STATISTIQUES

Vocabulaire

Définition

Une étude statistique s'effectue sur un ensemble appelé **population** dont les éléments sont appelés **individus**, et consiste à observer et étudier un même aspect sur chaque individu, appelé **caractère.**
Le caractère étudié prend un certain nombre de **valeurs** appelées **modalités**; si ces valeurs sont numériques, le caractère est dit **quantitatif** ; dans le cas contraire, il est dit **qualitatif**.

Exemple : Nombre de frères & sœurs des élèves de la classe :

- la population : la classe ;
- individus : élèves ;
- le caractère : nombre de frères et sœurs ;
- type : quantitatif.

Définition

On distingue 2 types de caractères quantitatifs ;

- si les valeurs prises sont peu nombreuses, le caractère est dit quantitatif **discret** ;
- si les valeurs sont nombreuses, le caractère est dit quantitatif **continue**.

Exemples :

- Nombre de frères & sœurs des élèves de la classe : caractère quantitatif discret ;
- Temps, en mn, passé devant la TV par les élèves de la classe : caractère quantitatif continu.

Définition

Lorsque le caractère est quantitatif continue, les valeurs du caractère étant très nombreuses, on regroupe les valeurs dans des intervalles de la forme $[a\,;\,b[$ appelés **classes** d'**amplitude** $b - a$ et de **centre** $\dfrac{a+b}{2}$.

Remarque : on perd ainsi en précision mais en gagne en simplicité !

Définition

→ L'**effectif** n_i d'une valeur x_i du caractère est le nombre d'individus de la population présentant cette valeur ;

→ L'**effectif total** n est le nombre total d'individus de la population; on a : $n = n_1 + n_2 + \ldots + n_p$;

→ La **fréquence** f_i d'une valeur x_i est le rapport de n_i à l'effectif total n de la population : $f_i = \dfrac{n_i}{n}$;

→ **L'effectif cumulé croissant** en x_i est le nombre $n_1 + n_2 + \ldots + n_i$; c'est donc le nombre d'individus dont la valeur du caractère est inférieure ou égale à x_i, (donc vaut au plus x_i).

Exemple : Nombre d'enfants dans les 30 familles des élèves d'une classe :

Nombre d'enfants	1	2	3	4	5
Effectifs	8	14	5	2	1
Effectifs cumulés croissants	8	22	27	29	30

- Effectif total : 30 ;
- Effectif de la valeur 2 : 14 ;
- Fréquence de la valeur 4 : $\dfrac{2}{30} \approx 0{,}0667$ soit 6,7% ;
- 27 familles ont au plus 3 enfants.

Exercice 1 :

Dans un lycée il y a 115 élèves en seconde; Compléter le tableau suivant :

Ages	14	15	16	17	18	TOTAL
Effectifs (à l'unité près)	8			26		
Fréquence (au centième)		0,28	0,38			
Effectifs cumulés croissants						

Présentation des données

Les données sont la plupart du temps présentées sous forme de **tableaux** d'effectifs ou de fréquences

- Variable discrète :

Valeurs de X	x_1	x_2	...	x_n
Effectifs	n_1	n_2	...	n_n
Fréquences	f_1	f_2	...	f_n

Exemple : Voici les notes à un devoir de 23 élèves d'une Seconde :

2.5 – 6 – 14 – 16.5 – 10.5 – 9 – 8 – 6 – 3 – 6 – 14 – 12 – 9 – 3 – 9 – 12 – 2.5 – 3 – 6 – 8 – 14 – 0 – 3

Notes	0	2,5	3	6	8	9	10,5	12	14	16,5	Total
Effectifs	1	2	4	4	2	3	1	2	3	1	23
Fréquences (en %)	4,3	8,7	17,4	17,4	8,7	13,0	4,3	8,7	13,0	4,3	100

Exercice 2 :

Ce tableau donne le nombre d'interventions d'un service de dépannage ; le compléter.

