

Statistique pour l'ingénieur

Cours n° 2

Statistique descriptive unidimensionnelle

Statistique descriptive

Tableaux de
données brutes



objective	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11
1	0	0	0	0	0	4854	0	0	0	0	0
1	1	0	0	0	0,94	5504	0	0	0,99	0	0
1	0,97	0	1	0	0	11178	0	0	0	0	0
1	0,97	0	0	0	0	2628	0	0	0	0	0
0	0	0	0,99	0	0	3712	0	0	0	0,97	0
1	0,96	0,99	1	0	0	5213	0	0	0	0,97	0,97
1	0	0	0	0	0	5483	0,05	0,41	0	0	0
0	0	0	0,99	0	0	7926	0	0	0	0,96	1
1	0,93	0,98	1	0	0	6743	0	0	0	0	0
1	1	0	0	0	0	6449	0	0	0	0	0
1	0	0	0	0	0,94	3547	0	0	0	0	0
0	0	0	0	0	0	2324	0	0	0	0,96	0
1	0	0	0	0	0	1924	0	0	0	0	0
1	0,99	0	1	0	0	6292	0,01	0,05	0	0	0,99
0	0	0	1	0	0	1353	0	0	0	0	0
1	0	0	1	0,97	0	11079	0	0	0	0	0
1	0	0	0	0	0	4810	0	0	0	0	0
0	0	0	0	0	0	1991	0	0	0	0	0
1	0,96	0,99	0	0	0	1592	0	0	0	0,98	0
1	0	0	0	0	0	3572	0	0	0,92	0	0

Présentation

Tableaux
statistiques

Visualiser

Représentations
graphiques

Résumer

Résumés
numériques

Les méthodes vont dépendre
du type de variable :

- Variable quantitative discrète
- Variable quantitative continue
- Variable qualitative

Variable quantitative discrète

- Exemple: age (arrondi) des étudiants d'ASI3

21	19	20	21	20	20	20	21
20	19	20	20	20	20	20	20
21	20	21	21	21	23	20	20
19	24	20	21	20	20	21	21
20	22	20	20	20	22	20	20
20	19	22	21	21			

Variable quantitative discrète

- Présentation des données:

- Tableau statistique:

pour chaque valeur x_i de la variable X , calcul de:

- n_i : nb d'occurrences de x_i
 - f_i : fréquence correspondante (%)
 - N_i : effectif cumulé
 - F_i : fréquence cumulée (%)

x_i	n_i	N_i	f_i	F_i
19	5	5	11.11	11.11
20	24	29	53.33	64.44
21	11	40	24.44	88.88
22	3	43	6.67	95.56
23	1	44	2.22	97.78
24	1	45	2.22	100.00

$$P_i = F_i / 100$$

$N=45$ $I=6$

Variable quantitative discrète

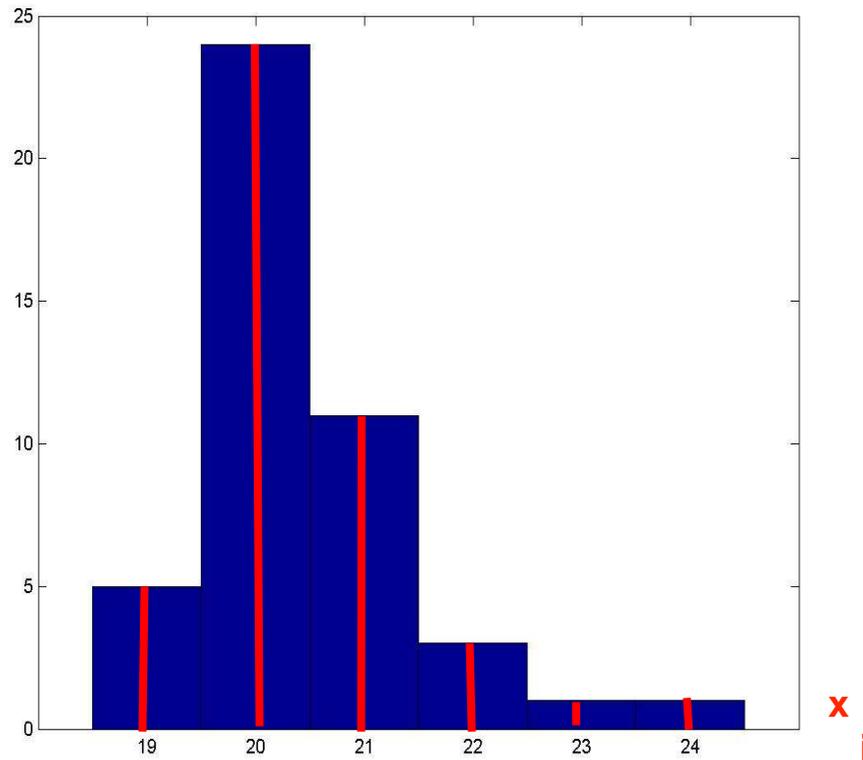
- Présentation des données:
 - Présentation tige et feuille (*stem and leaf*)

