

Chapitre II : Statistique descriptive (Tableaux de données écologique)

Cette partie a pour but la description des données recueillies lors de l'échantillonnage (sous forme de données écologique) ou de l'expérimentation.

Les **statistiques descriptives** décrivent les données, résumer les données, les analysent et suggèrent des relations. Les statisticiens proposent des méthodes avancées d'analyses des données. Comme classer et subdiviser les populations en groupes homogènes (partitionnement, CAH). Ou encore résumer les caractéristiques d'une population en composantes synthétiques potentiellement pertinentes (méthodes d'analyse factorielle : ACP, AFCM...).

L'objectif des statistique descriptive c'est d'analyser et à présenter sous forme synthétique les caractéristiques des données écologique.

Un tableau de distribution de fréquences est un mode synthétique de présentation des données numériques, montrant comment les résultats enregistrés sur une variable se distribuent dans ses différentes classes. Pour préparer ce tableau, il faut déterminer des classes et dénombrer le nombre d'éléments appartenant à chacune d'elles.

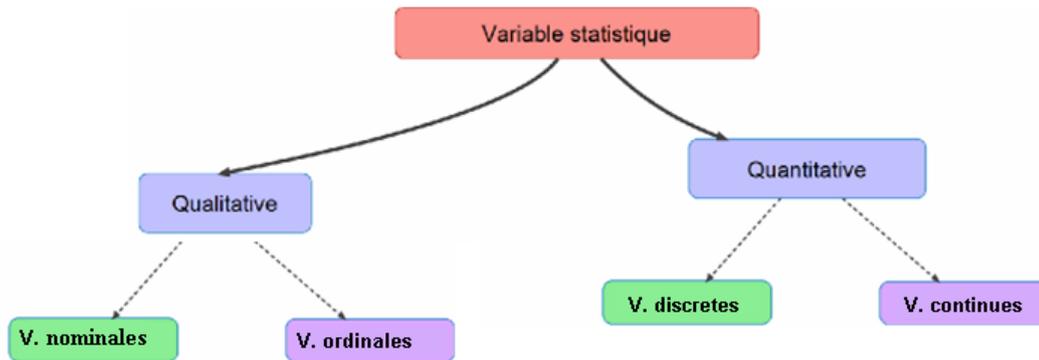
1. Variable statistique

La statistique « descriptive », comme son nom l'indique cherche à décrire une population donnée. Nous nous intéressons au caractéristique des unités qui peuvent prendre différentes valeurs.

nous obtenons différentes données pour chaque individu. Mais tout dépend de ce que nous étudions. Les données associées à un individu sont appelées **variables** ou caractères statistiques. Il s'agit des séries de données d'un tableur, par exemple. Nous distinguons ainsi deux grands types de variables : qualitatives et quantitatives.

Les **variables qualitatives** sont des termes, noms, identités, descriptions non-chiffrées... Elles peuvent être **nominales** (indications) ou **ordinales** (notion de hiérarchisation).

Les **variables quantitatives** sont des chiffres. Elles peuvent être **discrètes** (l'ensemble des valeurs possibles est fini) ou **continues** (l'ensemble des valeurs possibles est infini).



2. Présentation de tableaux des données d'une série statistique

2.1. Variable quantitative

Dans le cas des variables quantitatives d'une série statistique. La plus petite valeur admise dans la classe s'appelle alors **la borne inférieure** de la classe et la plus grande valeur, **la borne supérieure**. La gamme des valeurs admissibles constitue **l'intervalle de la classe**. La valeur centrale de la classe représente **l'indice de classe**.

Le nombre de classes à considérer peut être recherché à partir de l'une des formules suivantes :

Règle de Sturge

$$\text{Nombre de classes} = 1 + (3.3 \log n)$$

L'intervalle de chaque classe se calcule par la méthode suivante :

$$\text{Intervalle de classe} = \frac{\text{Valeur maximum} - \text{Valeur minimum}}{\text{Nombre de classes}}$$

Application : Dans le cadre d'une étude sur le rendement des exploitations agricoles de la wilaya de Mila, 20 vergers d'olivier répartis sur l'étendue de la région de mila sont considérés et les résultats en tonnes/hectare sont les suivants.

Verger	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Rendement	35	46	30	26	34	42	20	15	32	28	50	42	31	27	36	34	18	25	24	43

1. Transformez la série statistique en une distribution de fréquences.

2. Représentez graphiquement la distribution obtenue en utilisant les fréquences relatives.
3. Donnez la valeur du mode et calculez la médiane et la variance de l'échantillon.
4. Distribution de fréquences :

(L'utilisation de la calculatrice en mode stat est indispensable. Si non, il est impossible de faire vos calculs en un temps raisonnable).

Solution

+ Nombre de classes = $1 + (3.3 \log 20) = 1 + (3.3 * 1.30) = 5.29$

Nous arrondissons le résultat à **5 classes**.

