

Statistique descriptive

Chapitre 1 : Population, variable, modalité

Jaouad Madkour

12 octobre 2026

Plan du chapitre

Les trois mots de base	2
Population, individu, effectif	2
Définir la population est déjà une décision	2
La variable statistique	2
Variable et modalités	2
Des modalités mal construites	3
Une ligne, une colonne, une case	3
Le vocabulaire, sur le fil rouge	3
La typologie des variables	4
L'arbre à connaître	4
Les variables qualitatives	4
Le piège de la variable ordinale codée	5
Ce qu'une variable ordinale autorise	5
Les variables quantitatives	5
Discret ou continu : une question de pratique	6
Ce que le type autorise	6
Deux frontières à retenir	6
La variable binaire	7
Une proportion est une moyenne	7
S'entraîner	7
Exercice 1 : classer les variables d'une enquête	7
Exercice 1 : solution	8
Exercice 2 : cinq entreprises	8
Exercice 2 : solution	8
Ce qu'il faut retenir	9
Le résumé du chapitre	9

Les trois mots de base

Population, individu, effectif

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

Le champ de l'enquête

« 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 le premier endroit où regarder quand deux chiffres officiels se contredisent.

Avec ou sans les intérimaires

Notre entreprise emploie aussi 60 intérimaires, dont 45 femmes.

- Sans eux : $n = 500$ et la part des femmes vaut $215/500 = 43,0\%$.
- Avec eux : $n = 560$ et elle vaut $(215 + 45)/560 = 260/560 = 46,4\%$.

Même entreprise, même date : **trois points et demi d'écart**, dus à la seule définition.

La variable statistique

Variable et modalités

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.

Des modalités mal construites

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[.

Et l'oubli d'une modalité

Une question « Votre statut : salarié, indépendant, chômeur » laisse sans réponse l'étudiant et le retraité : les modalités ne sont **pas exhaustives**. On ajoute alors une modalité « autre », et l'on vérifie qu'elle reste marginale.

Une ligne, une colonne, une case

id	salaire	anciennete	diplome	region	sexe
1	8 540	10	Bac+5	Rabat	H
2	7 100	15	Bac	Rabat	F
3	4 380	4	Bac	Autres	F
4	11 790	19	Bac+2	Casablanca	F
5	13 430	19	Bac+5	Tanger-Tétouan	H
6	6 650	7	Bac+2	Tanger-Tétouan	H
⋮	⋮	⋮	⋮	⋮	⋮

une ligne : **un individu**
(le salarié n° 3)

une case : **une modalité**
(H, pour le salarié n° 5)

en tout : 500 lignes, 6 colonnes

une colonne : **une variable**
(le salaire)

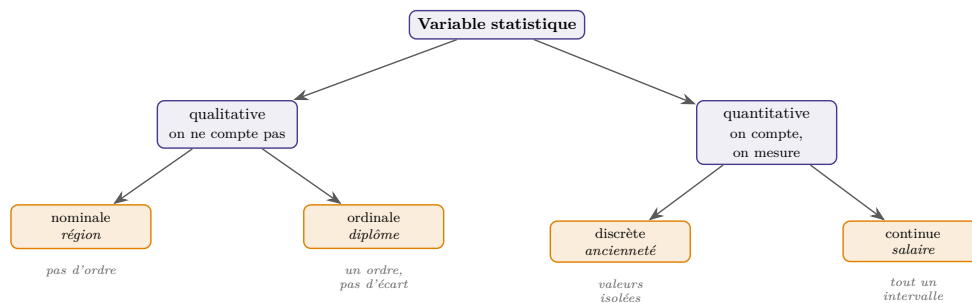
Lire un fichier de données

Un fichier statistique est toujours rangé de la même façon : **un individu par ligne, une variable par colonne**, et dans chaque case la modalité prise par cet individu pour cette variable.

Le vocabulaire, sur le fil rouge

Sur nos 500 salariés

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



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 ordinaire

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.

Une « moyenne de diplôme »

On code souvent $\text{Bac} = 1$, $\text{Bac}+2 = 2$, $\text{Bac}+5 = 3$, et l'on obtient une « moyenne de diplôme ». Ce nombre suppose que l'écart $\text{Bac} \rightarrow \text{Bac}+2$ vaut l'écart $\text{Bac}+2 \rightarrow \text{Bac}+5$, alors que rien ne le dit.

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

Trois codages, trois moyennes

Codage (Bac, Bac+2, Bac+5)	Calcul	« Moyenne »
1, 2, 3	$(219 + 2 \times 174 + 3 \times 107)/500$	1,78
1, 2, 5	$(219 + 2 \times 174 + 5 \times 107)/500$	2,20
1, 3, 6	$(219 + 3 \times 174 + 6 \times 107)/500$	2,77

Le codage était arbitraire, la moyenne l'est donc aussi.

Le diplôme médian

On range les 500 salariés par diplôme et l'on cherche celui du milieu, aux rangs 250 et 251.

