Etude statistique

Définition

La **statistique** est la branche des mathématiques appliquées qui a pour objet l’étude des phénomènes mettant en jeu un grand nombre d’éléments (valeurs numériques, notes, noms, couleurs …).

(*Hachette dictionnaire encyclopédique)*

Population et individu

- La **population** est l’ensemble sur lequel porte l’étude.
- Les **individus** sont les éléments qui composent la population.

Exemple : Si on fait une étude sur le nombre de kilomètres des voitures garées sur le parking du lycée ; la population est l’ensemble des voitures garées sur ce parking et un individu est une voiture garée sur ce parking.

Caractère

Le **caractère** est l'aspect ou la propriété observée et analysée.

Exemple : Dans l'exemple précédent, le caractère est le nombre de kilomètres des voitures garées sur le parking.

Il y a deux types de caractères :

Le caractère peut être quantitatif (du mot quantité) c'est-à-dire mesurable ou qualitatif (du mot qualité) c'est-à-dire non mesurable.

Exemples

Caractère quantitatif (mesurable)	Caractère qualitatif (non mesurable)
• Taille d'un élève.	• Couleur de cheveux d'un élève.
• Nombre de voix obtenues à une élection.	• Marque des voitures garées sur le parking du lycée.
• Pointure de chaussure d'un élève de la classe de seconde.	• Destination préférée des français pour partir en vacances.

Caractères quantitatifs

Les caractères quantitatifs se divisent eux même en deux types :

- **Caractère quantitatif continu :** Le caractère est mesurable et peut prendre toutes les valeurs d'un intervalle.
- **Caractère quantitatif discret :** Le caractère est mesurable mais ne peut pas prendre de valeurs intermédiaires.

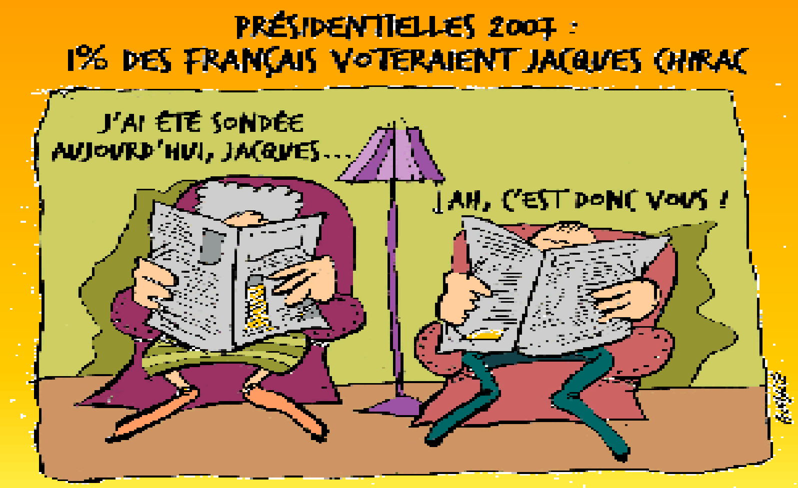
Exemples

Caractère quantitatif continu	Caractère quantitatif discret
• Poids d’un élève. • Montant de la dépense occasionnée à l’occasion de la rentrée des classes. • Temps d’un coureur pour parcourir 100 m.	• Nombre de voix obtenues à une élection. • Pointure de chaussure d’un élève de la classe seconde. • Nombre de voitures par foyer.

Un **echantillon** est une partie de la population.

Lorsque la population est trop grande, pour faire un sondage, on utilise un échantillon.

Par exemple, pour savoir qui du candidat N ou S va devenir président(e) on appelle 1000 français inscrits sur les listes électorales mais on ne peut pas appeler tous les électeurs.



Echantillon représentatif ou biaisé

Pour que le sondage soit valable, il faut que l'échantillon soit représentatif c'est-à-dire considéré comme le modèle, le type de la population.

Exemple : 1000 personnes choisis selon la méthode des quotas (de différents sexe, âge, revenus, origines, situation géographique ….).

Quand l'échantillon n'est pas représentatif ; on dit qu'il est biaisé.

Exemple : 1000 personnes habitant à Paris et dont le revenu mensuel est supérieur à 5000 €.