1	99999
2	000000000000000000000000000000
2	11111111111
2	222
2	3
2	4

Variable quantitative discrète

- Représentation graphique:
 - Diagramme en bâtons

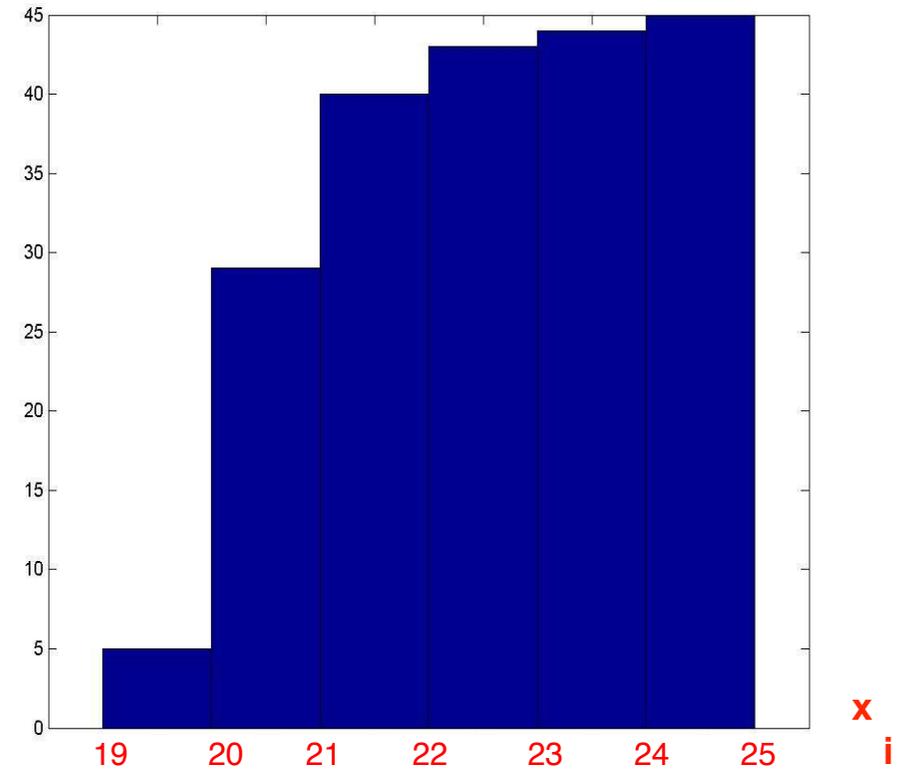
n ou f ou p
 i i i



A. ROGOZAN

- Diagramme cumulatif

N ou F ou P
 i i i



Si P => fonction de répartition empirique
 i

Variable quantitative discrète

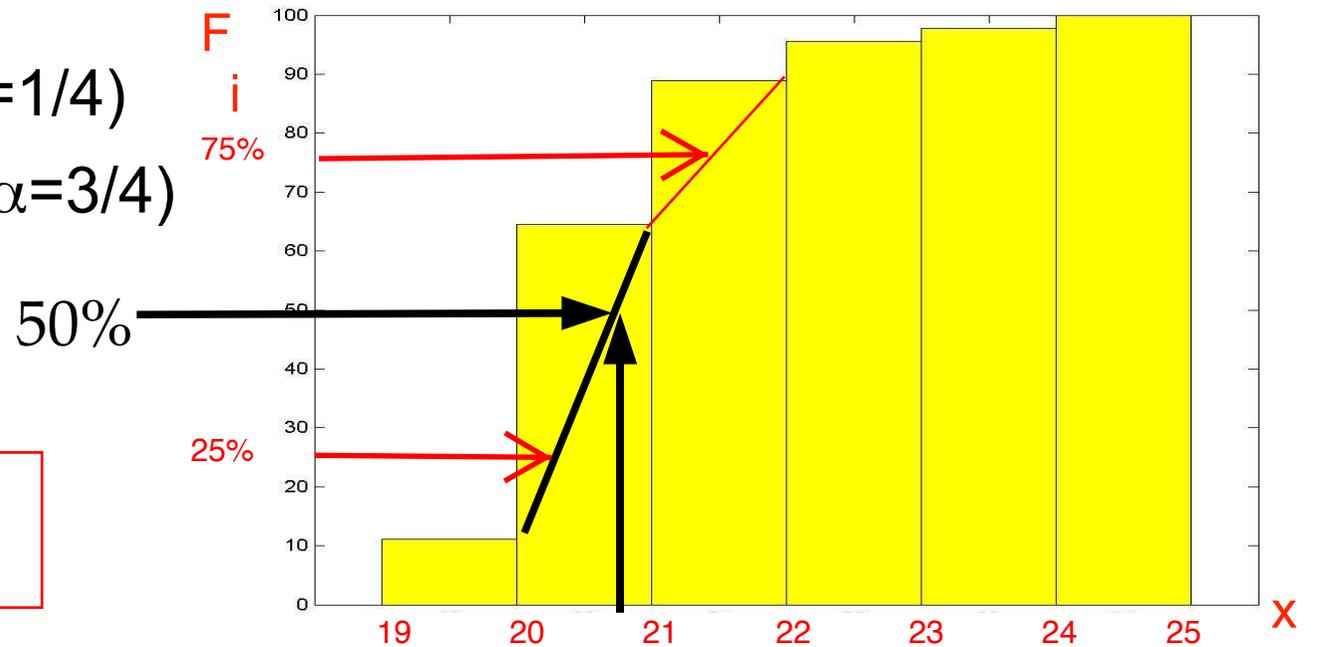
- Quantile d'ordre α

$$x_\alpha \text{ tel que } F(x_\alpha) = \alpha$$

- Quantiles courants

- Médiane ($\alpha=1/2$) : valeur qui sépare la série des observations en deux ensembles d'effectifs égaux

