

LA STATISTIQUE DESCRIPTIVE

1. HISTORIQUE ET DÉFINITION

Aussi loin que l'on remonte dans le temps et dans l'espace (en Chine et en Égypte, par exemple), les États ont toujours senti le besoin de disposer d'informations sur leurs sujets ou sur les biens qu'ils possèdent et produisent. Mais les recensements de population et les recensements de ressources, les statistiques (du latin *status* : état) sont restées purement descriptives jusqu'au 17^{ème} siècle.

Puis s'est développé le calcul des probabilités et des méthodes statistiques qui sont apparues en Allemagne, en Angleterre et en France. Beaucoup de scientifiques de tous ordres ont apporté leur contribution au développement de cette science 'LA STATISTIQUE' : PASCAL, HUYGENS, BERNOULLI, MOIVRE, LAPLACE, GAUSS, MENDEL, PEARSON, FISCHER etc....

Actuellement, beaucoup de domaines utilisent les méthodes statistiques (médecine, agronomie, sociologie, industrie etc....).

Définition de la statistique :

La **statistique** est d'un point de vue théorique une science, une méthode et une technique. La statistique comprend : la collecte des données, le traitement des données collectées, l'interprétation des données, la présentation afin de rendre les données compréhensibles par tous.

Ainsi la statistique est un domaine des mathématiques qui possède une composante théorique ainsi qu'une composante appliquée. La composante théorique est proche de la théorie des probabilités et forme avec cette dernière, les sciences de l'aléatoire. La statistique appliquée est utilisée dans presque tous les domaines de l'activité humaine¹ : ingénierie, management, économie, biologie, informatique, etc. Ces distinctions ne consistent pas à définir plusieurs domaines étanches. En effet, le traitement et l'interprétation des données ne peuvent se faire que lorsque celles-ci ont été collectées. La statistique possède des règles et des méthodes sur la collecte des données, pour que celles-ci puissent être correctement interprétées.

2. Domaine d'application

En 1982, le statisticien Pierre Dagnelie propose trois grandes tendances de la statistique⁴ :

- la statistique qualifiée d'« administrative » ou « gouvernementale » faite dans les instituts de statistique à propos de grands ensembles de données,
- la statistique dite « mathématique » ou « universitaire » faite avec peu de données et qui a pour but la novation,
- enfin la statistique « appliquée » ou « de terrain » faite dans les instituts de sondage d'opinion ou les facultés de médecine pour des problèmes concrets.

Dans la pratique, les méthodes et outils statistiques sont utilisés dans des domaines très variés :

1. en [géophysique](#), pour les prévisions [météorologiques](#), la [climatologie](#), la [pollution](#), les études des [rivières](#) et des [océans](#) ;
2. en [démographie](#) : le [recensement](#) permet de faire une photographie à un instant donné d'une [population](#) et permettra par la suite des sondages dans des échantillons représentatifs ;
3. en sciences [économiques](#) et [sociales](#), et en [économétrie](#) : l'étude du comportement d'un groupe de population ou d'un [secteur économique](#) s'appuie sur des statistiques. C'est dans cette direction que travaille l'[Insee](#). Les questions environnementales s'appuient également sur des [données](#) statistiques ;
4. en [sociologie](#) : les sources statistiques constituent des matériaux d'enquête, et les méthodes statistiques sont utilisées comme techniques de traitement des [données](#) ;
5. en [marketing](#) : le [sondage d'opinion](#) devient un outil pour la [décision](#) ou l'[investissement](#) ;
6. dans les [jeux de hasard](#) et les paris tels que le [loto](#) ou les paris équestres, pour "prévoir" les résultats ;
7. en [physique](#) : l'étude de la mécanique statistique et de la thermodynamique statistique (cf [Physique statistique](#)) permet de déduire du comportement de particules individuelles un comportement global (passage du microscopique au macroscopique) ;
8. en [métrologie](#), pour tout ce qui concerne les systèmes de mesure et les mesures elles-mêmes ;
9. en [médecine](#) et en [psychologie](#), tant pour le comportement des [maladies](#) que leur fréquence ou la validité d'un traitement ou d'un dépistage ;
10. en [archéologie](#), appliquée aux vestiges (céramologie, archéozoologie...)
11. en [écologie](#), pour l'étude des communautés végétales et des écosystèmes.
12. en [assurance](#) et en [finance](#) (calcul des risques, [actuariat](#)...)

3. Les principales branches de la statistique

Les principales branches de la statistique sont :

- la **statistique descriptive** qui propose des résumés graphiques et numériques des données à des fins de première analyse (application directe : tableaux de bord marketing)
- l'**analyse factorielle** qui cherche à réduire le nombre de variables décrivant un phénomène en produisant des indicateurs numériques de synthèse (combinaison des caractères observés)
- la **classification** qui produit des typologies ou des groupes homogènes dans une population disparate
- la **modélisation** qui vise à expliquer un phénomène en fonction des caractères observés, en permettant également une prédiction de valeurs sur des individus nouveaux ou inconnus
- les **tests** qui vérifient un certain nombre d'hypothèses sur des informations numériques (homogénéité, dépassement d'un seuil, ...)

