

# Statistiques

## Exercice 1

4 candidats se présentent à une élection. Les résultats sont donnés dans le tableau :

| Candidat                   | A      | B      | C      | D     |
|----------------------------|--------|--------|--------|-------|
| Nombre de voix obtenues    | 51 210 | 43 821 | 23 212 | 8 597 |
| Fréquence (en pourcentage) |        |        |        |       |

Compléter le tableau en donnant la fréquence en pourcentage pour chacun des candidats. Représenter ces données par un diagramme circulaire.

### Rappel

La fréquence est donnée par le quotient :  $\frac{\text{effectif}}{\text{effectif total}}$  (l'effectif correspond ici au nombre de voix)

## Exercice 2

Le tableau ci-dessous donne le nombre moyen de longs métrages réalisés par les dix plus gros producteurs mondiaux, entre 1990 et 1995 (source Unesco).

| Brésil | Chine | Etats-Unis | France | Inde | Italie | Japon | Philippines | Royaume-Uni | Thaïlande |
|--------|-------|------------|--------|------|--------|-------|-------------|-------------|-----------|
| 86     | 154   | 420        | 141    | 838  | 96     | 251   | 456         | 78          | 194       |

Représenter ces données par un diagramme à barres (ou diagramme en bâtons).

### Rappel

Dans un diagramme à barres (ou en bâtons), la hauteur de la barre représente l'effectif.

## Exercice 3

On a relevé le prix de vente d'un CD et le nombre de CD vendus chez différents fournisseurs. Les résultats sont donnés dans le tableau suivant :

|                        |    |    |    |    |    |
|------------------------|----|----|----|----|----|
| Prix de vente en euros | 15 | 16 | 17 | 18 | 19 |
| Nombre de CD vendus    | 97 | 34 | 43 | 20 | 6  |

Donner le prix de vente moyen p de ce CD.

## Définitions

On considère une série statistique donnée par le tableau suivant :

|                  |       |       |       |     |     |     |       |
|------------------|-------|-------|-------|-----|-----|-----|-------|
| Valeur : $x_i$   | $x_1$ | $x_2$ | $x_3$ | ... | ... | ... | $x_p$ |
| Effectif : $n_i$ | $n_1$ | $n_2$ | $n_3$ | ... | ... | ... | $n_p$ |

La **moyenne** de cette série est :  $\bar{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + n_3 \times x_3 + \dots + n_p \times x_p}{n_1 + n_2 + n_3 + \dots + n_p}$

La moyenne permet d'avoir une idée du "centre" de la série, c'est une mesure de **tendance centrale**.

La **variance** de cette série est :  $V = \frac{n_1 \times (x_1 - \bar{x})^2 + n_2 \times (x_2 - \bar{x})^2 + n_3 \times (x_3 - \bar{x})^2 + \dots + n_p \times (x_p - \bar{x})^2}{n_1 + n_2 + n_3 + \dots + n_p}$

L'**écart-type** de cette série est :  $\sigma = \sqrt{V}$

L'écart-type permet d'avoir une idée de la façon dont les valeurs de la série s'écartent par rapport à la moyenne. C'est une **mesure de dispersion**.

Un écart-type faible correspond à une série concentrée autour de la moyenne.

Les calculs de moyenne, de variance et d'écart-type sont, pour des séries prenant un grand nombre de valeurs, des calculs compliqués. Les calculatrices utilisées en mode statistique et les ordinateurs rendent alors de grands services pour ces calculs.