- Premier quartile Q_1 ($\alpha=1/4$)
- Troisième quartile Q_3 ($\alpha=3/4$)



Les quantiles ne sont pas nécessairement UNIQUES!

$$Q_1=20 \quad Me = 20 \quad Q_3=21$$

Variable quantitative discrète

- Résumés numériques : tendance centrale
 - Médiane :
(insensible aux variations des valeurs extrêmes)
 - Moyenne : $\bar{x} = \sum_{i=1}^I p_i x_i$
(sensible aux valeurs extrêmes \Rightarrow peu robuste)
 - Mode :
valeur la plus fréquente pour une distribution discrète

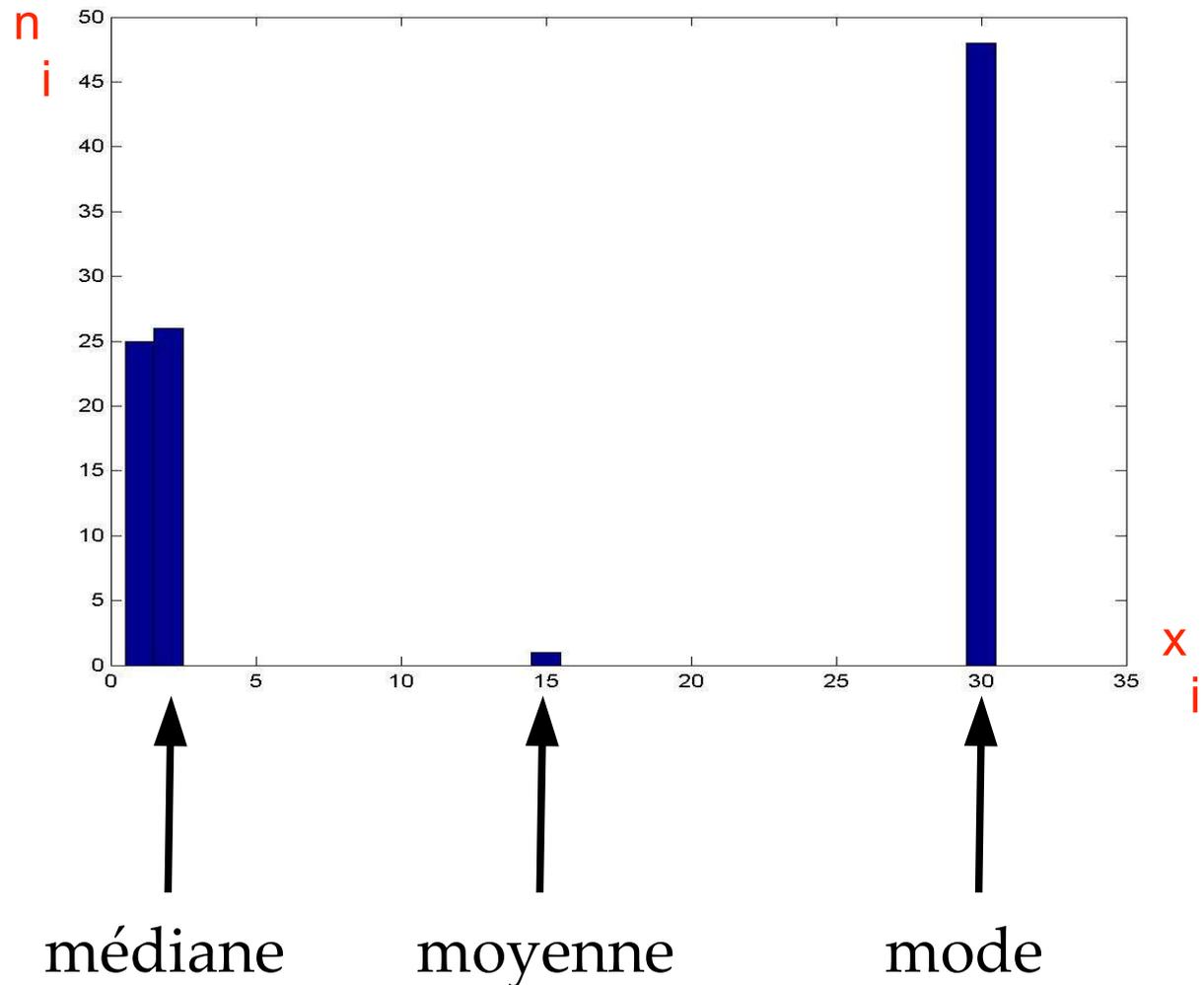
Dans notre exemple, médiane = mode = 20 \approx moyenne mais...

