FORMATIONS AUX STATISTIQUES : MANUEL DU FORMATEUR DE NIVEAU 1

Janvier 1991

SCÉNARIO 4

Représentations d'une série de nombres

INRA

DOCUMENTS "FORMATEUR"

DÉPARTEMENT DE BIOMÉTRIE

formation permanente

On dispose de 5 séries de données.

Pour chaque série, on a représenté :

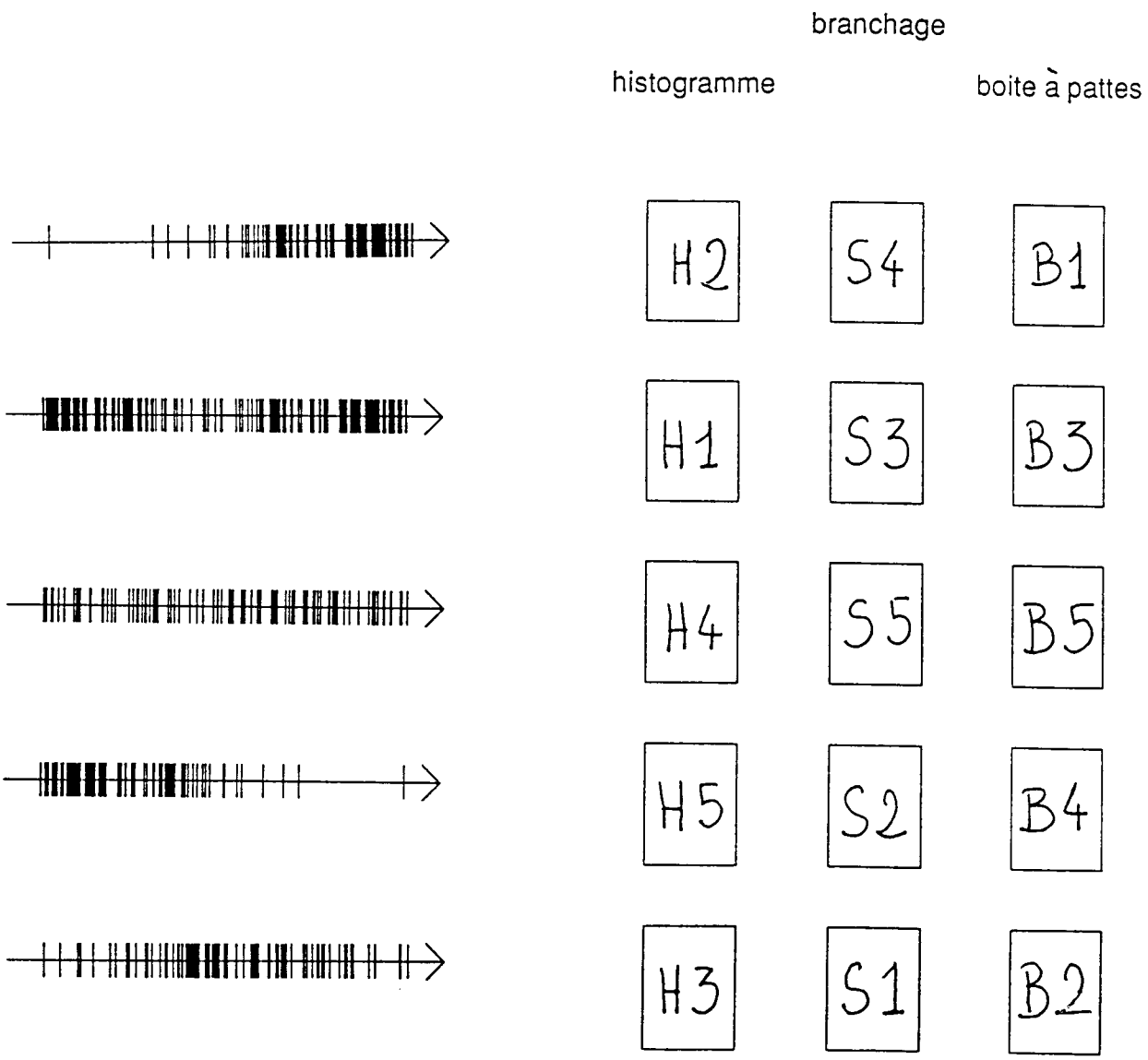
* sur un axe orienté toutes les valeurs de la série (chaque donnée étant représentée par un petit trait vertical),

* l'histogramme associé,

* le branchage associé,

* la boite à pattes associée.

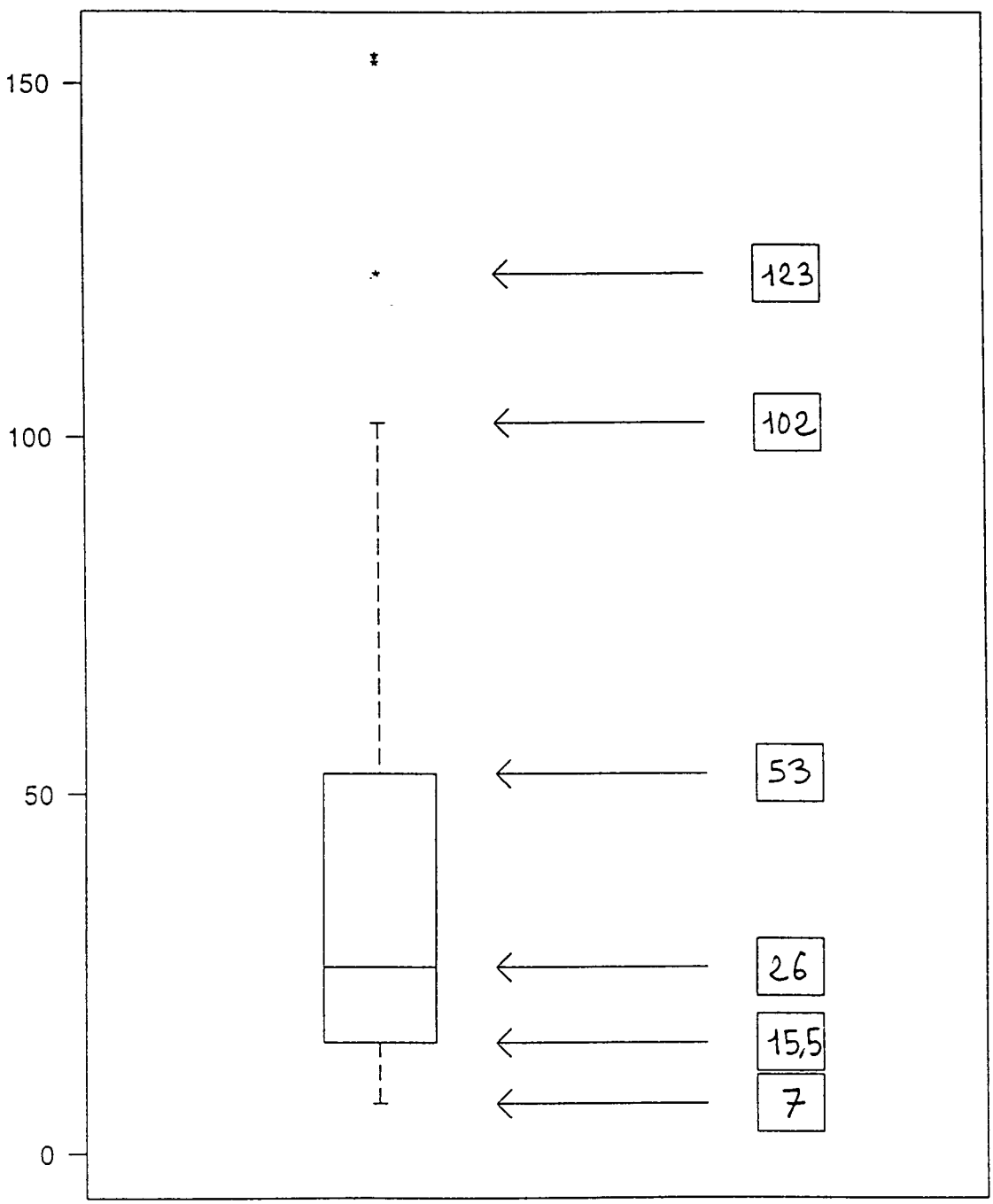
Or on a mélangé tous les histogrammes, les branchages, et les boites à pattes. Le but de l'exercice est d'indiquer, en face de chaque série de valeurs, les numéros d'histogramme, de branchage et de boite à pattes qui lui correspond.
A chaque série de valeurs, il correspond un seul histogramme, un seul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