Nombre d'interventions	x_i	3	5	6	7	8	9	Total
Nombre de jours	n_i	2	4	9	6	3	1	
Fréquences (en %)	f_i							

- Variable continue : on regroupe les valeurs en **classes** :

Classes	$[a_0 ; a_1[$	$[a_1 ; a_2[$	...	$[a_{n-1} ; a_n[$
Effectifs	n_1	n_2	...	n_n

Exemple : On regroupe les notes du devoir précédent en 4 classes : [0 ; 5[, [5 ; 10[, [10 ; 15[et [15 ; 20] :

Notes	[0 ; 5[	[5 ; 10[	[10 ; 15[	[15 ; 20]	Total
Effectifs	7	9	6	1	23

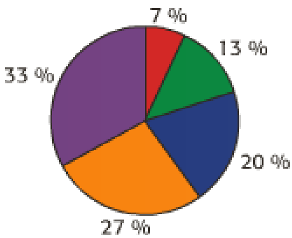
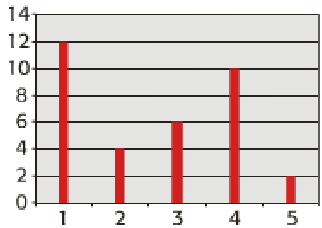
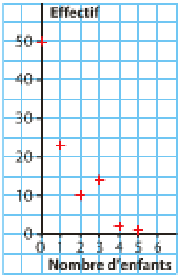
Représentations graphiques

On peut visualiser une série statistique par le biais de représentations graphiques.

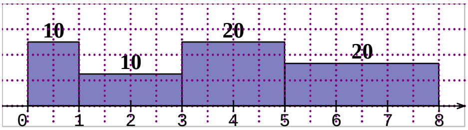
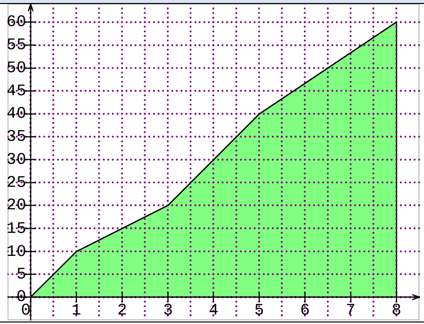
Selon la nature de la série statistique, on peut avoir recours à différents types de représentations graphiques.

Il existe de très nombreuses possibilités ; en voici quelques unes :

Caractères qualitatifs ou quantitatifs discrets

diagrammes à secteurs	diagramme en bâtons	nuage de points
secteurs angulaires d'angles au centre proportionnels aux effectifs	**bâtons** d'abscisses x_i et de hauteurs n_i (ou f_i)	**points** de coordonnées (x_i ; n_i) (ou (x_i ; f_i)
		

Caractères quantitatifs continues

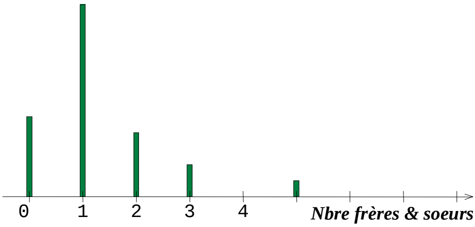
Histogramme	Courbe (ou polygone) des effectifs cumulés croissants
rectangles d'aires proportionnelles aux effectifs (ou fréquences)	**ligne brisée** reliant les points (x_i ; $n_1 + n_2 + \ldots + n_i$)
	

Exemples :

→ Nombre de frères et sœurs des élèves d'une classe :

Nombre de frères et sœurs	0	1	2	3	5
Effectifs	5	12	4	2	1

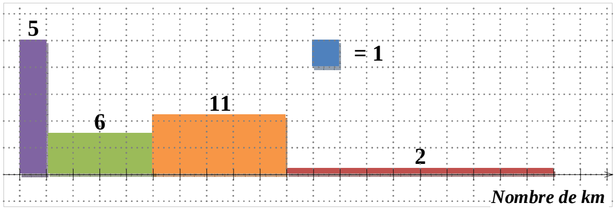
Diagramme en bâtons



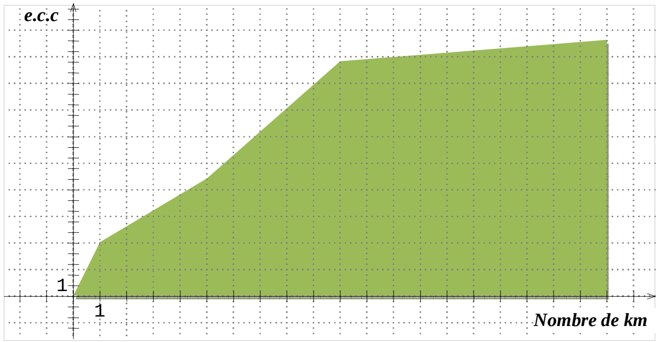
→ Nombre de km du domicile au lycée pour les élèves d'une classe :

