

1- Quelques terminologies en statistique

Comme toute science, la statistique possède son propre vocabulaire, qu'il convient de définir avant d'en aborder l'étude

1-1. Population et recensement

Une population est l'ensemble de toutes les personnes, de tous les objets, de tous les faits sur lesquels porte une étude. Chaque élément d'une population est appelé unité statistique

Un recensement est une étude réalisée sur toutes les unités statistiques d'une population

1-2. Échantillon et sondage

Un échantillon est un sous ensemble d'unités de la population sur lesquelles on effectue une étude. Si l'échantillon est choisi au hasard, on peut généraliser certains résultats à l'ensemble de la population. Dans le cas contraire, on ne peut le faire

Un sondage est une enquête menée auprès de l'échantillon de la population que l'on désire étudier.

Exemple1

Dans un sondage, on interroge 1000 électeurs, choisis au hasard parmi tous les électeurs de l'Algérie, afin de connaître leur intention de vote

Population : tous les électeurs de l'Algérie

Unité statistique : un électeur

Échantillon : 1000 électeurs choisis au hasard

Exemple2

On veut vérifier si, au cours de la dernière de production, la quantité moyenne de jus versé par une machine dans des contenants est bien conforme au volume désiré. Pour ce faire, on prélève au hasard 20 contenants remplis durant la dernière heure de production et on mesure la quantité de jus dans chacun.

Population : l'ensemble des contenants remplis durant la dernière heure de production

Unité statistique: un contenant de jus

Échantillon : les 20 contenants prélevés au hasard

1-3. La variable

Chaque individu, ou unité statistique, appartenant à une population est décrit par un ensemble de caractéristiques appelées variables. Sa valeur peut différer d'une unité statistique à l'autre. Il est possible d'associer plus d'une variable à une même unité statistique. On emploie une lettre majuscule (X,Y ou Z) pour désigner une variable, et une lettre minuscule (x, y ou z) pour désigner sa valeur.

Exemple

- a- Si l'unité statistique étudiée est un travailleur, on peut choisir comme variable : le salaire, le sexe, l'âge
- b- Si l'unité statistique étudiée est une plante, on peut choisir comme variable : la longueur de la tige, la couleur des feuilles, l'aspect de la feuille
- c- Si l'unité statistique étudiée est un étudiant, on peut choisir comme variable : le nombre d'étudiants, la langue d'enseignement, le nombre de modules enseigné

2- Classification des variables

On distingue deux types de variables : les variables quantitatives et les variables qualitatives

- Une variable est dite quantitative si les valeurs qu'elle prend sont des nombres (quantités)
- Une variable est dite qualitative si les valeurs qu'elle prend sont des catégories

Exemple

Q1- « quel est votre sexe ? »

Variable : sexe

Catégories : féminin, masculin

Type de variable : variable qualitative

Q2 -« combien d'enfants avez-vous ? »

Variable : le nombre d'enfants

Valeurs : 0 enfant, 1 enfant, 2 enfants, 3 enfants.

Type de variable : variable quantitative

2-1. Types de variables

Une variable qualitative peut être classée en variable ordinale, si il y a une relation d'ordre entre les catégories de la variable, si non elle est dite nominale

Exemple

L'intensité de la douleur : faible, moyenne, forte « variable qualitative ordinale »

La nationalité : Algérienne, française, canadienne, nigérienne « variable qualitative nominale »

- Une variable quantitative est dite discrète (discontinue) si les valeurs qu'elle peut prendre sont isolées les unes des autres, par contre elle est dite continue si la variable peut prendre n'importe qu'elle valeur contenue dans un intervalle donné de nombre réels

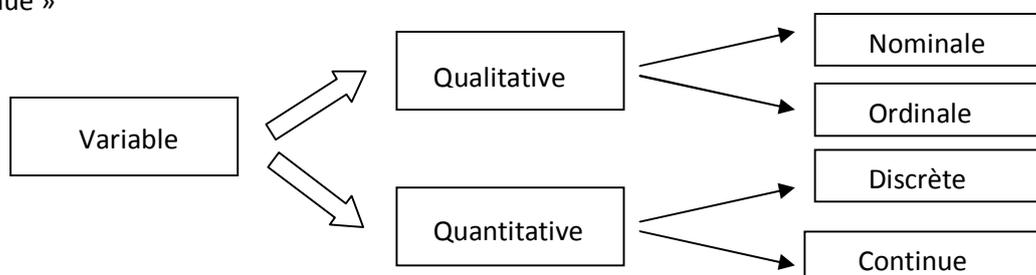
Exemple

Le nombre d'arbres par hectare : 100, 200, 500, 700. « Variable quantitative discrète »

Le nombre de passagers par bus : 20, 50, 80, 120. « Variable quantitative discrète »

Âge des agriculteurs d'une région : âge compris entre 25 et 65 ans « variable quantitative continue »

La taille des étudiants : la taille peut être comprise entre 165 cm et 185cm « variable quantitative continue »



2-2. Les échelles de mesure

On distingue plusieurs types d'échelles de mesure, soit nominale, ordinale, d'intervalle et de rapport