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

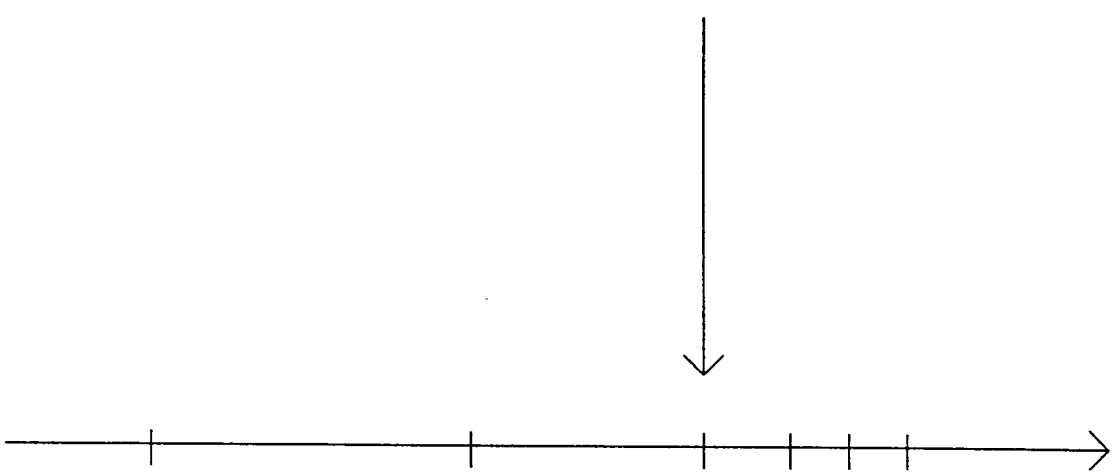
La boite à pattes asscciée :



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

valeur de la donnée manquante = 36

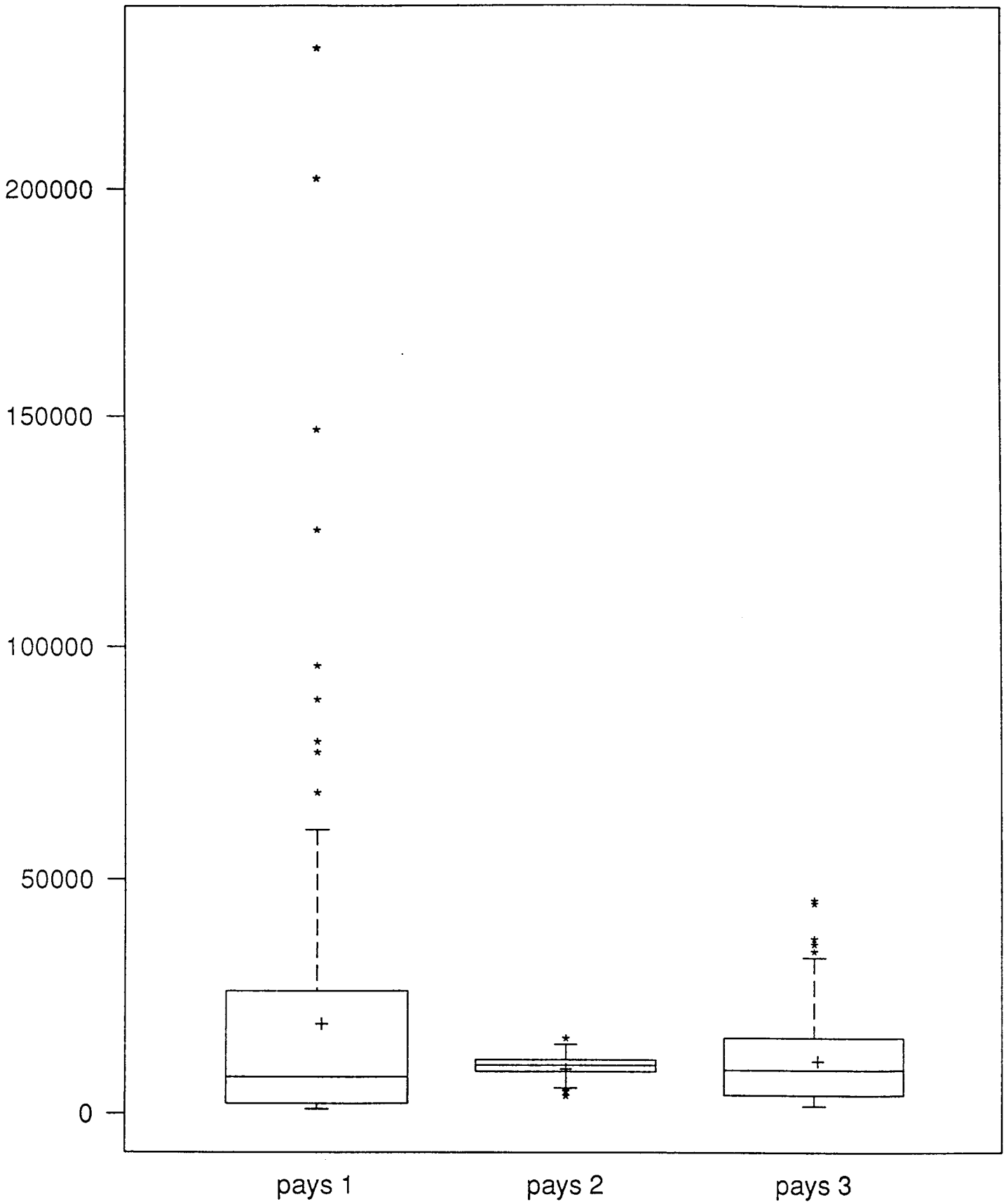
moyenne = 24



médiane = 27

valeur de la donnée manquante = toutes valeurs ≥ 31

Exemples de distributions possibles



Données sur les âges des ingénieurs 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 :

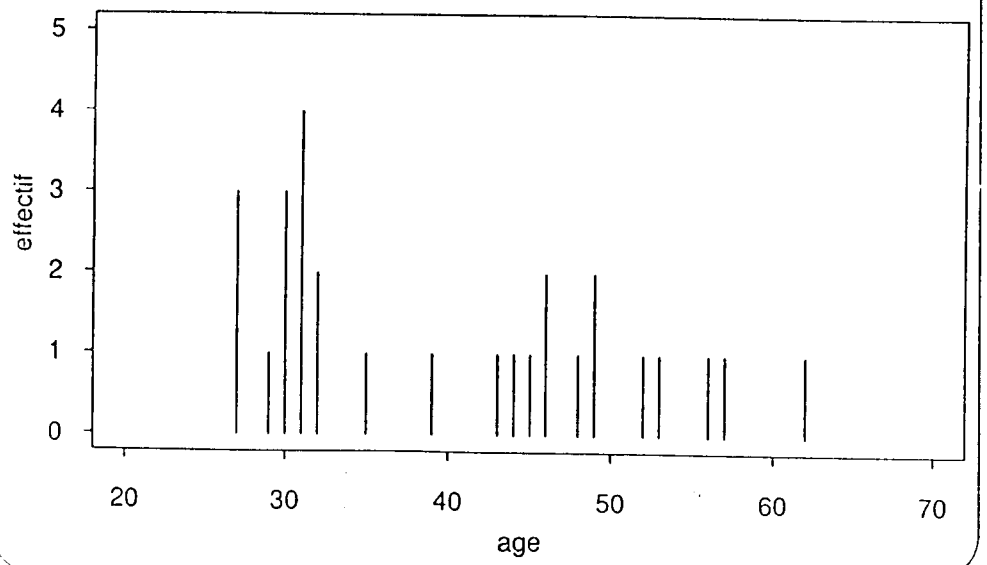
27 27 27 29 30 30 30 31 31 31 31 32 32 35

39 43 44 45 46 46 48 49 49 52 53 56 57 62

* Les effectifs par age :

age	:	27	29	30	31	32	35	39	43	44	45	46	48	49	52	53	56	57	62
effectifs	:	3	1	3	4	2	1	1	1	1	1	2	1	2	1	1	1	1	1

* Le diagramme en batons :