Nombre de km	[0 ; 1[	[1 ; 5[	[5 ; 10[	[10 ; 20]
Effectifs	5	6	11	2
Effectifs cumulés croissants	5	11	22	24

Histogramme

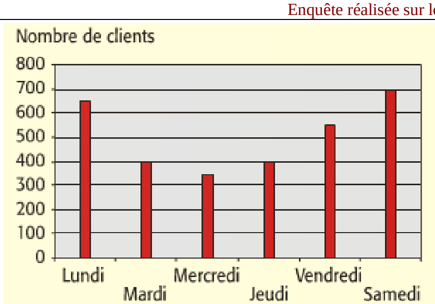


Polygone des effectifs cumulés croissants

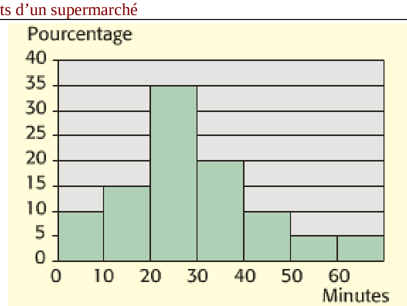


Exercice 3 :

Enquête réalisée sur les clients d’un supermarché



Nombre de clients selon le jour d’une semaine donnée



Temps passé par les clients cette même semaine

1. Nombre total de clients pour la semaine ?
2. Le jour où il y a le plus de clients ?
3. Pourcentage de clients restant moins de 30 mn ?

Exercice 4 :

En 2009, d’après une étude réalisée par le ministère de la Culture et de la Communication, la répartition des visiteurs dans les musées par groupes sociaux a été la suivante :

Etudiants	Classe populaire	Classe moyenne inférieure	Classe moyenne supérieure	Classe supérieure
9 %	32 %	16 %	22 %	21 %

Construire un diagramme à secteurs angulaires représentant cette répartition .

Exercice 5 :

Une machine remplit automatiquement des sachets de médicaments en poudre.
On a pesé un échantillon constitué de 100 sachets.
Les résultats sont les suivants :

Masse (en g)	[98 ; 98,5[	[98,5 ; 99[	[99 ; 99,5[	[99,5 ; 100[	[100 ; 100,5[	[100,5 ; 101]
Nombre	9	17	31	26	13	4

Représenter cette série par un histogramme.

Exercice 6 :

Pour mieux gérer les demandes de crédits de ses clients, le directeur d'une agence bancaire réalise une étude relative à la durée de traitement des dossiers. Une étude sur 50 dossiers a donné :

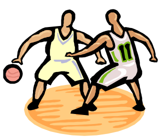
Durée (en mn)	[0 ; 10[	[10 ; 20[	[20 ; 30[	[30 ; 40[	[40 ; 50[
Nombre	5	10	17	12	6

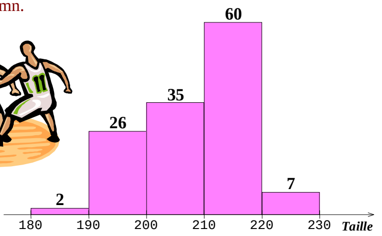
1. Tracer le polygone des effectifs cumulés croissants de cette série.
2. Calculer le pourcentage de dossiers don t l'étude est strictement inférieure à 15 mn.

Exercice 7 :

L'histogramme ci-contre représente la répartition par taille, en cm, de basketteurs de la NBA (le championnat professionnel des USA).

1. Dresser le tableau des effectifs cumulés croissants.
2. Représenter la courbe des effectifs cumulés croissants.
3. Combien mesurent au moins les 50 plus grands basketteurs de NBA ?





Paramètres d'une série statistique

Définition

On cherche à **résumer** une série à l'aide de quelques nombres significatifs, appelés **paramètres** ;
Il y en a de 2 sortes : les paramètres **de position** et les paramètres **de dispersion**.

Remarque : Pour résumer une série de mesures x_1, ..., x_n, on donne souvent un couple de nombres : un paramètre de position(ou valeur centrale) et un paramètre de dispersion qui permet d'évaluer à quel points les valeurs sont éloignées de la valeur centrale.

Paramétres de position

Moyenne

Définition

La **moyenne** d'une série statistique est le nombre réel, noté $\overline{x}$, défini par :

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + \cdots + n_px_p}{n} = f_1x_1 + \cdots + f_px_p \, .$$

Preuve : $\overline{x} = \dfrac{n_1x_1 + n_2x_2 + \cdots + n_px_p}{n} = \dfrac{n_1x_1}{n} + \dfrac{n_2x_2}{n} + \cdots + \dfrac{n_px_p}{n} = \dfrac{n_1}{n}x_1 + \dfrac{n_2}{n}x_2 + \cdots + \dfrac{n_p}{n}x_p = f_1x_1 + \cdots + f_px_p$.