### Exercice 4

A un test de mathématique, noté sur 10 points, les élèves de 2 classes ont obtenu les notes suivantes :

Classe A : 4 5 3 4 5 4 3 4 6 7 7 6 3 6 4 5 8 7 6 5 4 9 3 5 6 4 1 5 4 7 5

Classe B : 2 4 3 2 5 6 5 1 3 8 3 6 8 7 2 3 7 10 5 7 7 2 10 0 1 0 4 7 8 9 10

1°) Compléter les tableaux suivants :

Pour la classe A :

|                  |   |   |   |   |   |   |   |   |   |   |    |
|------------------|---|---|---|---|---|---|---|---|---|---|----|
| Note : $x_i$     | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Effectif : $n_i$ |   |   |   |   |   |   |   |   |   |   |    |

Pour la classe B :

|                  |   |   |   |   |   |   |   |   |   |   |    |
|------------------|---|---|---|---|---|---|---|---|---|---|----|
| Note : $x_i$     | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Effectif : $n_i$ |   |   |   |   |   |   |   |   |   |   |    |

2°) Représenter pour chaque classe la série de notes par un diagramme en bâtons.

3°) Calculer la moyenne de chacune des classes.

4°) Compléter pour chacune des classes le tableau suivant :

Pour la classe A :

|                                |   |   |   |   |   |   |   |   |   |   |    |
|--------------------------------|---|---|---|---|---|---|---|---|---|---|----|
| Note : $x_i$                   | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Effectif : $n_i$               |   |   |   |   |   |   |   |   |   |   |    |
| $(x_i - \bar{x})^2$            |   |   |   |   |   |   |   |   |   |   |    |
| $n_i \times (x_i - \bar{x})^2$ |   |   |   |   |   |   |   |   |   |   |    |

Pour la classe B :

|                                |   |   |   |   |   |   |   |   |   |   |    |
|--------------------------------|---|---|---|---|---|---|---|---|---|---|----|
| Note : $x_i$                   | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Effectif : $n_i$               |   |   |   |   |   |   |   |   |   |   |    |
| $(x_i - \bar{x})^2$            |   |   |   |   |   |   |   |   |   |   |    |
| $n_i \times (x_i - \bar{x})^2$ |   |   |   |   |   |   |   |   |   |   |    |

Calculer la variance et l'écart-type de chacune des séries.

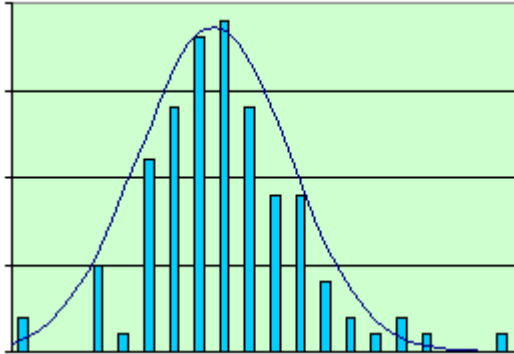
5°) A partir des résultats précédents, faire une comparaison des deux classes.

6°) Pour chacune des classes, donner la note du 16<sup>ème</sup> élève

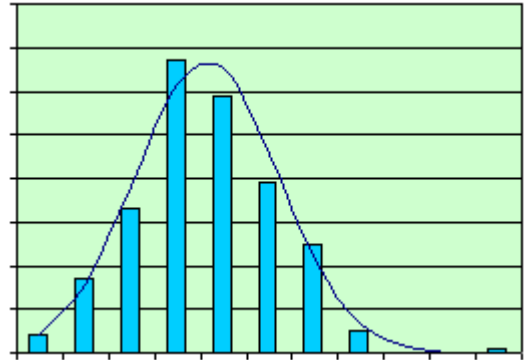
## Remarque

Le recueil d'un grand nombre de données dans certains domaines (données industrielles, données biologiques ...) conduit souvent à des diagrammes (diagrammes à barres, histogrammes) ayant sensiblement la même forme dite "en cloche".

Exemples : sur chacun des diagrammes a été superposé la courbe dite "courbe en cloche"



Fréquence des battements cardiaques



Circonférence du biceps

Dans un tel cas, les données recueillies sont qualifiées de «**données gaussiennes**».

Si l'on calcule pour de telles données la moyenne  $\mu$  et l'écart-type  $\sigma$ , on peut noter que :

La série est à peu près symétrique autour de la moyenne  $\mu$ .

Environ 95% des données se trouvent dans l'intervalle  $[\mu - 2\sigma ; \mu + 2\sigma]$ .