Effectif et fréquence

• Une **série statistique** représente l'ensemble des valeurs collectées.

• L'**effectif** c'est le nombre d'individus de la population ayant une valeur donnée (pour le caractère étudié).

• La **fréquence** c'est le quotient de l'effectif de la valeur par l'effectif total.

$$\boxed{\text{fréquence} = \frac{\text{effectif}}{\text{effectif total}}}$$

• Les valeurs extrêmes sont : la valeur maximale $x_{\max}$ et la valeur minimale $x_{\min}$.

• L'étendue *e* est la différence entre les valeurs extrêmes.

$$\boxed{e = x_{\max} - x_{\min}}$$

• Le mode est la valeur la plus fréquente, c'est-à-dire, celle ayant le plus grand effectif.

Si les valeurs sont regroupés en classe (intervalles), le mode est en fait une classe modale.

Moyenne

La moyenne de la série statistique suivante :

Valeur	x_1	x_2	$\dots$	x_p
Effectif	n_1	n_2	$\dots$	n_p

est le nombre noté $\bar{x}$ défini par :

$$\boxed{\bar{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + ... + n_k \times x_k}{n_1 + n_2 + ... + n_k}}$$

Si les valeurs sont regroupés en classe (intervalles), on calcule la moyenne en choisissant comme valeurs du caractère <u>les centres des classes</u>.

Exemple : Soit la série statistique suivante :

1 ; 100 ; 98 ; 101 ; 101 ; 100 ; 106 ; 990

La moyenne de cette série est 199,625.

Les deux valeurs extrêmes (1 et 990) sont des valeurs exceptionnelles ; on peut calculer la moyenne de la série privée de ces deux valeurs ; on dit qu'il s'agit d'une moyenne élaguée.

Dans cet exemple, la moyenne élaguée est :

$$\frac{100+98+101+101+100+106}{6}=101$$

La médiane Me d'une série statistique partage cette série en deux parties de telle sorte que :

- Au moins la moitié des valeurs sont inférieures ou égale à la médiane.

- Au moins la moitié des valeurs sont supérieures ou égale à la médiane.

Exemples :

- La médiane de la série : 2 ; 3 ; 5 ; 10 ; 12 ; 19 ; 20 est 10.
- La médiane de la série : 2 ; 3 ; 5 ; 10 ; 12 ; 19 est : $\frac{5+10}{2} = 7,5.$

Calcul de la médiane

Si la série contient *n* valeurs rangées dans l'ordre croissant :

- Si *n* est pair ; la médiane est la $\frac{n+1}{2}$ iéme valeur de la série.
- Si *n* est impair ; la médiane est la demi somme des $\frac{n}{2}$ ième et $\frac{n}{2}+1$ ième valeurs de la série.

Exemples :

Valeur	1	2	3	7
Effectif	3	5	4	9

$$n = 3+5+4+9 = 21 \text{ impair}$$

$$\frac{21+1}{2} = 11$$

La médiane est la 11e valeur donc : 3

Valeur	1	2	3	7
Effectif	3	5	4	12

$$n = 3+5+4+12 = 24 \text{ pair}$$

$$\frac{n}{2} = \frac{24}{2} = 12 \quad \frac{n}{2}+1 = 13$$

$$Me = \frac{12^e \text{ valeur} + 13^e \text{ valeur}}{2} = \frac{3+7}{2} = 5$$

Exercice 1 : énoncé

Calculer la médiane des séries statistiques suivantes :

1.

Valeur	4	7	9	13	14	18
Effectif	2	5	6	8	6	4

2.

Valeur	1	2	4	5	6	7
Effectif	3	7	10	12	7	1

Exercice 1 : aide

Commencer par calculer l'effectif total puis appliquer la méthode de la diapositive 15.

Exercice 1 : solution

1.

Valeur	4	7	9	13	14	18
Effectif	2	5	6	8	6	4

$n = 31$ impair

$$\frac{n+1}{2} = \frac{32}{2} = 16$$

La médiane est la valeur en 16ème position : 13.

2.

Valeur	1	2	4	5	6	7
Effectif	3	7	10	12	7	1

$n = 40$ pair

$$\frac{n}{2} = 20 \qquad \frac{n}{2} + 1 = 21$$

La médiane est la moyenne des valeurs en 20ème et 21ème position : 4 et 5 donc 4,5.