Le **DataMining** peut être vu comme un prolongement de la statistique qui s'attache au traitement de gros volumes de données, dans un but décisionnel, en utilisant des méthodes issues de la statistique "classique" (catégories ci-dessus) et de l'informatique décisionnelle (intelligence artificielle), telles que les réseaux de neurones ou les raisonnements basés sur la mémoire.

Donc cette science se divise en deux grandes branches :

Recueil de données, classement et réduction de ces données (**statistique descriptive**)

Analyse de ces données visant à la déduction de prévisions (**statistique inférentielle**).

Dans ce cours, Nous allons nous contenter de faire de la statistique descriptive.

Il s'agit de décrire les données statistiques par les tableaux statistiques et les différents graphiques. Et aussi synthétiser les données par Les mesures de

tendance centrale (quantiles, moyennes, mode et médiane...) et de dispersion (Étendue, Intervalle inter-quartile Écart moyen absolu, variance et écart-type)

4. Concepts et vocabulaire statistiques

4.1 La population statistique : la population statistique (notée Ω) est l'ensemble sur lequel on effectue des observations. On appelle donc la population statistique, tout ensemble d'unités statistiques constituant les unités observées. Elle est bien spécifiée s'il n'y a pas d'ambiguïté sur la définition de l'ensemble.

On peut distinguer deux types de population statistique.

La population statistique finie: quand le nombre d'unités observées est limité par exemple : le nombre d'étudiants dans une section, nombre d'enfants dans un ménage ...

La population statistique infinie: quand le nombre d'unités observées est illimité ou très grand par exemple : nombre d'étoiles dans le ciel, nombre de graines de blé collectées, nombre de pièces de monnaie en libre circulation...

4.2 L'unité statistique : On appelle unité statistique (ou individu statistique), tout élément de la population statistique étudiée. On la note ω .

4.3 L'échantillon statistique : C'est un sous ensemble de la population statistique considérée. Le nombre d'individus dans l'échantillon est **la taille** de l'échantillon.

En statistique prélever un **échantillon** consiste à extraire un ou plusieurs individus d'une population. Les renseignements obtenus sur un échantillon permettent de mieux connaître la population. Le recours à un échantillon répond en général à la nécessité pratique (manque de temps, de place, évaluation destructive d'une production...) ou économique (coût trop élevé) de s'abstraire de l'étude exhaustive de la population.

Un échantillon doit être représentatif.

4.4 Le caractère statistique (variable statistique):

C'est la propriété ou l'aspect singulier que l'on se propose d'observer dans la population ou l'échantillon. Un caractère qui fait le sujet d'une étude porte aussi le nom de **variable statistique**. Ce caractère doit caractériser tout individu appartenant à la population statistique étudiée.

Exemple : couleur des yeux, taille des filles, ...

Le caractère quantitatif :

Lorsque la variable peut être exprimée numériquement, donc lorsqu'on peut la mesurer, elle est dite quantitative (ou mesurable). Dans ce cas, elle peut être discrète (discontinue) ou continue. Les valeurs d'une variable quantitative sont des nombres exprimant une quantité sur lesquels les opérations arithmétiques sont possibles et ont un sens (somme, division, ...).

La variable peut alors être discrète ou continue selon la nature de l'ensemble des valeurs qu'elle est susceptible de prendre (valeurs isolées ou intervalle de \mathbb{R}).

- ✓ Elle est discrète si elle ne prend que des valeurs entières, des valeurs isolées les unes des autres (exemple : nombre d'enfants d'une famille).
- ✓ Elle est dite continue lorsqu'elle peut prendre toutes les valeurs d'un intervalle fini ou infini (exemple : diamètre de pièces, poids, salaires...).

Le caractère qualitatif :

Lorsque la variable ne se prête pas à des valeurs numériques, elle est dite qualitative (exemple : opinions politiques, nationalité, couleurs des yeux...). Elle peut être ordonnée ou non, dichotomique ou non. Ses valeurs sont des modalités (ou catégories) exprimées sous forme littérale ou par un codage numérique sur lequel des opérations arithmétiques n'ont aucun sens.

On distingue des variables qualitatives ordinales ou nominales, selon que les modalités peuvent être naturellement ordonnées ou pas.

Une variable est dichotomique si elle n'a que 2 modalités.

Exemple :

Etat du temps constaté à un endroit donné chaque jour (pluvieux, neigeux, beau, venteux, ...)

