

# Indicateurs statistiques

1



Les indicateurs statistiques permettent de résumer un grand nombre de données, trop nombreuses pour être « lisibles », afin d'en dégager l'information utile.

**D**u recensement de l'INSEE, on a extrait les indicateurs suivants, concernant la distance moyenne parcourue chaque jour, en kilomètres, par les actifs ayant un emploi hors de leur commune de résidence, pour les 96 départements de France métropolitaine.

Min	$Q_1$	Me	$Q_3$	Max
8,5	13,7	14,9	16,1	23,3

– La distance moyenne dans les Bouches-du-Rhône est 18,1 km par jour. Comment peut-on situer ce département par rapport aux autres ?

## Est-ce que je sais... ?

### 1. Calculer une moyenne pondérée

Au cours d'un trimestre, les notes de maths de Mehdi sont : 8, 11 et 10 en contrôles et 13, 15, 11 et 16 en travaux pratiques. Les contrôles sont comptés coefficient 2 et les travaux pratiques coefficient 1.

Quelle est la moyenne trimestrielle de Mehdi ?

### 2. Interpréter la médiane et les quartiles

D'après les données de l'INSEE concernant le nombre d'habitants des communes françaises (2009), on obtient les indicateurs suivants :

a) Est-il exact que la moitié des communes de France ont moins de 422 habitants ?

b) Interpréter le troisième quartile.

$Q_1$	Me	$Q_3$
191	422	1 041

## Activité 1 Résumer par moyenne et écart type



### Quels repères pour le taux de criminalité ?

Les données suivantes correspondent au « taux de criminalité » (nombre moyen de crimes et délits constatés par la police et la gendarmerie pour 1 000 habitants) en France de 1989 à 2008 (source : Direction centrale de la police judiciaire).

89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08
58,3	61,7	65,8	67	67,5	67,8	63,2	61,1	59,7	60,7	61	64,1	68,5	68,9	66,1	63,2	61,9	60,7	58,1	57,3

### 1 Détermination de la moyenne et de l'écart type à l'aide de la calculatrice

Entrer les taux de criminalité dans la calculatrice, puis afficher l'écran de calcul des statistiques à une variable.

a) Quel est le taux de criminalité moyen pour 1 000 habitants en France durant cette période ?

b) Selon les calculatrices, l'écart type est noté  $\sigma x$  ou  $\sigma n$ .

Quelle est la valeur, à  $10^{-2}$  près, de l'écart type fourni par votre calculatrice ?

```
1-Var Stats
x̄=63.13
Σx=1262.6
Σx²=79961.26
Sx=3.651402038
σx=3.558946473
n=20
```

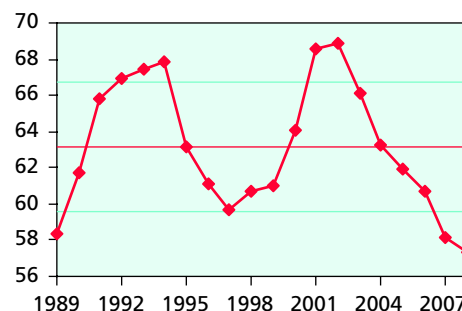
### 2 Interprétation des résultats

Le graphique ci-contre donne une illustration de l'évolution dans le temps du taux de criminalité.

a) Pourquoi peut-on dire que le taux de criminalité moyen est un indicateur de tendance centrale ?

b) À quel indicateur correspond l'écart vertical entre la droite rouge et chaque droite bleue ?

c) Quel est le pourcentage des données situées entre les deux lignes bleues ?

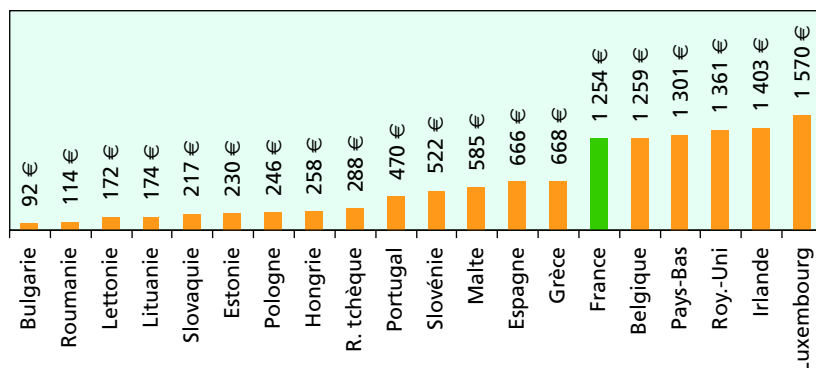


## Activité 2 Résumer par médiane et écart interquartile



### Les inégalités du salaire minimal

Le graphique suivant indique le salaire mensuel brut minimal, en euros, en vigueur (lorsqu'il existe) dans les pays de l'Union européenne début 2007 (source : Eurostat).



### 1 Médiane et quartiles

- Déterminer le salaire minimal médian. Interpréter la réponse.
- Déterminer le premier et le troisième quartile. Quels sont les pays correspondants ? Interpréter la position de la France.

### 2 Écart interquartile

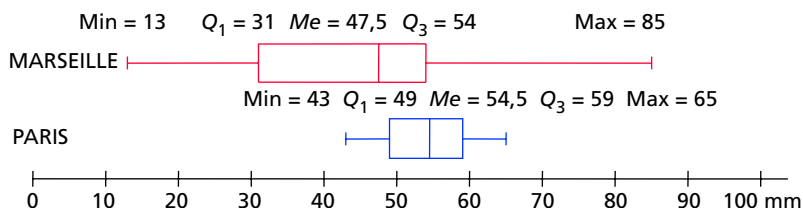
L'écart interquartile est celui qui sépare le premier et le troisième quartile.

