

Cours de statistique descriptive

Séance 1 : Les caractères et la mise en forme des données

Les statistiques et la géographie

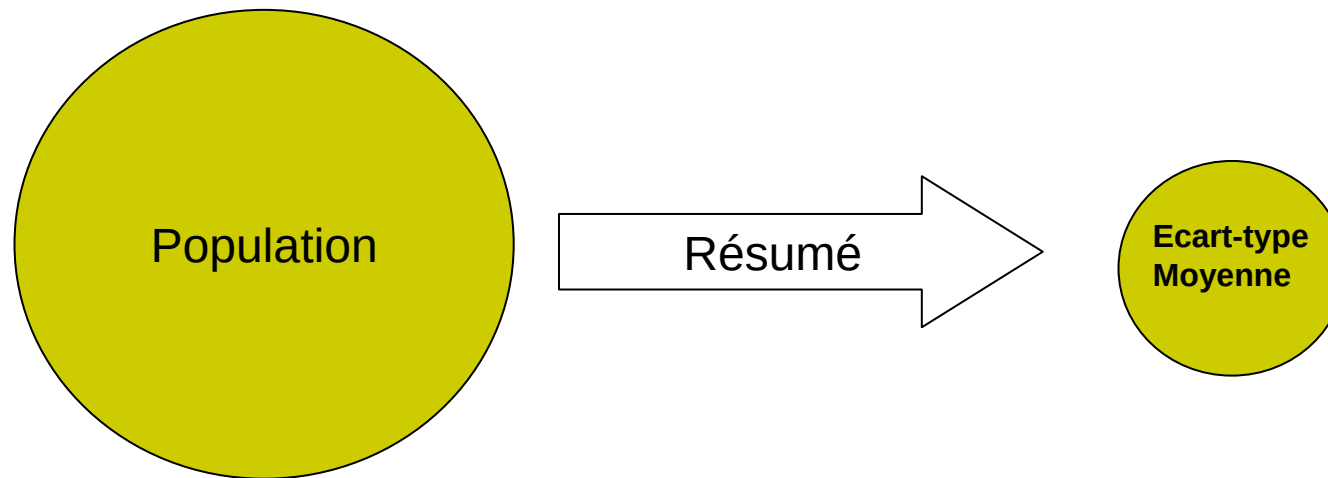
- ❑ La géographie est une science sociale

- ❑ Les statistiques sont un instrument de mesure des phénomènes géographiques
 - ➡ Approche numérique d'un problème
 - ➡ Raisonnement basé sur un échantillon

Organisation de la statistique

La statistique descriptive et la statistique inférentielle

La statistique descriptive

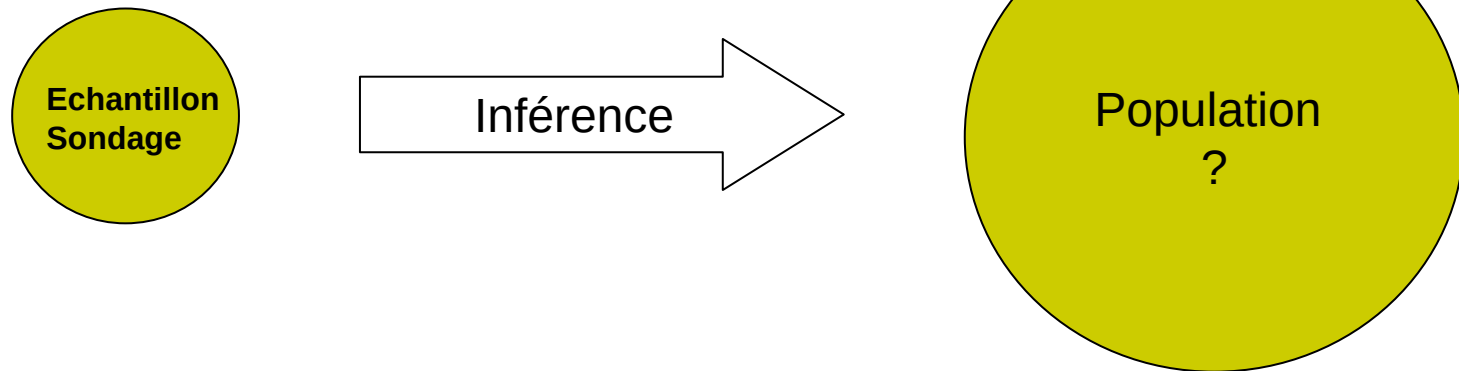


Elle permet de résumer quantitativement l'information recueillie sur un ensemble d'éléments au moyen d'une investigation exhaustive. C'est la statistique descriptive, qui fait l'objet de ce cours.

