

Les représentations numériques des distributions

Licence : 2004/2005
MLH

INTRODUCTION

REPRÉSENTATION VISUELLE / NUMÉRIQUE :

- : **RÉSUMÉ** LES RÉSULTATS
 - GRAPHIQUEMENT
 - AVEC DES INDICES (NOMBRES)

DEUX GROUPES D'INDICES

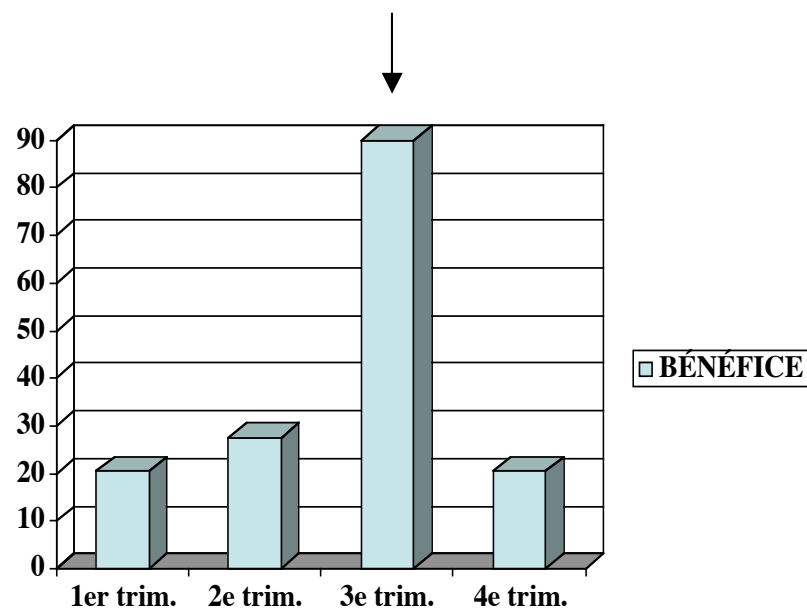
- DE CENTRALITÉ (DE POSITION, DE TENDANCE CENTRALE)
- DE DISPERSION