* Les effectifs par classe d'age :

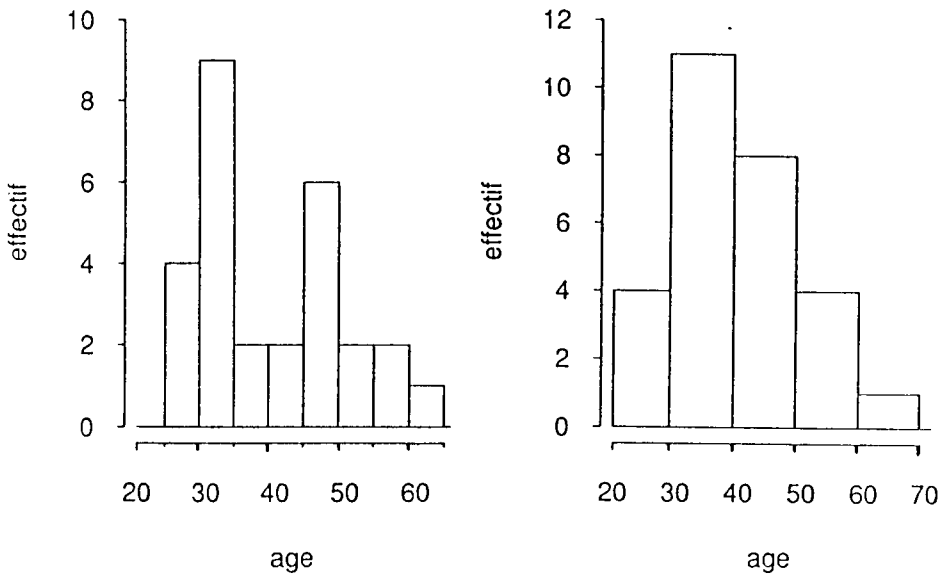
- amplitude de 5 ans :

classes d'age :	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64
effectifs :	4	9	2	2	6	2	2	1

- amplitude de 10 ans :

classes d'ages :	20-29	30-39	40-49	50-59	60-69
effectifs :	4	11	8	4	1

* Les histogrammes :



Données sur les ages des ingénieurs INRA

du centre Antilles-Guyane

moyenne : 39.7

variance : 110.5

ecart-type : 10.5

médiane : 37

quartile inferieur : 30.5

quartile superieur : 48.5

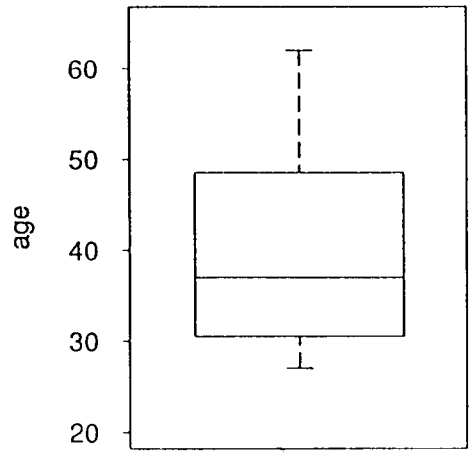
écart inter-quartile : 18

Branchage :

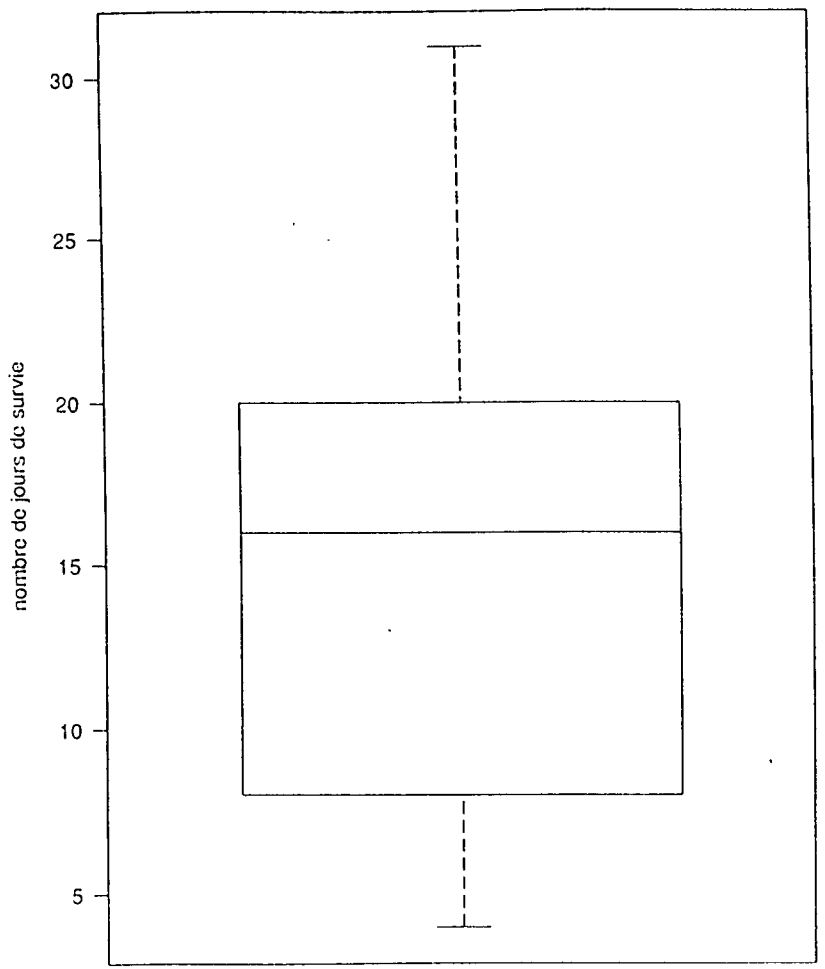
N = 28 Median = 37
Quartiles = 30.5, 48.5

```
2 : 7779
3 : 000111122
3 : 59
4 : 34
4 : 566899
5 : 23
5 : 67
6 : 2
```

Boite à pattes :



Boite à pattes des temps de survie de souris



Les ages des techniciens de recherche du centre

INRA Antilles-Guyane.

* Le paquet de données :

37 44 45 61 48 52 44 44 22 44 22 24 39 40 45
43 45 30 39 49 49 55 41 42 42 48 26 41 43

* Les données classées par ordre croissant :

22 22 24 26 30 37 39 39 40 41 41 42 42 43 43
44 44 44 44 45 45 45 48 48 49 49 52 55 61

Branchage des ages des techniciens de recherche

du centre INRA Antilles-Guyane

* Les données classées par ordre croissant :

22 22 24 26 30 37 39 39 40 41 41 42 42 43 43
44 44 44 44 45 45 45 48 48 49 49 52 55 61

```
2 : 2246
3 : 0799
4 : 011223344445558899
5 : 25
6 : 1
```

```
2 : 224
2 : 6
3 : 0
3 : 799
4 : 01122334444
4 : 5558899
5 : 2
5 : 5
6 : 1
```

* Les données classées :

22 22 24 26 30 37 39 39 40 41 41 42 42 43 43

44 44 44 44 45 45 45 48 48 49 49 52 55 61

* Les effectifs par age :

age	22	24	26	30	37	39	40	41	42	43	44	45	48	49	52	55	61
effectif	2	1	1	1	1	2	1	2	2	2	4	3	2	2	1	1	1