Diplôme	Effectif	Effectif cumulé
Bac	219	219
Bac+2	174	393
Bac+5	107	500

La lecture

Les rangs 1 à 219 ont le Bac, les rangs 220 à 393 un Bac+2. Les rangs 250 et 251 tombent dans ce second groupe : **le diplôme médian est Bac+2**. Ce résultat ne dépend d'aucun codage — seulement de l'ordre.

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 s'exprime en dirhams et centimes : il est discret au sens strict. Mais la liste de ses valeurs ne résumerait rien : on le traite comme continu. **Règle pratique** : si le nombre de valeurs distinctes est du même ordre que n , on traite la variable comme continue.

Compter les valeurs distinctes

Variable	Valeurs distinctes sur 500	Traitement
salaire	381	continue : classes
anciennete	32	discrète, ou classes
region	4	qualitative nominale
diplome	3	qualitative ordinale
sexe	2	binaire

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.

Sur le fil rouge

- Région : on peut dire « la plus représentée est Tanger-Tétouan », jamais « la région médiane ».
- Diplôme : « le diplôme médian est Bac+2 », jamais « le diplôme moyen ».
- Salaire et ancienneté : tout est permis.

Deux modalités, codées 0 et 1

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.

En effet, la somme des x_i compte les individus codés 1 :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{\text{nombre de 1}}{n}$$

Vérification

En codant 1 pour les femmes : la somme vaut 215, et la moyenne $215/500 = 0,43$ — exactement la part des femmes.

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

La part des Bac+5, par sexe

On code 1 un salarié titulaire d'un Bac+5, 0 sinon, et l'on fait la moyenne dans chaque groupe :

	Effectif	Bac+5	Moyenne de la variable 0/1
Femmes	215	48	$48/215 = 0,223$
Hommes	285	59	$59/285 = 0,207$
Ensemble	500	107	$107/500 = 0,214$

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. Un taux de chômage, un taux de défaut, une part de marché en sont des exemples.

Une enquête auprès des ménages

Une enquête relève, pour chaque ménage d'un quartier de Tanger :

1. le nombre de personnes du ménage ;
2. le statut d'occupation du logement : propriétaire, locataire, logé gratuitement ;
3. le revenu mensuel du ménage, en dirhams ;
4. la satisfaction à l'égard du quartier : faible, moyenne, forte ;
5. le milieu de résidence d'origine : urbain, rural ;
6. la dépense alimentaire mensuelle, en dirhams.

Pour chaque variable, donner son type, puis dire si l'on peut calculer un mode, une médiane,

une moyenne. Laquelle peut-on coder 0/1 ?

Le classement

Variable	Type	Mode	Médiane	Moyenne
1. Nombre de personnes	quantitative discrète	oui	oui	oui
2. Statut d'occupation	qualitative nominale	oui	non	non
3. Revenu mensuel	quantitative continue	oui	oui	oui
4. Satisfaction	qualitative ordinale	oui	oui	non
5. Milieu d'origine	qualitative nominale	oui	non	non
6. Dépense alimentaire	quantitative continue	oui	oui	oui

La variable binaire

Le **milieu d'origine** n'a que deux modalités : codé 1 pour rural, sa moyenne donne la **proportion** de ménages d'origine rurale. Revenu et dépense, qui prennent presque autant de valeurs que de ménages, se traitent comme continus.

Un fichier d'entreprises

Entreprise	Secteur	Salariés	Exportatrice	Taille
A	textile	180	oui	moyenne
B	agroalimentaire	45	non	petite
C	textile	620	oui	grande
D	services	12	non	petite
E	automobile	310	oui	moyenne

1. Quels sont la population, l'individu et l'effectif total ?
2. Donner le type de chacune des quatre variables.
3. Coder « exportatrice » en 0/1 et calculer sa moyenne. Que représente-t-elle ?
4. Déterminer la taille médiane. Peut-on parler de « secteur médian » ?

Vocabulaire et types

1. Population : ces cinq entreprises ; individu : **une entreprise** — pas une personne ; $n = 5$.
2. Secteur : qualitative nominale. Salariés : quantitative discrète. Exportatrice : qualitative nominale **binaire**. Taille : qualitative ordinale (petite < moyenne < grande).

Moyenne binaire et médiane ordinale

3. Codage 1, 0, 1, 0, 1 : moyenne $3/5 = 0,6$, soit 60 % **d'entreprises exportatrices**.
4. On range les tailles : petite, petite, moyenne, moyenne, grande. Avec $n = 5$, la médiane est la valeur de rang 3 : **la taille médiane est « moyenne »**. Le secteur n'a pas d'ordre : un « secteur médian » n'a aucun sens ; seul le mode existe ici, le textile.

Sept points

1. **Population, individu, effectif total** n : à définir par écrit avant tout calcul ; changer la définition change les chiffres.
2. Une **variable** prend des **modalités, exhaustives et incompatibles**.
3. Un fichier : un individu par ligne, une variable par colonne.
4. Quatre types : nominale, ordinale, discrète, continue.
5. Le **cumul** exige un ordre ; la **moyenne** exige des écarts qui aient un sens.
6. Coder une variable qualitative en nombres ne la rend pas quantitative ; une ordinale a une médiane, pas de moyenne.
7. 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 2.