Statistiques Cours

1. Définitions

Une ***série statistique*** est la donnée d'objets (***items***) auxquels sont associés des nombres (par exemple dans une élection le candidat A (item A) a obtenu 51210 voix, le candidat B a obtenu 43821 voix, etc.)

Les résultats sont souvent présentés sous forme de tableau.

Candidat	A	B	C	D
Nombre de voix obtenues	51 210	43 821	23 212	8 597

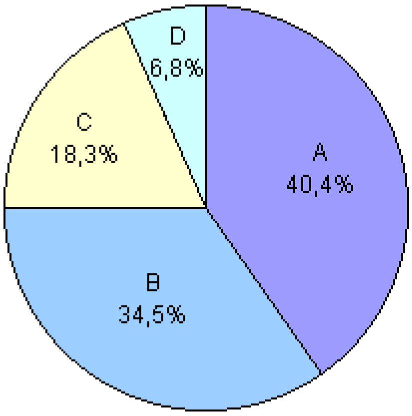
A partir de ce tableau on a l'***effectif total*** : 51 210 + 43 821 + 23 212 + 8 597 = 126 840

Et on peut calculer les ***fréquences*** associées à chaque terme de la série : $fréquence = \dfrac{effectif\ item}{effectif\ total}$.

Ces fréquences peuvent être mises sous forme de pourcentage en multipliant par 100.

On rajoute souvent une colonne ***total*** qui permet de vérifier en partie les calculs.

Candidat	A	B	C	D	Total
Nombre de voix obtenues	51 210	43 821	23 212	8 597	126840
Fréquence (en décimal)	0,404	0,345	0,183	0,68	1
Fréquence (en pourcentage)	40,4%	34,5%	18,3%	6,8%	100



On représente parfois les données dans un ***diagramme circulaire*** (communément appelé ***camembert***…) où la mesure angulaire de chaque secteur correspond à la fréquence. C'est une simple question de proportionnalité.

Le cercle complet correspond à un angle au centre de $360°$,

le secteur du candidat A aura pour mesure angulaire $360° \times 0{,}404 = 145°$

le secteur du candidat B aura pour mesure angulaire $360° \times 0{,}345 = 124°$

le secteur du candidat C aura pour mesure angulaire $360° \times 0{,}183 = 66°$

le secteur du candidat D aura pour mesure angulaire $360° \times 0{,}68 = 25°$

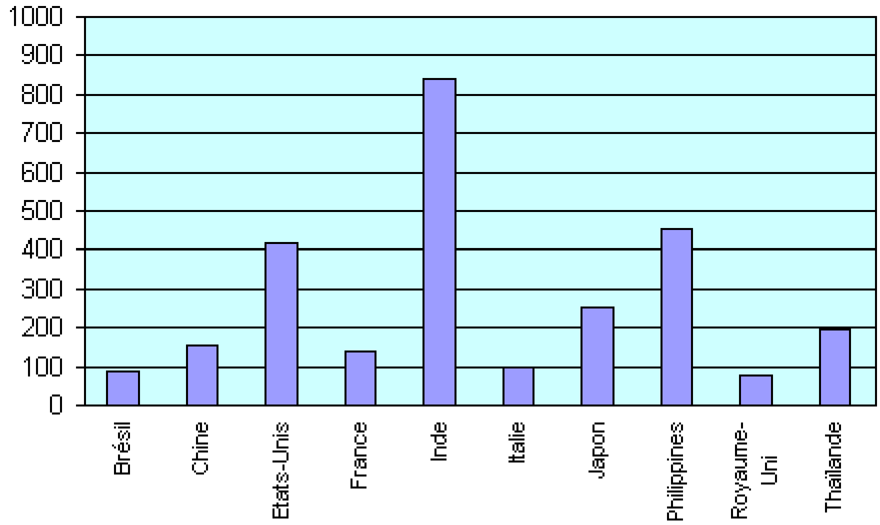
Remarque : pour représenter les données sur un diagramme semi-circulaire, il suffira de calculer les mesures des secteurs angulaires par rapport à 180°.

Plus fréquemment on représente les données dans un ***diagramme à barres ou en batons*** ou ***histogramme*** : dans un diagramme à barres, la hauteur de la barre représente l'effectif.

Exemple

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



Si les données d'une série sont discrètes, le ***mode*** est la ou les valeurs qui ont le plus grand effectif.

Si les données ont été réparties en classes, on parle alors plutôt de ***classe modale***.

L'***étendue*** d'une série est la différence entre la plus grande valeur et la plus petite.

Exemple

Données discrètes : 9, 11, 8, 10, 13, 12, 10, 11, 10.

Faisons le tableau des effectifs :

valeur	8	9	10	11	12	13
effectif	1	1	3	2	1	1

Le mode est la valeur qui a le plus gros effectif, c'est à dire 10. L'étendue est $13 - 8 = 5$.

Ici, vu le petit nombre de données, faire un tableau des effectifs est un peu artificiel. Par contre, dès que l'on travaille sur un nombre important de données, il devient vite très utile pour mettre en évidence le mode et l'étendue de la série.