LES INDICES DE CENTRALITÉ

1) LE MODE

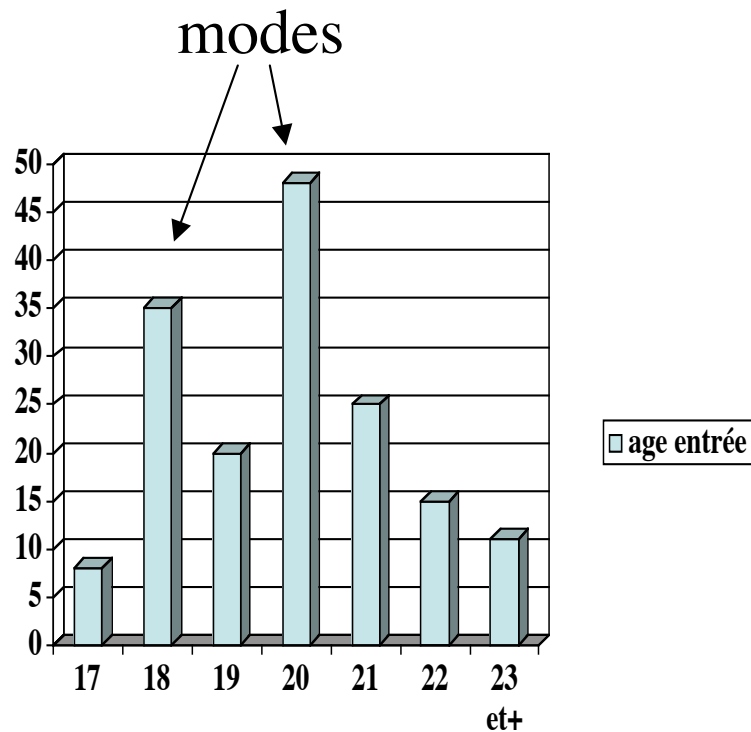
1 - 1 : DÉFINITION

1 - 2 : DÉTERMINATION

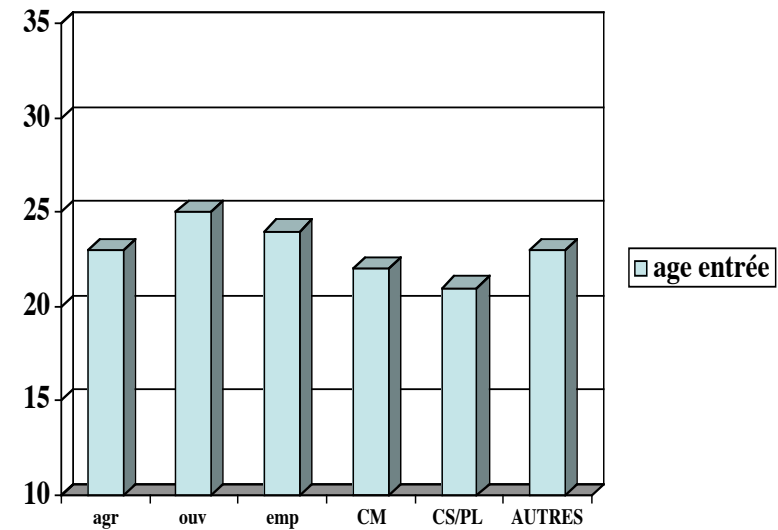


LES INDICES DE CENTRALITÉ

1 - 3 : distributions particulières



Bi-modale

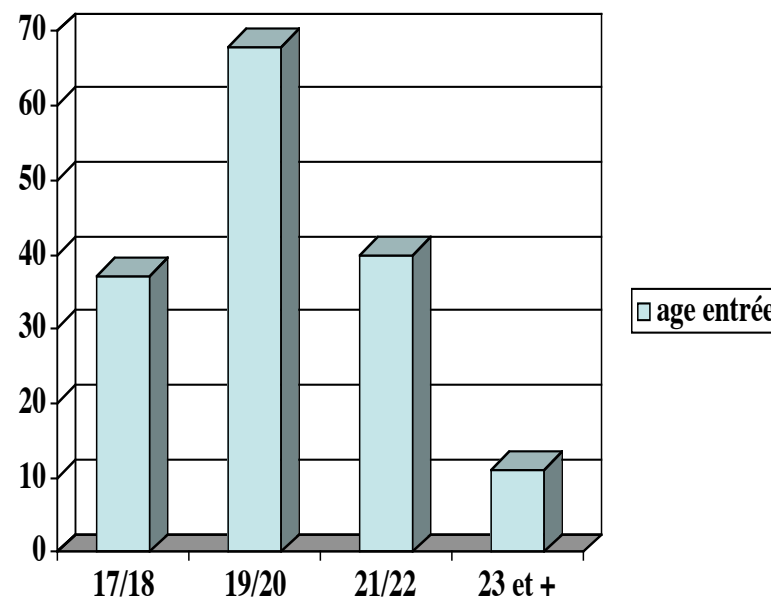
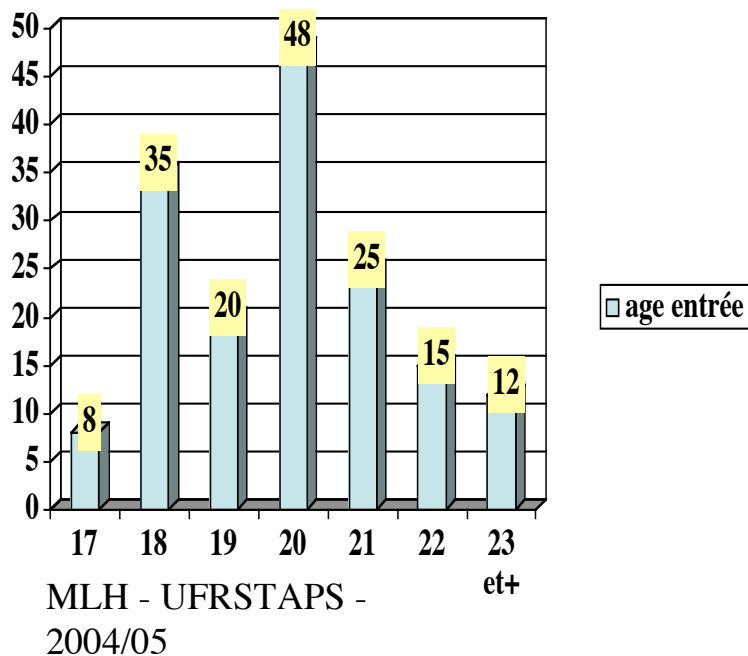