- Calculer l'écart interquartile.
- Combien de pays ont-ils un salaire minimal inférieur à l'écart interquartile ?
- Comparer l'écart interquartile des salaires minimaux dans l'Union européenne au salaire minimal en Bulgarie.

## Activité 3 Lire des boîtes à moustaches

### Comment pleut-il à Paris et à Marseille ?

Les diagrammes ci-dessous, nommés boîtes à moustaches, correspondent aux précipitations mensuelles moyennes, en millimètres, à Paris et à Marseille.



- Les « moustaches » correspondent aux valeurs extrêmes. Combien de millimètres tombe-t-il en moyenne à Paris durant le mois le plus pluvieux ?
- Les « boîtes » sont limitées par le premier et le troisième quartile et contiennent la médiane. Interpréter le fait que la boîte pour Marseille se termine avant la médiane de Paris.
- Comment les diagrammes montrent-ils que, la moitié des mois, il pleut moins à Marseille qu'à Paris ?
- Comment les diagrammes montrent-ils le lieu où les précipitations sont le plus dispersées ?



## 1 Indicateurs de tendance centrale

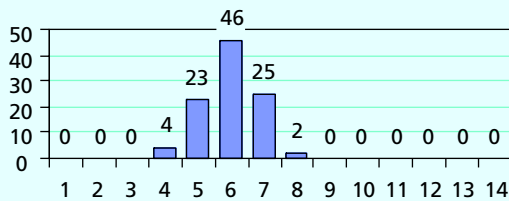
- Le **mode** (ou les modes) d'une série statistique est la valeur la plus fréquente.
- La **médiane**  $Me$  d'une série statistique de  $n$  valeurs classées par ordre croissant est :
  - la valeur du milieu si  $n$  est impair ;
  - la demi-somme des deux valeurs du milieu, si  $n$  est pair.

*Interprétation* : 50 % des valeurs de la série sont inférieures ou égales à la médiane.

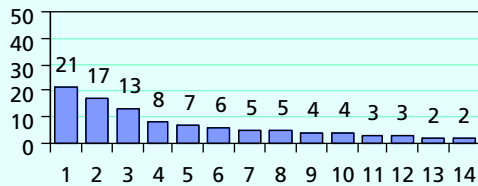
- La **moyenne**  $\bar{x}$  d'une série statistique est obtenue en divisant la somme des valeurs par l'effectif total  $n$ .

*Interprétation* : en remplaçant toutes les valeurs de la série par la moyenne, la somme totale est la même.

### Exemple



Série 1 : mode 6 ;  $Me = 6$  ;  $\bar{x} = 5,98$ .



Série 2 : mode 1 ;  $Me = 3$  ;  $\bar{x} = 4,71$ .

## 2 Indicateurs de dispersion

- L'**étendue**  $e$  d'une série statistique est la différence entre la plus grande et la plus petite valeur de la série.
- Le **premier quartile**  $Q_1$  et le **troisième quartile**  $Q_3$  sont les deux plus petites valeurs de la série telles qu'au moins 25 %, pour  $Q_1$ , et 75 %, pour  $Q_3$ , des valeurs leur soient inférieures ou égales.

L'**écart interquartile**  $Q_3 - Q_1$  est un indicateur de dispersion associé à la médiane.

*Interprétation* : c'est l'écart maximal entre les valeurs de la moitié centrale de la série ; plus l'écart interquartile est grand plus la dispersion est importante.

- L'**écart type**  $\sigma$  (sigma), fourni par la calculatrice ou le tableur, est un indicateur de dispersion associé à la moyenne.

*Interprétation* : plus l'écart type est grand plus la dispersion est importante.

### Exemple

Série 1 :  $e = 4$  ;  $Q_3 - Q_1 = 2$  ;  $\sigma \approx 0,85$ .

Série 2 :  $e = 13$  ;  $Q_3 - Q_1 = 5$  ;  $\sigma \approx 5,95$ .

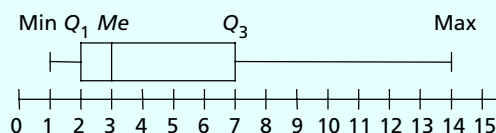
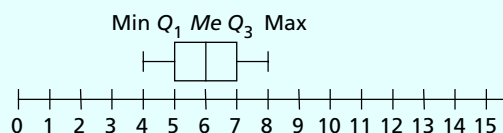
## 3 Diagramme en boîte à moustaches

Le diagramme en boîte à moustaches représente certains indicateurs d'une série.

La « boîte » est limitée par  $Q_1$  et  $Q_3$  et contient la médiane.

Les « moustaches » sont limitées par les valeurs extrêmes.

### Exemple



## 1 Comment déterminer moyenne et écart type ?

### Énoncé

On a effectué 100 simulations de 10 lancers d'une pièce équilibrée. Le tableau suivant indique le nombre de « pile » obtenu par simulation de 10 lancers.

Nombre de « pile »	0	1	2	3	4	5	6	7	8	9	10
Nombre de simulations	0	2	4	6	18	30	27	10	2	1	0

Déterminer la moyenne et l'écart type de cette série statistique (arrondir à  $10^{-2}$ ).

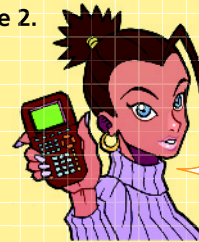
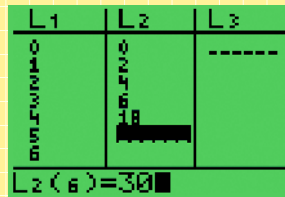
### Solution

La calculatrice donne :

- pour la moyenne :  $\bar{x} = 5,07$  « pile » sur 10 lancers ;
- pour l'écart type :  $\sigma \approx 1,45$  « pile ».

```
1-Var Stats
x=5.07
Σx=507
Σx²=2781
Sx=1.458205622
σx=1.45896275
n=100
```

- ... Dans la calculatrice, on entre les valeurs en liste 1 et les effectifs en liste 2.