Données réparties par classes :

classe	[0 ; 5[	[5 ; 10[	[10 ; 15[	[15 ; 20]
effectif	0	5	14	2

La classe modale est la classe qui a le plus gros effectif, c'est à dire la classe [10 ; 15[.

L'étendue de cette série est 20 – 5 = 15. Par simplification, on dira que l'étendue est 15 mais c'est un abus de langage. En effet, dans le tableau des données ci dessus, rien ne permet d'affirmer que les valeurs extrêmes sont 5 et 20.

Dans une série statistique lorsqu'on a des valeurs numériques pour les items (ou qu'on leur attribue des valeurs numériques), on peut calculer une ***moyenne*** (coefficientée, comme au bac) :

Exemple

Par exemple 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, P	15	16	17	18	19
Nombre de CD vendus	97	34	43	20	6

Le prix de vente moyen de ce CD est alors $\overline{P} = \dfrac{15\times 97+16\times 34+17\times 43+18\times 20+19\times 6}{97+34+43+20+6} \approx 16{,}02$ (la moyenne se note avec une barre au-dessus). Dans ce calcul, les notes (du bac) sont représentées par les prix, les coefficients par les nombres de CD.

Les écarts de la série par rapport à la moyenne sont mesurés à l'aide de deux nombres :

* la ***variance*** : on élève au carré la valeur des items et on refait une moyenne de ces carrés :

P	15	16	17	18	19
P^2	225	256	289	324	361
effectifs	97	34	43	20	6

$$\overline{P^2}=\frac{225\times 97+256\times 34+289\times 43+324\times 20+361\times 6}{97+34+43+20+6}\approx 258{,}01\,,$$

puis on soustrait le carré de la moyenne $\overline{P}^2\approx 256$: $\overline{P^2}-\overline{P}^2\approx 1{,}37$, c'est la ***variance*** : $\mathrm{var}(P)=1{,}37$.

* Enfin on calcule l'***écart-type*** $\sigma(P)=\sqrt{\mathrm{var}(P)}\approx 1{,}17$.

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 1

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	0	1	2	3	4	5	6	7	8	9	10	Total
Effectif : n												
Fréquences												

Pour la classe B :

Note : Y	0	1	2	3	4	5	6	7	8	9	10	Total
Effectif : m												
Fréquences												

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	0	1	2	3	4	5	6	7	8	9	10	Total
Effectif : n												
X^2n												

Pour la classe B :

Note : Y	0	1	2	3	4	5	6	7	8	9	10	Total
Effectif : m												
Y^2m												

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

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

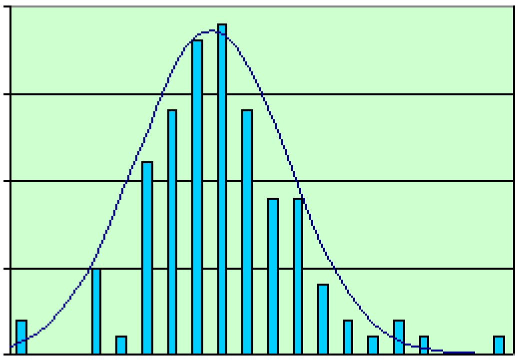
6. Pour chacune des classes, donner la note du 16ème élève.

2. Données Gaussiennes

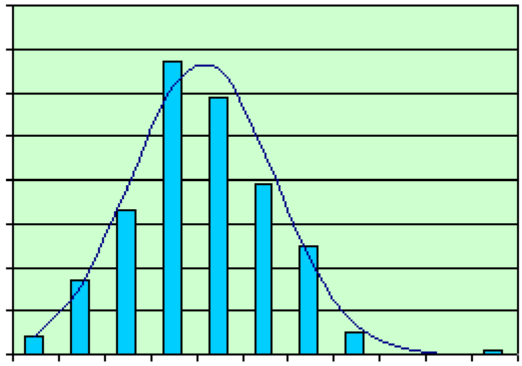
(du nom de Karl Friedrich Gauss, grand mathématicien allemand du 19[ème] siècle)

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"



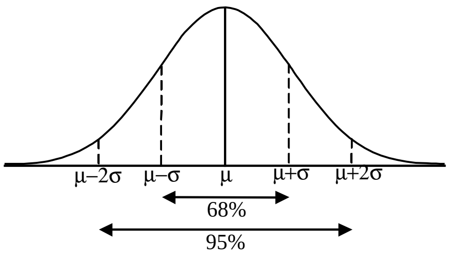
Fréquence des battements cardiaques



Circonférence du biceps

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

Si on calcule pour de telles données la moyenne μ (*mu*) et l'écart-type σ (*sigma*), on peut noter que :



- la série est à peu près symétrique autour de la moyenne μ ,

- environ 68 % des données se trouvent dans l'intervalle $[\mu-\sigma\,;\mu+\sigma]$; cet intervalle est appelé ***plage de normalité*** pour le ***niveau de confiance*** 0,68 ;

- environ 95% des données se trouvent dans l'intervalle $[\mu-2\sigma\,;\mu+2\sigma]$; cet intervalle est appelé ***plage de normalité*** pour le ***niveau de confiance*** 0,95 ;

- environ 99% des données se trouvent dans l'intervalle $[\mu-3\sigma\,;\mu+3\sigma]$; cet intervalle est appelé ***plage de normalité*** pour le ***niveau de confiance*** 0,99.

les observations ci-dessus n'ont aucun sens pour les séries qui traduisent des phénomènes non gaussiens ainsi que pour les séries gaussiennes où l'échantillon est trop petit (effectif < 30).

