FORMATIONS AUX STATISTIQUES : MANUEL DU FORMATEUR DE NIVEAU 1

Janvier 1991

SCÉNARIO 4

Représentations d'une série de nombres

INRA

DOCUMENTS "STAGIAIRE"

DÉPARTEMENT DE BIOMÉTRIE

formation permanente fp

Donnees sur les ages des ingenieurs INRA

du centre Antilles-Guyane

* Les valeurs :

53 27 46 57 31 35 48 27 43 29 31 39 44 32

27 31 52 45 30 46 49 30 31 49 56 62 30 32

* Les valeurs après classement :

* Les effectifs par age :

age : 27 29 30 31 32 35 39 43 44 45 46 48 49 52 53 56 57 62

effectifs :

* Le diagramme en batons :

* Les effectifs par classe d'age :

- amplitude de 5 ans :

classes d'age : 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64

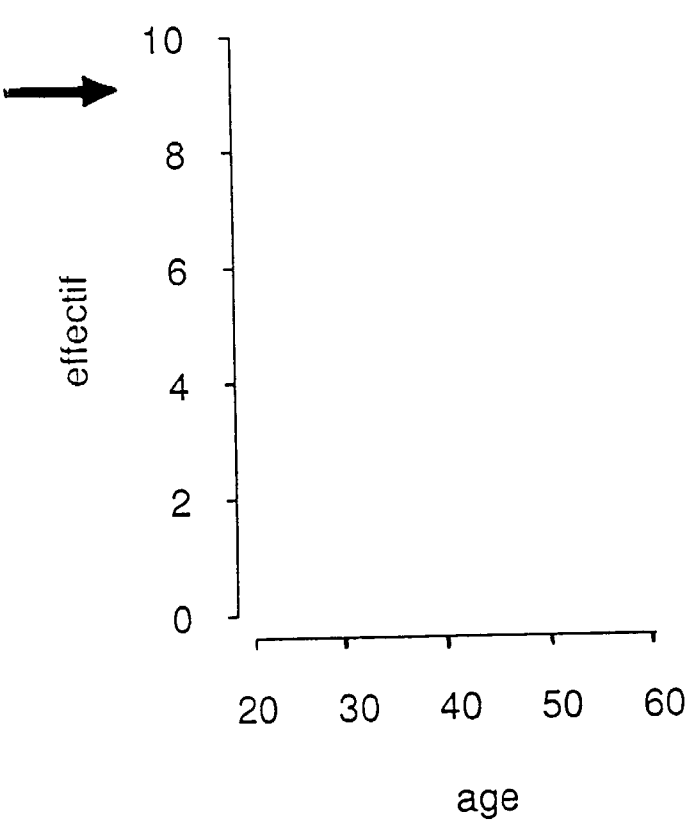
effectifs :

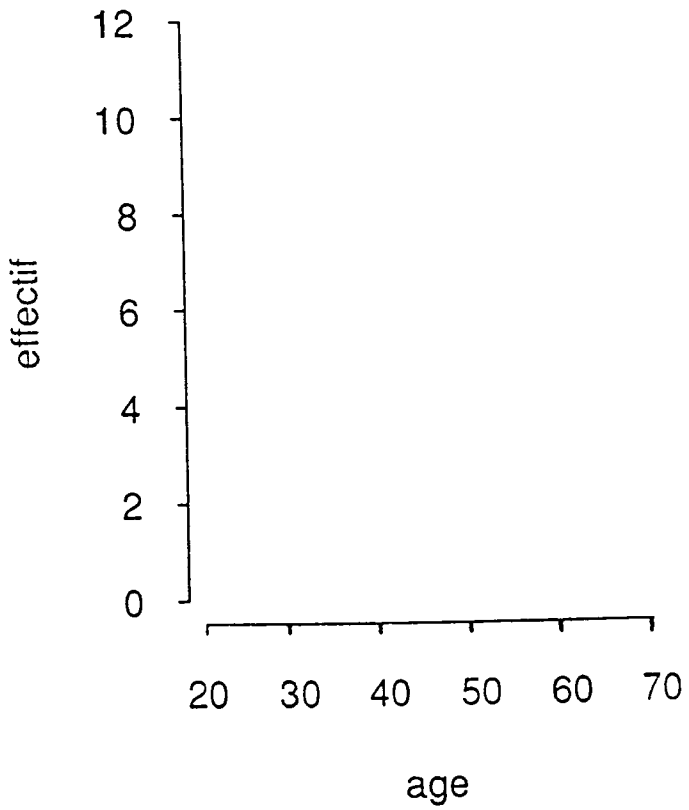
- amplitude de 10 ans :

classes d'ages : 20-29 30-39 40-49 50-59 60-69

effectifs :

* Les histogrammes :





du centre Antilles-Guyane

moyenne : 39.7

variance :

écart-type :

médiane : 37

quartile inferieur :

quartile superieur :

écart inter-quartile :

Branchage :

N = 28 Median = 37

Boite à pattes :