Bien distinguer les « valeurs » (correspondant à ce que l'on étudie) et les effectifs !

- ... La moyenne correspond à  $\bar{x}$  et l'écart type à  $\sigma x$  ou  $\sigma n$ .

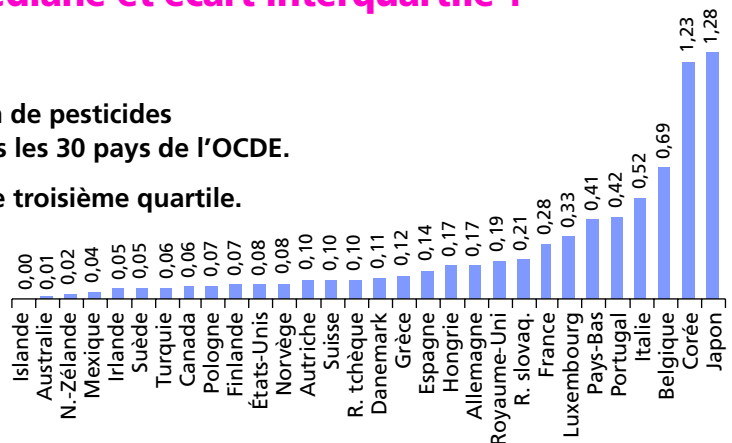
## 2 Comment déterminer médiane et écart interquartile ?

### Énoncé



Le graphique ci-contre fournit l'utilisation de pesticides (en tonnes par  $\text{km}^2$  de terre agricole) dans les 30 pays de l'OCDE.

- Déterminer la médiane, le premier et le troisième quartile.
- Calculer l'écart interquartile.



### Solution

- La médiane est la demi-somme des valeurs

centrales :  $Me = \frac{0,10 + 0,11}{2} = 0,105 \text{ t/km}^2$ .

Le premier quartile est au rang 8 et vaut

$Q_1 = 0,06 \text{ t/km}^2$ . Le troisième quartile est

au rang 23 et vaut  $0,28 \text{ t/km}^2$ .

2. L'écart interquartile vaut :

$Q_3 - Q_1 = 0,28 - 0,06 = 0,22 \text{ t/km}^2$ .

Comme il y a un nombre pair de valeurs ( $n = 30$ ), la médiane est donnée à l'aide des 2 valeurs centrales.

Pour déterminer le rang du premier et troisième quartiles, on prend l'entier directement supérieur ou égal à  $\frac{n}{4}$  et  $\frac{3 \times n}{4}$ .

Si ça n'est pas le cas, il faut d'abord ranger les valeurs dans l'ordre croissant !

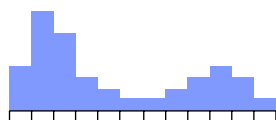


### 3 Comment choisir des indicateurs adaptés ?

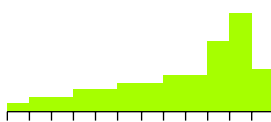
#### Énoncé

À la différence de la médiane  $Me$  et de l'écart interquartile  $Q_3 - Q_1$ , la moyenne  $\bar{x}$  et l'écart type  $\sigma$  sont sensibles aux valeurs extrêmes.

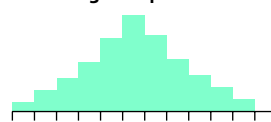
Pour les séries suivantes indiquer si le couple  $(\bar{x}, \sigma)$  et le couple  $(Me, Q_3 - Q_1)$  sont adaptés.



Série 1



Série 2



Série 3

#### Solution

Série 1 et série 2 : le poids des valeurs extrêmes est important, le couple  $(Me, Q_3 - Q_1)$  est sans doute préférable au couple  $(\bar{x}, \sigma)$ .

Série 3 : le couple  $(\bar{x}, \sigma)$  et le couple  $(Me, Q_3 - Q_1)$  sont adaptés.

Lorsque le poids des valeurs extrêmes est important, il peut être préférable de résumer la série par la médiane et l'écart interquartile.

Dans le cas d'une distribution assez symétrique, le couple  $(\bar{x}, \sigma)$  et le couple  $(Me, Q_3 - Q_1)$  conviennent.

### 4 Comment interpréter des indicateurs pour comparer des séries statistiques ?

#### Énoncé



Le tableau suivant fournit la durée moyenne journalière, en heures, d'écoute de la télévision en 2008 pour deux catégories de téléspectateurs (source : Médiamétrie).

Mois (2008)	Jan.	Fév.	Mar.	Avr.	Mai	Juin	Juil.	Août	Sep.	Oct.	Nov.	Déc.
15-34 ans	2,8	2,7	2,6	2,6	2,6	2,6	2,4	2,4	2,5	2,7	2,9	2,8
Ménagères - de 50 ans	3,9	3,7	3,8	3,6	3,4	3,3	3	3	3,4	3,7	3,9	3,8

a) Déterminer la moyenne et la médiane de chaque série, puis comparer la tendance centrale.

b) Déterminer l'étendue, l'écart type et l'écart interquartile, puis comparer la dispersion.

#### Solution

a)

	$\bar{x}$	$Me$
15-34 ans	2,633	2,6
Ménagères	3,542	3,65

Les ménagères ont tendance à davantage regarder la télévision que les 15-34 ans.

b)

	$e$	$\sigma$	$Q_3 - Q_1$
15-34 ans	0,5	0,149	0,2
Ménagères	0,9	0,307	0,5

La série des ménagères est plus dispersée.

La tendance centrale est indiquée par la moyenne et la médiane.