Organisation de la statistique

La statistique descriptive et la statistique inférentielle

La statistique inférentielle



Elle permet de généraliser à de grands ensembles d'éléments les conclusions tirées des résultats obtenus avec des ensembles beaucoup plus restreints appelés échantillons. C'est la statistique inférentielle ou probabiliste, qui n'est pas abordée dans cette UE

La statistique univariée, bivariée, multivariée

□ Univariée

➔ A chaque élément d'un ensemble correspond un seul caractère

□ Statistique bivariée ou multivariée

➔ A chaque élément d'un ensemble correspond plusieurs caractères

➔ L'objet de cette statistique sera de repérer les distributions homogènes et celles qui ne le sont pas

Plan des séances

- ❑ TD 1 : Le Tableau élémentaire
- ❑ TD 2 : La valeur centrale (moyenne, médiane, mode)
- ❑ TD 3 : Réflexion sur la valeur centrale : calcul et utilisation des paramètres de dispersion

(moyenne pondérée, moyenne sur des classes).
- ❑ TD 4 : Représentation graphique :

(Diagramme en barre ; histogramme et courbe de fréquence cumulée)
- ❑ TD 5 : Tableau de contingence avec standardisation

II – Le tableau élémentaire

Le tableau élémentaire : définition

- ❑ C'est un tableau à simple entrée où les lignes correspondent aux **éléments** de l'ensemble étudié et les colonnes aux **caractères** (ou variables) décrivant ces éléments.

La première colonne est en principe réservée à la liste nominale des éléments.

Les États du Proche-Orient

Nom	N°	superficie en milliers de km²	place pour la superficie	population en millions	densité hab/km²	PIB en \$	monnaie	fécondité en nbre enf/fem	façade littorale	consommation d'énergie kgep/h
					Elément 1					
Ar Saoudite	1	2 150	1	21.4	10	7 230	riyal	5.7	Mer Rouge	4 204
					Elément 2					
Bahreïn	2	0.700	14	0.7	1 035	9 370	dinar bahr	2.8	Golfe Persique	9 000
Em Ara Unis	3	84	9	3	35.6	-	dirham	3.5	Golfe Persique	9 977
Egypte	4	1 001	3	65.2	65	1 530	livre égy	3.5	Mer Méditer	709
Iran	5	1 633	64	64	40	1 750	rial iranien	2.6	Golfe Persique	1 651
Irak	6	438	23	23	54.3	-	dinar irak	5.3	Golfe Persique	1 263
Israël	7	21	10	6.4	309	16 710	shekel	3	Mer Méditer	3 029
Jordanie	8	89	8	5	57	1 750	dinar jor	3.6	Mer Méditer	1 028
Koweït	9	18	11	2	115	18 030	dinar kow	4.2	Golfe Persique	8 984
Liban	10	10	13	4.4	429	4 010	livre lib	2.5	Mer Méditer	1 280
Oman	11	212	6	2.4	11.5	-	ryal oman	6.1	Mer d'Oman	3 607
Qatar	12	11	12	0.6	54.3	-	rial qat	3.9	Golfe Persique	28 262
Syrie	13	185	7	16.6	90	1 000	livre syr	4.1	Mer Méditer	1 143
Yémen	14	528	4	18	34	460	rial	7.2	Golfe d'Aden	184

Ensemble

- ❑ C'est la collection (le plus souvent finie en géographie) d'unités, ou d'éléments, sur laquelle porte l'observation.
- ❑ Pour que cet ensemble soit correctement défini, il faut lui donner une définition précise de façon à ce que deux personnes différentes aboutissent toujours à la même liste d'éléments.
- ❑ L'ensemble des éléments observés sera appelé E
- ❑ Le titre du tableau élémentaire défini l'ensemble observé sans ambiguïté