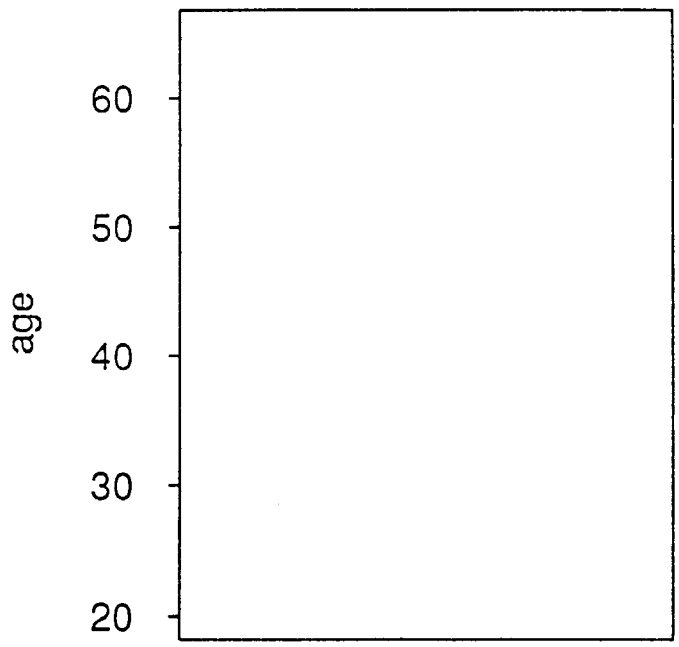
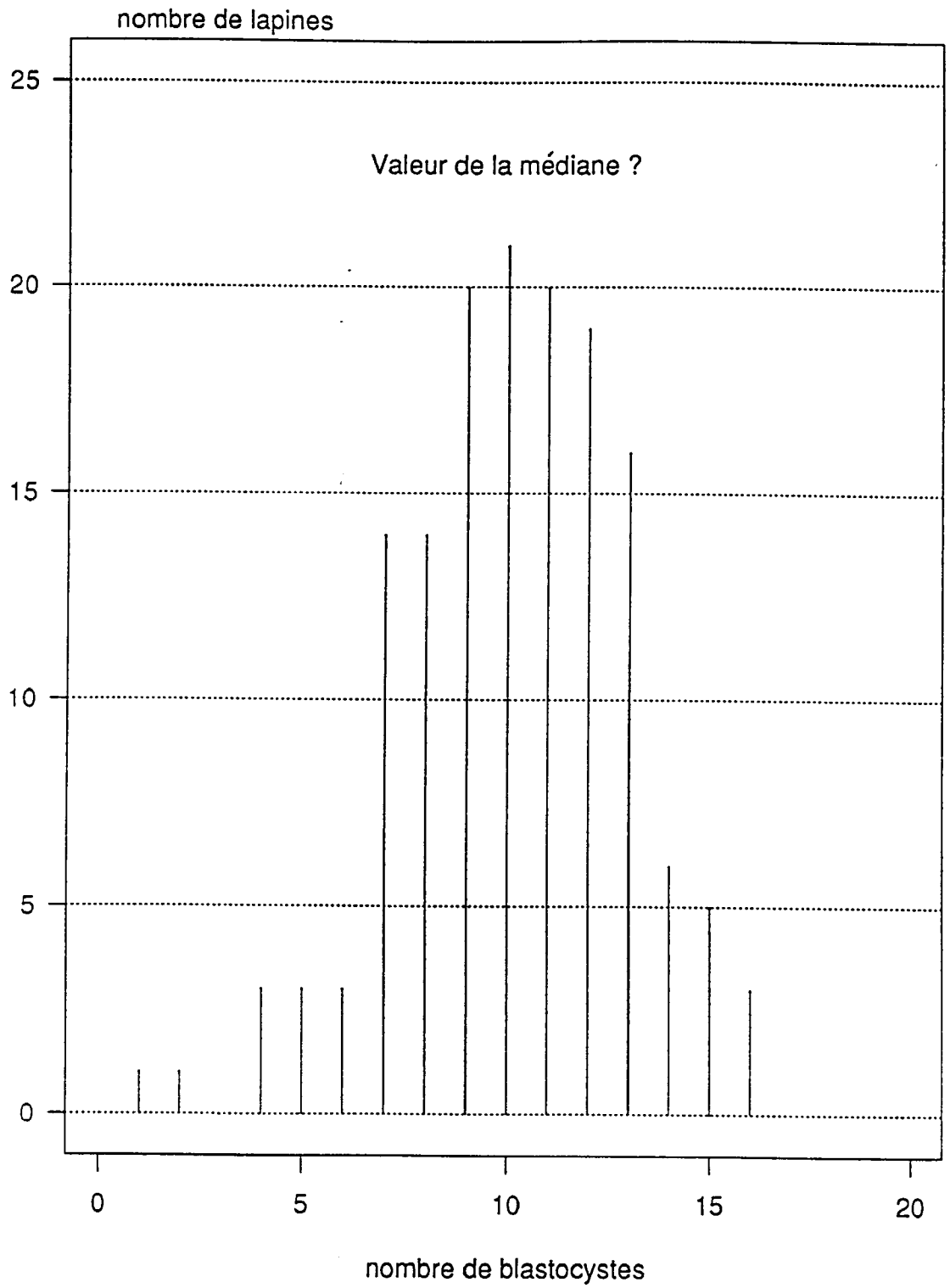
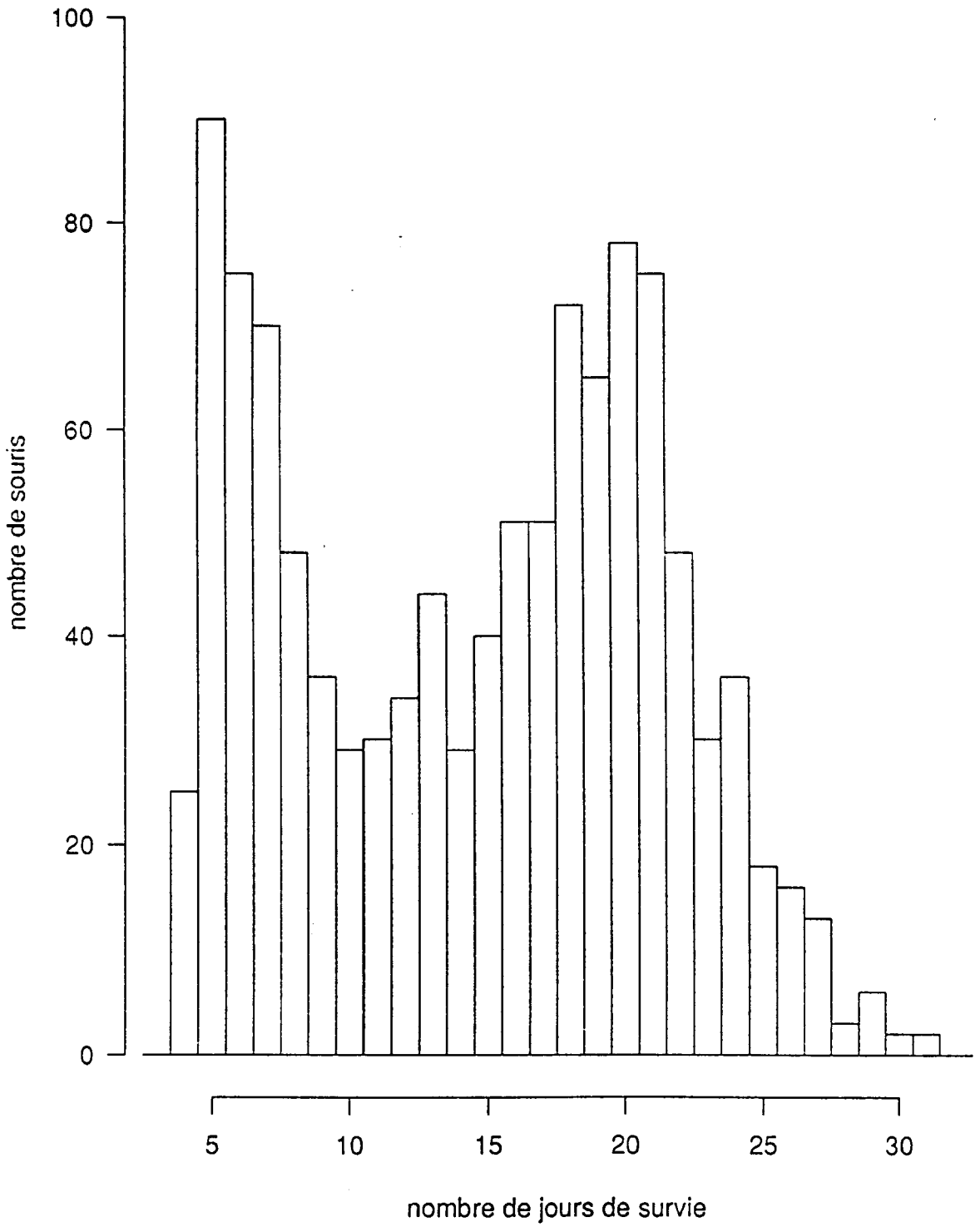