On effectue des essais sur un échantillon de 220 lampes électriques afin de tester leur durée de vie, exprimée en heures ; voici les résultats :

Durée	Effectifs	Fréquences
$[1000;1200[$	6	
$[1200;1300[$	14	
$[1300;1400[$	25	
$[1400;1500[$	75	
$[1500;1600[$	80	
$[1600;1700[$	10	
$[1700;1800[$	8	
$[1800;2100[$	2	

1. Compléter la colonne des fréquences.

2. Déterminer la moyenne de cette série.

3. Déterminer la classe modale.

Exercice 2 : aide

1. Fréquence = effectif / effectif total
2. Remplacer chaque classe (intervalle) par son centre.
3. La classe modale est la classe la plus fréquente, c’est-à-dire celle qui a le plus grand effectif

Milieu	Durée	Effectif	Fréquences
1100	$[1000;1200[$	6	0,03
1250	$[1200;1300[$	14	0,06
1350	$[1300;1400[$	25	0,11
1450	$[1400;1500[$	75	0,34
1550	$[1500;1600[$	80	0,36
1650	$[1600;1700[$	10	0,04
1750	$[1700;1800[$	8	0,04
1950	$[1800;2100[$	2	0,01

2. $$\frac{1100\times 6+1250\times 14+1350\times 25+1450\times 75+1550\times 80+1650\times 10+1750\times 8+1950\times 2}{6+14+25+75+80+10+8+2}=1477$$

3. Classe modale : $[1500;1600[$

Exercice 3 : énoncé

Voici les résultats d'une enquête auprès de 55 mères de famille portant sur l'âge de la mère à la naissance du premier enfant :

15	17	18	22	23	17	21	19	21	25	23	26	21
20	24	15	25	21	19	22	23	22	20	21	32	23
25	21	18	22	24	22	18	25	26	30	22	28	27
24	21	20	19	21	20	24	26	20	28	30	27	28
						38	35	32				

1) Quel est l'étendue de la série statistique ?

2) Compléter le tableau suivant :

âges	15	17	18	19	20	21	22	23	24	25	26	27	28	30	32	35	38
effectifs																	

3) Quelle est le mode de la série ?

4) Calculer l'âge moyen de la mère à la naissance du premier enfant.

5) Calculer la valeur de la médiane de la série.

Exercice 3 : aide

En comptant, le nombre d'occurrences de chaque âge, pour être sur de ne pas en oublier, on pourra barrer chaque âge comptabilisé et inscrit dans le tableau.

Exercice 3 : solution

1) L'étendue *e* de la série est l'écart maximal entre deux âges soit :

$$e = 38 - 15 = 23 \text{ ans}$$

2)

âges	15	17	18	19	20	21	22	23	24	25	26	27	28	30	32	35	38
effectifs	2	2	3	3	5	8	6	4	4	4	3	2	3	2	2	1	1

3) Le mode de la série est l'âge le plus fréquemment cité soit 21 ans.

4) Age moyen de la mère à la naissance du premier enfant :

$$\frac{15\times2+17\times2+18\times3+19\times3+20\times5+21\times8+.....+35\times1+38\times1}{2+2+3+3+5+8+....+1+1} = 23,2$$

Ce qui fait : 23 ans ; 2 mois et 11 jours.

(0,2 années représentent 0,2×365 soit 73 jours)

5) Médiane de la série : $n = 55$ impair $\dfrac{n+1}{2} = \dfrac{56}{2} = 28$

La médiane est la 28ème valeur de la série soit 22 ans.

Exercice 4 : énoncé

On reprend l’exercice précédent et on a groupé l'âge des mères par classe d'amplitude 6 ans, la première classe étant [15 ; 21[.

âges	[15 ; 21[	[21 ; 27[	[27 ; 33[	[33 ; 39[
effectifs	15	29	9	2

1) Représenter ces données dans un diagramme circulaire.

2) Représenter ces données dans un histogramme.