Exemple

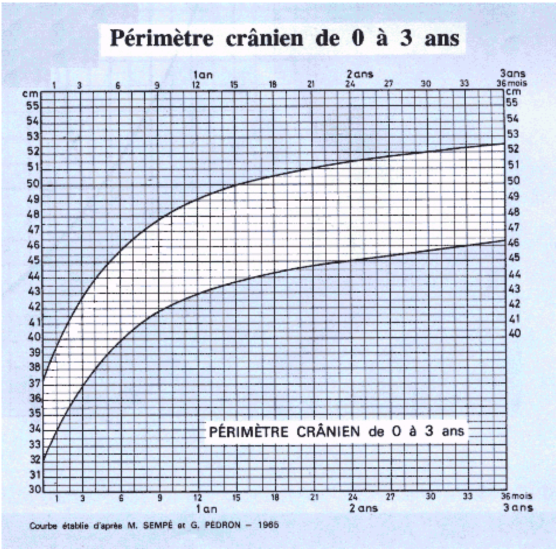
Les études statistiques portant sur un grand nombre d'enfants ont conduit à des données gaussiennes.

Ces études ont permis d'établir des courbes de croissance (taille, poids, périmètre cranien …) 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.

3. Médiane et quartiles

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). C'est le 50/50…

Exemple 1

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 4ème et du 5ème terme, c'est-à-dire 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 6ème terme, c'est-à-dire 92.

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.

Le deuxième quartile est la ***médiane***.

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 2

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 4ème 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 11ème terme de la série c'est-à-dire 19.

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

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 3

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 3ème terme de la série c'est-à-dire 5.

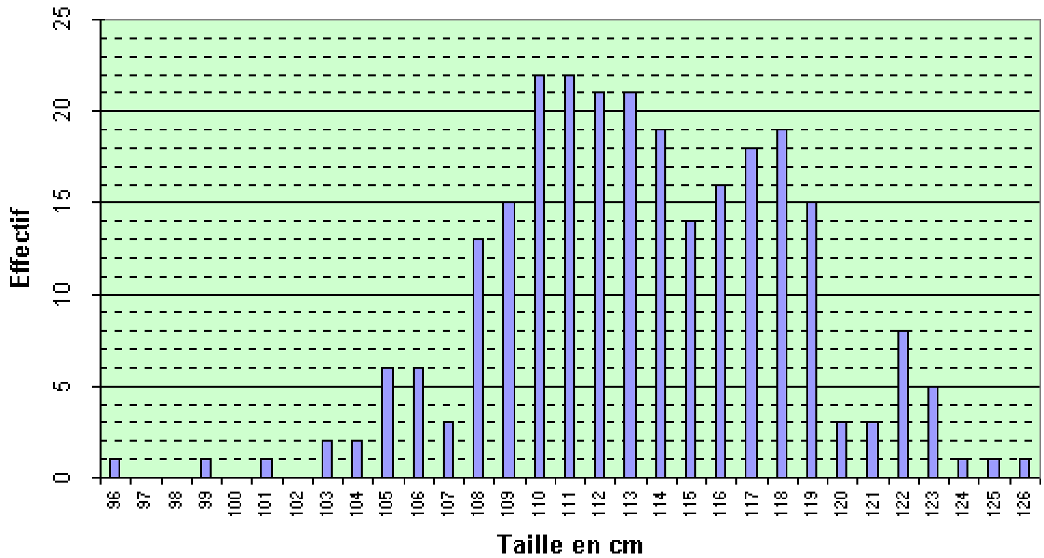
Neuf dixièmes (90%) des données correspondent à : $27 \times 0{,}9 = 24{,}3$. Le neuvième décile est alors, par définition, la plus petite valeur d' pour laquelle les valeurs de 25 termes de la série sont inférieurs ou égales à d'. Le neuvième décile est donc la valeur du 25ème terme c'est-à-dire 15.

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

Exercice 2

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 1er quartile et le 3ème quartile de cette série. Donner l'intervalle interquartile et l'écart interquartile.

5. Déterminer le 1er décile et le 9ème décile de cette série. Donner l'intervalle interdécile et l'écart interdécile.

6. Les données semblent presque gaussiennes : calculer moyenne et écart-type de cette série. La plage de normalité est-elle respectée ?

