

Statistique descriptive

Chapitre 2 : Population, variable, modalité



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Les trois mots de base

La variable statistique

La typologie des variables

Ce qu'il faut retenir

Les trois mots de base

Les trois définitions

- La **population** (ou univers) est l'ensemble des éléments sur lesquels porte l'étude.
- Un **individu** (ou unité statistique) est un élément de cette population.
- L'**effectif total**, noté n , est le nombre d'individus observés.

Un individu n'est pas forcément une personne

C'est le point que les étudiants oublient le plus vite. Un individu statistique peut être :

une entreprise, une commune, un ménage, une transaction, un jour de cotation, un produit en rayon, un contrat d'assurance.

L'individu est ce qu'on compte. Rien de plus.

Définir la population est déjà une décision

« Les salariés de l'entreprise » : à quelle date ? Les intérimaires en font-ils partie ? Les salariés à temps partiel ? Ceux en congé ?

Ces choix ne sont pas des détails : ils changent n , donc toutes les fréquences, donc toutes les moyennes.

Une étude statistique commence toujours par une définition écrite de la population. C'est ce qu'on appelle le champ de l'enquête, et c'est le premier endroit où regarder quand deux chiffres officiels se contredisent.

La variable statistique

Variable statistique et modalité

Une **variable statistique** est le caractère observé sur chaque individu.

Les **modalités** sont les valeurs que cette variable peut prendre.

Deux exigences sur les modalités

Les modalités doivent être :

- **exhaustives** : chaque individu en possède au moins une ;
- **incompatibles** : chaque individu n'en possède qu'une seule.

Autrement dit, elles forment une **partition** de la population.

Une exigence qui n'est pas théorique

Un questionnaire propose comme réponses : « moins de 5 000 », « de 5 000 à 10 000 », « plus de 10 000 ». Un salarié à 5 000 dirhams coche laquelle ?

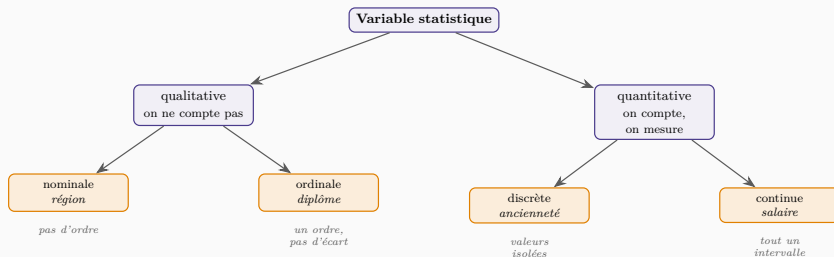
Les deux premières conviennent : les modalités ne sont **pas incompatibles**. Le tableau sera faux, quelle que soit la qualité des calculs qui suivront.

La convention à retenir : les classes s'écrivent **fermées à gauche, ouvertes à droite**, $[5\,000 ; 10\,000[$.

Sur nos 500 salariés

Élément	Ce que c'est ici
Population	les salariés de l'entreprise au 31 décembre
Individu	un salarié
Effectif total	$n = 500$
Variable	le salaire, l'ancienneté, le diplôme, la région, le sexe
Modalités du diplôme	Bac, Bac+2, Bac+5
Modalités du sexe	F, H

La typologie des variables



Qualitative nominale

Les modalités sont des **catégories sans ordre**.

Exemples : la région, le secteur d'activité, la forme juridique, le sexe.

On peut **compter** les individus de chaque modalité. On ne peut ni les ordonner, ni faire de moyenne.

Qualitative ordinale

Les modalités sont **ordonnées**, mais les écarts entre elles n'ont pas de sens numérique.

Exemples : le diplôme, un niveau de satisfaction, une notation financière (AAA, AA, A...), une catégorie socio-professionnelle hiérarchisée.

On peut compter **et** ordonner — donc calculer une médiane. On ne peut pas faire de moyenne.