amodale

LES INDICES DE CENTRALITÉ

1 - 3 : remarques

- mode : - **seul** indice de centralité des **échelles nominales**
- Dépend du mode de regroupement effectué



LES INDICES DE CENTRALITÉ

2) La médiane

2-1) Définition

C'est la classe à laquelle appartient l'individu médian

2-2) Détermination

Si n impair

- rechercher l'observation de rang $(n+1)/2$
- la médiane = la modalité à laquelle appartient cette observation

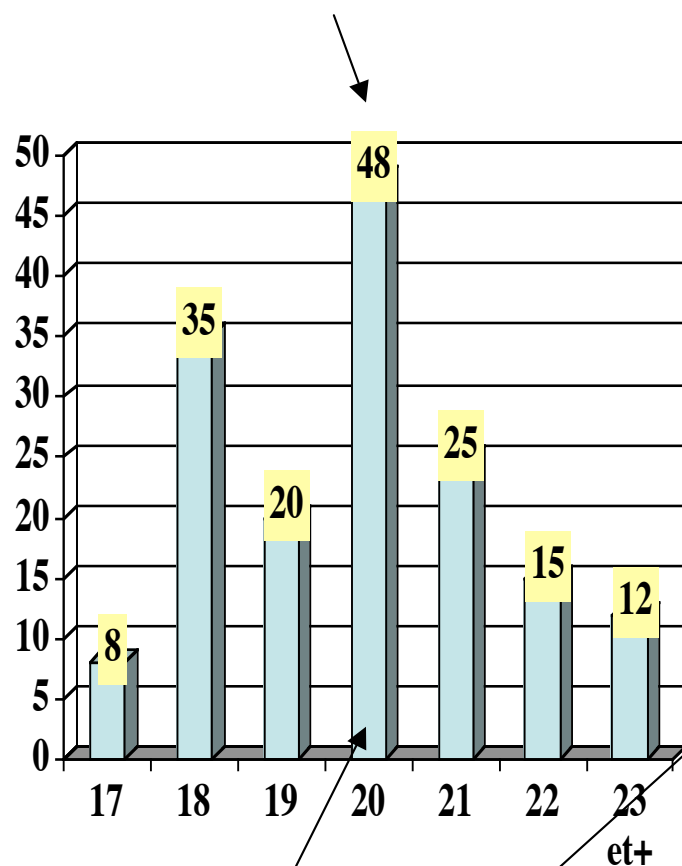
ex. $n = 163$ (nbre étudiants D1)

$$(n+1)/2 = 82$$

médiane : modalité à laquelle appartient le 82ème étudiant. Soit modalité « 20 ans »

LES INDICES DE CENTRALITÉ

Médiane : 20 ans



âge	17	18	19	20	21	22	+23	Tot
eff	8	35	20	48	25	15	12	163
Eff Cum	8	43	63	111	136	151	163	