Exercice 3

Lors d'une étude sur 3023 personnes, on a noté le nombre de personnes infectées ou non par le VIH (sida). On a d'autre part classé ces mêmes personnes en fonction du nombre total de leurs partenaires. Les résultats sont donnés dans le tableau suivant :

Nombre de partenaires	0	1	2	3	4	5	6	7	8	9	10	11
Personnes infectées	19	110	174	165	156	30	29	13	12	9	23	79
Personnes non infectées	264	491	418	348	250	83	72	40	29	19	51	139

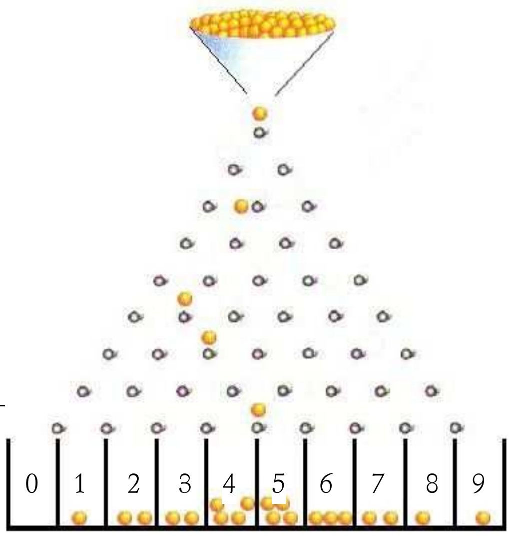
1. Calculez le nombre total de personnes infectées par le VIH et le nombre total de personnes non infectées. Que pensez-vous de la population étudiée ?

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

3. Expliquez pourquoi on ne peut pas conclure de ce graphique que les personnes ayant de nombreux partenaires sont moins infectés par le VIH que les personnes ayant un seul partenaire. Proposez une méthode permettant de répondre à cette interrogation.

Exercice 4 : Une machine de **Galton**

Sur le dessin ci-contre est schématisée une machine de Galton : le réservoir supérieur contient des billes qui en tombant, vont heurter un certain nombre de clous disposés de façon triangulaire.



On suppose que, sur l'ensemble des boules arrivant sur un clou, la moitié se dirige vers la gauche et la moitié se dirige vers la droite.

Ces boules arrivent finalement dans une des cases qui sont numérotées de 0 à 9.

1. On lance 512 boules. En schématisant la machine de Galton par un arbre, dénombrer le nombre de boules arrivant dans chacune des cases et vérifier le tableau ci-dessous.

Numéro de la case	0	1	2	3	4	5	6	7	8	9
Nombre de boules (Effectif)	1	9	36	84	126	126	84	36	9	1

2. Tracer un diagramme en arbre pour représenter ces données. Ces données semblent-elles être des données gaussiennes ?

3. Calculer la moyenne μ et l'écart-type σ de la série obtenue. Quel est la proportion de données se trouvant dans la plage de normalité ?

4. Quelle est la proportion des données se trouvant dans l'intervalle $[\mu - 3\sigma ; \mu + 3\sigma]$?

4. Diagramme en boîte

Ce type de diagramme est aussi appelé diagramme de Tuckey, boîte à moustaches ou boîte à pattes.

Il utilise le 1er et le 3ème quartile, les valeurs extrêmes, le 1er et le 9ème décile et éventuellement la médiane d'une série.

La construction ci-dessous 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

1er quartile : 110 | 3ème quartile : 117

1er décile : 108 | 9ème décile : 119

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

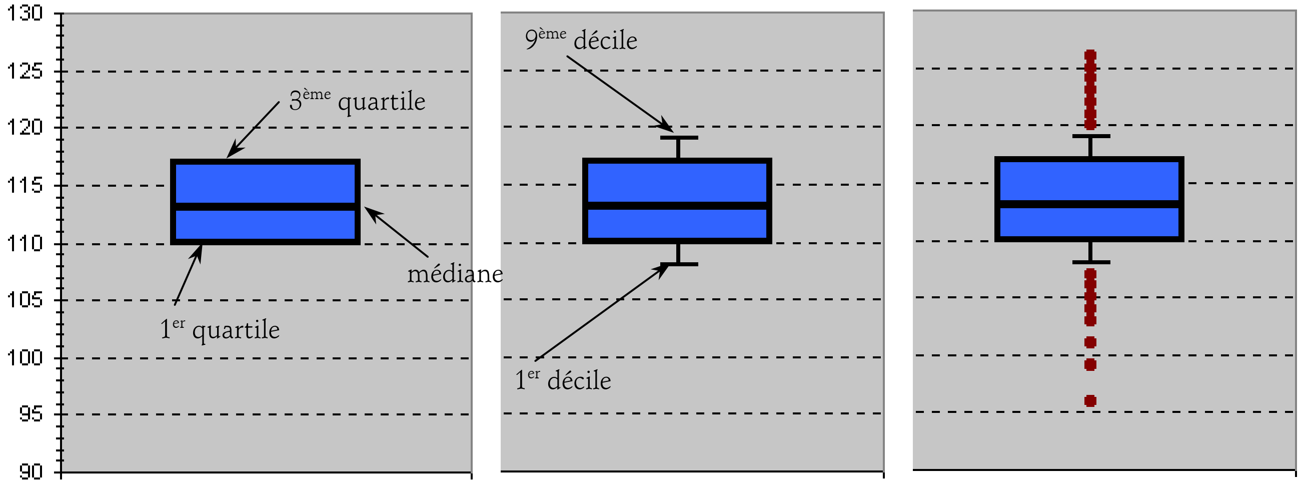
On pourra par exemple graduer entre 90 et 130 (si certaines valeurs sont manifestement hors normes, on n'en tiendra pas compte).