Les États du Proche-Orient

Nom	N°	superficie en milliers de km ²	place pour la superficie	population en millions	densité hab/km ²	PIB en \$	monnaie	fécondité en nbre enf/fem	façade littorale	consommatio n d'énergie kgep/h
Ar Saoudite	1	2 150	1	21.4	10	7 230	riyal	5.7	Mer Rouge	4 204
Bahreïn	2	0.700	14	0.7	1 035	9 370	dinar bahr	2.8	Golfe Persique	9 000
Em Ara Unis	3	84	9	3	35.6	-	dirham	3.5	Golfe Persique	9 977
Egypte	4	1 001	3	65.2	65	1 530	livre égy	3.5	Mer Méditer	709
Iran	5	1 633	2	64.7	40	1 750	rial iranien	2.6	Golfe Persique	1 651
Irak	6	438	5	23.7	54.3	-	dinar irak	5.3	Golfe Persique	1 263
Israël	7	21	10	6.4	309	16 710	shekel	3	Mer Méditer	3 029
Jordanie	8	89	8	5	57	1 750	dinar jor	3.6	Mer Méditer	1 028
Koweït	9	18	11	2	115	18 030	dinar kow	4.2	Golfe Persique	8 984
Liban	10	10	13	4.4	429	4 010	livre lib	2.5	Mer Méditer	1 280
Oman	11	212	6	2.4	11.5	-	ryal oman	6.1	Mer d'Oman	3 607
Qatar	12	11	12	0.6	54.3	-	rial qat	3.9	Golfe Persique	28 262
Syrie	13	185	7	16.6	90	1 000	livre syr	4.1	Mer Méditer	1 143
Yémen	14	528	4	18	34	460	rial	7.2	Golfe d'Aden	184

Quel est l'ensemble observé ?
En donner une définition précise.

Éléments

- ❑ Les éléments sont les objets constitutifs de l'ensemble.
- ❑ Ce sont des objets déterminés dont l'appartenance à tel ou tel ensemble E est sans ambiguïté.
- ❑ Les éléments peuvent être désignés par leur position dans le tableau de données : 1 pour le premier, i pour un élément quelconque, N pour le dernier élément.
 - ➡ $Nb : \text{éléments} = \text{individus}$
 - ➡ Exemple : stock de matériels, réponses à une question, population française.

- 
- ☐ *Combien d'éléments composent cet ensemble ?*
 - ☐ *Exemple d'éléments ?*
 - ☐ *Repérer le code, identificateur ?*
 - ☐ *Le nom ?*

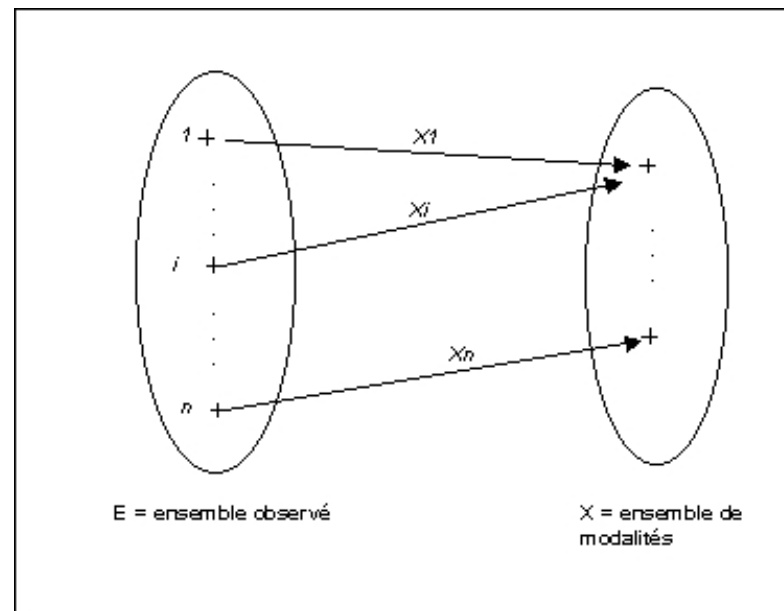
III – Les caractères dans le tableau élémentaire

Caractère (ou variable)

- ❑ Les éléments d'un ensemble sont décrits par un caractère.
- ❑ Le caractère est **une propriété que tous les individus d'un ensemble possède** même si ça modalité est différente (ex : le sexe d'une population : tout le monde à un sexe, mais la modalité est différente : F ou M, ex : l'âge etc).
- ❑ Cela revient à établir une correspondance entre chaque élément i de l'ensemble E et l'ensemble X des modalités ou des valeurs du caractère. **La fonction $f : E \rightarrow X$ $i \mapsto x_i$ est une application au sens mathématique**: chaque élément de E a une modalité (caractère qualitatif) ou une valeur (caractère quantitatif) et une seule dans X .