Diagramme en batons du nombre de blastocystes
par lapines (N=149)



Histogramme des temps de survie de souris



On dispose de 5 séries de données.

Pour chaque série, on a représenté :

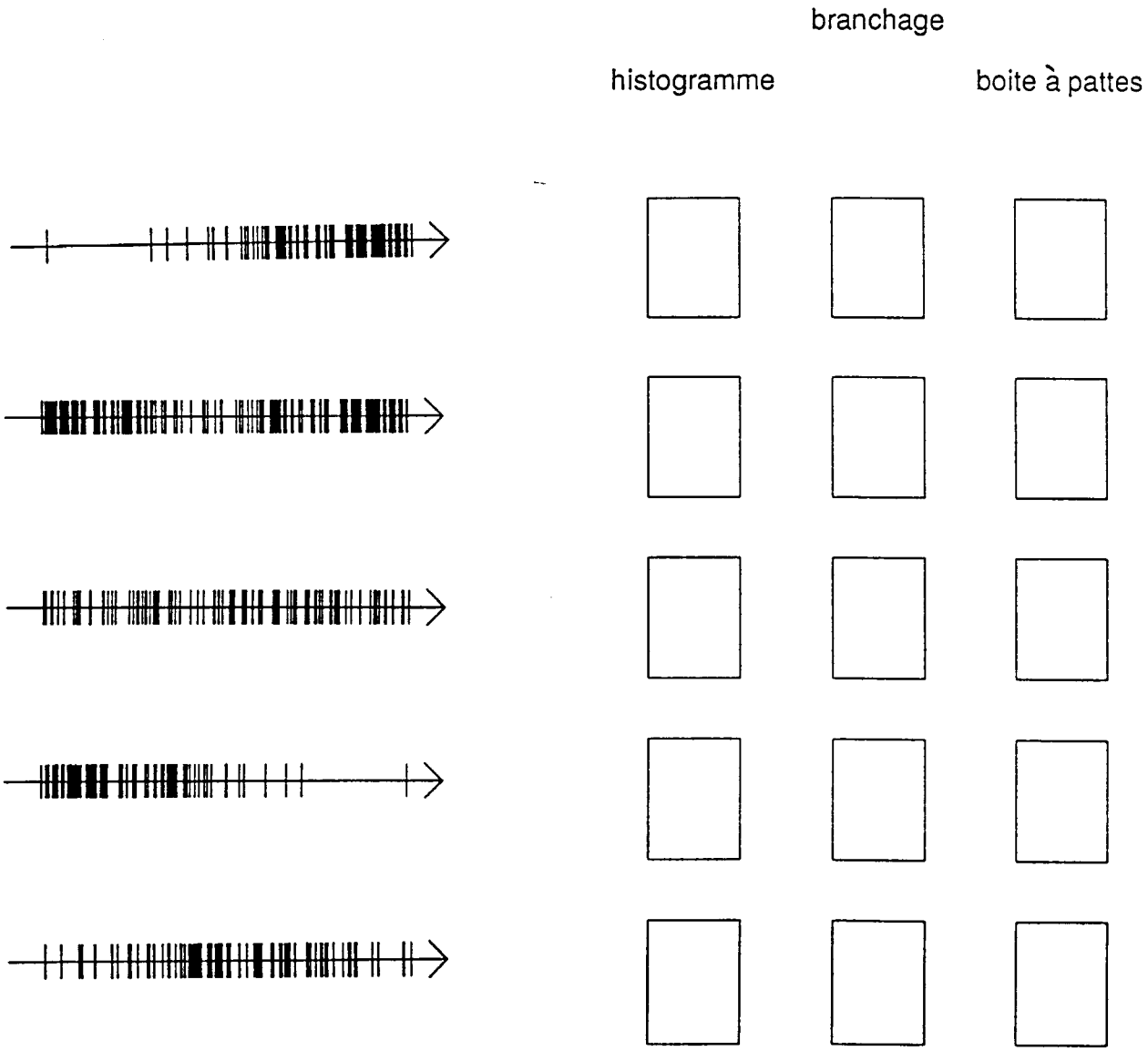
* sur un axe orienté toutes les valeurs de la série (chaque donnée étant représentée par un petit trait vertical),