age entrée

82ème sujet
MLH - UFRSTAPS
2004/05

LES INDICES DE CENTRALITÉ

Si n pair

- l'observation de rang $(n+1)/2$ n'est pas un nombre entier

- Rechercher les 2 individus qui entourent $(n+1)/2$

- la médiane = la modalité à laquelle appartiennent ces 2 observations

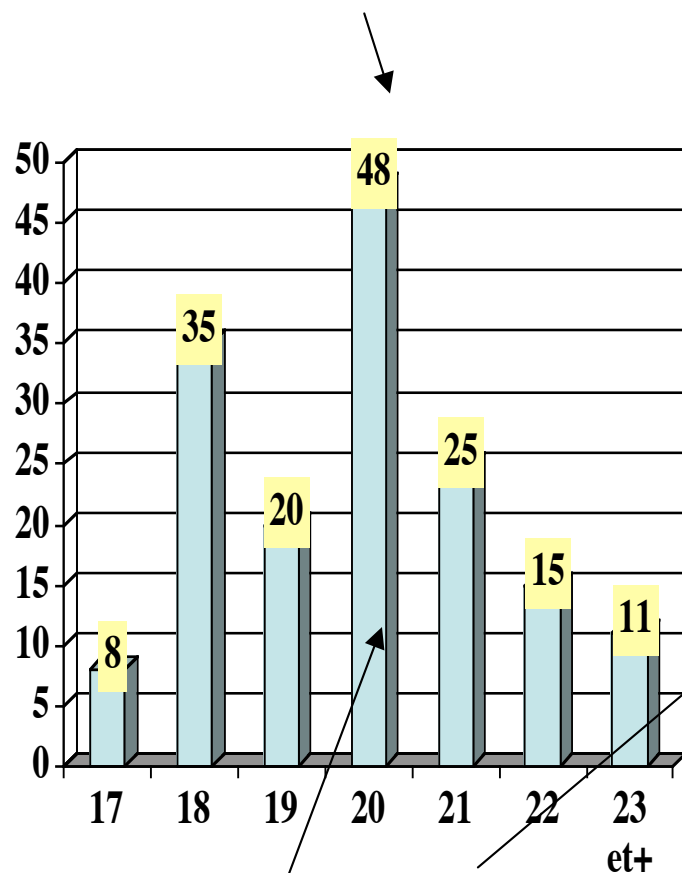
ex. $n = 162$ (nbre étudiants D1)

$$(n+1)/2 = 81,5$$

médiane : modalité à laquelle appartiennent les 81ème et 82ème étudiants. Soit modalité « 20 ans »