Remarque : Dans le cas d'une série continue, on prend le centre de chaque classe pour x_i.

Résultat : linéarité de la moyenne

La moyenne des images des valeurs du caractères par une fonction affine $f(x) = ax + b$ est l'image de la moyenne : $f(\overline{x}) = a\overline{x} + b$.

Preuve :

$$\overline{x'} = \frac{n_1(ax_1+b)+\cdots+n_p(ax_p+b)}{n} = \frac{a(n_1x_1+\cdots+n_px_p)+b(n_1+\cdots+n_p)}{n} = a\frac{n_1x_1+\cdots+n_px_p}{n} + b\cancel{\frac{n_1+\cdots+n_p}{n}} = a\overline{x}+b = f(\overline{x}).$$

Résultat : moyenne & sous-groupes

Si une série z est composée de deux sous-séries x et y d'effectifs p et q, on a : $\overline{z} = \dfrac{p\overline{x}+q\overline{y}}{p+q}$.

Preuve :

$$\overline{z} = \frac{n_1x_1+\cdots+n_px_p+m_1y_1+\cdots+m_qy_q}{p+q} = \frac{p\times\dfrac{n_1x_1+\cdots+n_px_p}{p}+q\times\dfrac{m_1y_1+\cdots+m_qy_q}{q}}{p+q} = \frac{p\overline{x}+q\overline{y}}{p+q}.$$

✎ **Exercice 8:**

Calculer la moyenne dans chacun des cas suivants :

1. Notes d'un élève : A a obtenu : 6 - 5 - 10 - 13 - 9 - 5 - 13 - 12.
2. Répartition par âge des élèves de secondes dans un lycée :

Ages	14	15	16	17	18
Fréquences	0,07	0,28	0,38	0,23	0,04

3. Répartition des accidents corporels de la route en 2000 :

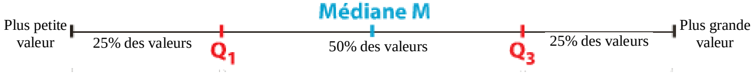
Tranche horaire	[0 ; 3 [	[3 ; 6 [	[6 ; 9 [	[9 ; 12 [	[12 ; 15[	[15 ; 18[	[18 ; 21[	[21 ; 24[
Nombre d'accidents	4 550	3 230	8 220	9 050	12 040	16 040	16 820	10 050

4. Moyenne à un contrôle de Maths sachant que le 1er groupe (de 14 élèves) a obtenu une moyenne de 8,3 et l'autre groupe (de 16 élèves) une moyenne de 10,5.

Médiane & Quartiles

Définition

- La **médiane M** est une valeur telle qu'il y a autant de données inférieures que supérieures à M.
- Le 1er **quartile** Q_1 est la plus petite donnée telle qu'au moins 25% des données lui soient inférieures ou égales ;
- Le 3ième **quartile** Q_3 est la plus petite donnée telle qu'au moins 75% des données lui soient inférieures ou égales .



Détermination de la médiane et des quartiles

→ d'une variable discrète

Résultat

Pour déterminer la médiane et les quartiles d'une variable discrète, on ordonne par ordre croissant les n valeurs alors nommées x_1, x_2, … et x_n; Puis :

→ $Q_1 = x_p$ où p est le 1er entier supérieur ou égale à $0{,}25n$;

→ $Q_3 = x_q$ où q est le 1er entier supérieur ou égale à $0{,}75n$;

→ M = valeur centrale si n impair et moyennes des 2 valeurs centrales si n pair, soit :

$$n=2p \Rightarrow M = \frac{x_p+x_{p+1}}{2} \quad \text{et} \quad n=2p+1 \Rightarrow M = x_{p+1}.$$

Exemple : Dans une maternité, on a référencé les périmètres crâniens à la naissance de 290 nouveaux nés :

Périmètre (en cm)	32	32,5	33	33,5	34	34,5	35	35,5	36	36,5	37	37,5
effectif	4	19	17	20	59	62	43	20	18	18	4	6
e.c.c.	4	23	40	60	119	181	224	244	262	280	284	290

$n = 290$;

$\rightarrow$ $0{,}25\times n = 72{,}5 \Rightarrow Q_1 = x_{73} = 34$;

$\rightarrow$ $0{,}75\times n = 217{,}5 \Rightarrow Q_3 = x_{218} = 35$;

$\rightarrow$ $290 = 2\times 145 \Rightarrow M = (x_{145} + x_{146})/2 = (34{,}5 + 34{,}5)/2 = 34{,}5$.

