

I/ POPULATION, INDIVIDUS ET VARIABLES

I.1/ Population

La **population** est l'ensemble des éléments qui forme le champ d'analyse d'une étude particulière.

Par exemple, dans une étude sur l'emploi, la population pourrait être l'ensemble des personnes en âge de travailler (→ exemple 1).

Malgré la connotation démographique, le concept de population en statistique est général et ne s'applique pas seulement aux êtres humains, mais aussi aux choses, aux événements, etc...

Dans une étude relative à l'assurance automobile, la population pourrait être l'ensemble des voitures assurées (→ exemple 2), ou bien l'ensemble des accidents survenus durant une période donnée (→ exemple 3)

I.2/ Individus

La population est donc constituée d'un ensemble d'éléments que l'on appelle **individus**. Dans l'exemple 1, les individus sont les individus de la population dans le sens courant du terme.

Dans l'exemple 2, les voitures sont les individus.

Dans l'exemple 3, les accidents sont les individus .

Il est fondamental de bien préciser la population et ses individus avant de s'engager dans les calculs et le traitement des données.

I.3/ Variable

Si les individus d'une population possèdent en commun le caractère d'être tous membres de la même population, ils varient cependant selon d'autres critères. Les voitures assurées par une compagnie d'assurance ont toutes les caractères communs d'être couvertes par la même assurance, mais elles varient selon leur couleur, leur marque, leur puissance, leur prix, le nombre de kilomètres parcourus, etc...

Ces caractéristiques sont appelées, en statistique, des **variables**.

Elles servent à décrire la population en question, c'est-à-dire, à préciser quels sont les aspects de cette population qui nous intéressent et qui seront analysés dans la présente étude.

Une variable, souvent représentée symboliquement par une lettre majuscule comme X ou Y, comprend d'une part, un **libellé** qui permet d'intituler la variable, et d'autre part un ensemble de **modalités** décrivant les différentes valeurs possibles de la variable.

Par exemple, la variable décrivant la couleur des voitures assurées aurait comme:

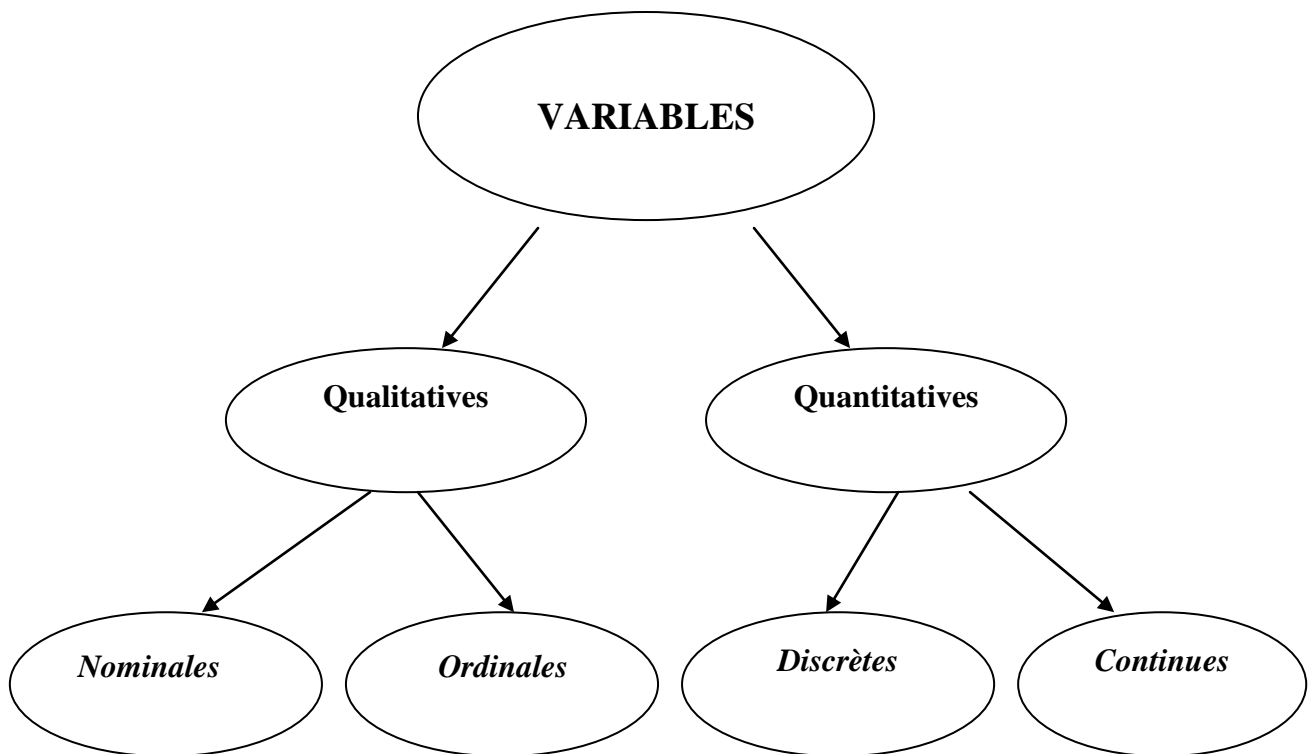
Libellé: « Couleur »

Modalités : « Rouge » « Orange » « Jaune » « Vert » « Bleu » « Indigo » et « Violet »

Le nombre de modalités selon lequel on définit une variable peut-être modifié suivant les besoins de l'enquête. Les couleurs des voitures peuvent être, par exemple, beaucoup plus nuancées et variées que les sept couleurs de l'arc-en-ciel.