Cet intervalle est appelé **plage de normalité**.

Environ 99% des données se trouvent dans l'intervalle  $[\mu - 3\sigma ; \mu + 3\sigma]$ .

## Exemple

Les études statistiques portant sur un grand nombre d'enfants ont conduit à des données gaussiennes.

Ces études ont permis d'établir des courbes de croissance (taille, poids, périmètre crânien ...) que l'on trouve sur le "Carnet de santé" remis aux parents à la naissance d'un enfant.

Ces courbes sont construites à partir de la plage de normalité  $[\mu - 2\sigma ; \mu + 2\sigma]$ .

Elles permettent aux parents et aux médecins de surveiller la croissance d'un enfant.

## Extraits d'un Carnet de santé

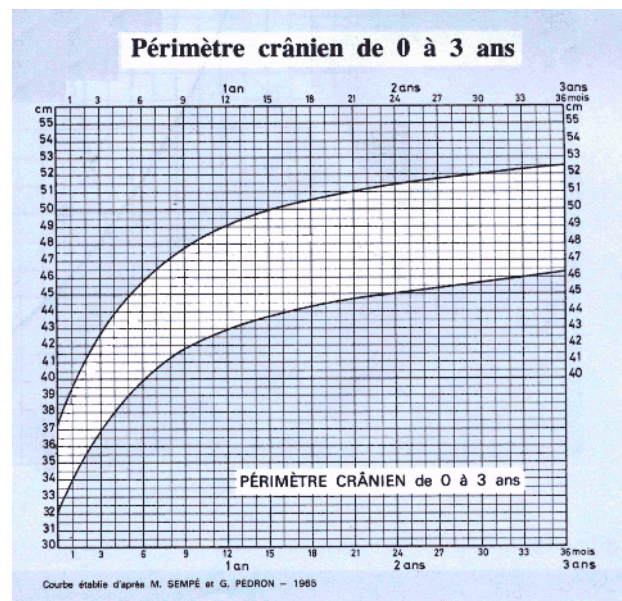
### Courbes de croissance

Une croissance physique rapide est une caractéristique essentielle des premières années de la vie.

Les courbes de taille, de poids et de périmètre crânien permettent de la surveiller.

Faites-les établir soigneusement ou établissez-les vous-même.

Sur ces courbes, les zones blanches correspondent à 2 écarts-type de part et d'autre de la moyenne.



## Définition

On considère une série, dont les valeurs sont **ordonnées** (rangées dans l'ordre croissant).

Si la série comporte un nombre pair  $2n$  de termes, la **médiane** de cette série est la demi-somme de la valeur du terme de rang  $n$  et de la valeur du terme de rang  $n+1$ .

Si la série comporte un nombre impair  $2n+1$  de termes, la **médiane** de cette série est la valeur du terme de rang  $n+1$  (c'est-à-dire le terme partageant la série en deux groupes de même effectif).

### *Exemple*

La série : 101 101 105 **105 107** 108 108 110 a un nombre pair (8) de termes.

La médiane de cette série est la demi-somme du 4<sup>ème</sup> et du 5<sup>ème</sup> terme, c'est-à-dire  $\frac{105 + 107}{2} = 106$ .

La série : 87 88 89 89 90 **92** 92 93 97 99 99 a un nombre impair (11) de termes.

La médiane de cette série est la valeur du 6<sup>ème</sup> terme, c'est-à-dire 92.

## Définition

On appelle **premier quartile** d'une série la plus petite valeur  $q$  des termes de la série pour laquelle au moins un quart (25%) des données sont inférieures ou égales à  $q$ .

On appelle **troisième quartile** d'une série la plus petite valeur  $q'$  des termes de la série pour laquelle au moins trois quarts (75%) des données sont inférieures ou égales à  $q'$ .

On appelle **intervalle interquartile** l'intervalle  $[q ; q']$ .

On appelle **écart interquartile** l'amplitude de l'intervalle  $[q ; q']$ , c'est-à-dire le nombre  $q' - q$ .