LES INDICES DE CENTRALITÉ

Médiane : 20ans



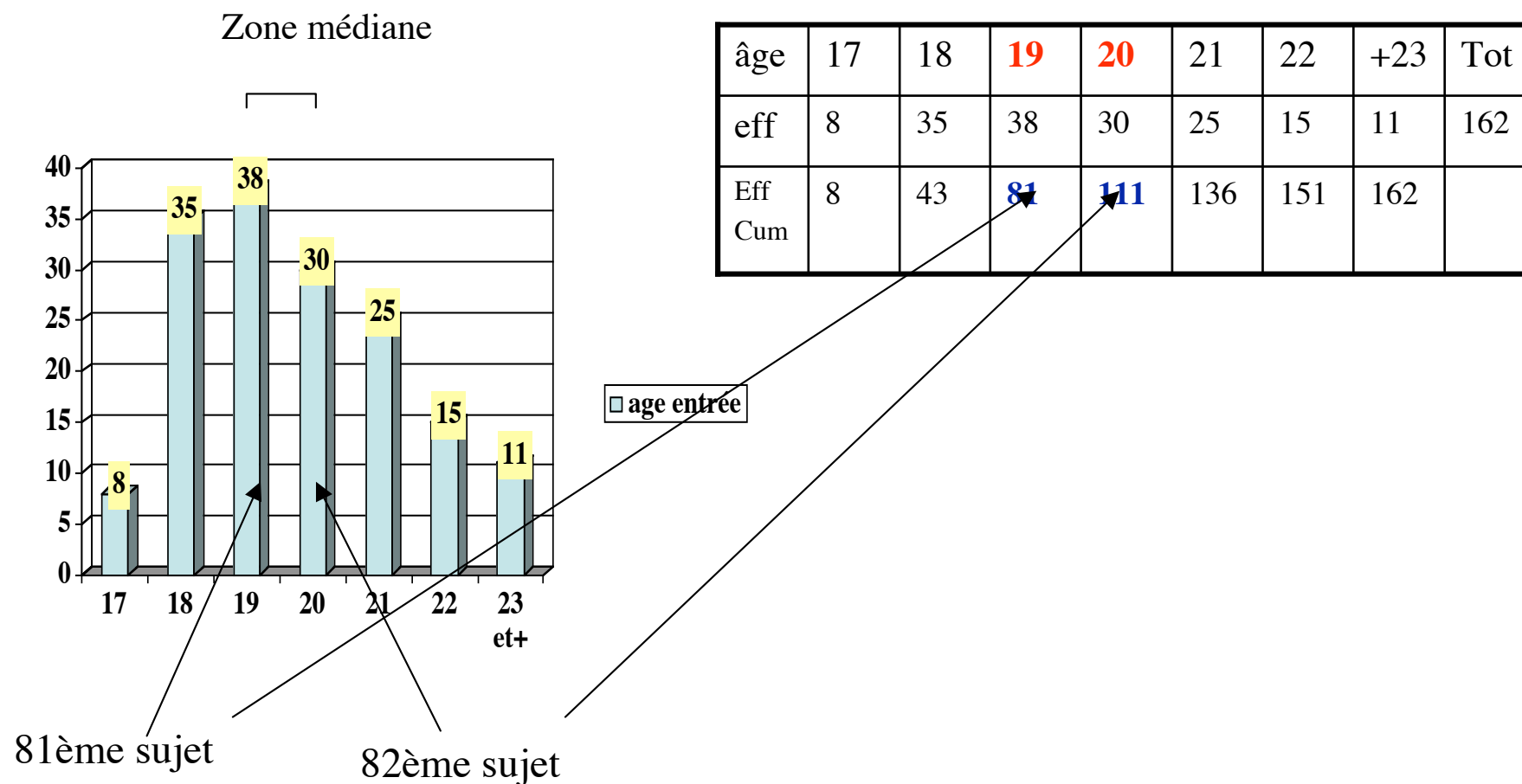
âge	17	18	19	20	21	22	+23	Tot
eff	8	35	20	48	25	15	11	162
Eff Cum	8	43	63	111	136	151	162	

age entrée

81ème et 82ème sujets

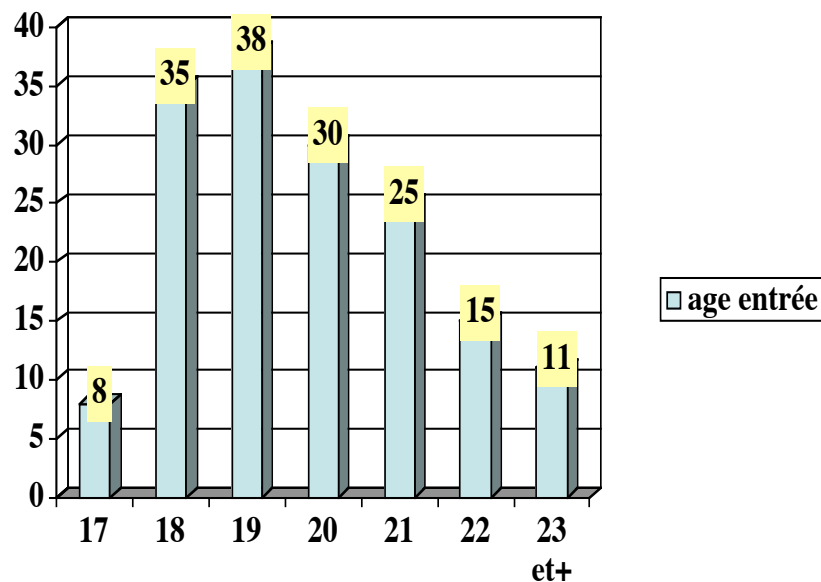
LES INDICES DE CENTRALITÉ

Cas particulier : les 2 observations sont dans 2 catégories différentes



LES INDICES DE CENTRALITÉ

2-3) Propriétés de la médiane



-Indépendante des valeurs extrêmes

Ex. salaire

- Donne les renseignements suivants :

