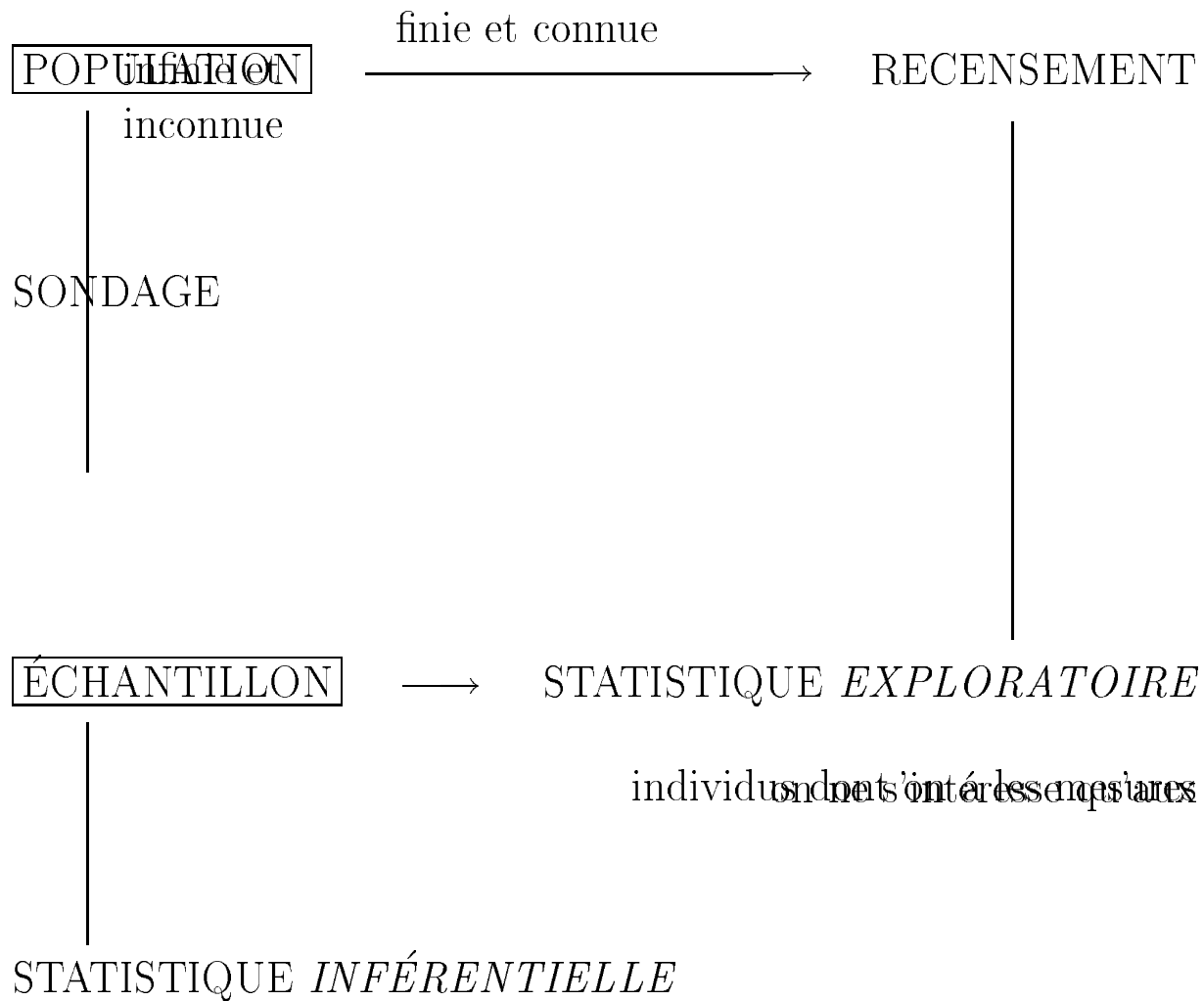


LES DEUX TYPES DE DÉMARCHES STATISTIQUES



LES ÉLÉMENTS DESCRIPTIFS D'UNE POPULATION

- Les représentations graphiques :
 - * histogramme
 - * courbe de densité
 - * fonction de répartition
- Les résumés numériques :
 - * tendance centrale :
 - ** moyenne
 - ** médiane
 - * dispersion :
 - ** variance (et écart-type)
 - ** étendue
 - ** écart inter-quartiles (*IQR*)
- Les résumés semi-graphiques :
 - * boîte à pattes
 - * branchage