Exercice 4 : aide

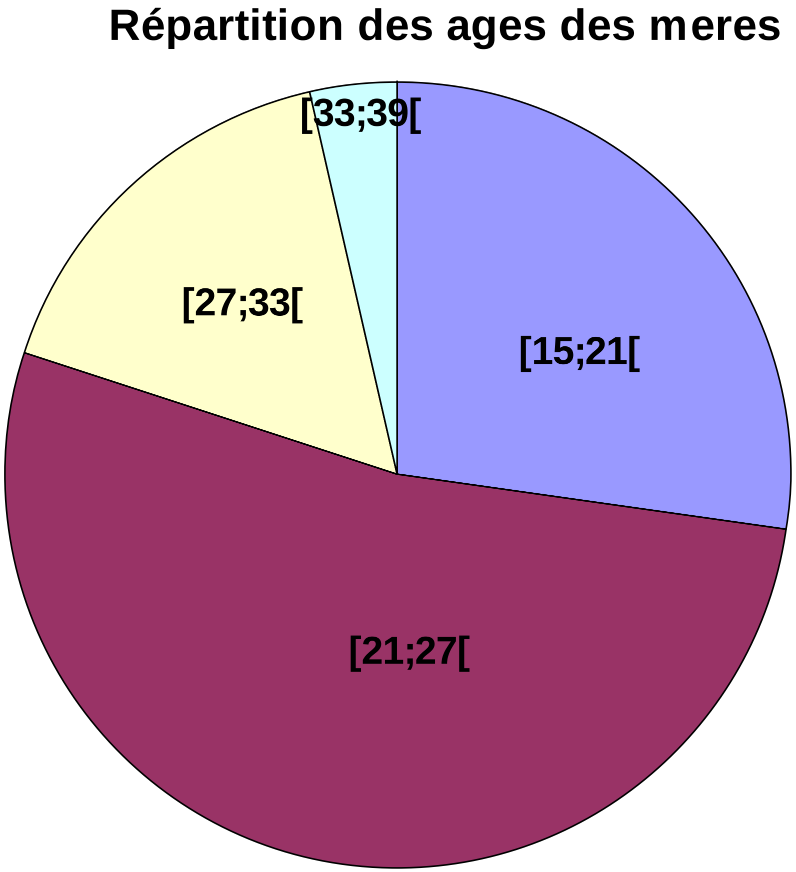
1) Pour le diagramme circulaire ; calculer l'angle de chaque classe en utilisant le fait que l'effectif total représente le tour complet soit 360° puis en faisant des produits en croix.

2) Pour l'histogramme, les classes sont de même amplitude ; choisir une échelle adaptée.

âges	[15 ; 21[	[21 ; 27[	[27 ; 33[	[33 ; 39[
effectifs	15	29	9	2

1) Calcul des angles de chaque classe :