* Les effectifs par classe d'age :

classe d'age	20-29	30-39	40-49	50-59	60-69
effectif	4	4	18	2	1

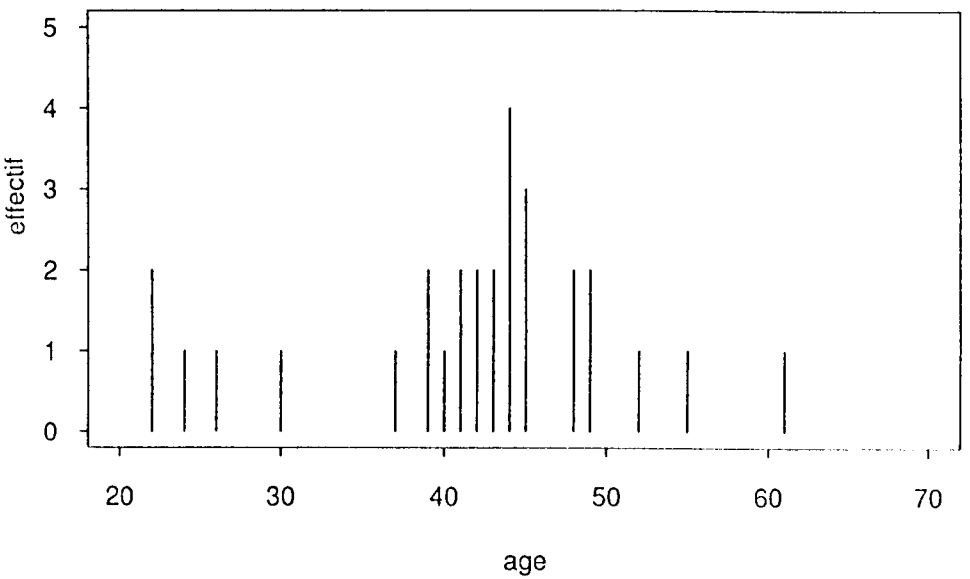
* Les fréquences par classe d'age :

classe d'age	20-29	30-39	40-49	50-59	60-69
effectif	4	4	18	2	1
fréquence	4 / 29	4 / 29	18 / 29	2 / 29	1 / 29
	0.14	0.14	0.62	0.07	0.03

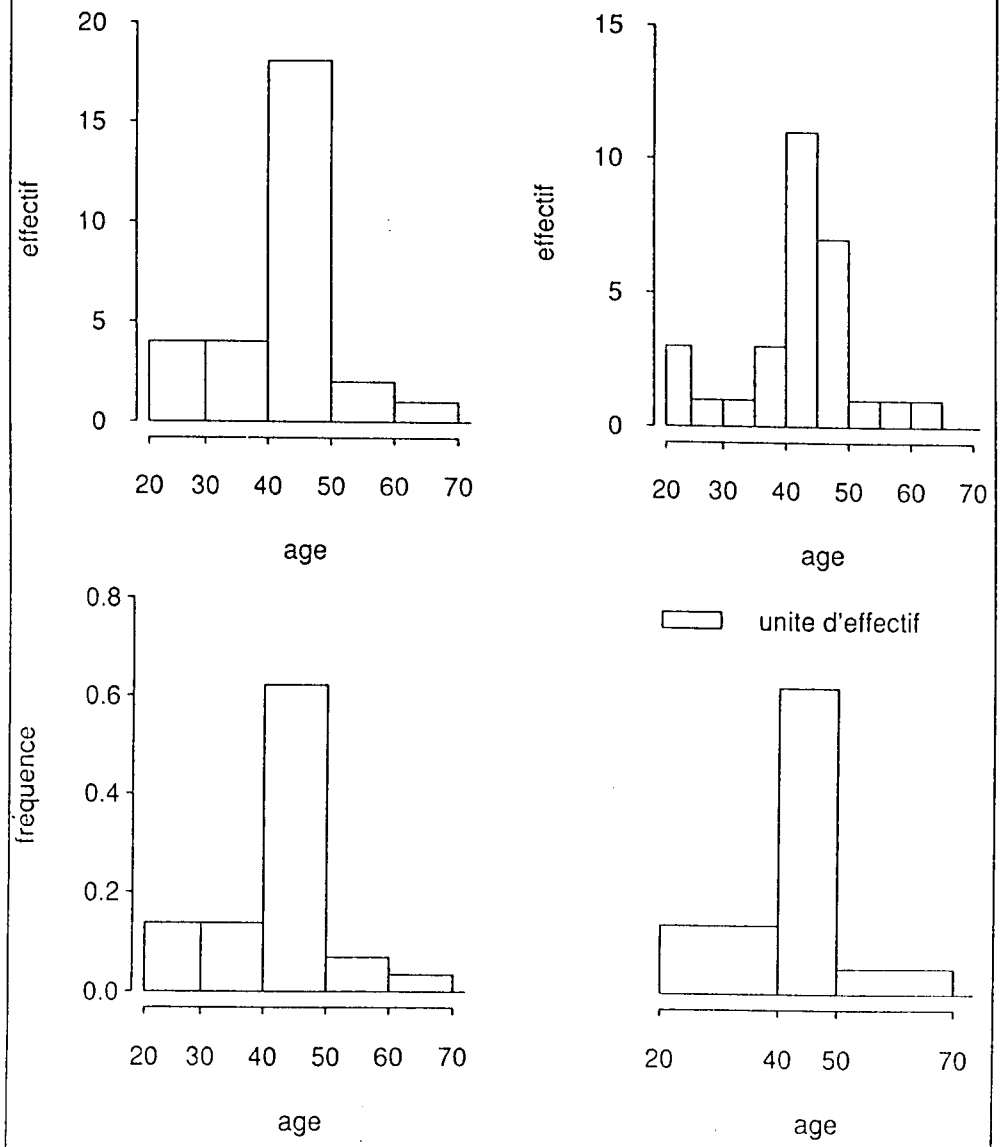
* Les effectifs par age :

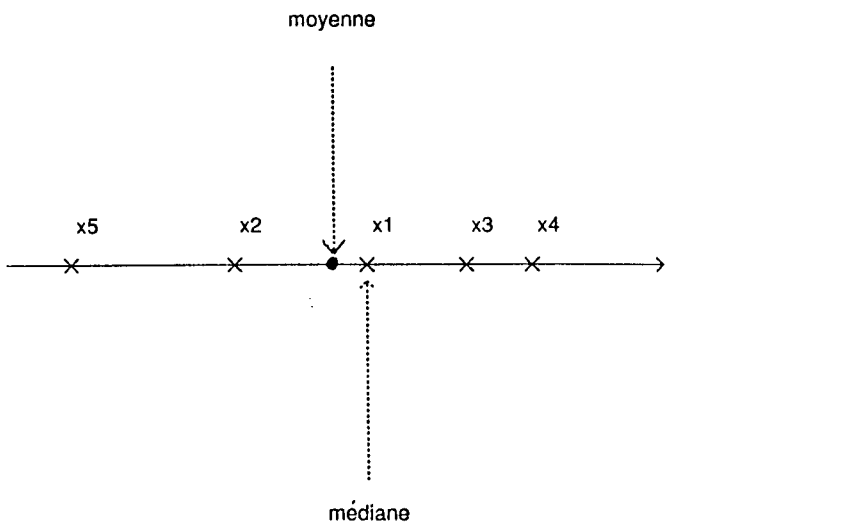
age	:	22	24	26	30	37	39	40	41	42	43	44	45	48	49	52	55	61
effectif	:	2	1	1	1	1	2	1	2	2	2	4	3	2	2	1	1	1

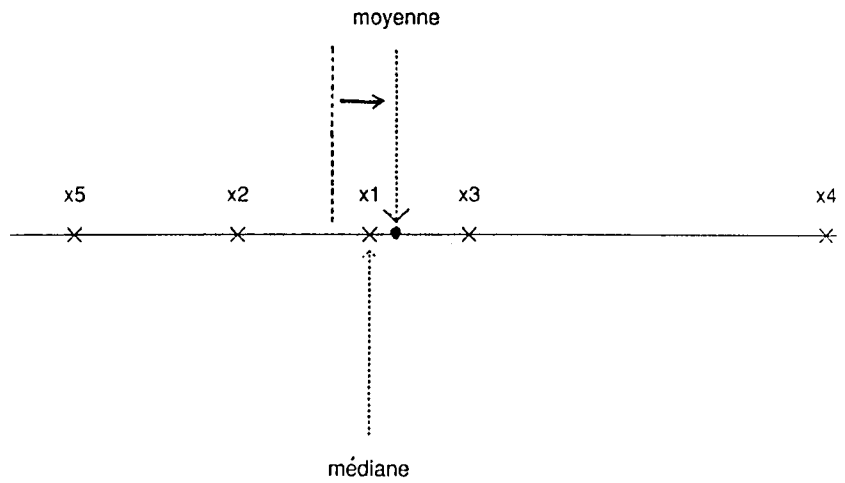
* Le diagramme en batons :