Âge	Effectif	Effectif cumulé	Fréquence	Fréquence cumulée
19	2	2	0.0002	0.0002
20	0	2	0.0000	0.0002
21	7	9	0.0008	0.0010
22	41	50	0.0049	0.0059
23	66	116	0.0079	0.0138
24	121	237	0.0145	0.0283
25	128	365	0.0153	0.0436
26	191	556	0.0230	0.0666
27	210	766	0.0252	0.0918
28	212	978	0.0255	0.1173
⋮	⋮	⋮	⋮	⋮
58	159	7840	0.0191	0.9421
59	152	7992	0.0183	0.9604
60	102	8094	0.0123	0.9727
61	66	8160	0.0079	0.9806
62	62	8222	0.0075	0.9881
63	52	8274	0.0063	0.9944
64	42	8316	0.0050	0.9994
65	5	8321	0.0006	1.0000
	8321		1.0000	

DIAGRAMME EN BÂTONS DE LA POPULATION INRA

Classe d'âges	<u>Effectif</u>	Effectif cumulé	<u>Fréquence</u>	Fréquence cumulée
]15 - 20]	2	2	0.0002	0.0002
]20 - 25]	363	365	0.0436	0.0438
]25 - 30]	1081	1446	0.1299	0.1737
]30 - 35]	1084	2530	0.1303	0.3040
]35 - 40]	1370	3900	0.1646	0.4687
]40 - 45]	1396	5296	0.1678	0.6365
]45 - 50]	1155	6451	0.1388	0.7753
]50 - 55]	916	7367	0.1101	0.8854
]55 - 60]	727	8094	0.0874	0.9728
]60 - 65]	227	8321	0.0272	1.0000
	<u>8321</u>		<u>1.0000</u>	

HISTOGRAMME DES ÂGES DE LA POPULATION INRA

10 classes d'âges

POPULATION INRA

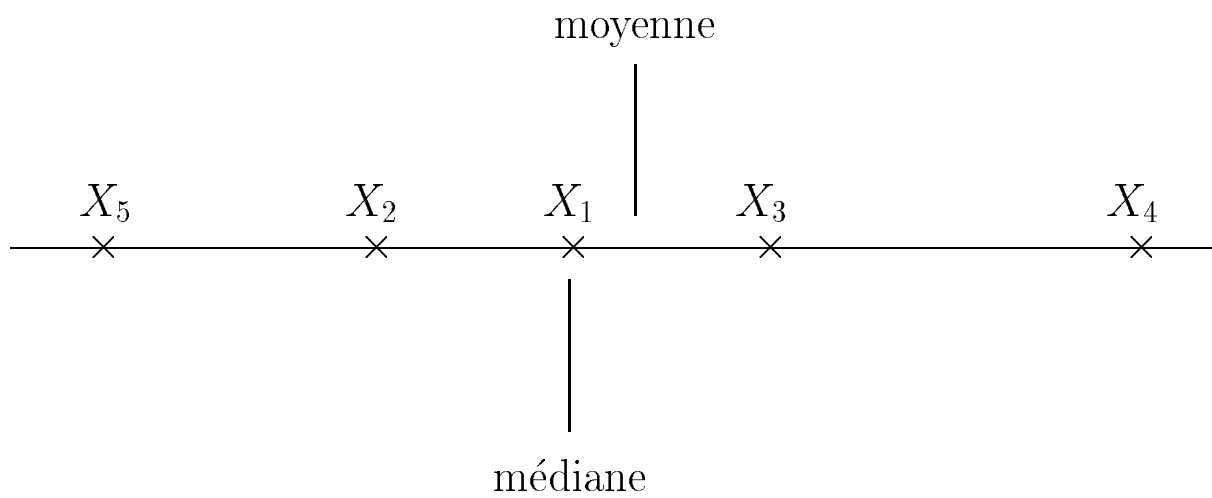
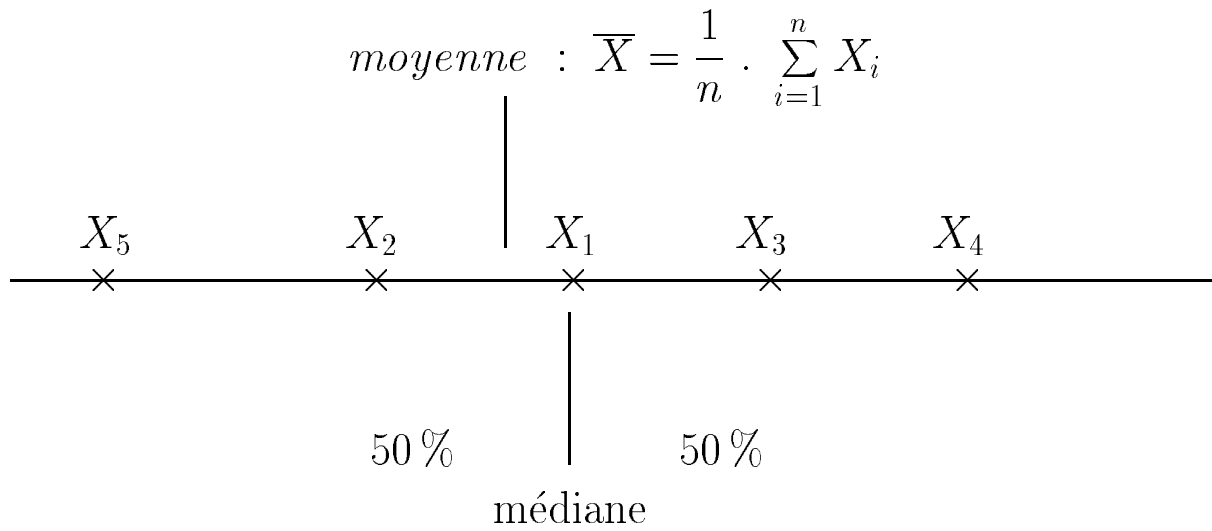
histogramme et polygone des fréquences d'âges