Caractère

❑ Exemple



E = Les Etats et territoires Americains avec $E = \{ \text{Honduras, Belize, Brésil, ...} \}$

X = Statut politique en 1989 avec $X = \{ \text{"dépendant"; "indépendant"} \}$

Pour chaque élément de E on peut établir une application dans l'ensemble du caractère Statut Politique

Différents types de caractère

1) Les caractères qualitatif : MODALITES

❑ Caractère qualitatif : deux types : **nominal** et **ordinal**

❑ Les **nominaux** ne sont pas mesurables, ce sont des **noms** ou ce qui revient au même des sigles ou des **codes**. Les différentes modalités ne sont pas ordonnables. Attention, même si les modalités sont des codes numériques, les opérations sur les modalités n'ont aucun sens.

- ➡ Exemple Ex 1 : type de relief avec trois modalités (plaine, montagne, plateau)
- ➡ Ex 2 : taille de la ville avec quatre modalités (petite, moyenne, grande, très grande)

❑ Les **caractères ordinaux**, ils s'expriment sur **une échelle ordinale**: chaque *modalité* est explicitement significative du rang pris par chaque individu pour le caractère considéré. Si E possède N éléments, les modalités seront 1er, 2eme, 3eme etc?. Comme on possède juste l'ordre des individus, on ne sait rien de l'intervalle des valeurs.

- ➡ Exemple: Rang des départements pour la population en 1999.

Différents types de caractère

2) Les caractères quantitatif : MESURES

- ❑ Les caractères quantitatifs expriment des quantités issus d'une mesure (ex : superficie) ou d'un dénombrement (ex : nombre d'enfant par ménage).
- ❑ Il est possible de les ordonner et de faire des calculs dessus
- ❑ Deux types de caractères quantitatifs
 - ➡ Les caractères quantitatifs repérable ou mesurable
 - ➡ Les caractères quantitatifs de stock ou de taux

Caractère quantitatif repérable ou mesurable

- ❑ Quantitatif repérable sur une échelle d'intervalle. Ces caractères permettent de repérer la position de chaque élément par rapport à une origine arbitraire. **La valeur 0 est donc conventionnelle et ne signifie pas l'absence du phénomène.**

➡ *Exemple: Latitude, longitude, température, altitude, ...*

- ❑ Quantitatif mesurable sur une échelle numérique. **Le 0 signifie bien l'absence du phénomène**

➡ *Exemple: population, taux de fécondité, précipitations*

Caractères quantitatifs de stock ou de taux

- ❑ Les caractères quantitatifs **de stock** expriment des **quantités concrètes** : la somme des valeurs prises par l'ensemble E des éléments a un sens.

➡ *Exemple: Population totale d'un département.*

- ❑ Les caractères quantitatifs **de taux** expriment le **rapport entre deux valeurs**, on les appelle parfois caractères de rapport. Leur total n'a pas de signification.

➡ *Exemple: Densité de population d'un département*

➡ *Exemple : proportion des actifs chômeurs à une date donnée.*

Caractère quantitatif discret ou continu

- ❑ Les caractères quantitatifs **discrets** sont des caractères dont les différentes situations où peuvent se trouver les éléments sont des **nombre isolés** dont la liste peut être établie a priori.

➡ *Exemple: Nombre de villes de plus de 100 000 h dans chaque département, Nombre d'enfants par foyer*

- ❑ Les caractères quantitatifs **continus** sont des caractères les différentes situations où peuvent se trouver les éléments ne sont pas dénombrables et qui sont définies sur **un intervalle (continu)** de valeur donnée.