L'effectif total (55) représente 360°



$$55 \rightarrow 360°$$

$$15 \rightarrow \frac{15 \times 360}{55} = 98°$$

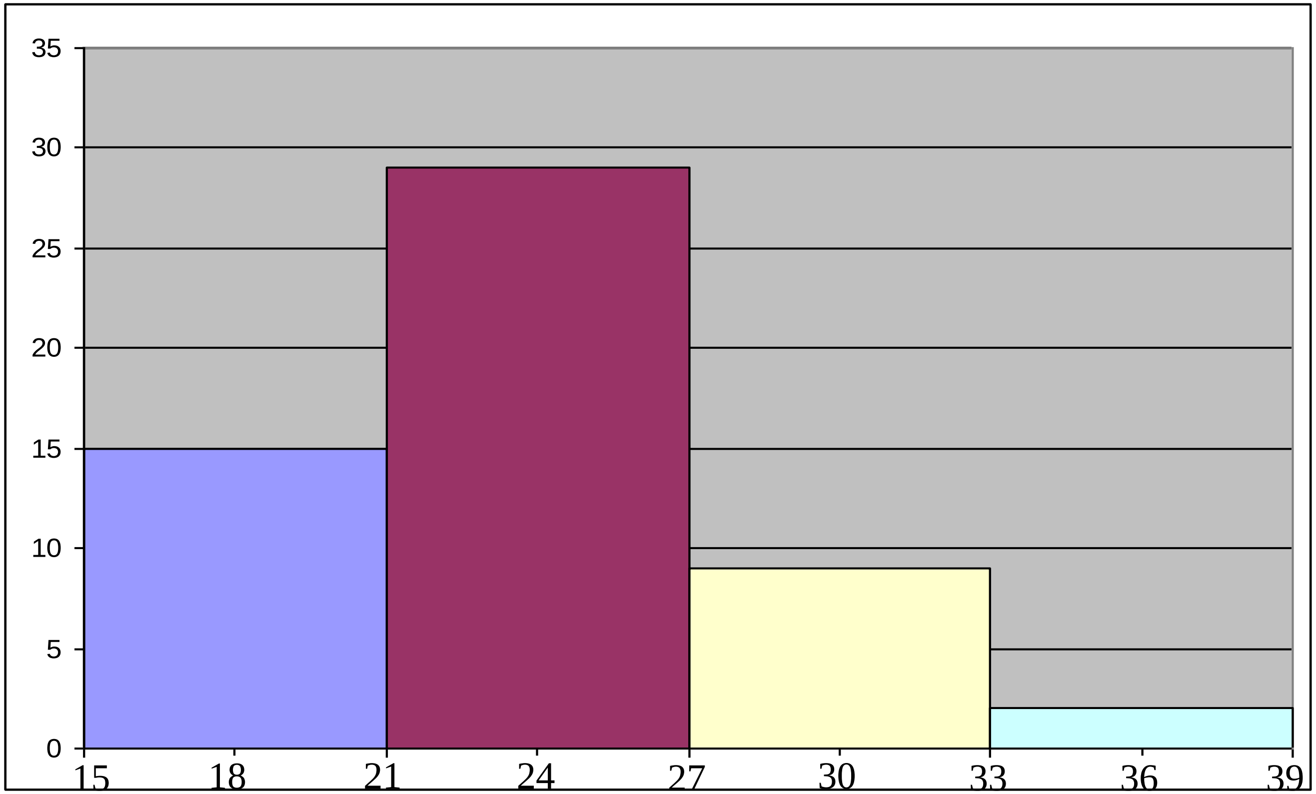
$$29 \rightarrow \frac{29 \times 360}{55} = 190°$$

$$9 \rightarrow \frac{9 \times 360}{55} = 59°$$

$$2 \rightarrow \frac{2 \times 360}{55} = 13°$$

2) Comme les classes ont le même effectif, il suffit de choisir une échelle ; par exemple :

- 1 cm pour 3 années en abscisses,
- 1 cm pour 5 effectifs en ordonnées.