La dispersion est indiquée par l'étendue et, plus précisément, par l'écart type et l'écart interquartile. Plus ces indicateurs sont élevés, plus la dispersion est importante.



# Trier et résumer un grand nombre de données



## Coupe du monde de football



Le fichier « 01\_coupes\_du\_monde.xls » ou « 01\_coupes\_du\_monde.ods » donne les scores des 708 matchs des coupes du monde de football de 1930 à 2006.

### 1 Résumés du nombre de buts marqués par match

Calculer en colonne I le nombre de buts marqués par match (hors tirs au but).

#### a) Valeurs extrêmes

Quel est le nombre minimal de buts marqués ? le nombre maximal ?  
(On peut utiliser les fonctions MIN et MAX du tableur.)

#### b) Résumés de tendance centrale et de dispersion

Déterminer la moyenne et l'écart type du nombre de buts marqués par match.  
(On peut utiliser les fonctions MOYENNE et ECARTYPEP du tableur.)

### 2 Nombre de buts marqués en finale

On souhaite extraire du fichier les scores des finales. Sélectionner la colonne B, créer un filtre en faisant Données/Filtre/Filtre automatique ou AutoFiltre, puis choisir la série désirée.

- Combien de fois l'équipe de France a-t-elle disputé la finale ?
- Lors de quelle finale a-t-il été marqué le plus de buts (hors tirs au but) ?

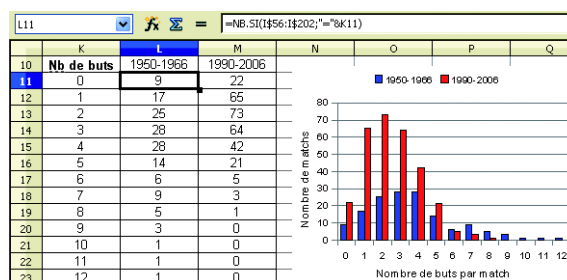
	A	B	C
	Année	Niveau de la compétition	Pays
1	1930	1er tour	France
2	1930	1er tour	Belgique
3	1930	1er tour	France
4	1930	1er tour	France
5	1930	1er tour	France
6	1930	1er tour	France
7	1930	1er tour	France

### 3 Comparaison des périodes 1950-1966 et 1990-2006

Retirer le filtre en sélectionnant Tous.

#### a) Résumé graphique

Pour les périodes 1950-1966 et 1990-2006, déterminer le nombre de matchs correspondant à chaque nombre de buts (on pourra utiliser la fonction NB.SI comme sur l'image d'écran ci-contre).  
Représenter les deux séries et analyser le graphique.



#### b) Résumés de tendance centrale et de dispersion

Pour chacune des deux séries précédentes, déterminer la médiane, l'étendue et l'écart interquartile (on pourra utiliser les fonctions MEDIANE, MIN, MAX et QUARTILE).

- Comparer les deux séries à l'aide de ces indicateurs.
- Prolongement : rechercher les raisons de ces différences statistiques avec l'aide du professeur d'EPS.



## Étudier une grande série en autonomie



### Nuits en hôtel



Ouvrir le fichier « 01\_nuities\_hotellerie.xls » ou « 01\_nuities\_hotellerie.ods » fournissant le nombre de nuitées dans l'hôtellerie par département, en France métropolitaine de 1997 à 2007 (source : Direction du tourisme).

**1** Représenter l'évolution du nombre total de nuitées dans l'hôtellerie en France métropolitaine de 1997 à 2007.



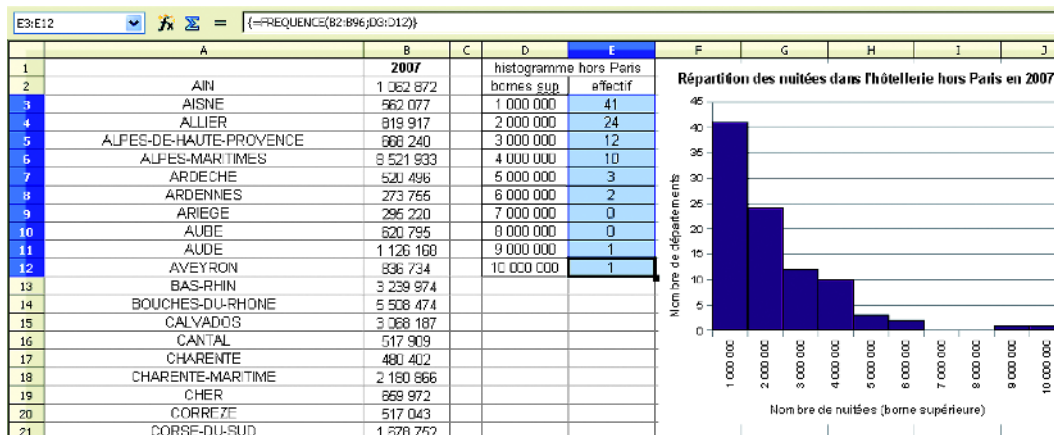
Appeler le professeur pour analyser le graphique obtenu.

**2** À l'aide de résumés graphiques et numériques, comparer le nombre de nuitées dans l'hôtellerie durant la période 1997-2007 dans les départements du Morbihan et de la Corse-du-Sud.



Appeler le professeur pour exposer votre démarche et vos résultats.

**3** Réaliser un histogramme montrant la répartition des nuitées dans les différents départements de France métropolitaine, hors Paris. On pourra procéder comme sur l'image d'écran ci-dessous : les bornes supérieures des intervalles vont de 1 à 10 millions, tous les millions, l'utilisation de la fonction matricielle **Fréquence** du tableur se fait en sélectionnant la plage de cellules (ici E3:E12) puis en validant en maintenant appuyées les touches **Ctrl** et **Majuscule** avant de faire **Entrée**.