### *Exemple*

La recherche des quartiles sera plus facile si les termes de la suite sont ordonnés.

La série 11, 12, 12, **13**, 15, 16, 16, 17, 17, 18, **19**, 20, 22, 23 a 14 termes.

Un quart (25%) des données correspond à :  $14 \times 0,25 = 3,5$ .

Le premier quartile est alors, par définition, la plus petite valeur  $q$  pour laquelle les valeurs de 4 termes de la série sont inférieurs ou égales à  $q$ .

Le premier quartile est donc la valeur du 4<sup>ème</sup> terme de la série c'est-à-dire 13.

Trois quarts (75%) des données correspondent à :  $14 \times 0,75 = 10,5$ .

Le troisième quartile est alors, par définition, la plus petite valeur  $q'$  pour laquelle les valeurs de 11 termes de la série sont inférieurs ou égales à  $q'$ .

Le troisième quartile est donc la valeur du 11<sup>ème</sup> terme de la série c'est-à-dire 19.

L'intervalle interquartile est  $[13 ; 19]$ . L'écart interquartile est  $19 - 13 = 6$ .

## Définition

On appelle **premier décile** d'une série la plus petite valeur  $d$  des termes de la série pour laquelle au moins un dixième (10%) des données sont inférieures ou égales à  $d$ .

On appelle **neuvième décile** d'une série la plus petite valeur  $d'$  des termes de la série pour laquelle au moins neuf dixièmes (90%) des données sont inférieures ou égales à  $d'$ .

On appelle **intervalle interdécile** l'intervalle  $[d ; d']$ .

On appelle **écart interdécile** l'amplitude de l'intervalle  $[d ; d']$ , c'est-à-dire le nombre  $d' - d$ .

### *Exemple*

La recherche des quartiles sera plus facile si les termes de la suite sont ordonnés.

La série 4, 5, **5**, 5, 6, 6, 7, 8, 8, 9, 9, 9, 9, 10, 10, 11, 11, 11, 12, 13, 13, 13, 14, 14, **15**, 15, 17 a 27 termes.

Un dixième (10%) des données correspond à :  $27 \times 0,10 = 2,7$ .

Le premier décile est alors, par définition, la plus petite valeur  $d$  pour laquelle les valeurs de 3 termes de la série sont inférieurs ou égales à  $d$ .

Le premier décile est donc la valeur du 3<sup>ème</sup> terme de la série c'est-à-dire 5.

Neuf dixièmes (90%) des données correspondent à :  $27 \times 0,9 = 24,3$ .

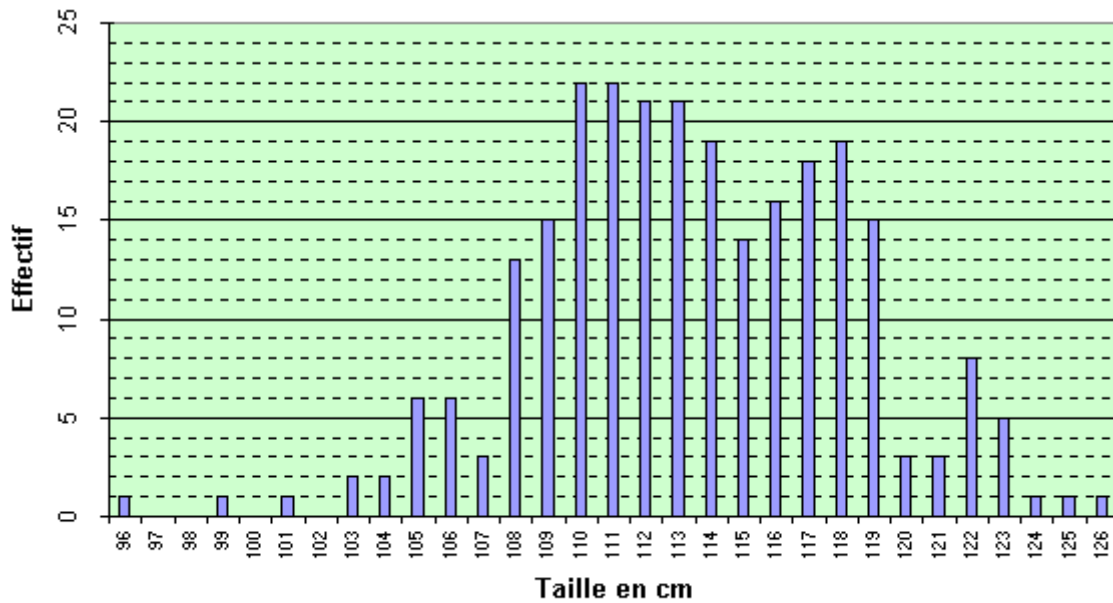
Le neuvième décile est alors, par définition, la plus petite valeur  $d'$  pour laquelle les valeurs de 25 termes de la série sont inférieurs ou égales à  $d'$ .