* l'histogramme associé,

* le branchage associé,

* la boite à pattes associeé.

Or on a mélangé tous les histogrammes, les branchages, et les boites à pattes. Le but de l'exercice est d'indiquer, en face de chaque série de valeurs, les numéros d'histogramme, de branchage et de boite a pattes qui lui correspond.
À chaque série de valeurs, il correspond un seul histogramme, un seul branchage et une seule boite à pattes.



S 1

```
 2 : 8
 3 : 4
 4 : 1237
 5 : 446
 6 : 01114889
 7 : 04669
 8 : 11122445667778889
 9 : 122233455666799
10 : 003369
11 : 01126699
12 : 012335
13 : 00135578
14 : 035689
15 : 57
16 : 77
17 : 0
```

S 2

```
 0 : 224455689999
 1 : 002335566667778999
 2 : 124446677888999
 3 : 11222334
 4 : 00111344667778
 5 : 347778
 6 : 012345678
 7 : 1233479
 8 : 135
 9 : 128
10 : 0
11 : 1
12 : 18
13 :
14 :
15 :
16 :
17 : 9
```

S 3

```
 0 : 2224455689999
 1 : 002335566667778999
 2 : 124446677888999
 3 : 11222334
 4 : 00111344667778
 5 : 347778
 6 : 012345678
 7 : 12334789
 8 : 1359
 9 : 128
10 : 04
11 : 19
12 : 138
13 : 23
14 : 368
15 : 258
16 : 11338
17 : 01356889
18 : 034459
19 : 18999
20 : 00445888
21 : 009
22 : 00223447777789
23 : 00234489
24 : 1122445566688
25 : 0125566679
26 : 1223477
```

S 4

```
 0 : 1
 1 :
 2 :
 3 :
 4 :
 5 : 29
 6 : 9
 7 :
 8 : 0289
 9 : 579
10 : 1367789
11 : 23456789
12 : 0233367
13 : 233334667999
14 : 0067788899
15 : 111222334466689
16 : 1112333444455778
17 : 00111124556688
```

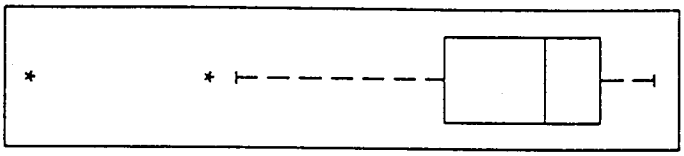
S 5

```
 0 : 7888
 1 : 1258
 2 : 34456
 3 : 12289
 4 : 035
 5 : 244679
 6 : 0135567
 7 : 24469
 8 : 48
 9 : 116799
10 : 15567
11 : 1133367
12 : 027789
13 : 00057899
14 : 4569
15 : 02378
16 : 01258
17 : 38
18 : 01367
19 : 005688
```