Variable quantitative discrète

- Résumés numériques : tendance centrale

x_i	n_i
1	25
2	26
15	1
30	48

$N_i=?$



Variable quantitative discrète

- Résumés numériques : dispersion

- Étendue ou intervalle de variation : $|x_{max} - x_{min}|$
(dépendant de valeurs extrêmes \Rightarrow instable)

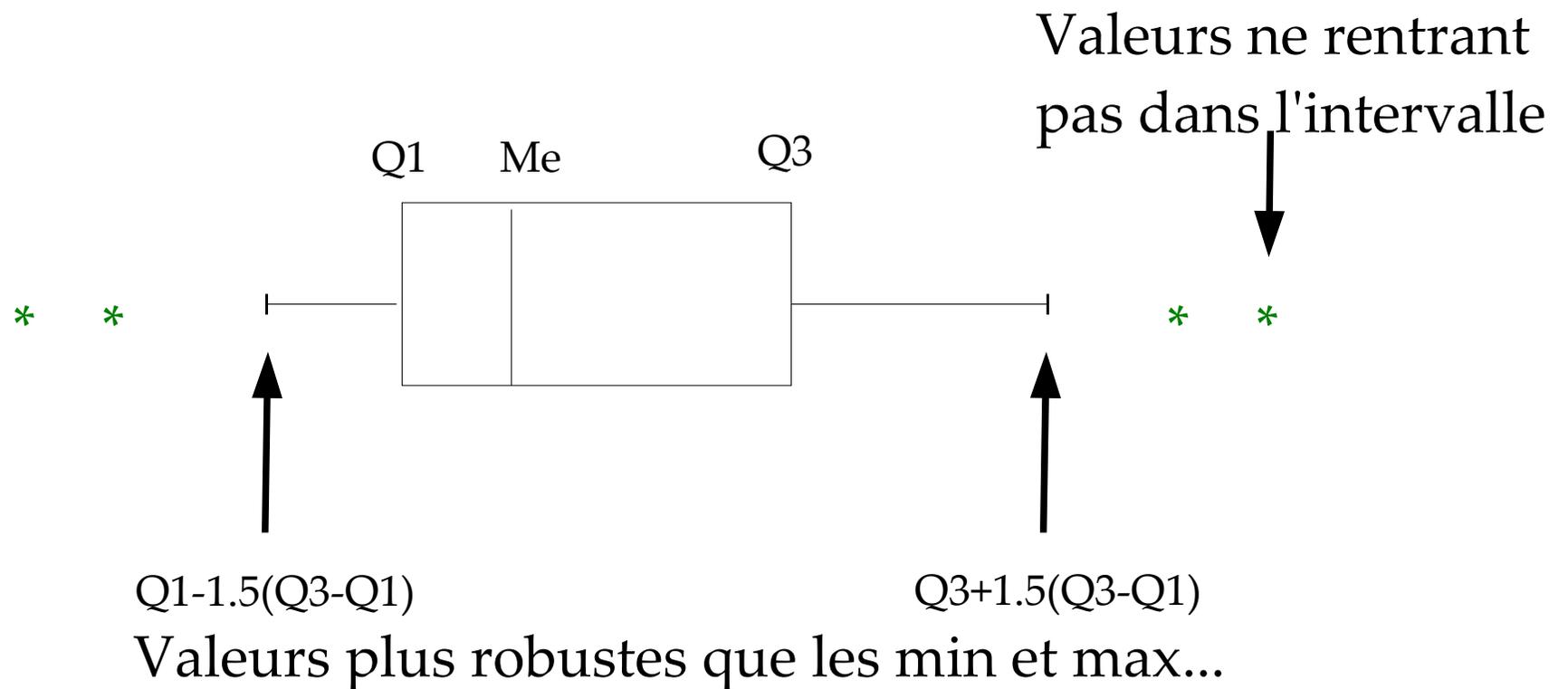
- Intervalle interquartile : $|Q_3 - Q_1|$
(plus robuste que l'étendue)