Le neuvième décile est donc la valeur du 25<sup>ème</sup> terme c'est-à-dire 15.

L'intervalle interdécile est  $[5 ; 15]$ . L'écart interdécile est  $15 - 5 = 10$ .

### Exercice 5

Le graphique ci-dessous représente la série des tailles en cm pour des enfants de 68 mois.  
(Source : Pr. M. Tauber, CHU Toulouse)



1°) En utilisant ce graphique, compléter le tableau suivant

|                 |    |    |    |    |     |     |     |     |     |     |     |     |     |     |     |     |
|-----------------|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Taille          | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 |
| Effectif        |    |    |    |    |     |     |     |     |     |     |     |     |     |     |     |     |
| Effectif cumulé |    |    |    |    |     |     |     |     |     |     |     |     |     |     |     |     |

|                 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
|-----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| Taille          | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 |  |
| Effectif        |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| Effectif cumulé |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |

2°) Vérifier que l'effectif total de la série est 259.

3°) Déterminer la médiane de cette série.

4°) Déterminer le 1<sup>er</sup> quartile et le 3<sup>ème</sup> quartile de cette série.  
Donner l'intervalle interquartile et l'écart interquartile.

5°) Déterminer le 1<sup>er</sup> décile et le 9<sup>ème</sup> décile de cette série.  
Donner l'intervalle interdécile et l'écart interdécile.

## Construction d'un diagramme en boîte

Ce type de diagramme est aussi appelé diagramme de Tuckey, boîte à moustaches ou boîte à pattes. Il utilise le 1<sup>er</sup> et le 3<sup>ème</sup> quartile, les valeurs extrêmes, le 1<sup>er</sup> et le 9<sup>ème</sup> décile et éventuellement la médiane d'une série.

La construction ci-contre est faite pour la série de l'exercice 5 (tailles en cm pour des enfants de 68 mois). Cette série était caractérisée par :

médiane : 113

1<sup>er</sup> quartile : 110

3<sup>ème</sup> quartile : 117

1<sup>er</sup> décile : 108

9<sup>ème</sup> décile : 119

On choisit une graduation verticale permettant de représenter les différentes valeurs de la série.

On pourra par exemple graduer entre 90 et 130.

(Si certaines valeurs sont manifestement hors normes, on n'en tiendra pas compte.)

Le "corps" du diagramme, c'est-à-dire la "boîte" est formée d'un rectangle ayant pour extrémité inférieure le 1<sup>er</sup> quartile et pour extrémité supérieure le 3<sup>ème</sup> quartile. A l'intérieur de ce rectangle on pourra tracer un segment représentant la médiane.

La largeur du rectangle n'est pas fixée, elle sera choisie de façon à obtenir un graphique "harmonieux".

Ce rectangle représente les données contenues dans l'intervalle interquartile.

On repère ensuite les hauteurs correspondant au 1<sup>er</sup> et au 9<sup>ème</sup> décile, et on trace deux pattes représentant les données contenues dans l'intervalle interdécile. (la largeur des pattes n'a pas d'importance).