Exercice 9 :

Les salaires mensuels en € des employés d'une petite entreprises sont :

1 000 ; 1 000 ; 1 200 ; 1 200 ; 1 200 ; 1 500 ; 1 500 ; 2 000 ; 2 500 ; 3 100.

1	L'effectif total est	11	6	10	20
2	Le nombre de salariés dont le salaire est au moins de 1 200€ est	2	8	5	10
3	La fréquence de salariés gagnant moins de 1 500 € est	0,50	0,33	0,30	0,70
4	Dans une représentation par un diagramme circulaire, l'angle associé à la valeur 1 200 est	60°	120°	108°	72°
5	Le salaire moyen est	1 620 €	1 883 €	1 350 €	1 200 €
6	Le salaire médian est	1 630 €	1 350 €	1 200 €	1 500 €
7	Le 1er quartile est	3	1 000 €	1 200 €	1 500 €

Exercice 10 :

Déterminer la médiane M et les quartiles Q_1 et Q_3 dans chacun des cas suivants :

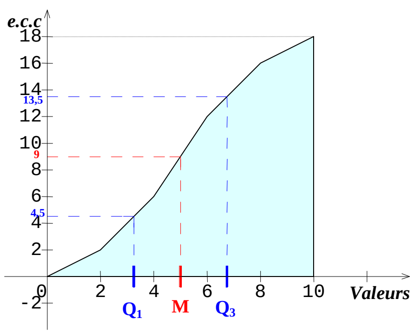
A	B	C
6 5 10 13 9 5 13 12	2 3 5 6 8 9 9 10 10	6 6 10 12 12 13 14 14 15 16

$\rightarrow$ d'une variable continue

Résultat

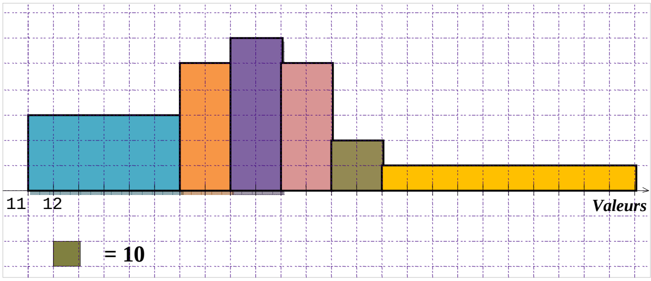
Pour déterminer la médiane et les quartiles d'une variable continue, on construit la courbe des e.c.c.;
Q_1 , M et Q_3 sont alors les abscisses des points de la courbe d'ordonnées respectives $0{,}25n$, $0{,}5n$ et $0{,}75n$.

Exemple :



Exercice 11 :

Voici l'histogramme d'une série :



1. Compléter le tableau suivant :

Classes	[11 ; 17[	[17 ; 19[	[19 ; 21[	[21 ; 23[	[23 ; 25[	[25 ; 35[
Effectifs						
Fréquences						
E.C.C.						

2. Quelle est la moyenne de cette série ?
3.
   a. Construire le polygone des e.c.c. associé à cette série ;
   b. En déduire sa médiane et ses quartiles.

Mode

Le **mode** est la valeur la plus souvent prise.

***Remarques* :**

- il peut y avoir plusieurs modes ou classes modales ;
- le mode n'a d'intérêt que si une valeur (ou plusieurs) est beaucoup plus souvent prise que les autres.

Mesures de dispersion

Etendue

Définition

L'**étendue** est l'écart entre la plus petite et la plus grande valeurs de la série : $e = x_{max} - x_{min}$.

Ecart interquartile

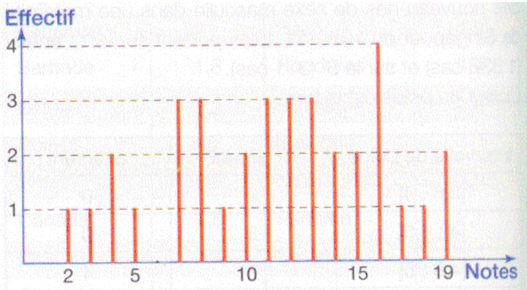
Définition

L'**écart interquartile** est l'écart entre le 1er et le 3ème quartiles : $I = Q_3 - Q_1$.

