de la population avec un niveau de confiance fixé à l'avance. Contrairement à l'intervalle de fluctuation, l'intervalle de confiance fournit un degré de certitude sur la précision de l'estimation.

En statistique, un estimateur est une fonction de données qui permet d'estimer un paramètre de la population. On dit qu'il est biaisé lorsque son espérance ne correspond pas à la valeur réelle du paramètre qu'il cherche à approcher.

Le recensement est le nom de statistique que l'on utilise lorsque l'on travaille sur la population entière. Elle repose sur une opération du dénombrement de la population étudiée. Les données massives permettent d'approcher la population totale sans échantillonnage classique.

Le choix d'un estimateur présente plusieurs enjeux. En effet, il repose sur le compromis entre le biais et la variance. Idéalement, un bon estimateur doit alors être consistant, efficace et, si possible, non biaisé.

On distingue plusieurs méthodes d'estimation d'un paramètre, à savoir :

- la méthode des moments, qui consiste à estimer les paramètres en égalisant certains moments théoriques ;
- le maximum de vraisemblance, qui cherche les valeurs des paramètres rendant l'observation de l'échantillon plus probable ;
- la méthode des moindres carrés, utilisée pour déterminer la droite ou la courbe qui s'ajuste le mieux à un ensemble de points.

Le choix entre ces méthodes dépend des propriétés que l'on souhaite obtenir, comme le biais, la variance, la simplicité d'application ou encore les hypothèses nécessaires.

On retient plusieurs types de tests en statistique, à savoir les tests paramétriques, comme le test t, le test Z, ou l'ANOVA, ainsi que les tests non paramétriques, tels que le test de Mann-Whitney, pour n'en citer qu'un.

Ces tests permettent de déterminer si une hypothèse statistique peut être rejetée ou non.

Pour construire un test, il faut définir les hypothèses H_0 et H_1 , choisir une statistique de test, fixer une région critique et un risque α .

L'inférence est critiquée pour la dépendance aux conditions d'échantillonnage, aux hypothèses fortes et à l'interprétation des p-values.