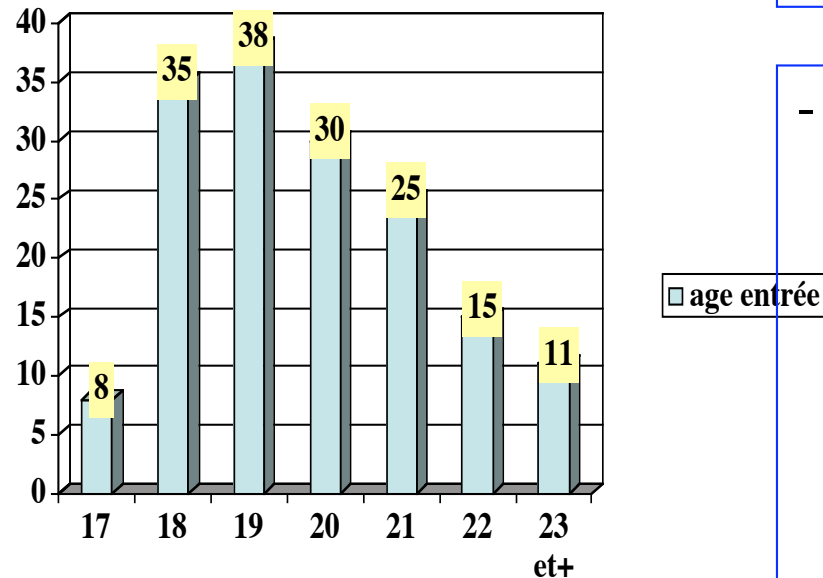
-si un individu appartient à la classe médiane, il a autant d'individus devant lui que derrière lui

-Si un individu appartient à une classe inférieure à la médiane il a plus d'individus devant lui que derrière lui

-Indice utilisé avec une **échelle ordinale au moins** (voir diapo 5)

LES INDICES DE CENTRALITÉ

3) La moyenne



- **L'indice de position des échelles numériques**

- À total égal, c'est la valeur que prendrait toutes les mesures si elles étaient égales

- Mode de calcul :

$$M = \sum x_i / n$$

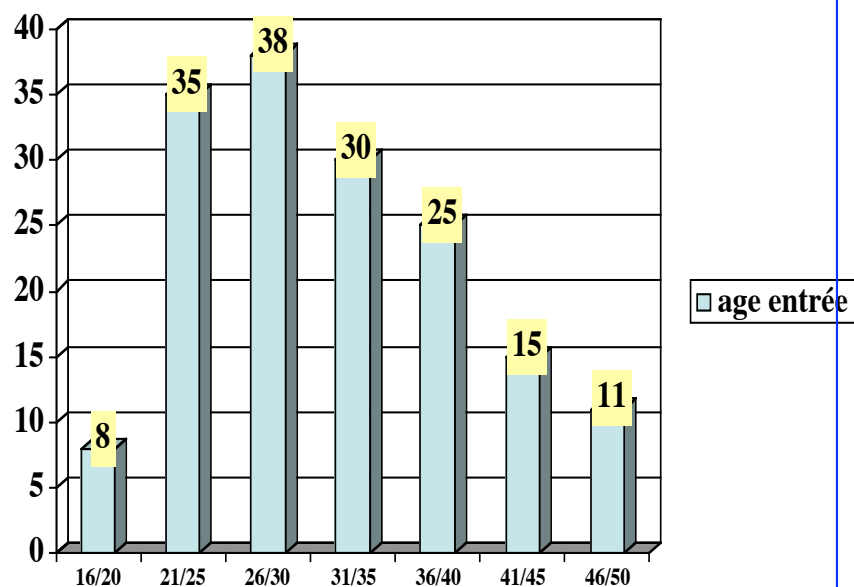
Ici :

$$M = [(8 \cdot 17) + (35 \cdot 18) + (38 \cdot 19) + (30 \cdot 20) + (25 \cdot 21) + (15 \cdot 22) + (11 \cdot 23)]$$

$$M = 19,72$$

LES INDICES DE CENTRALITÉ

3) Cas particulier : si intervalle de classe : 16/20 ans, 21/25ans, etc.