Le "corps" du diagramme, c'est-à-dire la "boîte" est formée d'un rectangle ayant pour extrémité inférieure le 1er quartile et pour extrémité supérieure le 3ème 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 1er et au 9ème 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



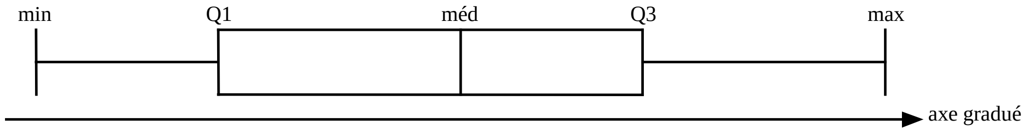
l'intervalle interdécile.

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 1er et au 99ème centile, ou même aux valeurs extrêmes.

Le diagramme en boîte d'une série aura l'allure suivante :



* Lorsque la série est trop importante, que l'on ne connaît pas les valeurs extrêmes ou qu'on les considère comme non significatives, on raccourcit souvent les moustaches au déciles D1 et D9.

* La boîte centrale représente l'intervalle interquartile et contient donc la moitié des données.

* Vous devez légender votre diagramme (min, max, nom de la série) et graduer l'axe.

* On emploie surtout ce type de diagramme pour comparer plusieurs séries entre elles.

Exemple

Deux classes de 1L comparent leurs résultats du trimestre et déclarent : "nos classes ont le même profil puisque dans les deux cas la médiane des résultats est 10". Qu'en pensez-vous ?

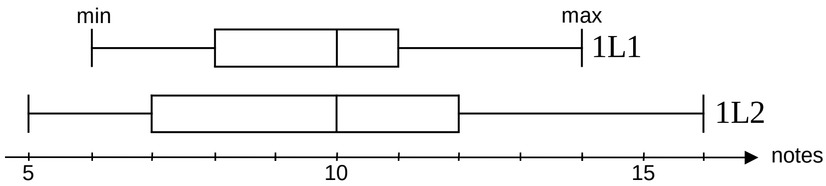
notes	5	6	7	8	9	10	11	12	13	14	15	16
effectifs 1L1	0	3	4	4	5	7	3	4	2	1	0	0
effectifs 1L2	2	4	3	3	3	4	3	2	2	3	1	2

1. Vérifier que les deux médianes valent 10 et déterminer les quartiles de chaque série.
2. Tracer côte à côte les diagrammes en boites de ces deux séries.

Pour la 1L1 : L'effectif total est 3+4+4+…+1 = 33 ; $\frac{1}{2}(33+1)=17$ donc la médiane est le 17ème terme de la série : méd = 10 ; $\frac{1}{4}\times 33=8,25$ donc le 1er quartile est le 9ème terme de la série : Q_1 = 8 ; $\frac{3}{4}\times 33=24,75$ donc le 3ème quartile est le 25ème terme de la série : Q_3 = 11.

Pour la 1L2 : L'effectif total est 2+4+3+…+2 = 32 ; $\frac{1}{2}(32+1)=16,5$ donc la médiane est la moyenne des 16ème et 17ème terme de la série : méd = 10 ; $\frac{1}{4}\times 32=8$ donc le 1er quartile est le 8ème terme de la série : Q_1 = 7 ; $\frac{3}{4}\times 32=24$ donc le 3ème quartile est le 24ème terme de la série : Q_3 = 12.

Diagrammes en boîtes :



Bilan : Le graphique ci-dessus met bien en évidence que l'écart interquartile et l'étendue sont plus resserrés en 1L1 qu'en 1L2 donc les élèves de 1L1 ont globalement un niveau plus homogène que ceux de 1L2.

Dans la pratique :

* On utilise très peu le mode et l'étendue (faciles à déterminer mais simplistes !).

* On utilise la médiane, quartiles, déciles et écart interquartile surtout pour les séries à grands effectifs (pas de calculs, il suffit d'ordonner la série ; peu sensible aux valeurs douteuses).

* On utilise souvent la moyenne et l'écart type pour des séries de tailles intermédiaires ou des séries gaussiennes (la moyenne reste l'indicateur le plus intuitif ; intérêt des plages de normalité).

Riche ou pauvre (Midi Libre du 15/01/07, Paul Villemus)

Que veut dire être riche ou être pauvre en France en 2007 ? La notion de richesse est éminemment ambiguë et subjective. En économie, la richesse d'un individu s'exprime en termes monétaires et est composée de ses revenus et de son patrimoine. D'après l'INSEE, le revenu moyen net d'impôt, en 2004, était de 29 000 euros par an et par ménage soit 2417 **€** par mois. L'ensemble des revenus disponibles était constitué des revenus du travail (69,7 % du total), des pensions (21,7 %), des revenus du patrimoine (loyers encaissés, intérêts des placements financiers, pour 3 %), des prestations familiales et sociales (3,7 %, y compris logement social) et des minima sociaux (1,2 %).