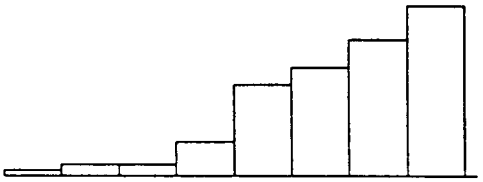
axe orienté commun

H 1

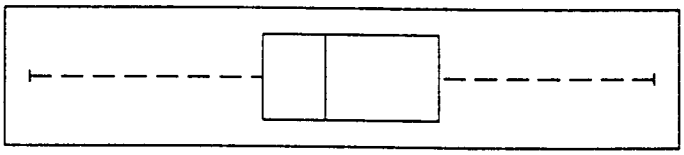
B 1



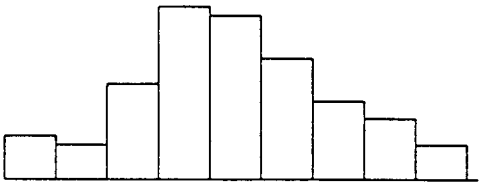
H 2



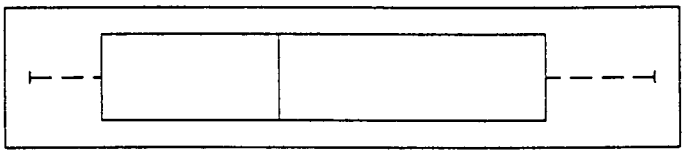
B 2



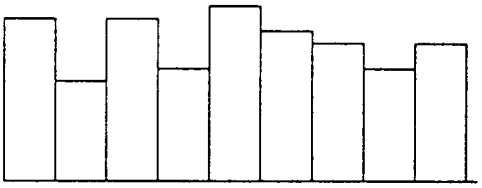
H 3



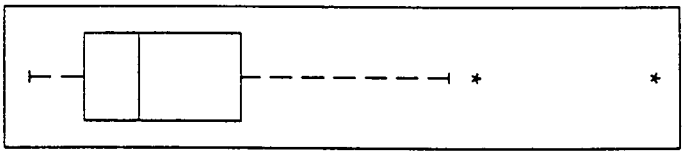
B 3



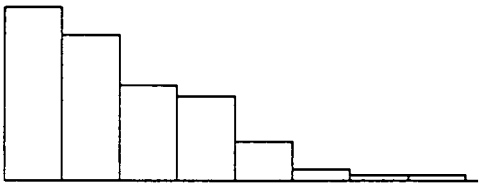
H 4



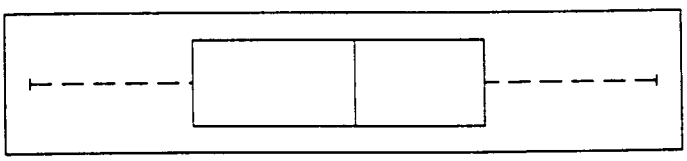
B 4



H 5



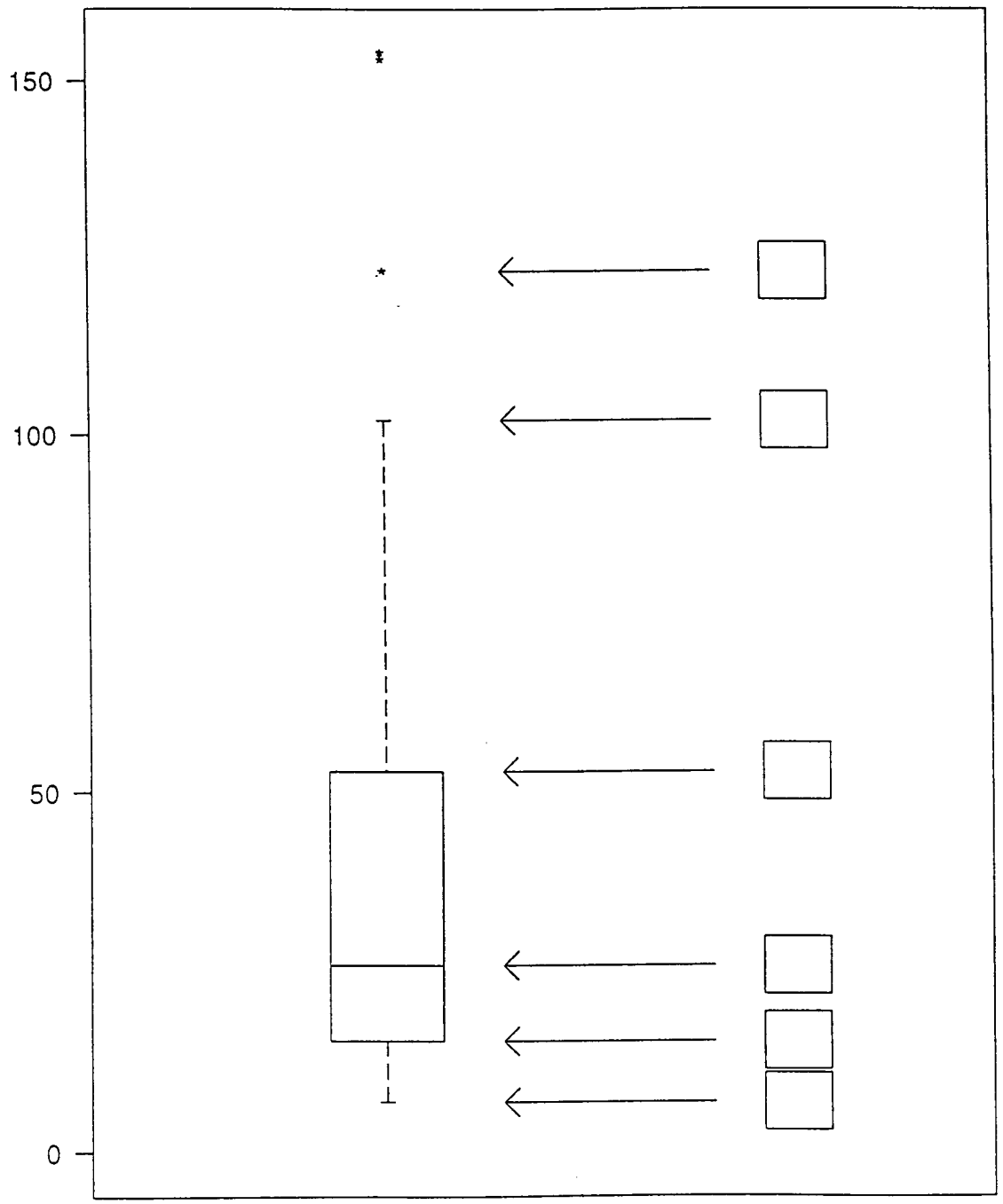
B 5



Les données ordonnées :

```
 7   8   8  10  10  11  11  13  13  15  16  16  18  20  20  21  22  23  23
24  28  28  31  33  35  44  45  46  50  50  56  60  64  68  89  93 102 123
153 154
```

La boite à pattes associée :



Indiquer dans les cases à droite, les valeurs correspondantes aux éléments fléchés de la boîte à pattes.

