

Statistique descriptive

But : Après obtention d'une base de données, on donne une description pratique :

- Par l'intermédiaire de représentation graphique
- Par l'intermédiaire d'indicateur statistique

Une base de données se présente sous la forme d'un tableau à n lignes et p colonnes

n nombre d'individus,

p nombre de variables

I. Un peu de vocabulaire

Différentes notions :

- population \neq échantillon
- population : l'ensemble total sur lequel porte l'étude statistique
- échantillon : partie de la population réellement questionné pour l'étude

Ex : étude statistique

On s'intéresse à la couleur des yeux des habitants de la région PACA. 400 habitants de cette région ont répondu à l'enquête :

- Population : tous les habitants de la région PACA
- Echantillon : les 400 personnes de la région ayant répondu

Pb : en théorie, l'échantillon doit être représentatif de l'ensemble de la population

➔ Tirage aléatoire ou au moins de hasard

- Variable ou caractère : c'est le sujet même de l'étude
Ex précédent : variable = couleur des yeux
- Individu : c'est une unité élémentaire de la population



Un individu, du sens statistique, n'est pas forcément une personne physique

II. Nature d'une variable

Les variables peuvent être de 4 natures ≠

Qualitatives		Quantitatives	
On attend quelque chose de non usuel ou de l'ordre du gagnant		On obtient un résultat chiffré	
catégorielle -> pas de hiérarchie	Ordinale -> une hiérarchie sur la réponse	Discrète -> les réponses possibles sont en nbr fini ou espacées	Continue -> pas d'énumération possible de toutes les valeurs possibles

Exemple :

- Catégorielle : - couleur des yeux
- catégorie socioprofessionnelle
- Ordinale : - mention au bac
- jugement pour la qualité d'un film
- Discrète : - nbr d'enfants par famille
- nbr de plaintes pour vol en une journée à Nice
- Continue : - taille
- poids

III. Représentations graphiques

Pour une variable catégorielle -> diagramme circulaire

Pour une variable ordinale -> diagramme en bâtons

Diagramme circulaire :

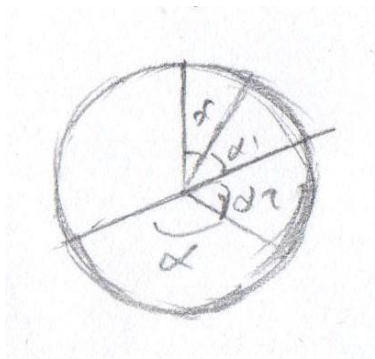
Tableau de représentation

Modalités	Effectifs	Angles
a_1	n_1	α_1
a_2	n_2	α_2
a_k	n_k	α_k

n_1 le nombre de fois où dans l'étude on a obtenu la modalité a_1 comme valeur

n effectif total

Remarque :



$$n = \sum_{\alpha=1}^k n_i$$

$$\alpha_1 = (n_1 / n) * 360 \text{ (degrés)}$$

Ex : bleu, vert, bleu, marron, bleu, noir, marron, marron, gris, marron

Modalités	Effectif	Angles
Bleu	3	$3/10 * 360 = 108$
Vert	1	$360/10 = 36$
Marron	4	$4/10 * 360 = 144$
Noir	1	$360/10 = 36$
Gris	1	$360/10 = 36$

Diagramme en bâtons

Tableau de représentation

Modalités	Effectifs	Fréquences
a_1	n_1	f_1
a_k	n_k	f_k

f_1 la fréquence d'apparition de la modalité a_1

$$f_i = n_i / n$$

Remarque : $\sum_i f_i = 1$

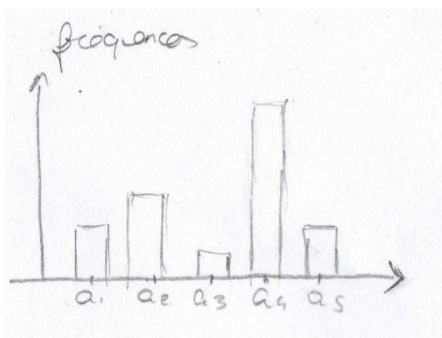
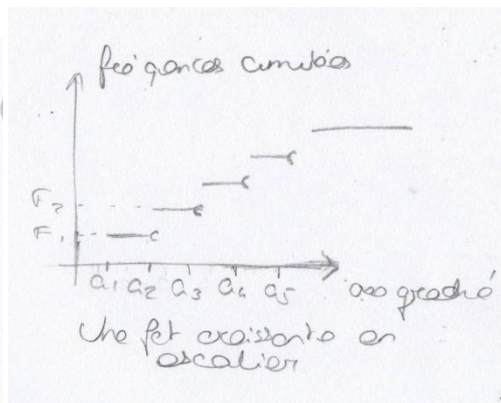
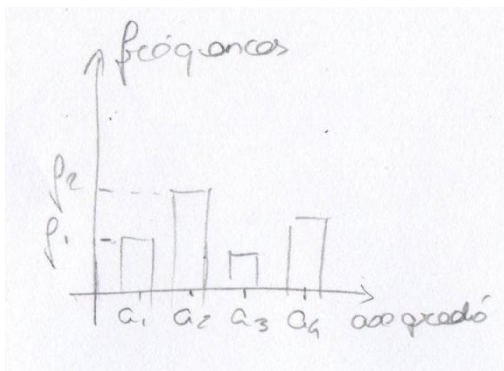