En économie on utilise un autre critère de mesure que la moyenne : la *médiane*. La médiane est la valeur qui partage la population en deux parties de même effectif. Ainsi le revenu disponible médian net d'impôt en France est de 2 042 **€** par mois : la moitié des ménages gagne plus et l'autre moitié gagne moins. Le seul salaire médian s'élève à 1 750 **€** par mois. D'ores et déjà on peut dire qu'un salaire de 4 000 **€** net par mois est plus de deux fois supérieur à la médiane nationale. Mais gagner 4 000 **€** si l'on vit seul ou à plusieurs ce n'est pas la même chose. Les statisticiens, jamais à court d'imagination, ont inventé la notion de niveau de vie.

Quand on parle de niveau de vie d'un individu, on désigne toutes les ressources perçues par le ménage divisées par le nombre d'unités de consommation (UC) du ménage : la première personne vaut 1 UC, les autres 0,5 UC chacune si elles ont moins de 14 ans et 0,3 UC sinon. Les revenus sont donc les sommes perçues par le ménage dans son ensemble (quel que soit le nombre de personnes du ménage) et le niveau de vie, les ressources par unité de consommation. Le niveau de vie moyen français est de 1 502 **€** par mois en 2004, la médiane étant de 1 314 **€** par mois.

Il n'y a pas en économie de « *seuil de richesse* », en revanche, il existe dans beaucoup de pays et en France en particulier, un « *seuil de pauvreté* ». Selon l'INSEE il est égal à 50 % du niveau de vie médian de la population, soit 657 **€** par personne. Il y a donc en France 6 % de pauvres dans la population totale, soit 3,6 millions de personnes et 1,6 million de ménages pauvres ! Si on appliquait le seuil de pauvreté utilisé par l'Union Européenne (60 % du niveau de vie médian), on aurait 6,9 millions de pauvres en France !!!

On pourrait considérer que les « riches » sont ceux qui ont un niveau de vie deux fois supérieur au niveau de vie médian ou qui ont deux fois plus de revenus disponibles que le revenu net médian. Seraient alors considérés comme riches une personne seule touchant plus de 2 400 **€** par mois, un couple gagnant 4 500 **€**, et une famille avec deux enfants de moins de 14 ans touchant 6 100 **€**.

Avec ces revenus disponibles on appartient aux 10 % des personnes les plus « aisées » de France. Mais pour évaluer la richesse d'une personne, il faut également connaître son patrimoine. Car si l'on est propriétaire ou locataire, cela induit de grosses différences de niveau de vie. Le patrimoine médian des ménages est de 165 000 **€** en 2004. En France, 450 000 ménages, soit 2 % des contribuables, ont aquittté l'ISF. Pour payer l'ISF il faut avoir un patrimoine net supérieur à 760 000 **€** (y compris la résidence principale), soit 4,5 fois le patrimoine médian. Le nombre d'assujettis a doublé depuis 1999. Les Français ont toujours tendance à surestimer le « *seuil psychologique et virtuel de richesse* ». Etre pauvre, en revanche, est bien une situation objective et socialement dégradante.

5. Exercices corrigés

Exercice 1 : traitement statistique d'un contrôle d'épaisseur

Le contrôle d'une production de tablettes en bois est effectué toutes les heures par prélèvement d'un échantillon de 20 éléments. On effectue un relevé de l'épaisseur de chaque élément.

L'épaisseur normale est de 18 mm. Une pièce sans défaut est une pièce dont l'épaisseur ne s'écarte pas de plus de 0,5 mm de l'épaisseur normale.

Les pièces ne correspondant pas à cette exigence sont classées en deux catégories :

Défaut mineur	L'épaisseur s'écarte de 0,5 à 1 mm de l'épaisseur normale.	Défaut qui ne nécessite pas d'opération de reprise.
Défaut majeur	L'épaisseur s'écarte de 1 à 2 mm de l'épaisseur normale.	Défaut qui permet d'atteindre le stade suivant sous réserve d'opérations particulières sur cet élément ou un autre.

Lors d'un contrôle on a relevé les épaisseurs suivantes (en mm) :

Epaisseur en mm	16,6	16,8	17,2	17,3	17,6	17,7	17,8	17,9	18	18,2	18,3	18,4	18,8
Effectif	1	1	1	1	3	1	2	2	1	2	1	3	1

1. a. Quel est le nombre de pièces présentant un défaut mineur ?

b. Quel est le nombre de pièces présentant un défaut majeur ?

c. Quel pourcentage de l'échantillon est constitué par des pièces présentant un défaut ?

2. a. Calculer l'épaisseur moyenne $\bar{x}$ de l'échantillon ; arrondir au centième de mm.

b. Calculer l'écart-type σ de l'échantillon.

c. Quel est le pourcentage de pièces dont l'épaisseur est dans l'intervalle $\left[\bar{x}-\sigma\,;\bar{x}+\sigma\right]$? Ces données semblent-elles gaussiennes ?