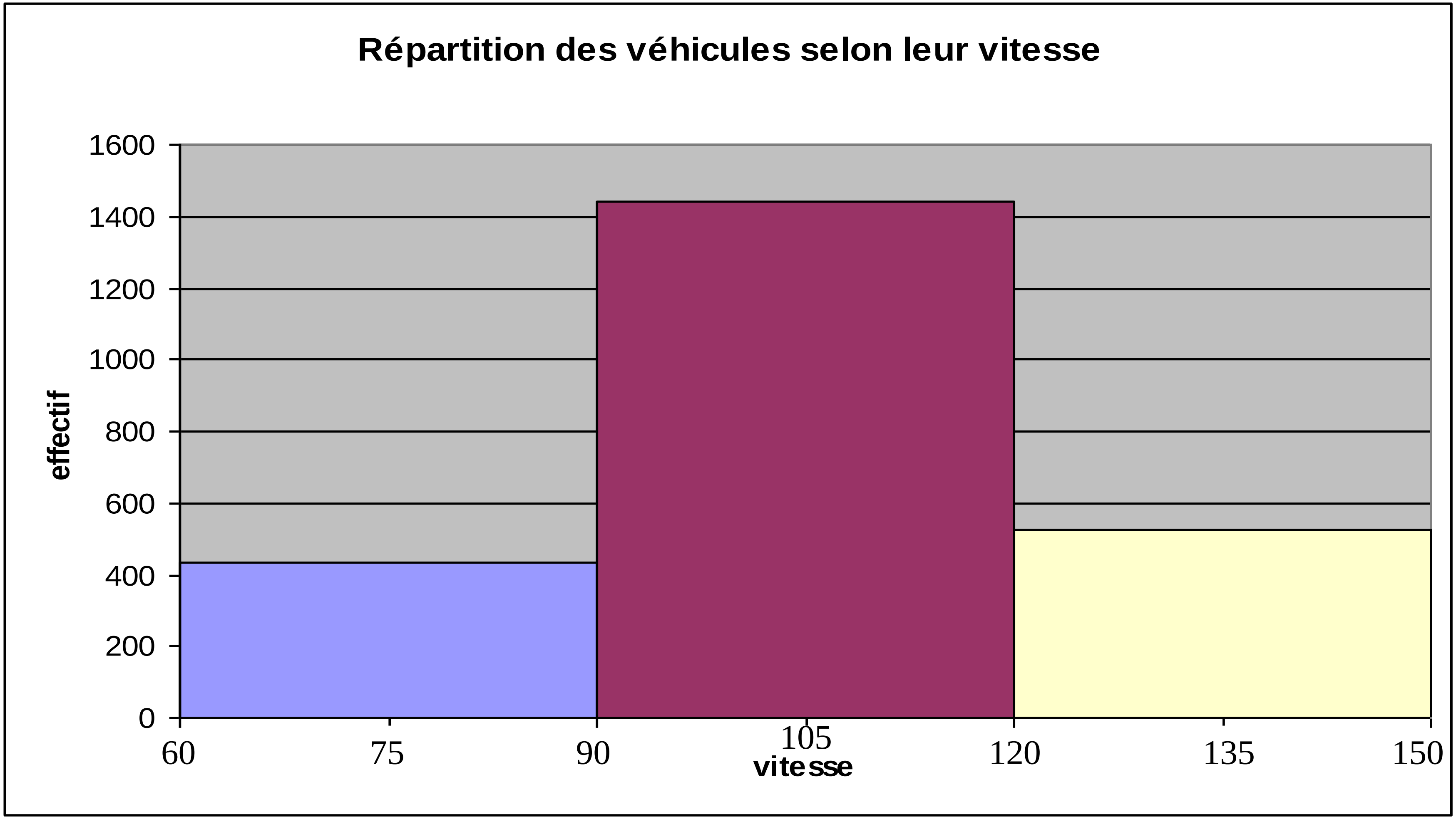
Sur une autoroute, des relevés de vitesse par radar, effectués sur 2400 véhicules ont permis de dresser le tableau ci-après. Les vitesses sont exprimées en km/h.

Vitesse	Nombre de véhicules
[60 ; 90[	432
[90 ; 120[	1440
[120 ; 150[	528

Représenter par un histogramme la répartition des véhicules suivant leur vitesse.

Exercice 5 : aide

Les classes sont de même amplitude ; choisir une échelle adaptée.



Échelle : une graduation représente 15 km/h en abscisses

une graduation représente 200 effectifs en ordonnées

Evaluons nous

Si l’on fait une étude sur la durée moyenne d’une communication téléphonique, alors, le caractère étudié est :

Note	5	10	12	15
Fréquence	0,2	0,5	0, 1	0,2

1) Qualitatif

2) Quantitatif discret

3) Quantitatif continu

Réponse :

3) Si l'on fait une étude sur la durée moyenne d'une communication téléphonique, alors, le caractère étudié est quantitatif continu.

QCM : Question 2

Quelle est la moyenne de la série suivante ?

Valeur	[0 ; 4]	[4 ; 6]	[6 ; 10]
Effectif	120	20	20

1) 53,33

2) 3,125

3) 5

Réponse :

2) 3,125

En effet, $$\bar{x} = \frac{2 \times 120 + 20 \times 5 + 20 \times 8}{(120 + 20 + 20)}$$

$$\bar{x} = \frac{500}{160}$$

$$\bar{x} = 3{,}125$$

Quelle est la médiane de la série suivante ?

Valeur	11	12	13	14
Effectif	5	50	45	20

1) 12

2) 13

3) 45

<u>Réponse :</u>

2) 13

En effet, l'effectif total est n = 5 + 50 + 45 + 20 = 120

n est pair , donc la médiane est la moyenne des valeurs en 60ème et 61ème position : $\frac{13+13}{2}=13$