Branchage pour la série 1

N = 500 Median = 40
Quartiles = 37, 44

Low: 21

```
26 : 0
27 :
28 : 00
29 : 000000000
30 : 0000000
31 : 00000
32 : 00000000
33 : 000000000000000000
34 : 00000000000000000000000
35 : 000000000000000
36 : 0000000000000000000000000000
37 : 0000000000000000000000000000000
38 : 000000000000000000000000000000000000000000000
39 : 00000000000000000000000000000000000
40 : 0000000000000000000000000000000
41 : 000000000000000000000000000000000000000000000000000
42 : 00000000000000000000000000000000
43 : 000000000000000000000000000000000000
44 : 0000000000000000000000000000000
45 : 000000000000000000000
46 : 00000000000000000
47 : 0000000000000000
48 : 000000000000
49 : 0000000000
50 : 000000000
51 : 00000
52 : 00
53 : 0
54 :
55 :
56 : 0
```

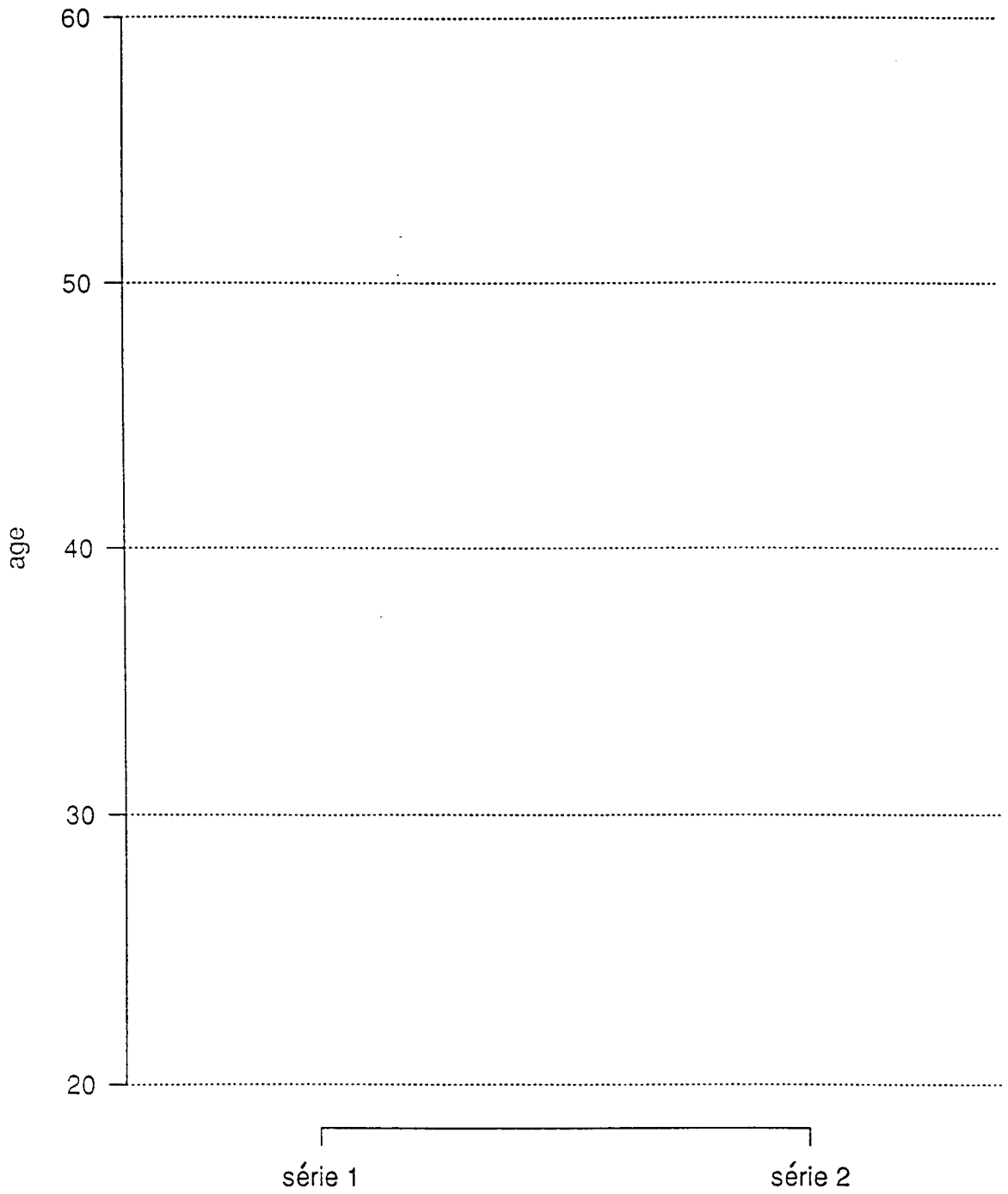
Branchage pour la série 2

N = 500 Median = 40
Quartiles = 38, 42

```
32 : 000
33 : 000000
34 : 00000000000
35 : 000000000000000
36 : 0000000000000000000000
37 : 00000000000000000000000000000000000000000000000
38 : 000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
39 : 0000000000000000000000000000000000000000000000000000000000000000000000000000000000000
40 : 000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
41 : 0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
42 : 00000000000000000000000000000000000000000000000000000000000000000000000000000000000000
43 : 000000000000000000000000000000000000000000000000000000000000000000000000000000000000
44 : 0000000000000000000000000000000
45 : 000000000000000000
46 : 0000000
47 : 0000000
48 : 00
```