- Mode de calcul :

On prend le centre de chaque classe et on considère que tous les individus de la classe ont cette valeur

$$M = [(8 \cdot 17) + (35 \cdot 23) + (38 \cdot 28) + (30 \cdot 33) + (25 \cdot 38) + (15 \cdot 43) + (11 \cdot 48)]$$

$$M = 30,88$$

Mais ...approximation

LES INDICES DE DISPERSION

RAPPEL : A QUOI CA SERT ?

-1) L'intervalle de variation

- Définition. C'est l'écart entre la plus petite et la plus grande modalités observées

$$- IV = \text{étendue} = x_{\max} - x_{\min}$$

- Exemple : soit les notes de 8 élèves :

-1, 2, 2, 5, 7, 7, 10, 12

$$- IV = 12 - 1 = 11$$

LES INDICES DE DISPERSION

2) l'indice de diversité

- Définition. Indique la **probabilité** que 2 observations tirées au hasard dans la distribution appartiennent à 2 modalités différentes

-NB : C'est l'indice de dispersion utilisé pour les **échelles nominales**

$$- D = 1 - \sum n_i^2 / N^2$$

Ou n_i = effectif de chaque modalité

N = effectif total

- Exemple : la distribution des âges de 220 personnes

modalité	20/24	25/29	30/34	35/39	40/44	45/49	Total
effectifs	n1 = 20	n2 = 40	n3 = 60	n4 = 50	n5 = 30	n6 = 20	N = 220

LES INDICES DE DISPERSION

$$\begin{aligned} - D &= 1 - \sum (n_i/N)^2 \\ &= 1 - [(20^2 + 40^2 + 60^2 + 50^2 + 30^2 + 20^2) / 220^2] \\ &= 0,81 \end{aligned}$$

- **Conclusion** : il y a une probabilité de .81 (81 chances sur 100) que 2 individus tirés au hasard dans l'échantillon appartiennent à 2 catégories différentes

- Exemple : la distribution des âges de 220 personnes

modalité	20/24	25/29	30/34	35/39	40/44	45/49	Total
effectifs	n1 = 20	n2 = 40	n3 = 60	n4 = 50	n5 = 30	n6 = 20	N = 220

LES INDICES DE DISPERSION

- Questions :

-- Dans quel cas a-t-on $D = 0$

- Exemple : la distribution des âges de 220 personnes

modalité	20/24	25/29	30/34	35/39	40/44	45/49	Total
effectifs	$n1 = 0$	$n2 = 0$	$n3 = 220$	$n4 = 0$	$n5 = 0$	$n6 = 0$	$N = 220$

Dans quel cas a-t-on D maximum (quelle valeur ?)

- Exemple : la distribution des âges de 220 personnes

modalité	20/24	25/29	30/34	35/39	40/44	45/49	Total
effectifs	$n1 = 36$	$n2 = 36$	$n3 = 37$	$n4 = 37$	$n5 = 37$	$n6 = 37$	$N = 220$

LES INDICES DE DISPERSION

3) Les quantiles

- Définition.

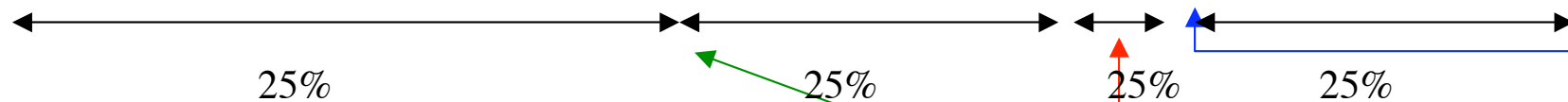
- La médiane découpe la **distribution (ordinaire)** en 2 parties égales.
- Les quantiles sont les n modalités qui découpent la distribution en n intervalles égaux (intervalles interquantiles)
 - Les **quartiles** sont les 3 modalités qui découpent la distribution en **4 intervalles égaux**
 - Les **déciles** sont les 9 modalités qui découpent la distribution en **10 intervalles égaux**