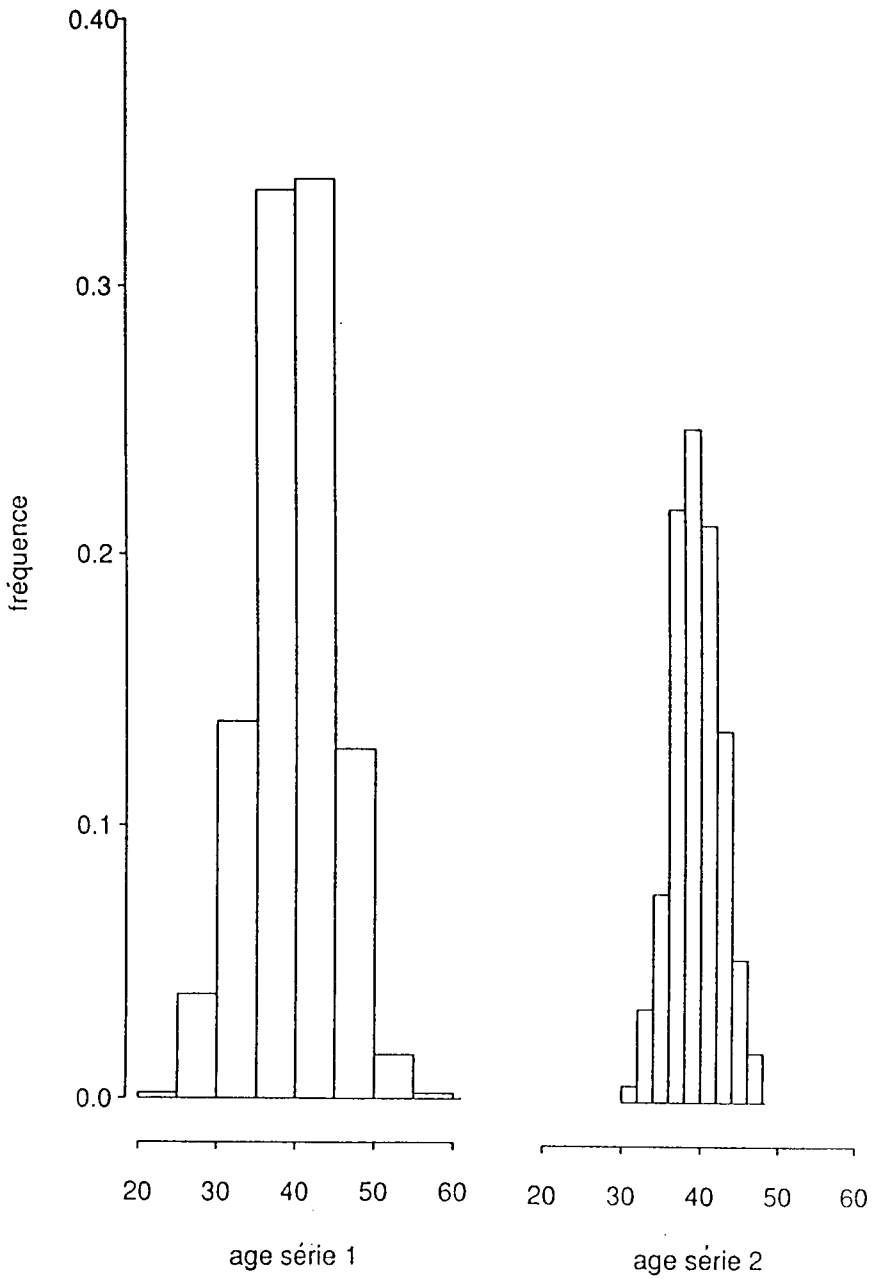
Histogrammes ages techniciens INRA Antilles-Guyane



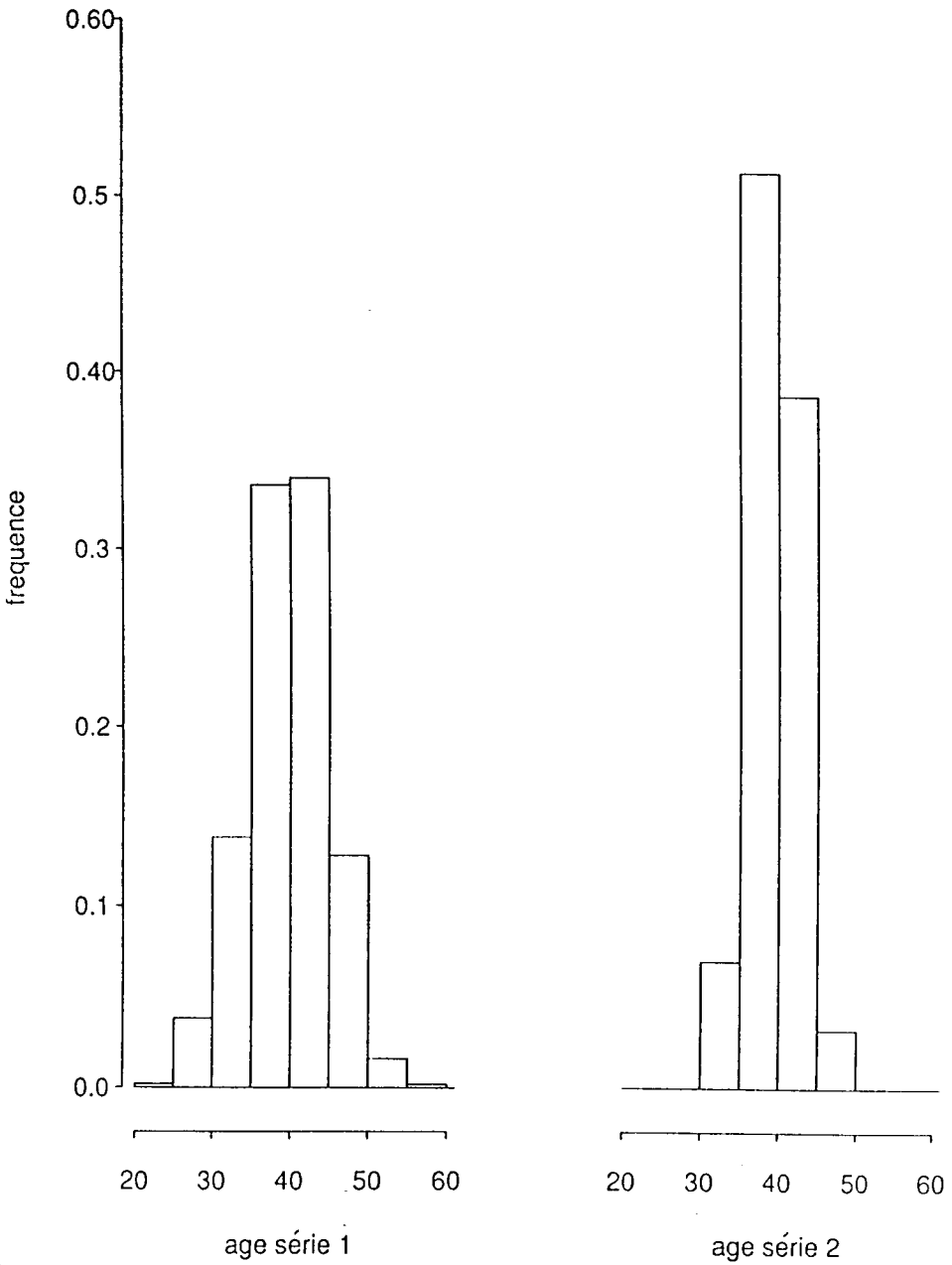

moyenne
x5
x2
x1
x3
x4
médiane

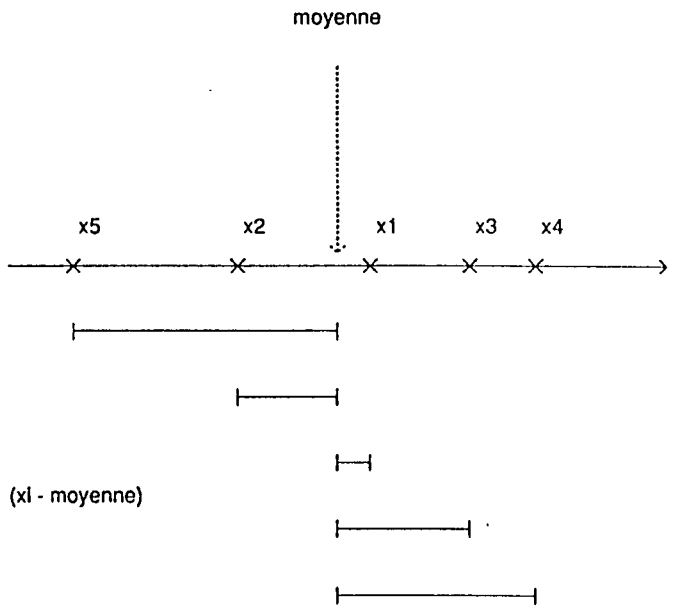

moyenne
x5
x2
x1
x3
x4
médiane

Histogrammes avec mème médiane et moyenne



Histogrammes avec mème médiane et moyenne





On dispose de 5 "individus", sur lesquels on a mesuré la valeur d'une variable X; on dispose donc de 5 valeurs, que l'on peut représenter graphiquement par 5 points sur un axe.