➡ *Exemples: Superficie des départements, Altitude moyenne des départements, salaire moyen des employés en 1999*

En résumé

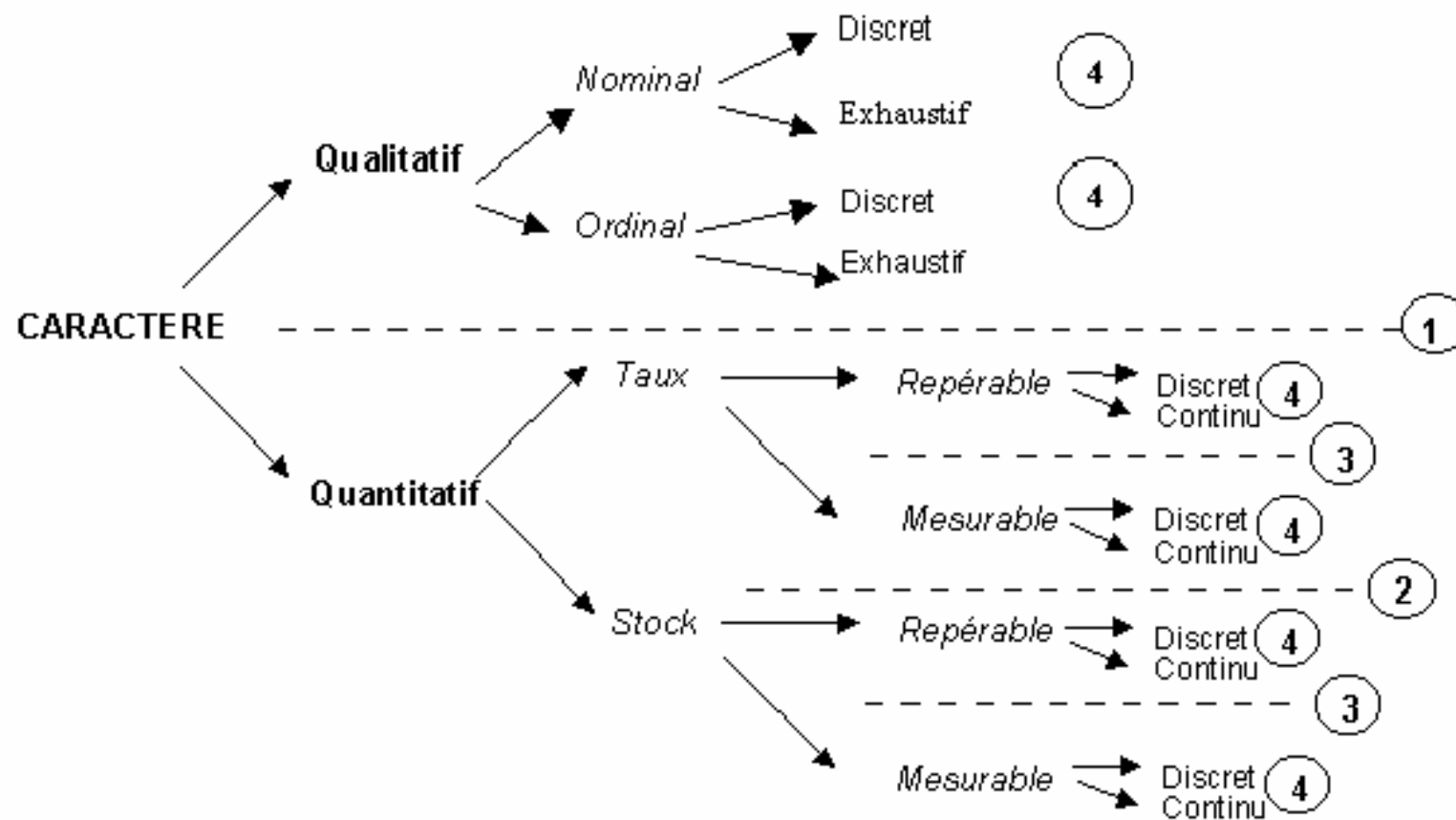
- ❑ **Un caractère quantitatif doit être caractérisé par un point**

- ➔ Exhaustif (autant de modalités que d'élément) ou discret (moins de modalités que d'éléments)

- ❑ **Un caractère qualitatif doit être caractérisé par trois points**

- ➔ Repérable ou mesurable
- ➔ De Stock ou de Taux
- ➔ Discret ou continu

- ❑ **Intérêt de la définition des caractères** : Il est très important de savoir définir le type d'un caractère car les outils statistiques ou les représentations graphiques ne sont pas les mêmes selon le type de caractère à étudier.