Quel couple d'indicateurs ( $\bar{x}$ ,  $\sigma$ ) ou ( $Me$ ,  $Q_3 - Q_1$ ) vous semble-t-il adapté ? Examiner l'effet de la prise en compte, ou non, de Paris.



Appeler le professeur pour exposer vos résultats.



# Exercices et problèmes

**R** Exercices avec réponses en fin d'ouvrage

**\*\*** Exercices plus difficiles

## Exercices

### Déterminer et interpréter mode, moyenne et écart type

- R 1** Les tableaux suivants donnent les notes obtenues à un devoir dans deux classes A et B de 24 élèves.

Notes	7	8	9	10	11	12	13	14	15
Effectifs dans la classe A	1	2	4	7	3	2	2	2	1

Notes	2	4	5	6	8	9	10	11	12	14	15	16	17	18	20
Effectifs classe B	1	1	1	4	2	1	2	1	1	4	1	2	1	1	1

- Déterminer le (ou les) mode(s) dans chaque classe.
- Calculer l'étendue des notes dans chaque classe.
- Calculer la moyenne et l'écart type des notes dans chaque classe.

- 2** Le tableau suivant correspond au volume des précipitations en milliards de m<sup>3</sup> en 2005 pour les 22 régions de France métropolitaine.

Précipitations	6	8	9	11	13	14	15	17
Nombre de régions	2	1	2	1	1	1	1	2

Précipitations	18	20	21	22	32	36	41
Nombre de régions	2	2	1	3	1	1	1

- Quelle est la valeur modale des précipitations en 2005 ?
- Déterminer le volume moyen des précipitations et l'écart type.
- Situer, par rapport à la moyenne des régions, la région Bourgogne, qui a reçu 21 milliards de m<sup>3</sup> de précipitations en 2005.
- Quel indicateur montre-t-il que la quantité de précipitation selon les régions est très variable ?

### **3** \* Salaires mensuels



La distribution des salaires mensuels des employés d'une société est présentée dans le tableau suivant.

Salaires (en euros)	Nombre d'employés
[1 000 ; 1 500[	8
[1 500 ; 2 000[	12
[2 000 ; 2 500[	9
[2 500 ; 3 000[	7
[3 000 ; 3 500[	4
Total	40

- Calculer la fréquence des employés qui gagnent moins de 2 000 € par mois.
- En prenant comme salaires les centres de chaque classe, calculer le salaire moyen et l'écart type (arrondir à la centaine d'euros).

### **4** \*\* La formule de l'écart type

Le tableau suivant donne les puissances  $x_i$  en CV des 80 véhicules d'une entreprise.

$x_i$	$n_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$n_i (x_i - \bar{x})^2$
5	20	- 2,4	5,76	115,2
7	35			
9	15			
11	10			
Total	80			

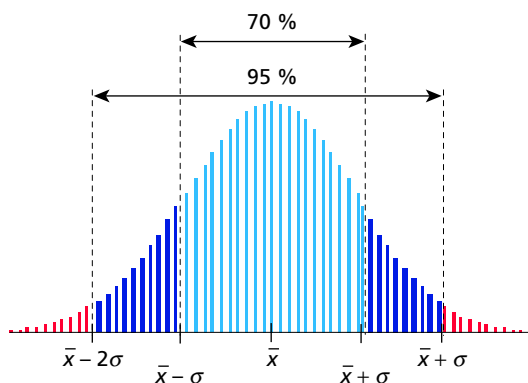
- Vérifier que la puissance moyenne des véhicules est  $\bar{x} = 7,4$  CV.
- Compléter la colonne  $(x_i - \bar{x})$  des écarts entre les  $x_i$  et la moyenne  $\bar{x}$ .
- Compléter la colonne  $(x_i - \bar{x})^2$  des écarts au carré.
- En complétant la dernière colonne, calculer la moyenne  $\bar{E}$  des écarts au carré.
- Calculer  $\sqrt{\bar{E}}$  et comparer avec les résultats affichés sur l'écran suivant d'une calculatrice.

```
1-Var Stats
x̄=7.375
Σx=590
Σx²=4640
Sx=1.911821982
σx=1.899835519
n=80
```

## 5 \* Courbe de Gauss

Dans de nombreuses situations, les données se répartissent selon une courbe de Gauss (ou « courbe en cloche »). Dans ces cas, on observe les résultats suivants :

- environ 70 % des valeurs sont comprises dans l'intervalle  $[\bar{x} - \sigma; \bar{x} + \sigma]$  ;
- environ 95 % des valeurs sont comprises dans l'intervalle  $[\bar{x} - 2\sigma; \bar{x} + 2\sigma]$ .



Les tailles, en cm, de 100 personnes sont reportées dans le tableau suivant.

Taille	Centre	Effectif
[145 ; 150]	147,5	3
]150 ; 155]	152,5	4
]155 ; 160]	157,5	7
]160 ; 165]	162,5	15
]165 ; 170]	167,5	20
]170 ; 175]	172,5	21
]175 ; 180]	177,5	14
]180 ; 185]	182,5	8
]185 ; 190]	187,5	5
]190 ; 195]	192,5	3

1. Construire un histogramme. Peut-on considérer que ces personnes se répartissent approximativement, pour la taille, selon une courbe de Gauss ?

2. En supposant que les valeurs de chaque classe sont situées au centre, déterminer la moyenne et l'écart type (arrondir à l'unité).

Estimer le pourcentage des personnes dont la taille est comprise dans l'intervalle  $[\bar{x} - \sigma; \bar{x} + \sigma]$  puis  $[\bar{x} - 2\sigma; \bar{x} + 2\sigma]$ .