On peut ensuite terminer le graphique, en faisant figurer par des points les données qui sont en dehors de l'intervalle interdécile.

Si certaines données, sont manifestement très éloignées, on ne les représentera pas, mais on écrira leurs valeurs au dessous du diagramme.

### Remarques

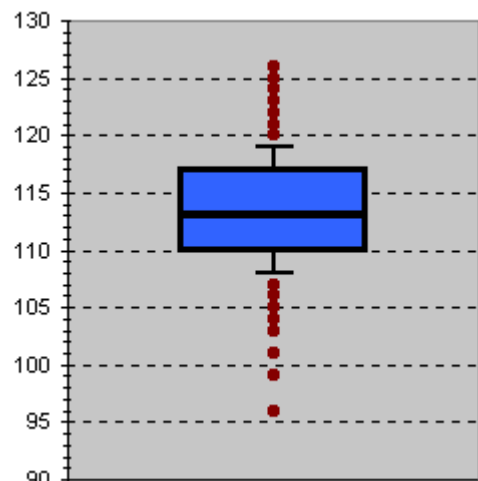
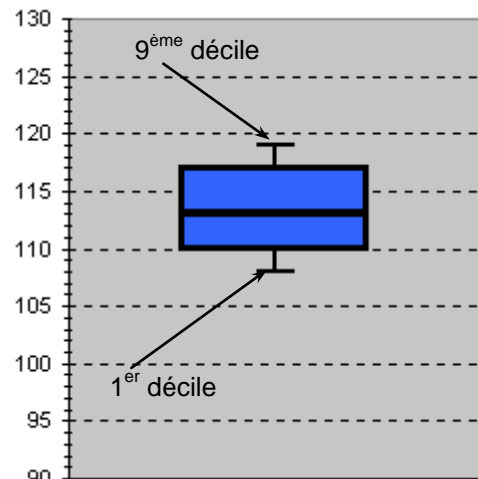
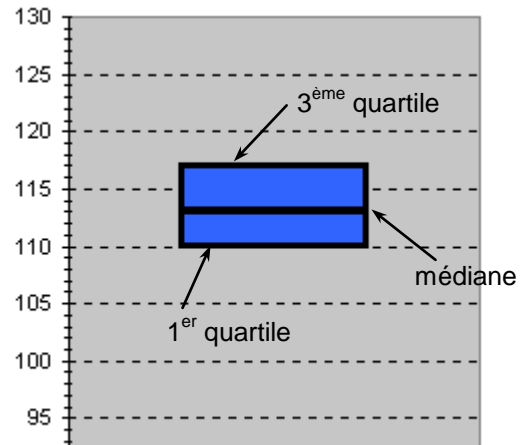
Une boîte avec des "pattes" courtes indique que la série est assez concentrée autour de sa médiane.

Au contraire des "pattes" longues indique que la série est assez dispersée.

Un des avantages de cette représentation, est qu'elle nécessite très peu de calculs.

La représentation peut aussi se faire horizontalement, la graduation se trouvant alors sur l'axe horizontal, d'où l'appellation de "boîte à moustaches".

Le graphique est parfois fait en dessinant des pattes correspondant au 1<sup>er</sup> et au 99<sup>ème</sup> centile, ou même aux valeurs extrêmes.



### Exercice 6

On a relevé les taux de cholestérol de 200 employés des hôpitaux de Los Angeles victimes de maladie cardiaque.