- Calcul :

- médiane : on recherche le sujet $(n+1)/2$: la médiane = la modalité à laquelle il appartient
- le **1er quartile (Q1)** : on recherche le sujet $1*(n+1)/4$. Le 1er quartile est la modalité à laquelle il appartient
- Le **2ème quartile (Q2)** : on recherche le sujet $2*(n+1)/4$ $(n+1)/2$médiane
- Le **3ème quartile (Q3)** : on recherche le sujet $3*(n+1)/4$

LES INDICES DE DISPERSION

notes	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
effectifs	2	5	2	2	0	0	1	10	0	1	9	4	6	4	1
Effectifs cumulés	2	7	9	11	11	11	12	22	22	23	32	36	42	46	47



$$Q1 : 1 \cdot (n+1) / 2 = 1 \cdot (47+1) / 2 = 24 \rightarrow Q1 : \text{moda } 6$$

$$Q2 : 2 \cdot (n+1) / 4 = 2 \cdot (47+1) / 4 = 24 \rightarrow Q2 : \text{moda } 10$$

$$Q3 : 3 \cdot (n+1) / 4 = 3 \cdot (47+1) / 4 = 36 \rightarrow Q3 : \text{moda } 11$$

LES INDICES DE DISPERSION

4) L'ÉCART MOYEN - LA VARIANCE - L'ÉCART TYPE

- Définition. Ce sont des indices de dispersion pour les échelles numériques

4 - 1) **L'écart moyen** : C'est la moyenne des écarts de chaque valeur par rapport à la moyenne, en valeur absolue

$$\text{- EM : } \sum |x_i - m| / N$$

- x_i = différentes valeur de x

- m = moyenne des valeurs de x

- N = nombre de valeurs

LES INDICES DE DISPERSION

4 - 2) L'écart type :

$$- \sigma = \sqrt{\sum (x_i - m)^2 / N}$$

- x_i = différentes valeur de x

- m = moyenne des valeurs de x

- N = nombre de valeurs

4 - 2) la variance :

$$- V = \sigma^2 = \sum (x_i - m)^2 / N$$

- x_i = différentes valeur de x

- m = moyenne des valeurs de x

- N = nombre de valeurs

LES INDICES DE DISPERSION

Notes :	Ecart à la moyenne :	Ecart à la moyenne en valeur absolue	Ecart à la moyenne élevé au carré
x_i	$x_i - m$	$ x_i - m $	$(x_i - m)^2$
1	$1 - 5,75 = -4,75$	4,75	22,56
2	$2 - 5,75 = -3,75$	3,75	14,06
2	$2 - 5,75 = -3,75$	3,75	14,06
5	$5 - 5,75 = -0,75$	0,75	0,56
7	$7 - 5,75 = 1,25$	1,25	1,56
7	$7 - 5,75 = 1,25$	1,25	1,56
10	$10 - 5,75 = 4,25$	4,25	18,06
12	$12 - 5,75 = 6,25$	6,25	39,06
	Somme des écart = 0	Somme des écart = 26 $\text{Ecart moyen} = 26/8 = 3,25$	Somme (des écart au carré) = 111,48 $\text{Variance} = 111,48/8 = 13,94$ $\text{Ecart type} = \sqrt{13,94} = 3,73$

LES INDICES DE DISPERSION

5) Le coefficient de variation

Si on veut éliminer l'unité de mesure et comparer des distributions avec des unités de mesure différente on utilise CV

Plus CV est faible, plus les données sont homogènes

Soit une distribution des poids de 100 étudiants

$$m_1 = 71 \text{ kg}$$

$$\sigma_1 = 6,5 \text{ kg}$$

$$CV = (\sigma_1 / m_1) * 100$$

$$CV = 8,73$$

Soit une distribution de la tailles de ces 100 étudiants

$$m_2 = 174 \text{ cm}$$

$$\sigma_2 = 8,5 \text{ cm}$$

$$CV = (\sigma_2 / m_2) * 100$$

$$CV = 4,88$$

Conclusion ???