## Déterminer et interpréter la médiane et l'écart interquartile

R 6 Dans une classe, la liste des notes obtenues à un devoir par les élèves classés par ordre alphabétique est la suivante :

8 – 6 – 9 – 19 – 9 – 11 – 13 – 7 – 13 – 14 – 7 – 10 – 10 – 10 – 7 – 13 – 14 – 10 – 13 – 15 – 5 – 16 – 13 – 9 – 10 – 7 – 12 – 5 – 12 – 2 – 9 .

1. Classer les notes dans l'ordre croissant.
2. Déterminer la note médiane. Quelle est sa signification ?
3. Déterminer le premier et le troisième quartile, puis l'écart interquartile.

## 7 Salaires européens

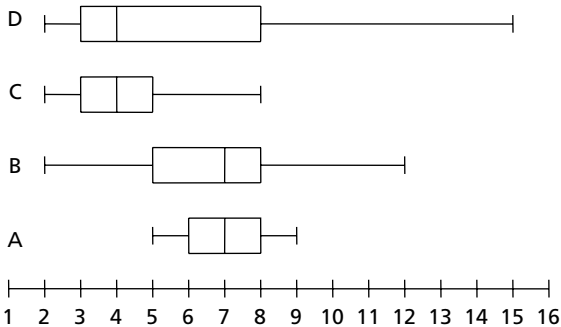
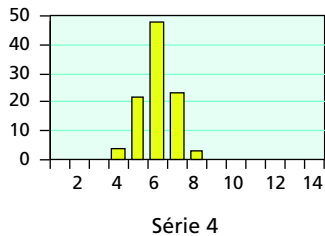
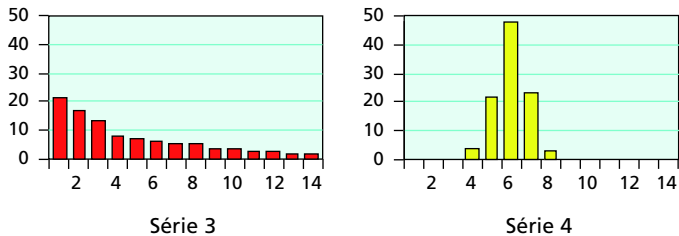
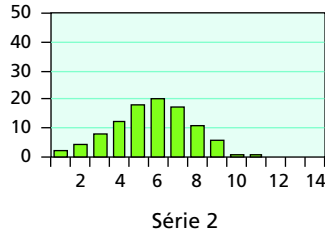
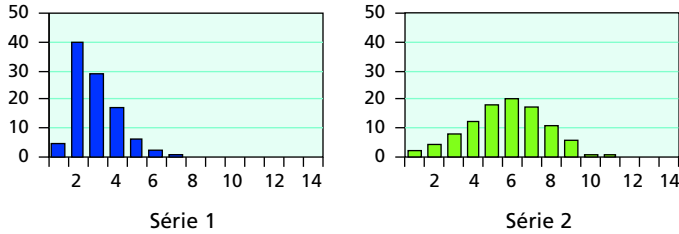
Le tableau suivant fournit le salaire moyen annuel en euros des hommes et des femmes, pour un temps plein (données Eurostat 2005, sauf Pologne, 2004, et Grèce, 2003).

	Femmes	Hommes	Rapport
Belgique	32 715	37 822	1,16
Pologne	5 506	6 663	1,21
France	26 586	32 316	1,22
Suède	29 052	35 770	1,23
Danemark	40 884	50 676	1,24
Grèce	14 376	17 889	1,24
Allemagne	34 522	43 945	1,27
Portugal	12 412	16 133	1,30
Pays-Bas	30 900	40 300	1,30
Royaume-Uni	33 562	46 518	1,39
Rép. tchèque	5 925	8 285	1,40
Hongrie	6 700	9 905	1,48
Autriche	26 514	40 022	1,51

1. Interpréter la valeur 1,51 obtenue pour l'Autriche.
2. Pour chacune des trois séries, déterminer la médiane et l'écart interquartile.
3. Situer la position de la France.

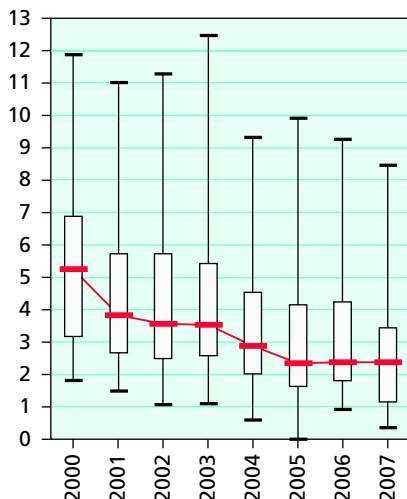
## Interpréter des boîtes à moustaches

- 8** Associer les boîtes à moustaches à chacune des séries représentées.



## 9 \* Qualité de l'air

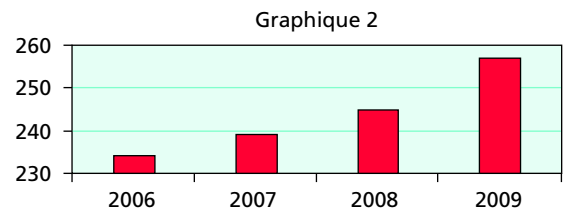
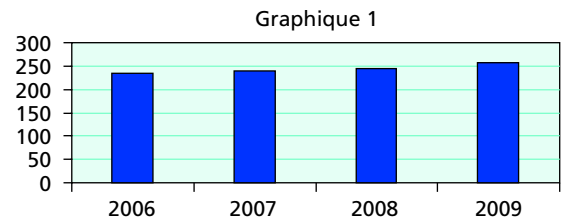
Le graphique suivant correspond aux moyennes annuelles des concentrations (en  $\mu\text{g}/\text{m}^3$ ) en dioxyde de soufre ( $\text{SO}_2$ ) mesurées entre 2003 et 2007 dans les stations des 21 régions de France métropolitaine (source : Banque de données sur la qualité de l'air).