X_i pour l'individu i, avec $1 \leq i \leq 5$.

Un paramètre de position semble insuffisant pour résumer efficacement une série de valeur; on cherche donc un paramètre de dispersion, lié d'abord à la moyenne.

La première idée est d'utiliser les segments :

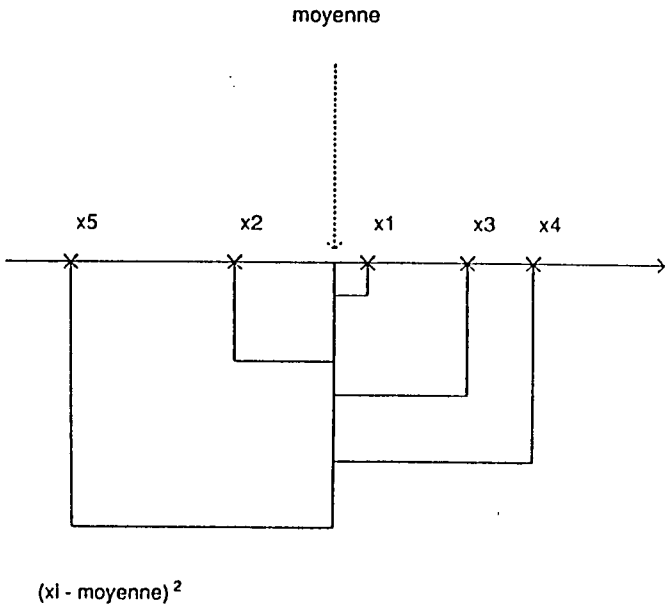
$$(X_i - \text{moyenne}) \Longrightarrow \sum_{i=1}^{n} (X_i - \text{moyenne}) = 0$$

soit 5 segments ici, et d'en faire la somme par exemple, puis la moyenne; or cette somme des écarts à la moyenne est nulle.

L'autre idée qui vient alors à l'esprit est d'utiliser les carrés de ces mêmes segments :

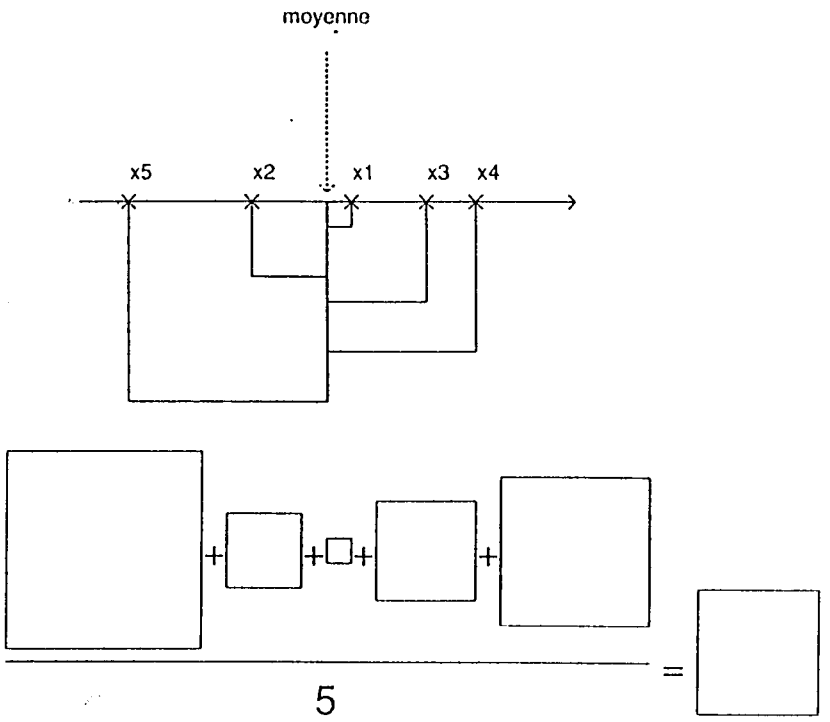
$$(X_i - \text{moyenne})^2$$

Or on peut facilement représenter ces "carrés" de façon géométrique, par des figures carrées justement, dont le côté est égal à $(X_i - \text{moyenne})$.



Transparent 10 :

On a donc représenté les carrés géométriques obtenus à partir des points, et de la longueur des segments correspondants.



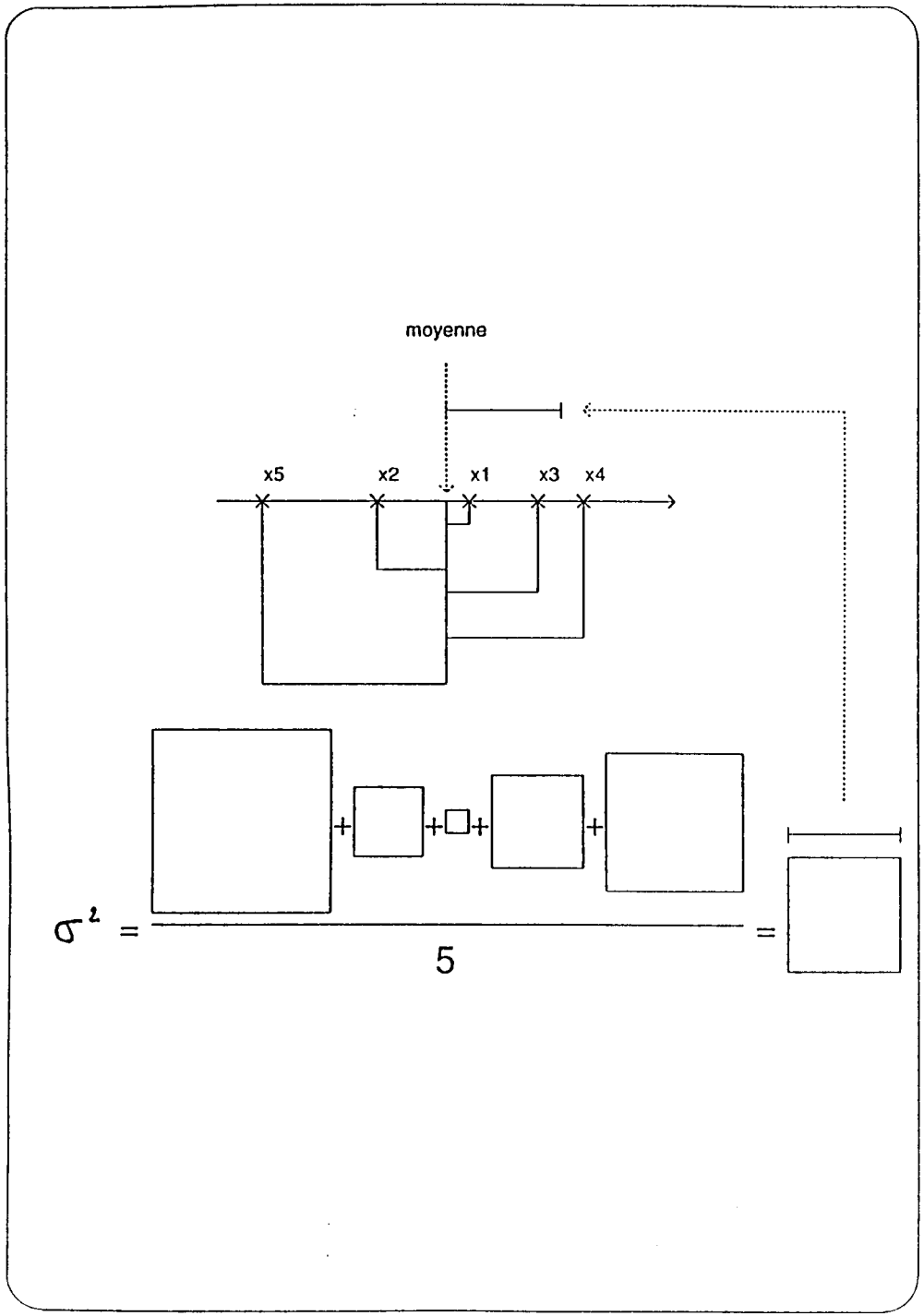
Transparent 11 :

Lidée suivante est donc de faire la somme de ces surfaces, puis d'obtenir une surface moyenne, que l'on peut facilement mettre sous la forme d'une surface carrée.