courbe de densité pour une variable continue

POPULATION INRA

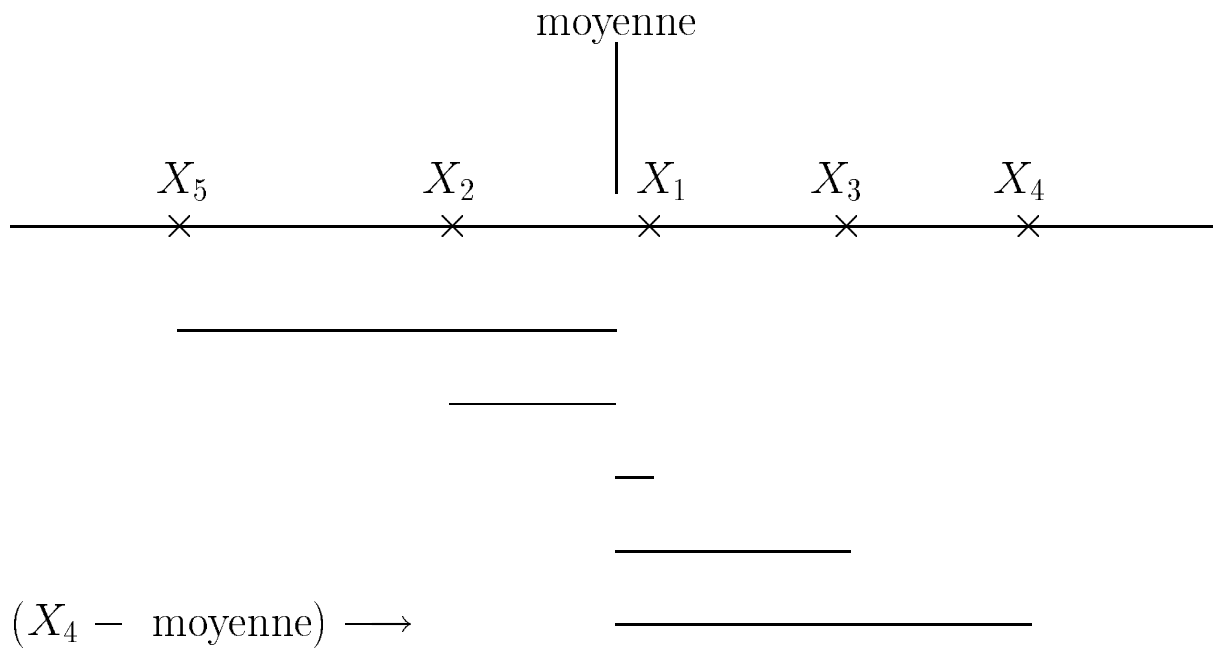
histogramme et polygone des fréquences cumulées d'âges