- Variance : $s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \sum_{i=1}^I p_i (x_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1} x_i^2 - (\bar{x})^2$

Écart-type : s

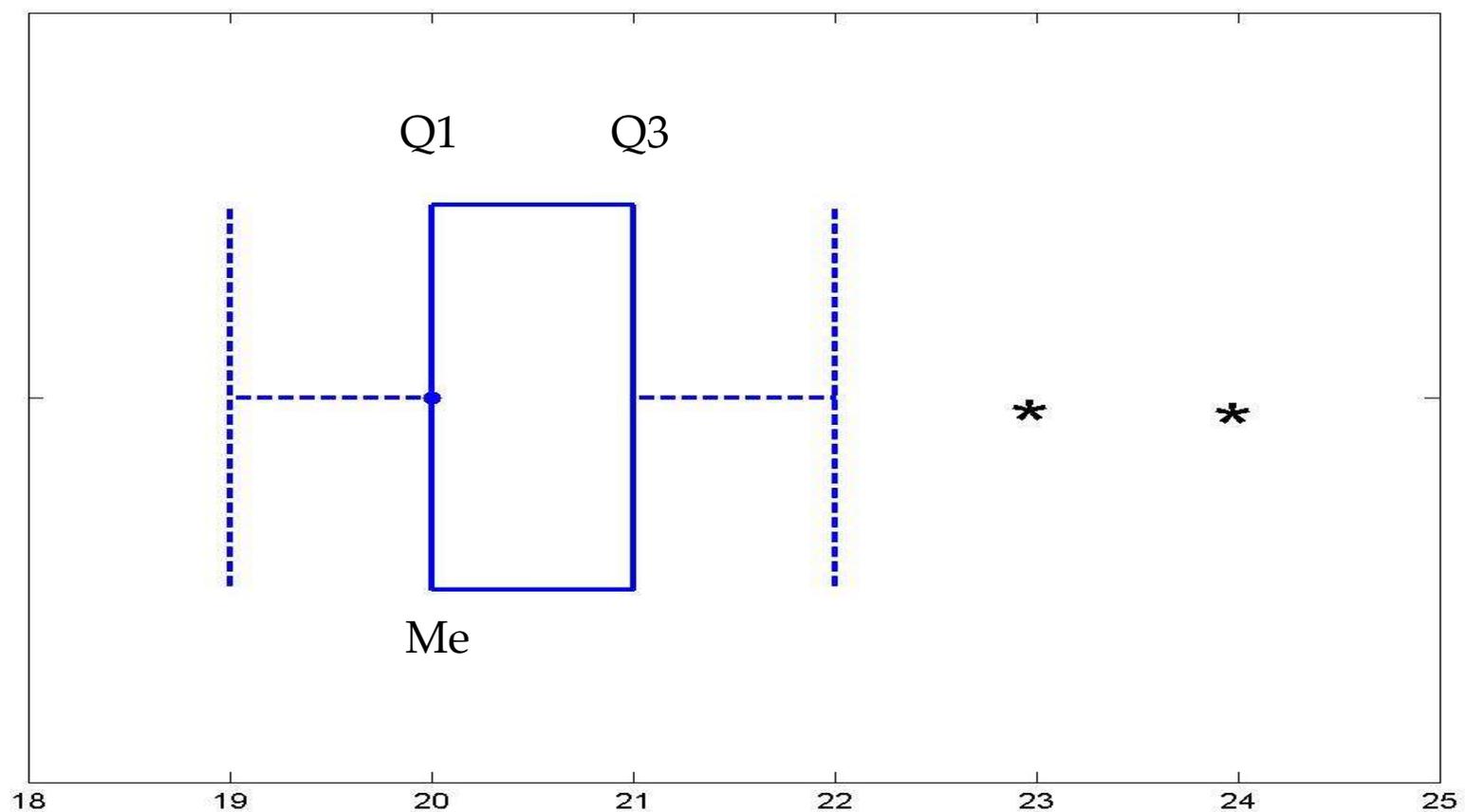
Variable quantitative discrète

- Récapitulatif : la boîte à moustaches (*box and whisker plot*)



Variable quantitative discrète

- Exemple:



Variable quantitative discrète

- **Résumés numériques : forme**
Pour vérifier le caractère normal d'une distribution
(en sachant que le coeff. d'aplatissement de la loi normale=3)

- Coeff. d'asymétrie : $\gamma_1 = \frac{m_3}{s^3}$ où $m_3 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3$

- Coeff. d'aplatissement : $\gamma_2 = \frac{m_4}{s^4}$ où $m_4 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4$

Variable quantitative continue

- Exemple : notes obtenues à un médian ...