Cette surface moyenne est la variance.

On voit tout de suite, que cette surface, n'est pas de la même dimension que les données initiales, et que l'on ne peut pas la superposer sur l'axe des valeurs X_i.

Or on vient de voir que l'on a pu mettre cette surface moyenne, sous la forme elle aussi d'un carré; le côté du carré correspond donc à un segment, et peut donc lui, être facilement superposé sur l'axe des valeurs X_i, donnant par la même une idée de la dispersion.



Ce segment, côté du carré de la surface moyenne, est appelé l'écart-type; c'est donc la racine carrée de la variance.

Il est bien de la même dimension que les données initiales.

$\Longrightarrow$ la formule de calcul de la variance, en utilisant le signe de sommation avec un indice :

$$\text{variance} = \frac{1}{n} \cdot \sum_{i=1}^{n} (Xi - X)^2$$

$$\text{écart-type} = \sqrt{\text{variance}}$$

étendue

50 % de l'effectif | 50 % de l'effectif

min. | médiane | max.

étendue

50 % | médiane | 50 %

min. | Q1 | Q2 | Q3 | max.

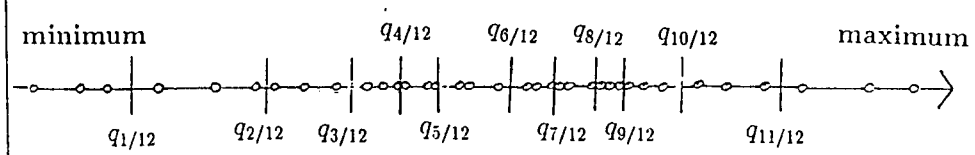
25 % | 25 % | 25 % | 25 %

IQR

IQR = Q3 - Q1

LES QUANTILES

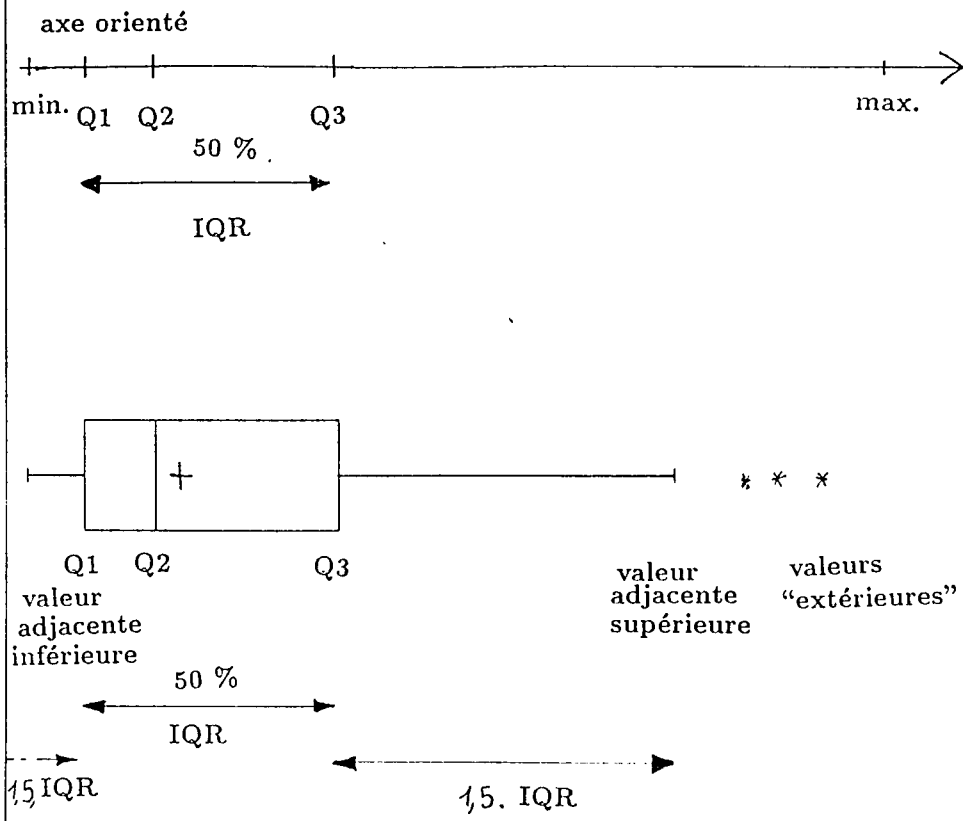
exemple : 12 parties d'effectifs égaux



$q_{7/12}$ est une valeur qui partage les données réordonnées dans l'ordre croissant en $7/12$ de l'effectif à gauche et $3/12$ de l'effectif à droite.

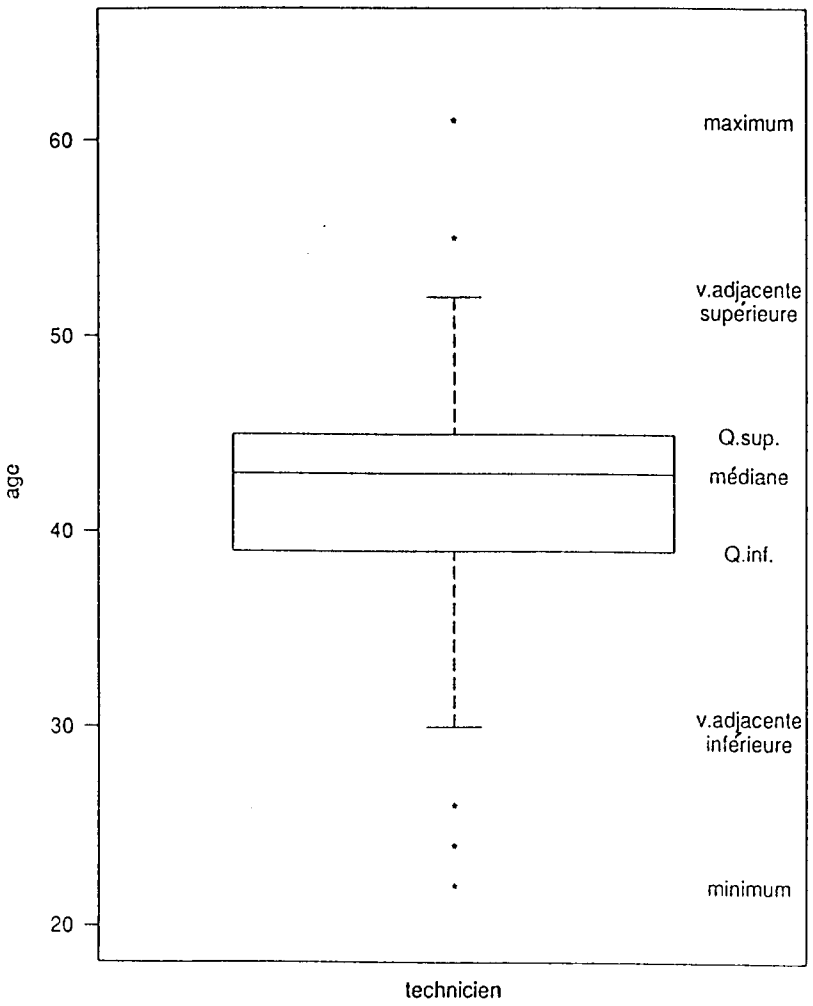
remarque : au lieu de découper en 12 parties on pourrait découper en un autre nombre.