- ① Un caractère est qualitatif si la moyenne des valeurs est impossible ou absurde
Un caractère est quantitatif si la moyenne des valeurs possède une signification
- ② Un caractère quantitatif est un taux si la somme des valeurs n'a pas de signification
Un caractère quantitatif est un stock si la somme des valeurs a une signification
- ③ Un caractère quantitatif est repérable si la valeur 0 est conventionnelle
Un caractère quantitatif est mesurable si la valeur 0 signifie concrètement l'absence
- ④ Un caractère qualitatif ou quantitatif est discret s'il a moins de modalités que d'éléments
Un caractère quantitatif est continu si ses modalités possibles sont en nombre infini
Un caractère qualitatif est exhaustif s'il y a autant de modalités que d'éléments

Exercice : à partir du tableau des États du moyen Orient définir la nature de chaque caractère

	Superficie	Rang	Population	Densité	Pib/hab	Monnaie	Fécondité	Façade	Cons°NRJ
Qual/quant									
Ord ou non									
Mesurable ou non									
Taux ou stock									
Discrète ou continu									

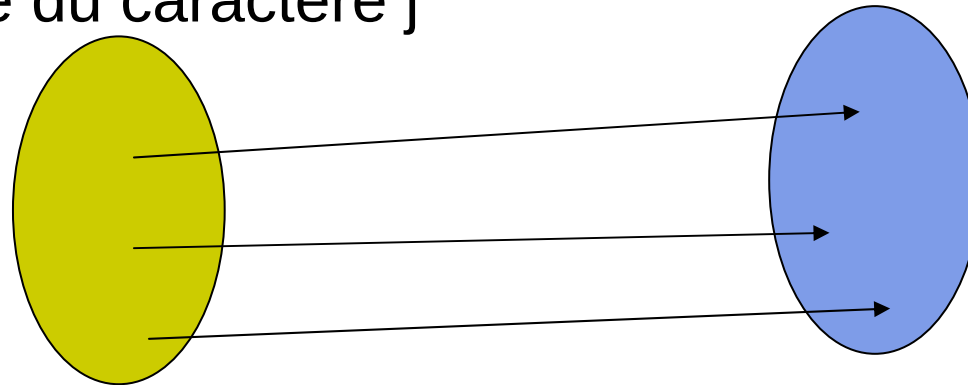
Exercice : à partir du tableau des États du moyen Orient définir la nature de chaque caractère

	Superficie	Rang	Population	Densité	Pib/hab	Monnaie	Fécondité	Façade	Cons°NR J
Qual/quant	Quant.	Qual.	Quant.	Quant.	Quant.	Qual.	Quant.	Qual.	Quant.
Ord ou non		Ord.				Non ord.		Non ord.	
Mesurable ou non	Mesurable		Mesurable	Mesurable	Mesurable		Mesurable		Mesurable
Taux ou stock	Stock		Stock	Taux	Taux		Taux		Taux
Discrète ou continu	Continue		Discrète.	Continue	Continue		Continue		Continue

IV – Spécificité du tableau élémentaire

Le tableau élémentaire

- ❑ Le tableau élémentaire possède deux caractéristiques
 - ➔ c'est un tableau **brut**, de collecte de l'information
 - ➔ à chaque élément (unité spatiale) correspond une modalité de la variable
- ❑ En fait c'est une **application mathématique** : à chaque élément de l'ensemble E correspond une modalité du caractère j



E = ensemble observé
Des unités géographiques

X = ensemble des modalités
De la variable

Le tableau de dénombrement

❑ Le tableau de dénombrement au contraire :

- ➔ est un tableau **élaboré** de construction de l'information : construit à partir du tableau élémentaire.
- ➔ à chaque modalité de la variable correspond un certain nombre d'unités statistiques appelées effectif* de la modalité

❑ En général,

- ➔ il y a autant de tableaux de dénombrement qu'il y a de variables
- ➔ les unités statistiques deviennent anonymes regroupées dans leur modalité

Caractère ou variable (regroupé en classe)	Effectif des classes
100 000-200 000 (nom des modalités)	Nombre d'éléments

**Exercice : à partir du document 2
reconstruire un tableau élémentaire**