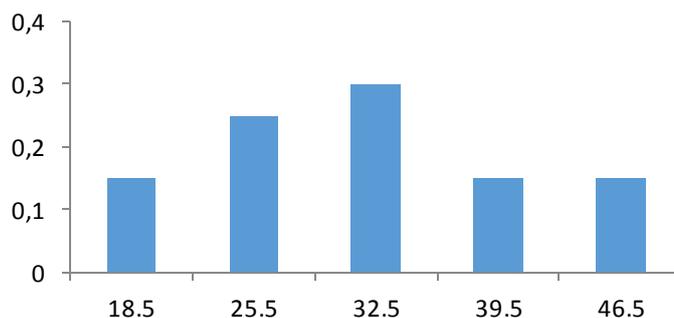
+ Intervalle de classe = $\frac{\text{val. Max.} - \text{val. Min.}}{\text{Nombre de classes}} = \frac{50 - 15}{5} = 7$

Nous aurons ainsi les classes suivantes : [15- 22[; [22- 29[; [29- 36[; [36- 43[; [43- 50].

+ Tableau de distribution des fréquences :

Classes	Indice de classe	Fréquences absolues	Fréquences relatives
[15- 22[18.5	3	0.15
[22- 29[25.5	5	0.25
[29- 36[32.5	6	0.30
[36- 43[39.5	3	0.15
[43- 50]	46.5	3	0.15

5. Représentation graphique :



Mode = **32.5** (centre de la classe modale).

Médiane (Me) = $(\text{val } 10 + \text{val } 11)/2 = (31 + 32)/2 = \mathbf{31.5}$

Variance de l'échantillon = $\delta^2 = \mathbf{82.5}$

2.2 Variables qualitatives

Généralement Le caractère étudié n'est pas mesurable il fait l'objet d'un classement. Donc chaque catégorie correspond à une classe. Le nombre d'éléments composant chacune d'elles s'appelle toujours l'effectif ou la fréquence absolue de la classe.

Application: On a pris un échantillon de 50 arbres prélevés de la forêt de zougba dans la

région de Mila.

CV= Chêne vert, CL= Chêne liège, PA= Pin d'alep , PP= Pin pignon, CA =chêne afares

**CV CL PA CV PA CV PA CL CV CA PP PA CL PA CV CA PP CL PA CV CL CV CV
 CL CV CV PP CV CL PA CV PA CA CV PA CL CVCA PP PA CV CA CV PP PA
 CV PA CV**

Alors ici la variable est X= Arbre, qui est une variable qualitative nominale. Pour présenter ces données sous forme de tableau, on dresse un tableau, dans la première colonne on énumère les cinq modalités de la variable, dans la seconde colonne on donne la fréquence absolue ou l'effectif de chacune des modalités (c'est-à-dire le nombre de fois que cette modalité se répète dans l'échantillon) et dans la troisième colonne, on donne la fréquence relative de chacune des modalités. La fréquence relative d'une modalité étant égale à sa fréquence absolue divisée par la taille de l'échantillon. Ce qui donne :

x	F absolu	F relative	F cumulé
CV	19	0.38	19
CL	8	0.16	27
PA	13	0.26	40
PP	5	0.1	45
CA	5	0.1	50
Total	N=50	1	50

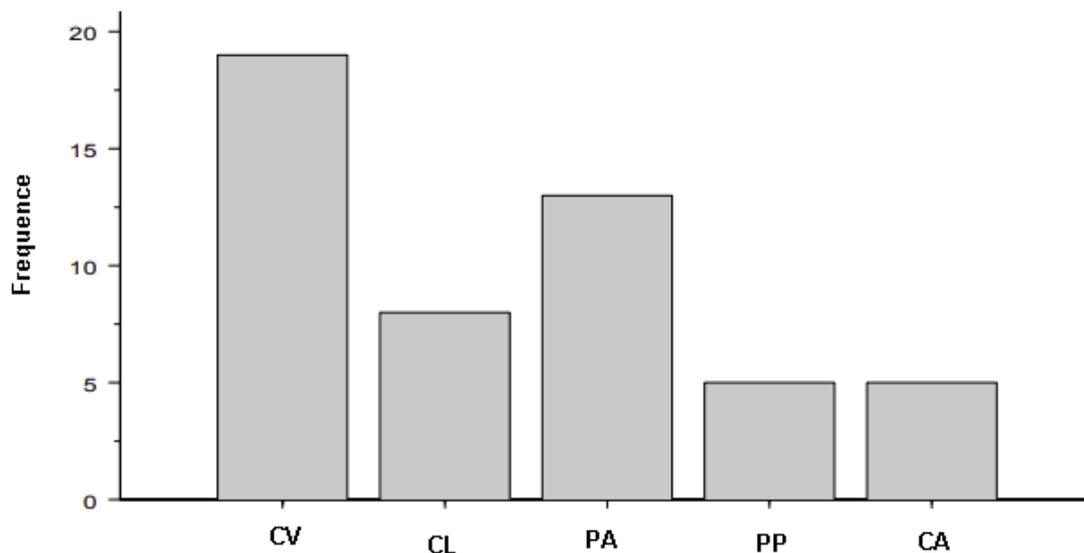


Diagramme en baton de la Distrubution des classe d'arbre