3. Si l'épaisseur moyenne s'écarte de plus de 0,25 mm de l'épaisseur normale, un réglage de la machine doit être effectué. Est-ce le cas ?

Correction

1. a. On a un défaut mineur si l'épaisseur est entre 17 et 17,5 mm ou lorsque l'épaisseur est entre 18,5 et 19 mm ; on a donc 3 tablettes qui ont ce défaut.

Epaisseur en mm	16,6	16,8	17,2	17,3	17,6	17,7	17,8	17,9	18	18,2	18,3	18,4	18,8
Effectif	1	1	1	1	3	1	2	2	1	2	1	3	1

b. On a un défaut majeur pour une épaisseur entre 16 et 17 mm ou entre 19 et 20 mm, soit 2 pièces.

c. Il y a donc 5 pièces présentant un défaut, soit 5 sur 20, soit 25 %.

2. a. b. Les calculs donnent :

moyenne $\bar{x}$	17,83

variance	0,29
ecart-type σ	0,54

c. Il y a 13 éléments dans l'intervalle $\left[\bar{x}-\sigma\,;\bar{x}+\sigma\right]=\left[17,28\,;18,37\right]$, soit 65 %. Il semblerait que les données soient gaussiennes.

3. $18-0,25=17,75$; or la moyenne est de 17,83, soit un écart de 0,17 mm. Il n'est donc pas nécessaire de faire de réglage.

Exercice 2 : Pluviométrie (bac)

On se propose dans cet exercice de comparer les régimes pluviométriques de différentes villes de Bretagne et de Provence : sur une période de 38 ans, on a mesuré, en millimètres d'eau par m^2, les quantités de pluie tombées chaque année sur chacune de ces villes (pour simplifier le langage, on donnera le nom de pluviométrie à ces quantités).

Partie I : dans la ville de Brest

On donne ci-dessous, les valeurs de pluviométrie de Brest (en millimètres par m^2), classées dans l'ordre croissant :

782,0	840	860,4	872,5	886,4	913,9	971,2	983,6	994,5	1029,7
1029,7	1031,8	1039,9	1045,7	1053,4	1061,1	1062,7	1097,8	1099,8	1101,0
1137,4	1140,2	1174,1	1180,9	1208,7	1209,6	1222,1	1224,1	1233,3	1238,9
1269,1	l269,4	1281,8	1297,2	1313,4	1383,3	I462,7	1603,6		

1. Calculer la moyenne de la pluviométrie à Brest pour les 6 années où la pluviométrie a été la plus faible.

2. Calculer, en détaillant, le premier quartile, la médiane et le troisième quartile de la série des 38 valeurs, puis compléter le tableau de l'annexe 1 à rendre avec la copie.

3. Construire la boîte à moustaches correspondant à la pluviométrie de Brest sur le diagramme figurant en annexe 2 à rendre avec la copie (comme pour les autres diagrammes, les extrémités des « pattes » seront constituées des premier et neuvième déciles, donnés dans le tableau de l'annexe 1).

Partie II : dans l'ensemble des villes

À l'aide des renseignements figurant dans les annexes 1 et 2, répondre aux questions suivantes :

1. Pour quelle ville, l'écart interquartile est-il le plus faible ? Combien cet écart vaut-il ?

2. Citer les villes dans lesquelles, pour au moins 50% des années, il est tombé plus de 900 mm d'eau par m^2.

3. Interpréter concrètement, en faisant une phrase, le fait que, pour la pluviométrie de Marseille, le premier quartile est égal à 447.

4. a. En observant les diagrammes de l'annexe 2, trouver la région dans laquelle se trouvent les deux villes ayant la pluviométrie la plus irrégulière. La réponse sera argumentée.

b. Quelles autres données, figurant dans l'annexe 1, permettent la même conclusion ?

Correction

Partie I : dans la ville de Brest

Pluviométrie de Brest (en millimètres par m^2), classées dans l'ordre croissant :

782,0	840	860,4	872,5	886,4	913,9	971,2	983,6	994,5	1029,7
1029,7	1031,8	1039,9	1045,7	1053,4	1061,1	1062,7	1097,8	1099,8	1101,0
1137,4	1140,2	1174,1	1180,9	1208,7	1209,6	1222,1	1224,1	1233,3	1238,9
1269,1	l269,4	1281,8	1297,2	1313,4	1383,3	I462,7	1603,6		