Diagramme en bâtons :

Tableau de représentation

Valeurs	Effectifs	Fréquences	Fréquences cumulées (croissante)
a_1	n_1	f_1	$F_1 = f_1$
a_2	n_2	f_2	$F_2 = f_1 + f_2$
a_k	n_k	f_k	$F_k = 1$



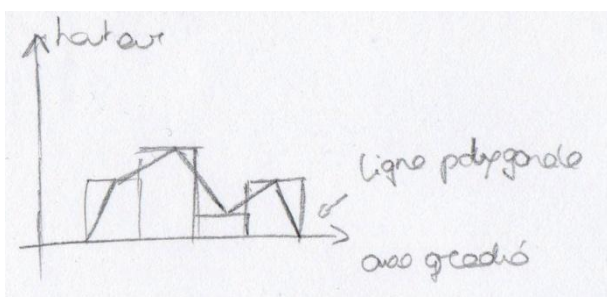
Histogramme :

Tableau de représentation

Classes	Effectifs	Fréquences	Fréquences cumulées	Hauteurs
$[a_1, a_2[$	n_1	f_1	$F_1 = f_1$	h_1
$[a_2, a_3[$	n_2	f_2	$F_2 = f_1 + f_2$	h_2
$[a_k, a_{k+1}[$	n_k	n_k	$F_k = 1$	h_k

$h_i = f_i / (a_{i+1} - a_i) =$ fréquences associés à la ième classe / amplitude de la ième classe

Histogramme :

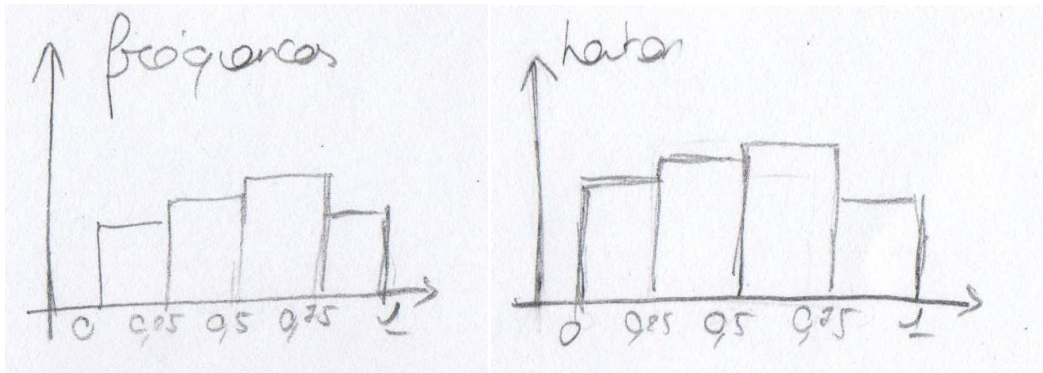


Remarque : Aire de l'histogramme

$$\sum_{i=1}^k (a_{i+1} - a_i) * h_i = \sum_{i=1}^k (a_{i+1} - a_i) * \frac{f_i}{(a_{i+1} - a_i)} = \sum_{i=1}^k f_i = 1$$

Ex : on suppose avoir 1000 observations qui sont ceux des réalisations d'une loi uniforme sur [0,1[

Classes	Effectifs	Fréquences	Hauteurs
[0, 0.25[240	0.24	0.96
[0.26, 0.50[250	0.25	1
[0.51, 0.75[264	0.262	1.048
[0.76, 1[248	0.248	0.992



Classes	Effectifs	Fréquences	Hauteurs
[0, 0.25[240	0.24	0.96
[0.25, 0.75[512	0.512	1.024
[0.75, 1[248	0.248	0.992