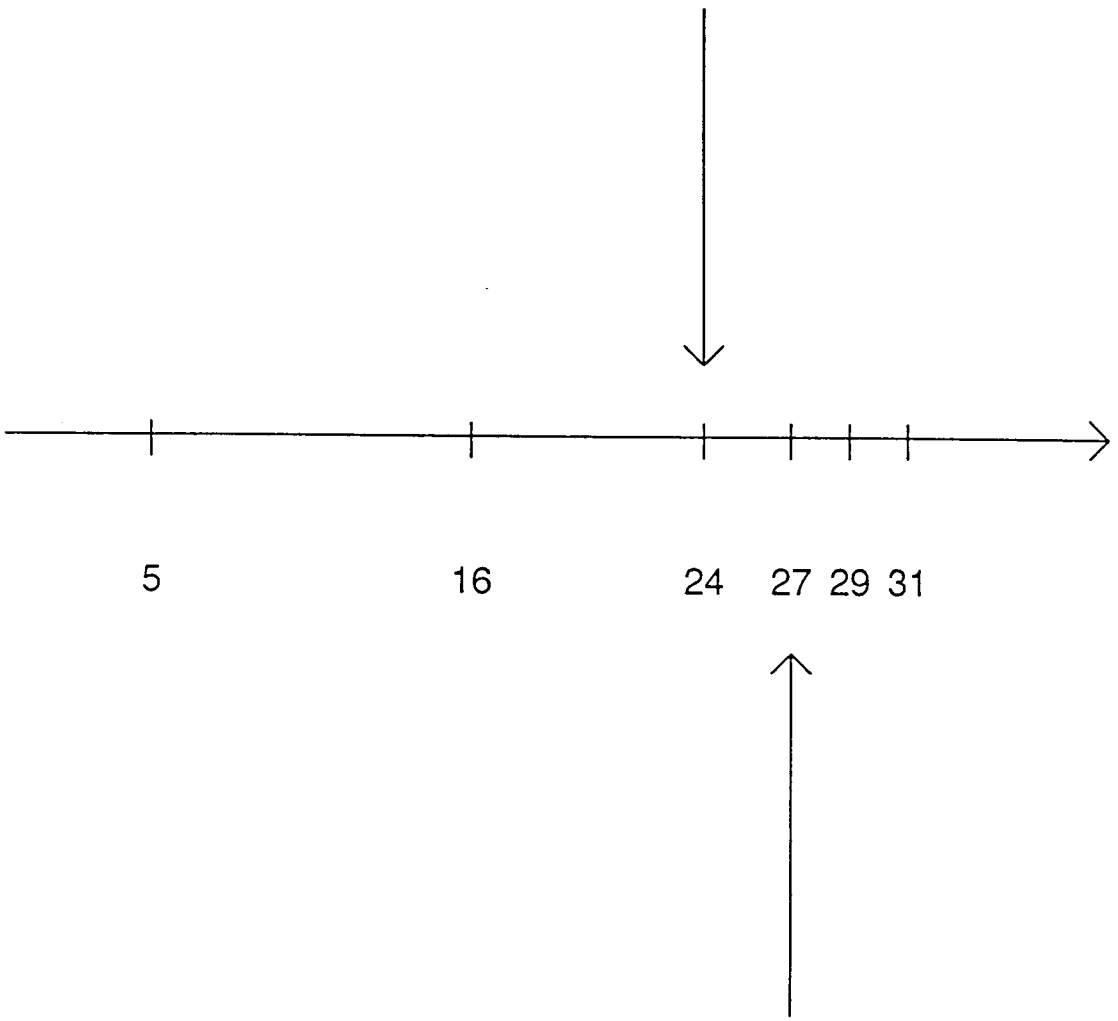
Boite à pattes des séries 1 et 2

Indiquer un paramètre differenciant bien les deux séries :



valeur de la donnée manquante =

moyenne = 24



médiane = 27

valeur de la donnée manquante =

Pour décrire les revenus d'une population d'un pays, on peut donner la médiane ou la moyenne des salaires mensuels :

	moyenne	mediane
pays 1	20 000	8 000
pays 2	10 000	10 000
pays 3	11 800	9 000

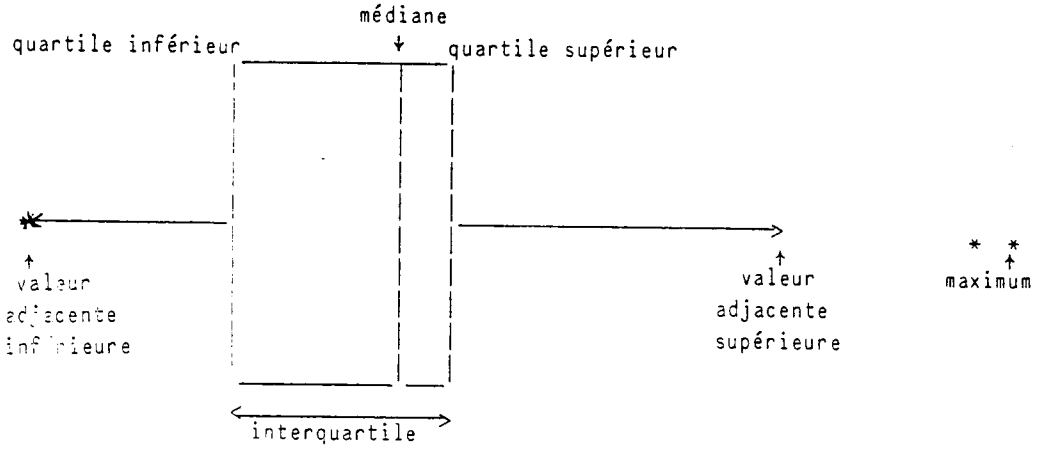
Quel autre paramètre voulez-vous connaitre avant de vous décider ?
à émigrer

Suggérer des distributions .