- Comment constate-t-on que la concentration moyenne en  $\text{SO}_2$  est supérieure à  $5 \mu\text{g}/\text{m}^3$  dans plus de la moitié des régions en 2000 ?
- Comment constate-t-on que la concentration moyenne en  $\text{SO}_2$  est inférieure à  $4 \mu\text{g}/\text{m}^3$  dans plus de 75 % des régions en 2007 ?
- Quelle est l'évolution de la tendance centrale de la pollution au dioxyde de soufre de 2000 à 2007 ?
- Quels indicateurs montrent-ils que la pollution au dioxyde de soufre est plus dispersée entre les régions en 2000 qu'en 2007 ?

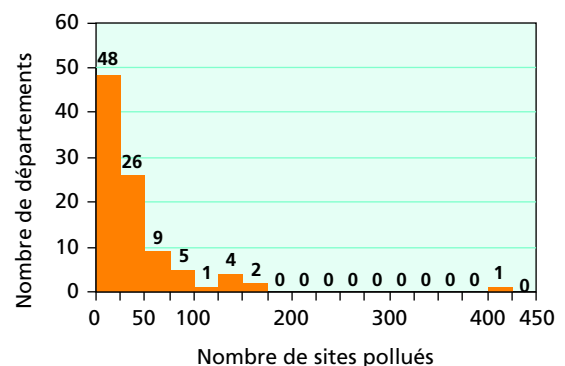
## Choisir des résumés adaptés

- 10** Chloé dirige une entreprise et a besoin d'impressionner son banquier. Les graphiques ci-dessous indiquent ses bénéfices pour les quatre dernières années. Quel graphique doit-elle montrer ?



## 11 Sites pollués

L'histogramme suivant correspond, pour les 96 départements de métropole, au nombre de sites pollués faisant l'objet d'une action publique (source : Institut français de l'environnement).



1. D'après la répartition des valeurs, vaut-il mieux résumer cette série par le couple moyenne et écart type, ou par le couple médiane et écart interquartile ?

2. En supposant que les valeurs sont toutes situées au centre des classes, estimer la médiane et le troisième quartile.

3. Le département des Landes compte 27 sites pollués. Situer ce département par rapport aux autres en utilisant les résultats de la question 2.

## Interpréter des indicateurs pour comparer des séries statistiques

12 Lors d'un examen, on souhaite comparer les résultats des candidats à trois épreuves A, B et C. Pour cela, on a prélevé un échantillon aléatoire de 30 candidats, dont voici les résultats.

Notes sur 20	Effectifs		
	Épreuve A	Épreuve B	Épreuve C
5	0	0	3
6	0	4	0
7	1	6	4
8	3	8	0
9	6	4	5
10	4	3	0
11	7	0	6
12	4	1	1
13	2	0	0
14	2	1	4
15	1	2	5
16	0	1	2

1. Calculer la moyenne de chacune des trois épreuves (arrondir à  $10^{-2}$ ). Quelle est l'épreuve qui vous semble la moins réussie ?

2. Calculer l'écart type pour l'épreuve A (arrondir à  $10^{-2}$ ).

3. Pour l'épreuve B l'écart type vaut 2,74 et pour l'épreuve C l'écart type vaut 3,74. Quelle épreuve a-t-elle les résultats les plus homogènes ? les plus hétérogènes ?

## R 13 Entrées au cinéma

Le tableau ci-contre donne le nombre d'entrées au cinéma (en millions de spectateurs) pour l'Union européenne et les États-Unis de 1998 à 2008 (source : Centre national de la cinématographie).

	1998	1999	2000	2001	2002	2003
UE	828	810	844	935	938	890
USA	1 438	1 440	1 385	1 437	1 597	1 520

	2004	2005	2006	2007	2008
UE	1 006	894	926	917	920
USA	1 484	1 378	1 395	1 400	1 364

1. Déterminer le nombre annuel moyen d'entrées durant cette période dans l'Union européenne et aux États-Unis.

2. Pour chacune des séries, déterminer l'étendue, la médiane et l'écart interquartile.

3. Comparer la tendance centrale et la dispersion du nombre d'entrées dans l'Union européenne et aux États-Unis.



## Problèmes

### Problème 1

#### Buts marqués

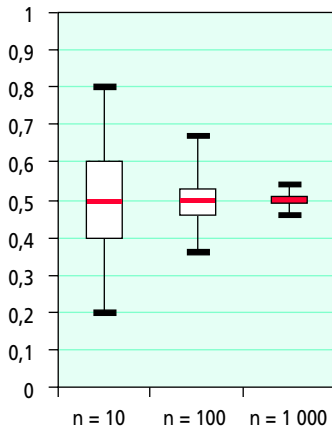
Le tableau suivant correspond au nombre de buts marqués durant une journée de ligue 1 de football, depuis la saison 2002/2003 jusqu'à la saison 2008/2009 (source : Ligue professionnelle de football).

	02/03	03/04	04/05	05/06	06/07	07/08	08/09
Min	16	14	14	13	8	13	11
$Q_1$	19	20,25	18	17,25	20	19,25	20
Me	22	22,5	22	20	22	22	22,5
$Q_3$	25	28	24,75	24,75	25	25	25
Max	31	33	36	38	35	43	33

1. Durant quelle saison a-t-on marqué le moins de buts durant une journée ?
2. Interpréter la médiane de la saison 2008/2009.
3. Comparer la tendance centrale de ces 7 saisons de ligue 1.
4. Calculer les écarts interquartiles, puis déterminer les saisons où le nombre de buts marqués par journée est :
  - le plus dispersé ;
  - le moins dispersé.