Les différentes modalités des variables doivent être à la fois **incompatibles** (un individu ne peut pas appartenir à la fois à deux ou plusieurs modalités) et **exhaustives** (tous les cas ont été prévus).

Comme certaines méthodes statistiques s'appliquent à quelques types de variables et pas à d'autres, il est nécessaire de distinguer les différentes sortes de variables :



**Variables
Qualitatives**



Une variable qualitative est une variable dont les modalités sont des mots ou des lettres

Exemple:

la variable « Sexe » a 2 modalités : « homme » et « femme »

Les modalités d'une variable qualitative peuvent être classées sous la forme d'une **échelle nominale** ou d'une **échelle ordinale**.

Echelle nominale : on dit d'une variable dont les modalités ne sont pas naturellement ordonnées, qu'elle est définie sur une échelle nominale.
Exemple: la variable « sexe »

Echelle ordinale : Si les modalités peuvent être ordonnées, on est en présence d'une échelle ordinale : les modalités représentent donc un ensemble de rapports ordonnés.
Exemple : la variable « qualification professionnelle » avec les 3 modalités : « non qualifié » « semi-qualifié » et « qualifié »

**Variables
Quantitatives**



Une variable quantitative est une variable dont les modalités sont des valeurs numériques

Exemples :

Les variables « Age », « Température », « Revenu »

Une distinction fondamentale concernant les variables quantitatives est celle effectuée entre les variables **discrètes** et les variables **continues**.

Variable discrète : Une variable quantitative est dite discrète si elle a un nombre fini ou dénombrable de valeurs.

Exemple : Nombre de personnes dans une famille, nombre d'accidents dans une journée.

Variable continue : Une variable quantitative est dite continue si toutes les valeurs d'un intervalle de \mathfrak{R} sont acceptables.

Exemple : Poids d'un nouveau né, taille d'un individu

I.4/ Observation

Les résultats observés d'une variable sur une population constituent les **observations**.

Par exemple, si on considère la variable « âge au dernier anniversaire observé » pour une population de 5 personnes (Jean, Marie, Thérèse, Luc et Thomas), les observations pourraient être : 19 ans, 21 ans, 20 ans, 19 ans et 22 ans. Les observations représentent, respectivement, l'âge des cinq personnes.

Par convention, on utilise les lettres minuscules de l'alphabet pour représenter les observations, indexées par un chiffre entier pour distinguer les observations correspondant aux différents individus.

Donc, les observations obtenues sur une variable quelconque X seront exprimées par :

$$x_1, x_2, x_3, \dots, x_n$$

où n représente le nombre d'individus et x_i la valeur de X pour le $i^{\text{ème}}$ individu observé.

I.5/ Série statistique

On parle aussi de série statistique :

Une série statistique est l'ensemble des valeurs observées pour une ou plusieurs variables sur les n individus de la population.

Une série (statistique) simple est l'ensemble des n valeurs observées pour une variable.

Une série simple est représentée par $(x_1, x_2, \dots, x_i, \dots, x_n)$

Une série (statistique) double est l'ensemble des valeurs observées pour deux variables : elle est représentée par une série de n couples:

$$(x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n)$$

(x_i, y_i) étant la valeur du couple de variables (X, Y) pour le $i^{\text{ème}}$ individu observé, $i = 1, \dots, n$.

I.6/ Données – Tableau de données individuelles

Si une enquête porte sur p variables et n individus, l'ensemble des observations récoltées peut se présenter sous forme de tableau de données individuelles :

TABLEAU DE DONNEES INDIVIDUELLES

Individus	Variables			
	X_1	X_2	X_j	X_p
1	X_{11}	X_{12}	X_{1j}	X_{1p}
2	X_{21}	X_{22}	X_{2j}	X_{2p}
...
i	X_{i1}	X_{i2}	X_{ij}	X_{ip}
...
n	X_{n1}	X_{n2}	X_{nj}	X_{np}

x_{ij} est la valeur prise par la j -ième variable pour le i -ième individu

Les x_{ij} correspondent aux observations relevées auprès de n individus de l'enquête. Les valeurs prises par ces observations forment ce qu'on appelle aussi **les données de base** (synonyme du mot observation).

Mais la notion de « données » a un sens beaucoup plus large que le terme « observation ».

Quand on parle de données, on ne fait pas seulement référence aux données de base, mais également à toutes les transformations que l'on a pu faire à partir des données initiales.

Exemples de transformation de données :

- Les catégories des variables qualitatives sont codées avec des valeurs numériques afin de faciliter leur traitement informatique. Dans le cas de la variable « Sexe », par exemple, on pourrait noter « 0 » pour la catégorie homme et « 1 » pour la catégorie femme.
- Pour l'analyse statistique de certains phénomènes, il peut être préférable d'utiliser les logarithmes des observations plutôt que les données de base.
- Etc...