12.25	9.0	12.5	0.0	9.5	9.25	15.0	11.75	12.75
14.0	11.0	12.75	11.75	12.5	12.5	13.0	10.63	14.0
10.0	12.7	11.0	13.0	11.5	12.5	14.25	16.08	10.25
12.75	11.0	13.0	10.75	0.0	13.75	11.63	7.75	12.5
12.0	10.5	11.25	10.0	13.0	11.0	12.5	0.0	10.5
17.0	12.63							

- Pour des variables quantitatives continues
regrouper les valeurs de la variable X en k classes d'extrémités e_0, \dots, e_k

Variable quantitative continue

- Présentation des données:
 - Tableau statistique pour chaque **intervalle** $[e_{i-1}, e_i[$

$[e_{i-1}, e_i[$	n_i	N_i	f_i	F_i
$[0,2[$	3	3	6.4	6.4
$[2,4[$	0	3	0.0	6.4
$[4,6[$	0	3	0.0	6.4
$[6,8[$	1	4	2.1	8.5
$[8,10[$	3	7	6.4	14.9
$[10,12[$	16	23	34.0	48.9
$[12,14[$	18	41	38.3	87.2
$[14,16[$	4	45	8.5	95.7
$[16,18[$	2	47	4.3	100.0
$[18,20[$	0	47	0.0	100.0

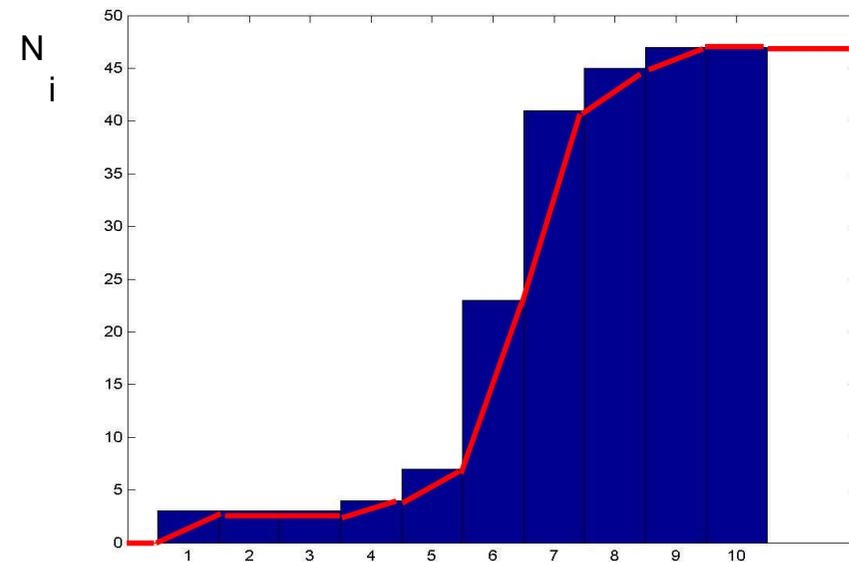
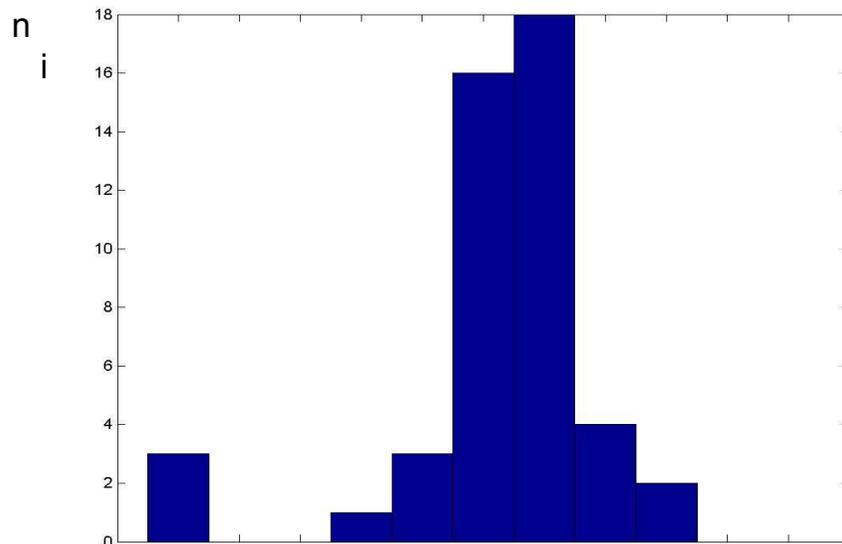
Bornes des classes e_i
Centres des classes : c_i
Amplitudes des classes : a_i
 $I = \text{nb des classes}$