***Remarque* :** l'écart interquartile est moins sensible que l'étendue aux données extrêmes.

Exercice 12 :

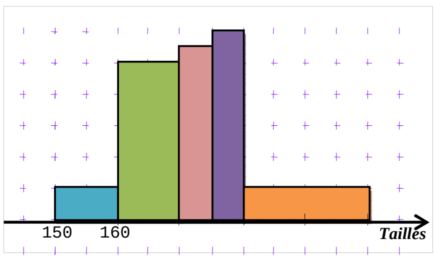
Voici le diagramme en bâtons des résultats d'une classe de 35 élèves :



1. Construire un tableau des effectifs de cette série discrète.
2. Déterminer la moyenne, la médiane, les quartiles, le mode, l'étendue et l'intervalle interquartile de cette série :
   a. à la main ;
   b. à la calculatrice.
3. On crée 4 classes de même amplitude [0 ; 5[, [5 ; 10[, [10 ; 15[et [15 ; 20] que l'on appelle respectivement « très faible », « insuffisant », « correct » et « très bien ». Etablir la distribution, puis construire l'histogramme obtenus à partir de ces classes.

Exercice 13 :

1375 élèves sont inscrits dans un lycée. Cet histogramme indique la répartition des élèves de ce lycée selon leur taille, en cm.



1. Etablir un tableau des effectifs, des fréquences et des e.c.c.
2. Calculer la moyenne.
3. Construire la courbe des fréquences cumulées croissante et calculer la médiane et les quartiles Q_1 et Q_3 et les déciles D_1 et D_9.
4. Déterminer l'étendue et l'écart interquartile.

Résumé d'une série statistique

On résume souvent les séries statistiques en donnant un indicateur de position et un de dispersion.

Les couples choisis sont en général :

- **moyenne** / **étendue** : ce couple a l'inconvénient d'être sensible aux valeurs extrêmes, lesquelles peuvent parfois masquer les caractéristiques principales des autres valeurs de la série.

ou

- **médiane** / **intervalle interquartile** : ce couple a l'avantage d'être peu sensible aux valeurs extrêmes.

ou

- **moyenne** / **écart type** (vue plus tard !).

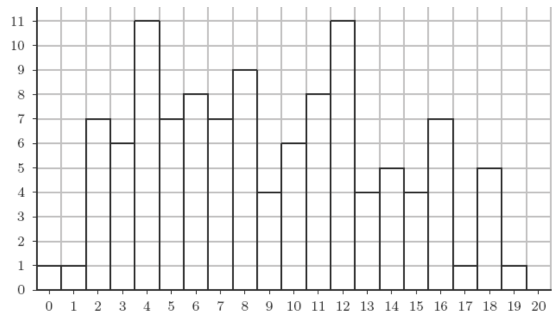
Remarque : Le couple médiane – intervalle interquartile est souvent représente graphiquement par un *diagramme en boîte*, appelé parfois *boîte à moustaches* ou *boîte à pattes*.

Diagramme en boîte :

en lieu et place de :

Exercice 14 :

1. L'histogramme ci-dessous représente les notes obtenues en mathématiques lors du brevet par une classe de collège. La hauteur de chaque barre représente le nombre d'élèves ayant obtenus la note correspondante.



1.

a. Remplir le tableau et en déduire les données statistiques qui suivent.

Note	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Effectif	1	1	7																		
Effectif cumulé	1	2	9																		

b. Compléter (grâce à la calculatrice) :

Min =..... ; Q1 = ... ; Med =..... ; Q3 = ... ; Max =..... ; Mode =..... ; $\overline{x}$ =. . . . .

2. Au niveau départemental, les résultats obtenus sont les suivants :

Notes	[0 , 4[	[4 , 8[	[8 , 12[	[12 , 16[	[16 , 20]
Pourcentage des candidats	25	20	35	15	5
Pourcentage cumulés					

a. Calculer la moyenne obtenue au brevet dans le département.
b. Construire la courbes des pourcentages cumulés croissants.
c. Déterminer graphiquement la médiane et les quartiles de la série départementale : Q1 = ... ; Med = ... ; Q3 = ...

3.

a. Construire sur un même graphique les diagrammes en boîte associés respectivement au collège et au département.
b. Comparer les notes du collège à celles du département.

☆ ☆ ☆ ☆ ☆ ☆ ☆