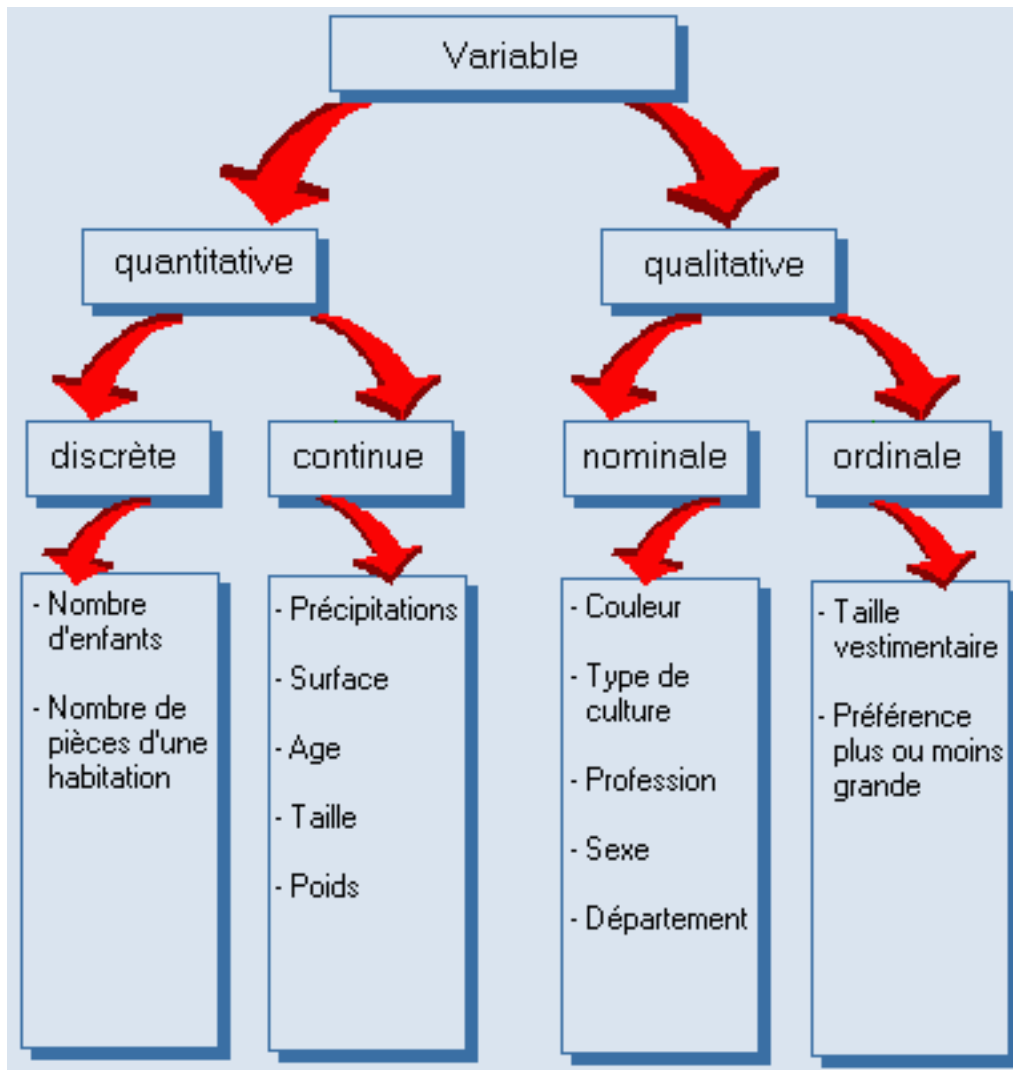


Fig. 1 – Résumé des différents types de variables

Exemple :

Population statistique = Population algérienne

Unité statistique = l'individu algérien

Echantillon = quelques algériens

Caractère statistique = la taille en mètre

Type de ce caractère = quantitatif

4.5 Les modalités statistiques

Chaque unité statistique occupe une position ou une place dans la population statistique par rapport au caractère étudié. Les individus se répartissent dans des modalités de la variable.

Exemple 1 :

On réalise une étude sur le niveau des étudiants de première année universitaire en informatique.

Population statistique = les étudiants de première année universitaire

Caractère statistique = le niveau scientifique

Type du caractère = qualitatif

Modalités = faible, moyen, bon, très bon, excellent

Exemple 2 :

On tire un échantillon de 20 ménages et on compte le nombre d'enfants dans chaque ménage, on obtient les résultats suivants :

0 1 3 1 4 5 5 5 6 7
2 1 4 4 4 5 2 1 1 2

Population statistique = 20 ménages

Caractère statistique = nombre d'enfants

Type du caractère = quantitatif

Modalités de ce caractère = 0, 1, 2, 3, 4, 5, 6, 7

Attention :

Avant chaque étude statistique, vous devez :

- 1. identifier avec précision la population statistique et l'unité statistique**
- 2. cibler le caractère étudié pour confirmer l'appartenance de chaque individu à la population étudiée**
- 3. Chaque individu doit appartenir à une modalité du caractère et un seul seulement et une modalité peut contenir plusieurs unités. La somme des individus des modalités = la somme des individus de la population étudiée**