2-1- Échelle nominale

Il est souvent pratique d'identifier les catégories d'une variable nominale par des codes numériques (variable couleur des yeux, yeux bleu=1, yeux verts=2). La variable nominale servant exclusivement à identifier la catégorie à laquelle chaque observation appartient, ses différentes valeurs ne représentent que des étiquettes, des codes numériques. Le chiffre qui code chaque valeur de la variable nominale est arbitraire. Par conséquent, les informations contenues dans les variables nominales n'ont aucune propriété mathématique. Par exemple, on peut attribuer le code 0 à la catégorie « masculin » et le code 1 à la catégorie « féminin » de la variable sexe. Dans ce contexte, il n'y a évidemment pas de relation d'ordre entre les codes : on ne peut pas écrire $0 < 1$, puisque 0 et 1 sont des chiffres attribués de façon arbitraire. De plus bien qu'on puisse techniquement calculer la moyenne des codes 0, 1, 1, 1, en les additionnant et en divisant la somme par 4, le résultat n'a aucun sens. Comment interpréter une moyenne de 0.75 pour « la variable sexe »

Une échelle nominale sert uniquement à différencier les catégories d'une variable au moyen de codes. Elle ne permet pas d'établir une relation d'ordre entre les codes, ni d'effectuer des opérations arithmétiques (+, -, X, ÷) sur ceux-ci.

2-2- Échelle ordinale

Une échelle ordinale permet de mesurer la position de chaque observation par rapport aux autres observations sur une variable. Cette position se nomme rang. Contrairement aux variables nominales, le chiffre numérique attribué à chaque observation n'est pas arbitraire, mais a une signification. Cette signification représente la position de chaque observation relative. Les codes numériques que nous assignons à chaque catégorie de la variable représentent une différence réelle. Par exemple, à la question « Aimez-vous les études ? », les catégories pourraient être codées ainsi :

1- Pas du tout 2- Un peu 3- Moyennement 4- Beaucoup

Dans ce cas, on peut établir une relation d'ordre entre les codes et affirmer que $1 < 2$ ou $3 < 4$.

On utilise aussi une échelle ordinale dans le cas d'une variable quantitative dont les valeurs sont présentées sous forme de catégories codées. Par exemple, à la question « quel est votre salaire ? »

1- 19 000 Da et moins 2- entre 20 000 et 49 000 Da 3- 50 000 et plus

Les codes numériques permettent d'affirmer que les revenus de code 1 sont inférieurs à ceux de code 2, mais on ne peut pas effectuer de calculs avec les codes $(1+2+2+4) \div 4 = 2.25$ ne peut correspondre à la moyenne des salaires des 4 personnes.

Une échelle ordinale sert à différencier les catégories d'une variable et à les ordonner selon des codes par contre, elle ne permet pas d'effectuer des opérations arithmétiques (+, -, X, ÷) sur ces codes

On se sert d'une échelle d'intervalle uniquement pour des variables quantitatives. Les échelles d'intervalles mesurent non seulement la position relative de chaque observation, mais indiquent aussi l'ampleur des différences entre elles. Les échelles d'intervalles contiennent plus d'informations que ne le font les échelles ordinales ou les échelles nominales.

Les échelles d'intervalles souffrent néanmoins d'une limite importante : elles n'ont pas de points zéro. L'absence d'un zéro absolu pour certaines mesures apparaît lorsque cette valeur est impossible. Les échelles d'intervalles sont souvent utilisées pour mesurer des phénomènes en sciences humaine (niveau d'intelligence « QI » des élèves, mesure de la personnalité, le degré de satisfaction de la clientèle, le degré de stress).

Par exemple, la mesure de la chaleur en degré Celsius (C^0) est une mesure d'intervalle. S'il fait 10^0C le lundi, 15^0C le mardi et 30^0C le mercredi, nous pouvons conclure que la température a davantage augmenté de mardi à mercredi qu'elle ne l'a fait entre lundi et mardi. Il serait faux de conclure qu'il fait deux fois plus chaud mercredi que mardi, car une température de zéro ne signifie pas une absence totale de chaleur

Une échelle d'intervalle sert à différencier les valeurs de la variable, et à les ordonner et à les comparer au moyen de la soustraction, mais non de la division

2-3- Échelle de rapport

On emploie une échelle de rapport avec des variables quantitatives. Dans cette échelle le zéro est absolu, c'est-à-dire que la valeur zéro signifie donc une absence de la valeur de la variable. Par exemple, on utilise une échelle de rapport pour mesurer le nombre d'échecs d'un étudiant : un étudiant qui a 6 échecs en 4 de plus de celui qui en a 2 ; on peut aussi dire qu'il en a 3 fois plus. La valeur zéro indique qu'un étudiant n'a aucun échec

Une échelle de rapport sert à différencier les valeurs de la variable, et à les ordonner et à les comparer au moyen de la soustraction et de la division. C'est le type d'échelle qui offre le plus d'information

Echelle de mesure et type de variable

Échelle de mesure	Type de variable
Échelle nominale	Qualitative nominale
Échelle ordinale	Qualitative ordinale Quantitative (discrète et continue)
Échelle d'intervalle	Quantitative (discrète et continue)
Échelle de rapport	Quantitative (discrète et continue)