Le piège de la variable ordinale codée

On code souvent $\text{Bac} = 1$, $\text{Bac}+2 = 2$, $\text{Bac}+5 = 3$, et l'on obtient une « moyenne de diplôme » de 1,78.

Ce nombre n'a **aucun sens** : il suppose que l'écart $\text{Bac} \rightarrow \text{Bac}+2$ est le même que $\text{Bac}+2 \rightarrow \text{Bac}+5$, alors que rien ne le dit. Le codage était arbitraire ; 1, 2, 5 aurait donné une autre moyenne.

Un codage numérique ne transforme pas une variable qualitative en variable quantitative. C'est l'erreur la plus fréquente dans les mémoires d'étudiants.

Quantitative discrète

Les valeurs sont **isolées**, généralement entières, et on peut les énumérer.

Exemples : l'ancienneté en années, le nombre d'enfants, le nombre de salariés, le nombre de transactions.

Quantitative continue

La variable peut prendre **toutes** les valeurs d'un intervalle. On est obligé de la **regrouper en classes**.

Exemples : le salaire, le chiffre d'affaires, la durée, le taux de rentabilité.

La frontière est une question de pratique, pas de principe

Le salaire est-il vraiment continu ? Il s'exprime en dirhams et centimes : c'est donc discret au sens strict.

Mais il prend **381** valeurs différentes sur **500** salariés. En dresser la liste ne résumerait rien. On le traite donc comme continu et on le regroupe en classes.

La règle pratique : si le nombre de valeurs distinctes est du même ordre que n , on traite la variable comme continue.

Le tableau des opérations permises

	Nominale	Ordinale	Discrète	Continue
Effectifs, fréquences	oui	oui	oui	oui
Mode	oui	oui	oui	oui
Effectifs cumulés	non	oui	oui	oui
Médiane, quantiles	non	oui	oui	oui
Moyenne, écart-type	non	non	oui	oui
Diagramme en bâtons	oui	oui	oui	non
Histogramme	non	non	non	oui

Comment lire ce tableau

Les colonnes sont **emboîtées** : chaque type autorise tout ce que le précédent autorise, plus quelque chose.

Il suffit donc de retenir les **deux frontières** :

- le **cumul** exige un ordre;
- la **moyenne** exige des écarts qui aient un sens.

Toutes les erreurs de ce cours viennent d'un franchissement de l'une de ces deux frontières.

La variable binaire

Une variable qualitative à **deux** modalités — le sexe, oui/non, en défaut ou non.

Codée 0 et 1, elle possède une propriété remarquable : sa **moyenne est la proportion** de 1.

Vérification

Sur nos données, en codant 1 pour les femmes : la somme vaut 215, le nombre de femmes, et la moyenne vaut $215/500 = 0,43$ — soit exactement la part des femmes dans l'effectif.

C'est la seule variable qualitative dont on puisse légitimement calculer la moyenne — parce que le codage 0/1 n'est pas arbitraire : il compte.

Pourquoi c'est important

Toute la statistique des proportions, en inférentielle, repose sur ce fait. Une proportion **est** une moyenne, et tous les résultats sur les moyennes s'y appliquent.

Ce qu'il faut retenir

Six points

1. **Population, individu, effectif total n** : à définir par écrit avant tout calcul.
2. Une **variable** prend des **modalités**, qui doivent être **exhaustives** et **incompatibles**.
3. Quatre types : nominale, ordinale, discrète, continue.
4. Le **cumul** exige un ordre ; la **moyenne** exige des écarts qui aient un sens.
5. Coder une variable qualitative en nombres ne la rend pas quantitative.
6. Une variable **binaire** codée 0/1 a pour moyenne la proportion.

La suite

Le vocabulaire est en place. On peut maintenant construire l'objet qui résume à lui seul toute une variable : le **tableau statistique**.

C'est le chapitre 3, et il ouvre la deuxième partie du cours.