Boite a pattes

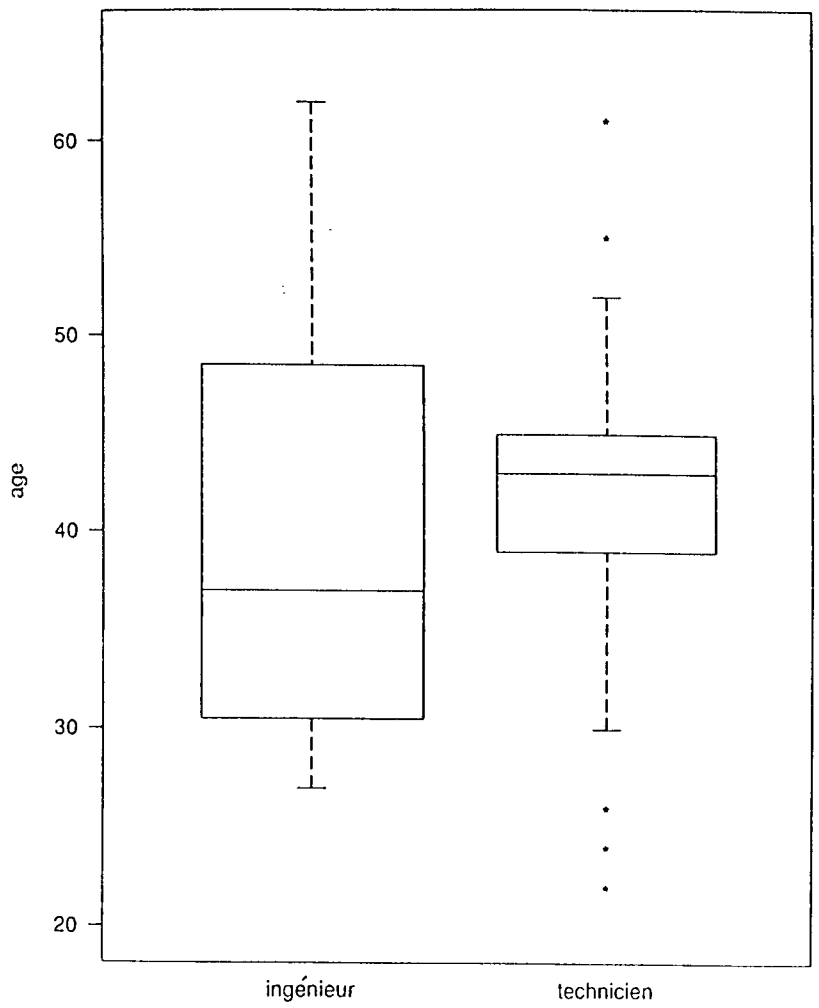


remarque : "+" pour la moyenne

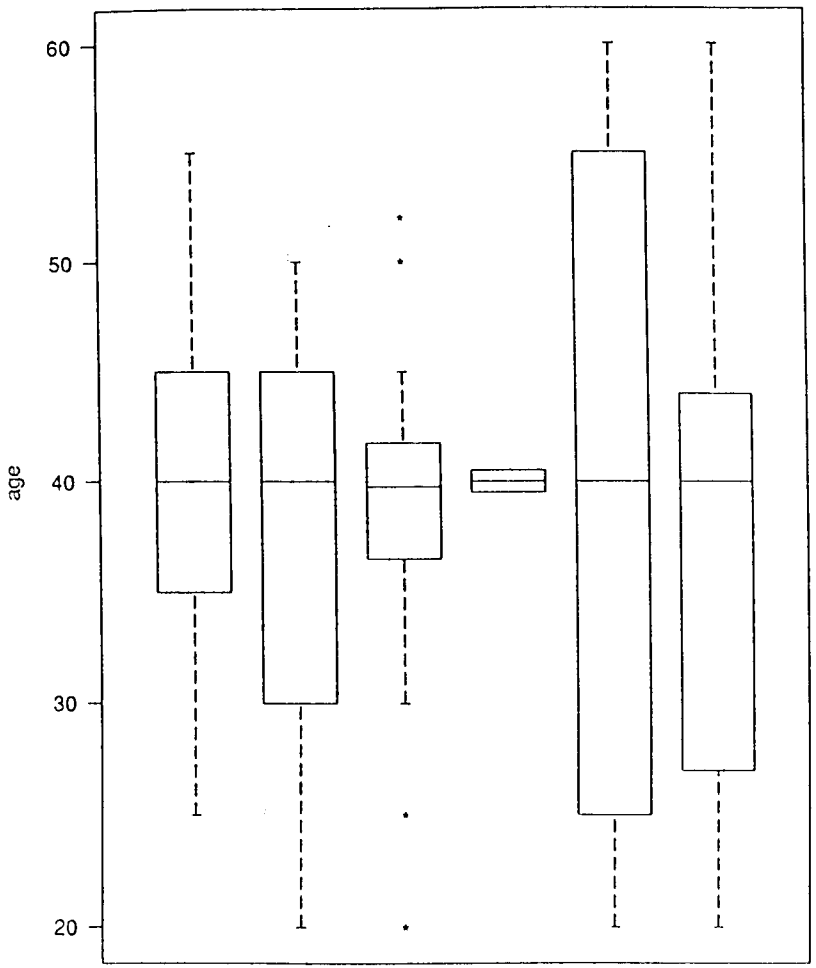
Boite à pattes : age / centre INRA Antilles-Guyane



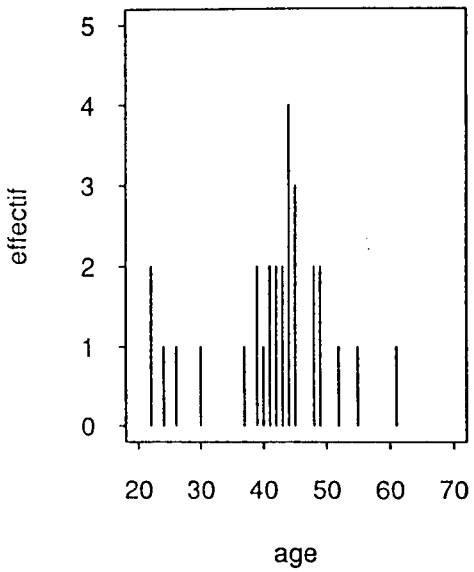
Boites à pattes : age / centre INRA Antilles-Guyane

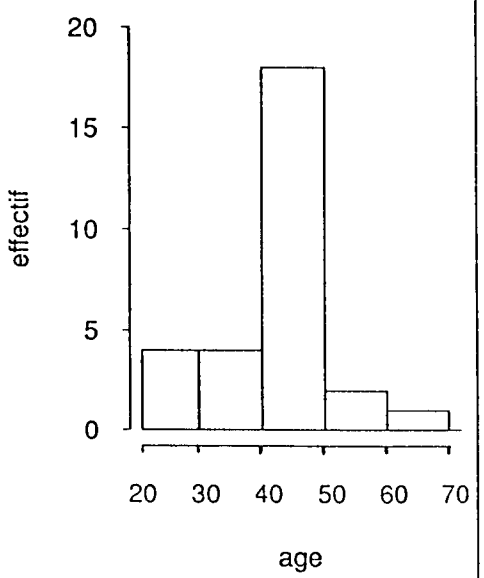


Boites à pattes



Age techniciens INRA Antilles-Guyane



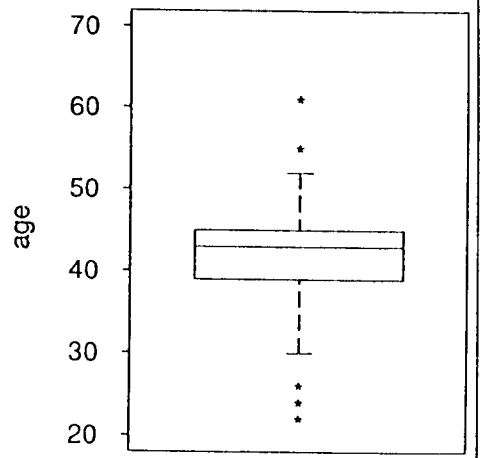



```
N = 29   Median = 43
Quartiles = 39, 45


Low:   22  22  24  26

   3 : 0
   3 : 799
   4 : 01122334444
   4 : 5558899
   5 : 2
   5 : 5

High:  61
```



LES OUTILS

ordre synthétique croissant	solutions numériques	solutions graphiques	solutions semi-graphiques
↓	classement		
	effectifs par age	diagramme en batons	
	effectifs par classe	histogramme	branchage
	fréquences		
	paramètres de position et de dispersion :		boite à pattes
	- minimum, maximum, médiane, moyenne		
	- étendue, écart inter-quartile, variance, écart-type		