courbe de fonction de répartition pour une variable continue

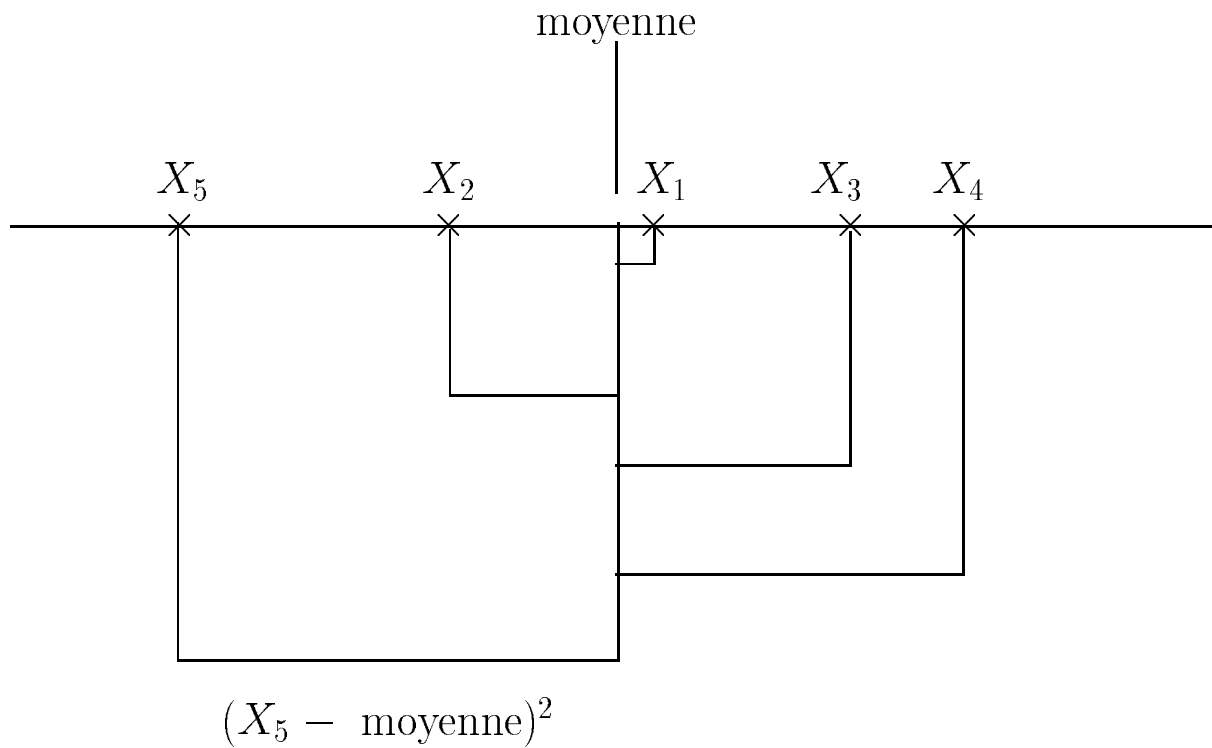


moyenne = médiane

moyenne = médiane



$$\sum_{i=1}^5 (X_i - \text{moyenne}) = 0$$



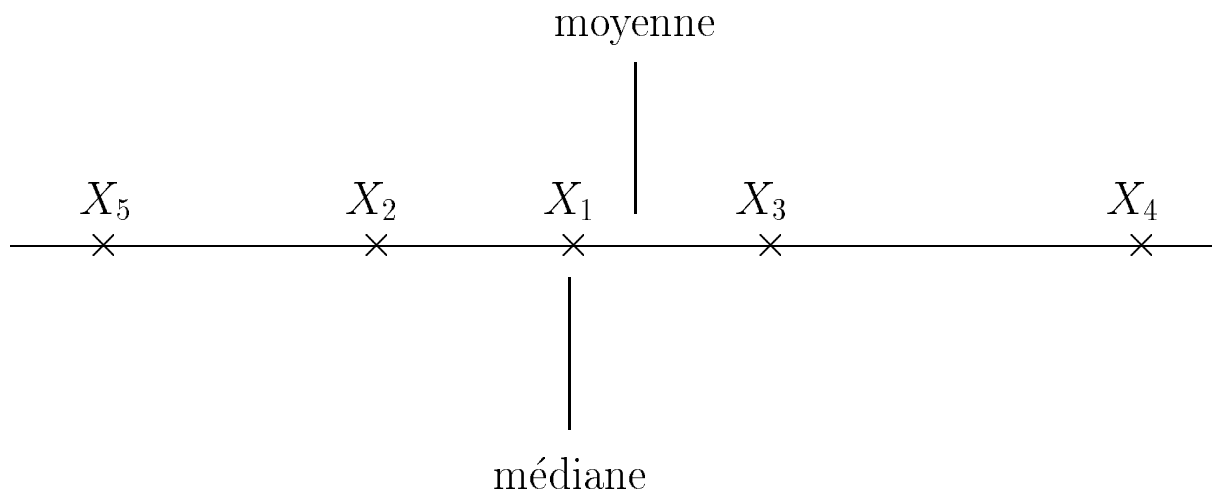
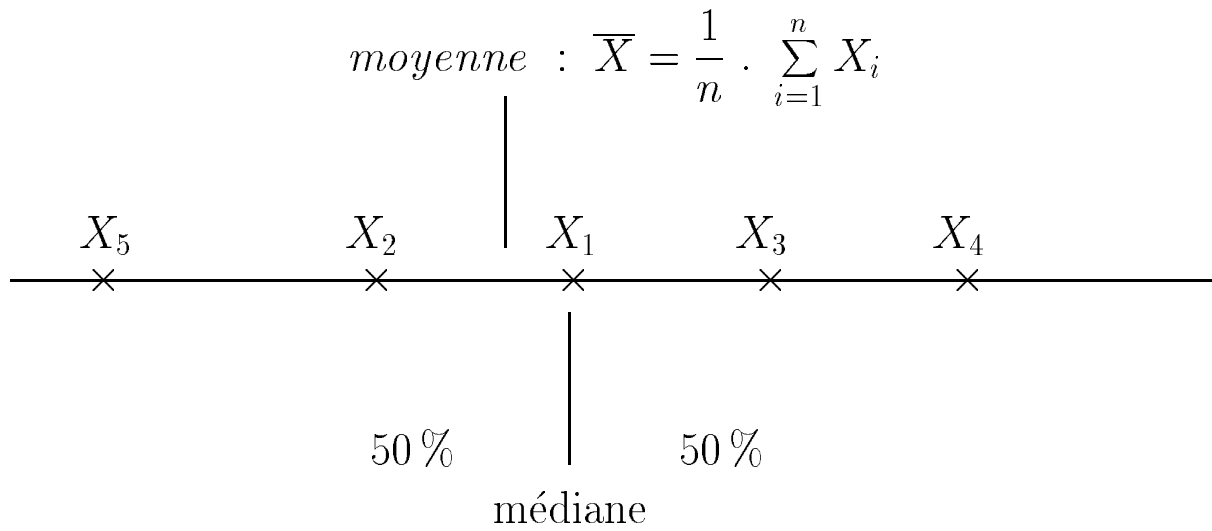
$$\sum_{i=1}^5 (X_i - \text{moyenne})^2$$

moyenne

moyenne

S^2

$$S^2 = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \text{moyenne})^2$$



MESURES DE DISPERSION

étendue

50 % 50 %

min médiane max

étendue

50 % 50 %

min Q_1 Q_2 Q_3 max

25 % 25 % 25 % 25 %

IQR

$$IQR = Q_3 - Q_1$$

LES QUANTILES

k parties :