Probabilités :
 $p_i = f_i/100$

Probabilités cumulées :
 $P_i = F_i/100$

Variable quantitative continue

- Représentation graphique:
 - Histogramme
 - Courbe cumulative
- Interpolation linéaire sur les $P_i \Rightarrow$ Fonction de répartition empirique $F(x)$



Attention !!! Représentation pour chaque classe de

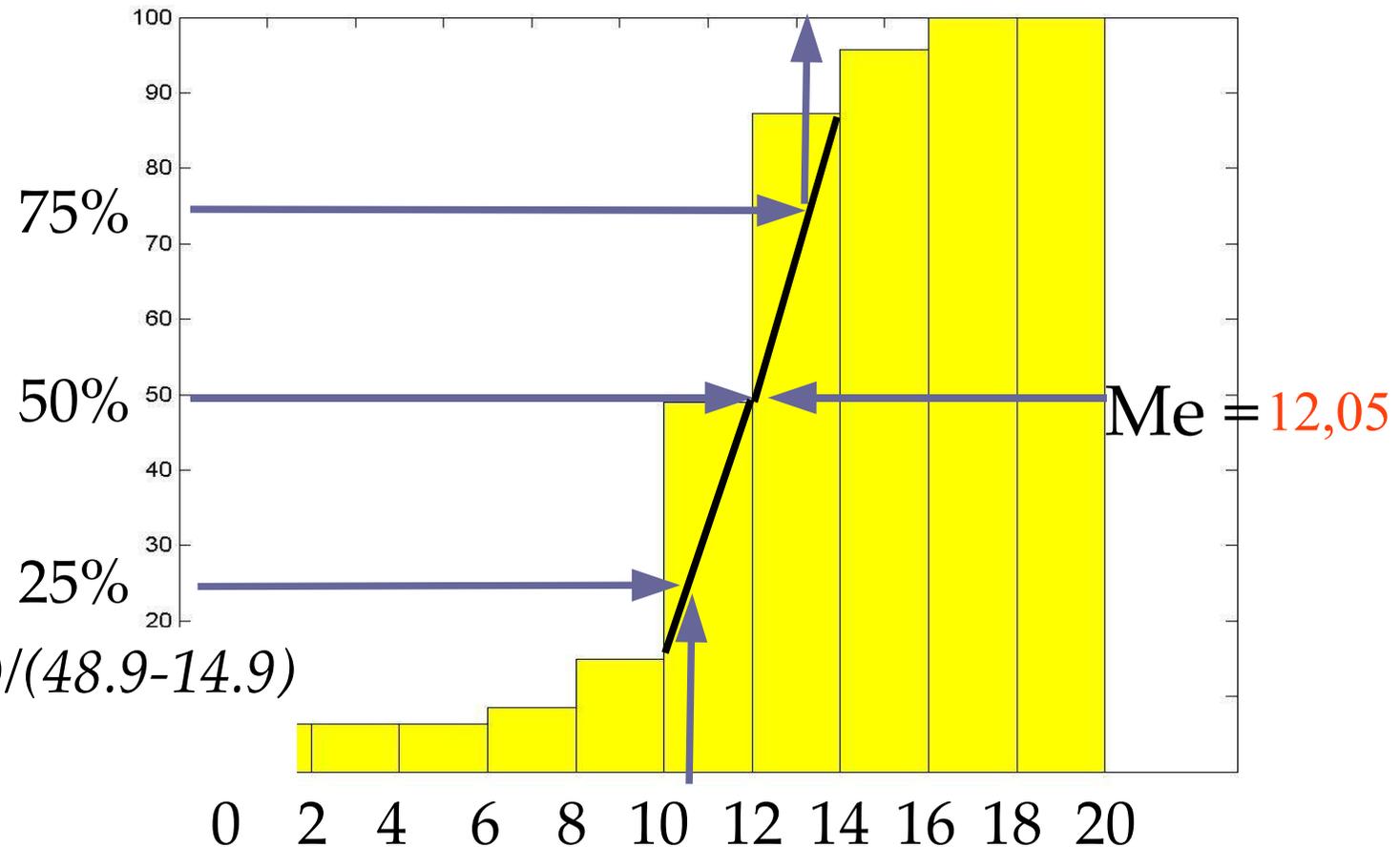
- n_i (dans le cas des classes d'égale amplitude - notée a_i)
- n_i/a_i (dans le cas contraire)

