

COURS 8 : VOCABULAIRE DE STATISTIQUES DESCRIPTIVES

Une **population** statistique est l'ensemble sur lequel on effectue des observations. Les éléments de cet ensemble sont appelés les **individus**.

Ce qui est observé ou mesuré sur les individus d'une population statistique est une **variable (ou caractère) statistique**. Les valeurs prises par une variable statistique sont appelées ses **modalités**.

L'**effectif** d'une modalité i est le nombre d'individus n_i ayant cette modalité, et sa **fréquence** f est la proportion d'individus ayant cette modalité, c'est-à-dire

$$f = \frac{n_i}{n},$$

où n est l'effectif total.

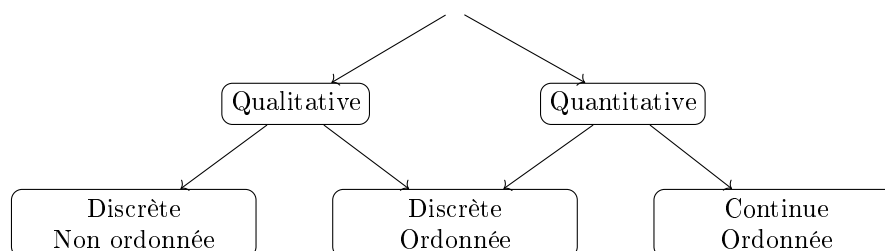
Exemple : Les clients d'une entreprise forment une population. On peut observer leur âge, leurs dépenses annuelles, etc. : ce sont des modalités.

TAXONOMIE

Nous allons faire trois distinctions importantes :

- Une variable est dite **quantitative** si ses modalités représentent une quantité. On peut calculer la moyenne et les autres paramètres statistiques. Une variable qui n'est pas quantitative est dite **qualitative**.
- Une variable est dite **ordonnée** si ses modalités ont un ordre naturel, et **non ordonnée** sinon.
- Une variable est dite **discrète** si elle ne peut prendre que des valeurs isolées, généralement entières. Une variable est dite **continue** si elle peut prendre n'importe quelle valeur dans un intervalle, en particulier des nombres décimaux.

Toute variable qualitative est discrète, et toute variable quantitative est ordonnée. La détermination de caractères d'une variable peut donc se faire à l'aide du diagramme suivant :



Attention : Une variable représentée par un numéro n'est pas quantitative ! Par exemple, un numéro de téléphone, un numéro de département ou une adresse sont des variables qualitatives : faire une moyenne de numéros de téléphones n'a aucun sens.

REPRÉSENTATION

De nombreux types de représentations graphiques existent. Dans les grandes lignes (des exceptions existent toujours!), nous recommandons :

- variables **discrètes, non ordonnées** : diagramme circulaire (“camembert”);
- variables **discrètes, ordonnées** : diagramme en bâtons;
- variables **continues** : histogrammes.

Cette recommandation n’est pas rigide, et on pourra par exemple utiliser un diagramme en bâtons au lieu d’un diagramme circulaire s’il y a trop de modalités (plus de 4 ou 5), ou un histogramme au lieu d’un diagramme en bâtons pour une variable quantitative discrète ayant beaucoup de modalités.

Bien entendu, cela n’aborde pas d’autres types de représentations (cartes, notamment).