1 2 3 i $k - 1$ k

min Q_1 Q_2 Q_3 Q_{i-1} Q_i Q_{k-2} Q_{k-1} max

Q_i est la valeur qui est supérieure à $100 \left(\frac{i}{k}\right) \%$

des observations classées dans l'ordre croissant

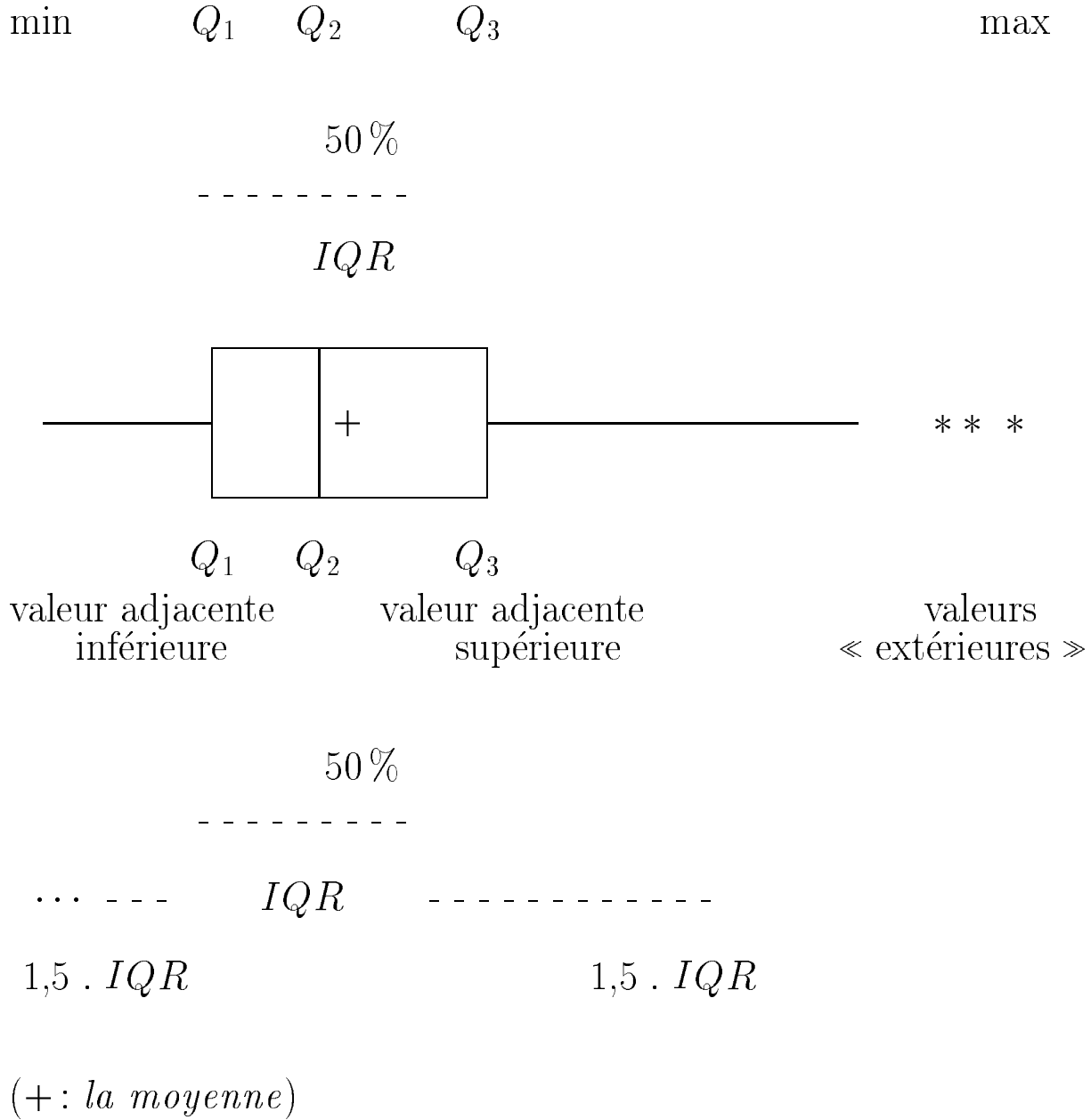
FONCTION DE RÉPARTITION

âges dans la population INRA

FONCTION DE RÉPARTITION

âges dans la population INRA

LA BOÎTE À PATTES



**BOÎTE À PATTES : ÂGES DES
AGENTS INRA DU CENTRE DE LILLE**

BOÎTE À PATTES : ÂGES DES AGENTS INRA

distribution
uniforme

distribution
« pointue »

distribution
« creuse »

distribution
bimodale

Histogramme des âges des 139 agents de Lille

Branchage des âges du Centre de Lille

N = 139 Median = 36

Quartiles = 31, 42

Decimal point is 1 place to the right of the colon

1 : 9

2 : 13444

2 : 5566666677788889999

3 : 00000111111111112222223333334444

3 : 5555566666677777788888899999999

4 : 00000112222222334444

4 : 556667889

5 : 0004

5 : 55566

6 : 012233

VARIABLE ALÉATOIRE

- variable à valeurs dans un ensemble fini ou infini,
- variable discrète ou continue,
- à chacune des valeurs de la variable est associée :
 - * une probabilité si variable discrète,
 - * une densité de probabilité si variable continue.
- Une variable aléatoire est une loi de probabilité.