270 320 310 250 250 300 250 270 270 190 200 260 260 330 280 280 250 240 330 250  
250 230 270 230 240 200 210 240 210 270 210 130 220 290 220 200 220 330 270 260  
300 150 350 230 210 250 230 250 220 310 180 280 300 290 190 220 250 230 220 220  
200 230 230 220 360 290 270 240 170 190 280 250 270 280 300 240 210 260 190 250  
260 240 290 230 270 250 360 190 180 260 350 180 250 280 270 240 220 230 220 220  
240 300 280 220 240 230 300 280 220 240 190 170 320 150 320 200 210 270 230 270  
270 250 230 290 220 220 310 260 260 230 250 300 200 160 230 270 280 180 300 270  
270 270 250 250 240 250 280 210 350 200 230 210 240 200 210 330 200 260 310 160  
290 300 320 340 350 170 290 200 140 310 260 260 240 220 180 320 220 300 310 250  
240 300 330 240 300 330 200 190 300 240 210 240 200 260 170 270 250 250 270 190

1°) Organiser ces données dans un tableau faisant apparaître les effectifs.

2°) Déterminer la médiane, le 1<sup>er</sup> quartile, le 3<sup>ème</sup> quartile, le 1<sup>er</sup> décile et le 9<sup>ème</sup> décile de la série.

3°) Construire un diagramme en boîte pour représenter la série.

# Statistiques

## Utilisation du tableur

Les exercices correspondent aux exercices de mathématiques.

### Exercice 1

Faire le tableau de valeurs, écrire les formules donnant les fréquences. Les résultats seront affichés en pourcentage avec 1 chiffre après la virgule.

Faire le diagramme circulaire. Vous enregistrerez votre travail sous le nom st01.xls

### Exercice 2

Faire le tableau et le diagramme à barres correspondant.

Vous enregistrerez votre travail sous le nom st02.xls

### Exercice 3

Faire le tableau et écrire la formule permettant d'obtenir le prix moyen d'un CD.

Vous enregistrerez votre travail sous le nom st03.xls

### Exercice 4

Ouvrir le classeur st04.xls

Sur la feuille "Classe A", Compléter la plage B2:L2, puis la plage B3:L3.

Le diagramme en bâtons a déjà été prévu et doit apparaître alors automatiquement.

Écrire dans la cellule B7 la formule pour le calcul de la moyenne.

Compléter ensuite les deux dernières lignes du tableau, puis écrire dans les cellules B8 et B9 les formules donnant la variance et l'écart-type. Les résultats seront affichés avec à 0,01 près.

La fonction racine carrée est obtenue en écrivant = racine( )

Faire de même sur la feuille "Classe B", et faire le diagramme en bâtons qui n'a pas été fait.

Sur la feuille "Données", vous trouverez les données brutes concernant la classe A et la classe B.

Calculer pour chaque classe la moyenne, la variance et l'écart-type en utilisant les fonctions : moyenne( : )  
var.p( : )    ecartypep( : ). Faire afficher les résultats à  $10^{-2}$  près et constater qu'ils sont corrects.

### Exercice 5

Ouvrir le classeur st05.xls

Sur la feuille Diagrammes, vous trouverez le diagramme à barres représentant la taille en cm pour des enfants de 68 mois. Ce diagramme comporte trois erreurs dans les effectifs des catégories 99 ; 110 et 122.

Les différents paramètres, Effectif Total, Médiane etc... sont donc faux, ainsi que le diagramme en boîte se trouvant plus bas. Corrigez les erreurs en agissant directement sur les barres du diagramme à barres.

Vérifiez alors que les divers paramètres de la série sont corrects ainsi que le diagramme en boîte.

### Exercice 6

Ouvrir le classeur st06.xls

Compléter la ligne "Effectif" par les valeurs trouvées, puis la ligne "Effectif cumulé" par les formules appropriées.

Vérifier que les différents paramètres de la série sont corrects ainsi que le diagramme en boîte.

### Exercice 7

Ouvrir le classeur st07.xls

Écrire dans les cellules B8 et B9 les formules donnant le nombre total de personnes infectées et le nombre total de personnes non infectées.

Faire, pour les personnes infectées par le VIH, le diagramme à barres représentant leur nombre de partenaires.

### Exercice 8

Ouvrir le classeur st08.xls

Compléter par des formules l'arbre schématisé sur la feuille de calcul.

Écrire dans les cellules B16 et B17 les formules donnant la moyenne et l'écart-type de la série. (on pourra utiliser des lignes de calculs intermédiaires).