(le rectangle construit pour chaque classe a une surface égale à l'effectif de la classe)

Variable quantitative continue

- Quantile d'ordre α
 - interpolation linéaire de chaque quantile

$$Q3 = 12 + (14 - 12) * (75 - 48.9) / (87.2 - 48.9) = 13.4$$



$$Q1 = 10 + (12 - 10) * (25 - 14.9) / (48.9 - 14.9) = 10.6$$

Variable quantitative continue

- Résumés numériques :
 - La moyenne: deux solutions ...
 - on utilise les n_i du tableau statistique et c_i = centre des classes
 - on utilise les données brutes (on ne les a pas toujours)

Variable qualitative

- Valeurs non numériques

- variable ordinaire s'il est possible d'ordonner les valeurs
- variable nominale sinon
- Taille = {petit, moyen, grand}
- EC = {SE, Stats, RNA, IHM}

H	F	H	H	H	H	H	H	H
H	H	F	H	F	H	H	H	H
H	H	H	H	F	H	H	H	H
F	F	H	H	H	H	H	F	H
H	H	H	H	H	H	H	F	H
H	H	H	H	H	H			

1erC SRC 1erC GTR ...

Variable qualitative

- Présentation des données:
 - Tableau statistique:

x_i	n_i	f_i
H	37	82.2
F	8	17.8

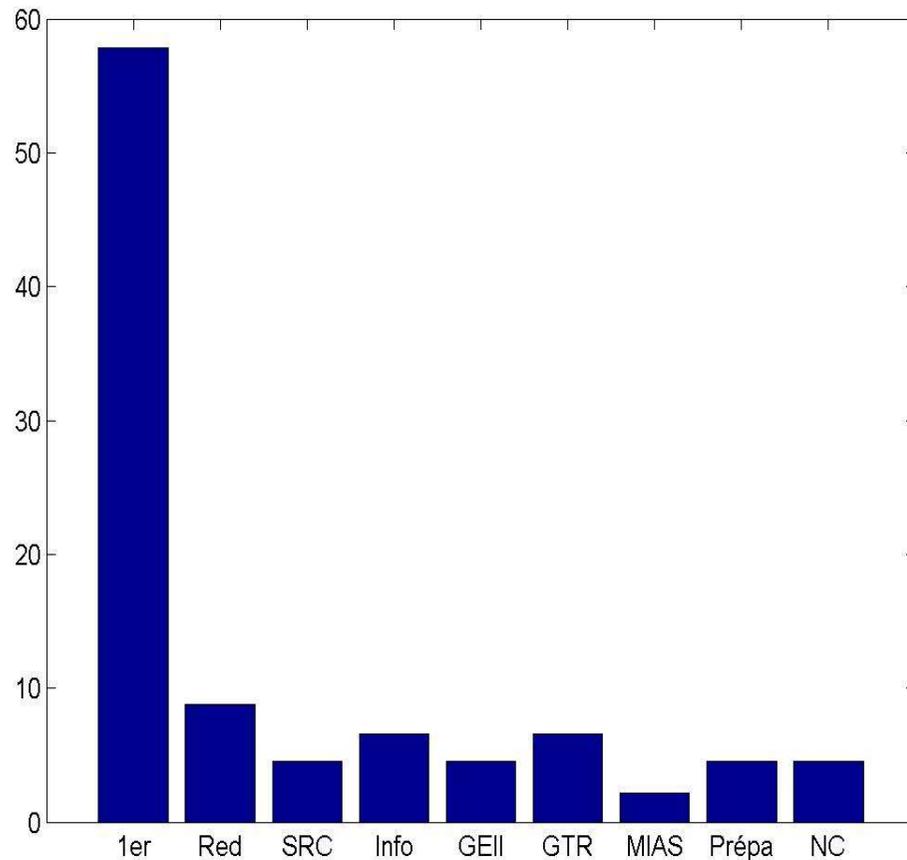
x_i	n_i	f_i
1erC	26	57.8
Red	4	8.8
SRC	2	4.5
Info	3	6.6
GEII	2	4.5
GTR	3	6.6
MIAS	1	2.2
Prépa	2	4.5
NC	2	4.5

Variable qualitative

- Résumés numériques :
 - Ils ont peu de sens ici ...
 - Exemple :
 - Mode: OK *les étudiants d'ASI3 viennent en majorité du 1er cycle*
 - Moyenne : *comment calculer la provenance moyenne ?*
 - Médiane : il faut au moins une relation d'ordre (variable ordinale)

Variable qualitative

- Représentation graphique:
 - Diagramme en colonnes



- Diagramme en secteur ...