UNE POPULATION INFINIE

courbe de fonction de répartition

courbe de densité

FONCTION DE RÉPARTITION: F

$$F(t) = \Pr\{X \leq t\}$$

$$\Pr\{a \leq X \leq b\} = F(b) - F(a)$$

DENSITÉ: f

$$F(t) = \int_{-\infty}^t f(x)dx$$

$$\text{soit : } \frac{dF(x)}{dx} = F'(x) = f(x)$$

$$\Pr\{a \leq X \leq b\} = F(b) - F(a) = \int_a^b f(x)dx$$

COURBE DE FONCTION DE RÉPARTITION : F

âge

$$0.25 = F(34, 8) = F(Q_1)$$

$$\text{donc } Q_1 = F^{-1}(0, 25)$$

soit pour $i \in \{1, 2, 3\}$

$$Q_i = F^{-1}(i/4)$$

OBJECTIF DE LA STATISTIQUE INFÉRENTIELLE

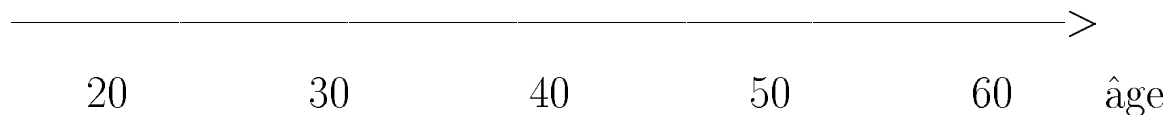
- Notions introduites :
 - * échantillon, n -échantillon
 - * variable aléatoire calculée à partir d'un n -échantillon :
 - une statistique,
 - * paramètre théorique et version empirique,
 - * un estimateur.

- *n*-échantillon :

c'est une suite de n variables aléatoires *indépendantes* ayant la même loi : F .

- Exemple de paramètre caractérisant F : la *médiane* m

- Exemple de *statistique*, estimateur version empirique (fonction d'un n -échantillon) : la *médiane empirique* : \widetilde{X}



\widetilde{X}_i ($i = 1, 2, \dots$, nombre de stagiaires)

\widetilde{X} est une statistique

POPULATION	ÉCHANTILLON
fixe	aléatoire
paramètres théoriques	versions empiriques = estimateurs
	↓
« tout est connu »	INFÉRENCE
	↓
	« information » sur les paramètres de la population inconnue

Version empirique estimant F : \widehat{F}

$$F(t) = \Pr\{X \leq t\}$$

$$\widehat{F}(t) = (1/n) \cdot (\text{nombre de } X_i \leq t)$$

$$\widehat{F}(t) = (1/n) \cdot \#\{X_i \leq t\}$$

ou

$$\widehat{F}(t) = (1/n) \cdot \sum_{i=1}^n \mathbf{1}_{\{X_i \leq t\}}$$

avec

$$\mathbf{1}_{\{X_i \leq t\}} = \begin{cases} 1 & \text{si } X_i \leq t \\ 0 & \text{sinon} \end{cases}$$

variable indicatrice

FONCTION DE RÉPARTITION EMPIRIQUE

âges

NOTATIONS

	paramètres théoriques (population)	version empiriques (échantillon)
médiane	m	\widetilde{X} (\widehat{m})
moyenne	μ	\overline{X} ($\widehat{\mu}$)
variance	σ^2	$\widehat{\sigma}^2$ (S^2)
fonction de répartition	F	\widehat{F}

ESTIMATEUR

Un estimateur est une fonction $T_n(X_i), i \in \{1, 2, \dots, n\}$, des valeurs observées sur un échantillon.

C'est une variable aléatoire dont on espère que la valeur sera « souvent proche » du paramètre que l'on cherche à estimer.

EXEMPLE :

$$\bar{X} = \left(\frac{1}{n}\right) \cdot \sum_{i=1}^n X_i$$

est un estimateur de μ