1. On fait la moyenne des six premiers termes du tableau : 859,2.

2. Q1 : 38/4=9,5 : on prend la moyenne du 9ème et du 10ème terme : 1012,1.

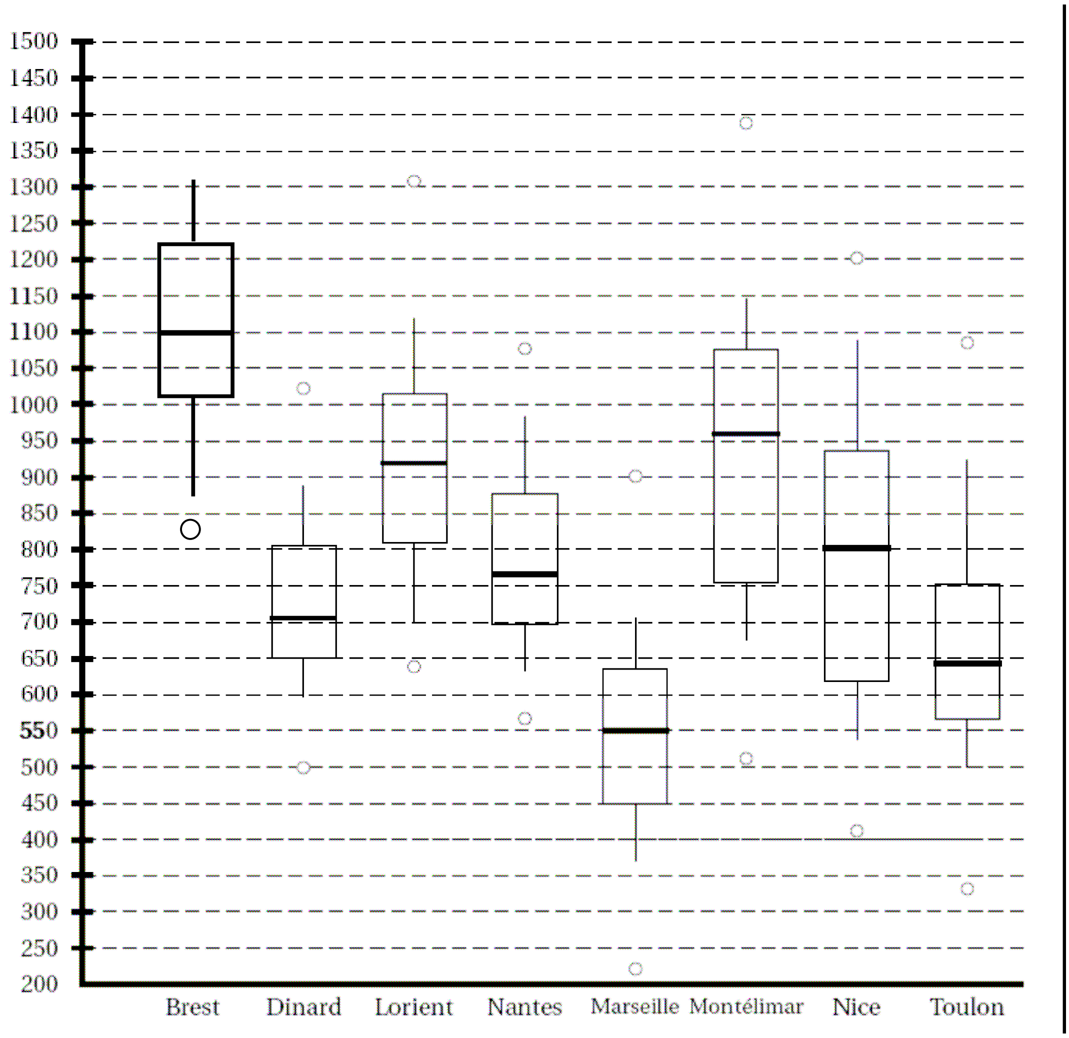
Médiane : 38/2=19 ; on prend le 19ème terme, soit 1099,8.

Q3 : 38*3/4=28,5 : on prend la moyenne du 28ème et du 29ème terme : 1228,7.

	Villes de Bretagne				Villes de Provence			
	Brest	Dinard	Lorient	Nantes	Marseille	Montélimar	Nice	Toulon
Minimum	**840**	499,7	638,1	567	221,7	512,1	412,4	331,2
D1	872,5	597,8	698,1	632,8	369,8	675,8	538	499,2
Q1	**1012,1**	649,6	808	696	447	753,8	617,2	564,6
Médiane	**1099,8**	705,35	919,2	765,55	549,65	959,45	801,3	642,9
Q3	**1228,7**	806,8	1016	878,6	636,6	1077,5	938,6	754,2
D9	1313,4	886,5	1119,4	982,8	705,8	1145,2	1089,3	922,4
Maximum	**1603,6**	1021,2	1307,5	1076,1	901,5	1389,2	1203,9	886,3

Moyenne	1121,2	725,6	919,8	791,3	543,7	927,6	798,9	673,6
écart- type	174,3	119,8	150,8	127,8	142,3	204,9	208,7	163,7

3.



Partie II : dans l'ensemble des villes

1. L'écart interquartile est le plus faible pour Dinard où il vaut 157,2.

2. Il faut que la médiane soit supérieure à 900 : cela concerne Brest, Lorient, Montélimar.

3. A Marseille le premier quartile est égal à 447, donc pendant 25 % des années il est tombé moins de 447 mm d'eau au m^2.

4. a. On a une pluviométrie irrégulière si l'intervalle interquartile est grand : les deux plus grand sont pour Nice et Montélimar.

b. L'écart-type donne le même type d'indication : c'est le plus grand dans ces deux mêmes villes.