Scénario 4

VOCABULAIRE

Boîte à pattes (boxplot) : la boîte à pattes est un graphique où les données sont représentées par une boîte qui contient la moitié des données, cette boîte est séparée en deux par la médiane. Elle a des pattes de longueur maximale 1 1/2 écart interquartile. Les valeurs au-delà des pattes sont dites extérieures.

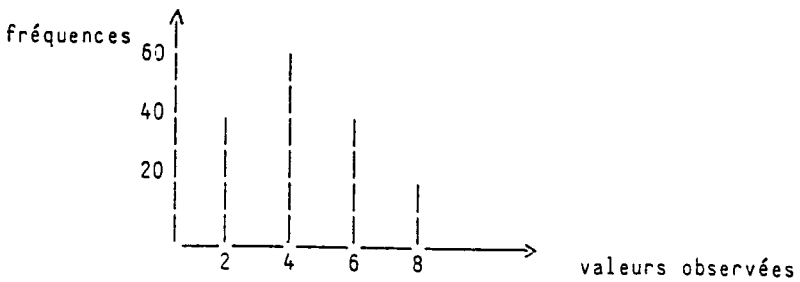


Branchage (stem and leaf) : cette représentation ne peut s'appliquer qu'à une variable quantitative. Pour construire un diagramme "stem and leaf", on écrit d'abord les premiers chiffres à gauche d'une ligne verticale. Ensuite on représente chaque valeur en inscrivant le (ou les) chiffre(s) restant(s) à droite de la verticale.

1	4 4
2	4 5 6 7 8 9
3	1 2 3 3 4 5 6 7 8 8 9 9 9 9 9
4	3 6 7 7
5	0 1 1 4 4 4 4 6 6 8
6	0 0 0 0 0 0 2 2 3
7	1 2 4 4 8
8	0 1 3 3 5
9	1 3 4 6
10	0

Exemple de lecture : la ligne 9 correspond aux valeurs 91, 93, 94 et 96.

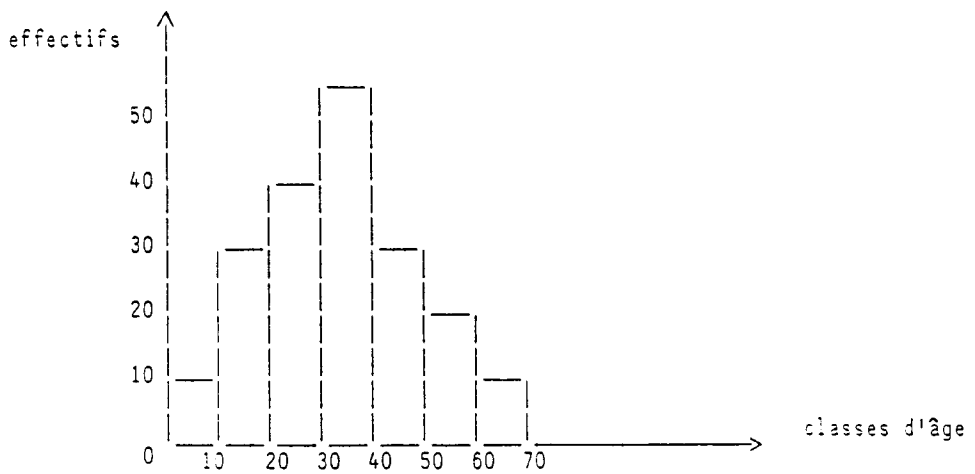
Diagramme en bâtons (bar diagram) : cette représentation s'utilise pour une variable qualitative. Un diagramme en bâtons est constitué de plusieurs traits verticaux. Chaque trait est parallèle à un axe vertical et coupe l'axe horizontal en une valeur qui représente une valeur observée. La longueur de ce trait est proportionnelle à la fréquence (ou effectif) de la valeur observée.



Exemple de lecture : la valeur 6 a été observée 40 fois

Ecart-type (standard deviation) : l'écart-type est la racine carrée de la **variance**. Il constitue une mesure de la dispersion des observations.

Histogramme (histogram) : l'histogramme se compose de rectangles contigus. Sur l'axe horizontal figurent les modalités ou les classes de la variable observée (classes d'âge par exemple) et sur l'axe vertical sont représentés les effectifs. La base des rectangles est choisie de même amplitude que la classe et leur hauteur est proportionnelle à l'effectif de la classe correspondante.



Exemple de lecture : la classe des âges 20-30 ans contient 40 individus

Médiane (median) : la médiane de l'échantillon est la valeur qui sépare les observations en deux moitiés de même effectif.

10 12 15 19 27 31 32

La médiane est 19 dans cet exemple.

Moyenne (mean) : la moyenne de l'échantillon est égale à la somme des valeurs observées divisée par le nombre d'observations.

Quartile (quartile) : un ensemble d'observations peut être divisé en quatre parties d'effectifs égaux. Ces parties sont délimitées par cinq valeurs particulières dont certaines sont appelées quartile :

- minimum
- quartile inférieur
- médiane
- quartile supérieur
- maximum

Ainsi le quartile inférieur est la valeur telle que le quart des observations lui sont inférieures et les trois-quarts supérieures.

Tandis que le quartile supérieur est la valeur telle que les trois quarts des observations lui sont inférieures et le quart supérieur.

Variance (variance) : la variance d'un échantillon est la **moyenne des carrés** des écarts à la moyenne des observations. Elle constitue une mesure de la dispersion des observations.