### Problème 2 \*\*

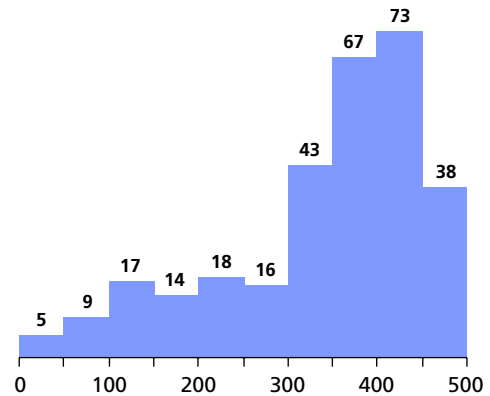
On simule des lancers de pile ou face et on s'intéresse à la fréquence des « pile » sur 10, 100 ou 1 000 lancers. Le diagramme suivant indique la répartition des fréquences des « pile » sur 200 échantillons de taille 10, 100 et 1 000 (d'autres simulations sont visibles sur le fichier « 01\_pile\_ou\_face.xls »).



1. Quelle est, pour les trois séries, la médiane des fréquences de « pile » par échantillon ? Interpréter le résultat.
2. Quelle est, des trois séries, celle qui est la plus dispersée et celle qui est la moins dispersée ?
3. Lire approximativement sur le graphique :
  - le nombre d'échantillons de taille 10 dont la fréquence de « pile » est supérieure à 0,6 ;
  - le nombre d'échantillons de taille 1 000 dont la fréquence de « pile » est supérieure à 0,6 ;
  - le nombre d'échantillons de taille 10 dont la fréquence de « pile » est comprise entre 0,4 et 0,6 ;
  - le nombre d'échantillons de taille 1 000 dont la fréquence de « pile » est comprise entre 0,4 et 0,6.

### Problème 3 \*

L'histogramme suivant correspond aux achats durant les six derniers mois d'un échantillon représentatif de 300 clients d'une entreprise de vente sur Internet.



1. Quelle est la classe modale ?
2. Quelle est la fréquence des clients ayant dépensé plus de 450 € ?
3. D'après la forme de l'histogramme, la dépense médiane est-elle inférieure ou supérieure à la dépense moyenne ?
4. Quel indicateur de tendance centrale choisir pour définir le « client type » ?
5. Quel indicateur de tendance centrale choisir pour le comptable de l'entreprise ?
6. En supposant que les valeurs de chaque classe sont situées au centre, estimer la moyenne et l'écart type du montant des achats.

### Problème 4 \*

#### Nombre de chômeurs



Le fichier « 01\_chomage.xls » ou « 01\_chomage.ods » fournit le nombre trimestriel de chômeurs en France de 1975 à 2008.

1. a) Choisir un type de graphique permettant d'illustrer l'évolution du chômage, puis le réaliser.  
b) Commenter le graphique obtenu.
2. Calculer la moyenne  $\bar{x}$ , l'étendue  $e$  et l'écart type  $\sigma$  du nombre de chômeurs, en milliers, durant cette période.
3. Le nombre de chômeurs au premier trimestre 2009 est  $n \approx 2\,455\,000$ . Comparer  $n - \bar{x}$  et  $\sigma$ .

Pour chaque énoncé, indiquer la ou les bonnes réponses.

	A	B	C												
1 On considère les 12 notes suivantes : 5, 7, 8, 9, 10, 10, 12, 13, 14, 17, 17, 19.	L'écart interquartile vaut 6.	L'écart interquartile vaut 8.	L'écart interquartile vaut 9.												
2 La proportion des valeurs comprises dans l'intervalle $[Q_1 ; Q_3]$ est environ :	25 %	50 %	75 %												
3 Pour cette série... <table><tr><td>Notes</td><td>4</td><td>8</td><td>10</td><td>12</td><td>16</td></tr><tr><td>Effectifs</td><td>1</td><td>3</td><td>5</td><td>4</td><td>2</td></tr></table>	Notes	4	8	10	12	16	Effectifs	1	3	5	4	2	l'écart type vaut $\sigma \approx 2,96$ .	l'écart type vaut $\sigma \approx 3,07$ .	l'écart type vaut $\sigma = 4$ .
Notes	4	8	10	12	16										
Effectifs	1	3	5	4	2										
4 Pour cette série...	la proportion des valeurs comprises dans l'intervalle $[\bar{x} - 2\sigma ; \bar{x} + 2\sigma]$ vaut environ 25 %.	la proportion des valeurs comprises dans l'intervalle $[\bar{x} - 2\sigma ; \bar{x} + 2\sigma]$ vaut environ 50 %.	la proportion des valeurs comprises dans l'intervalle $[\bar{x} - 2\sigma ; \bar{x} + 2\sigma]$ vaut environ 95 %.												
5	L'écart type de la série 1 est inférieur à l'écart type de la série 2.	L'écart type de la série 1 est égal à l'écart type de la série 2.	L'écart type de la série 1 est supérieur à l'écart type de la série 2.												
6 Je suis peu sensible aux valeurs extrêmes de la série. Je suis...	le couple médiane et écart interquartile.	le couple moyenne et écart type.	l'étendue.												
7 Je suis calculé en utilisant les valeurs numériques complètes de toute la série statistique. Je suis...	le couple médiane et écart interquartile.	le couple moyenne et écart type.	l'étendue.												
8 Pour le cas suivant, le couple médiane et écart interquartile est préférable au couple moyenne et écart type :															
9 Le diagramme ci-dessous  peut correspondre à la série :															
10	D'après le diagramme, la série la plus dispersée est la série 1.	D'après le diagramme, la série la plus dispersée est la série 2.	D'après le diagramme, on ne peut pas connaître la série la plus dispersée.												