

TD 4 : STATISTIQUES DESCRIPTIVES : PRÉSENTATION DES DONNÉES

SECTION 1 ET 2. VOCABULAIRE, EFFECTIFS ET FRÉQUENCES

Ex 1. Une agence propose des appartements meublés en location pour les vacances. Elle dispose d'un parc de 551 appartements sur toute la France. Pour chaque logement, on a les renseignements suivants :

- Nombre de **Couchages**
- **Département**
- **Surface** en m^2
- **Type** de vacances : "Ville" "Balnéaire" "Montagne" "Rurale"
- **Standing** : "Économique" "Confort" "Élégant" "Luxe"

a. Déterminer la nature des variables statistiques considérées :

Variable	Qualitative	Quantitative	Ordonnée	Non-ordonnée	Continue	Discrète
Nombre de Couchages						
Département						
Surface						
Type de vacances						
Standing						

b. En suivant les consignes présentées en cours, indiquez quel(s) type(s) de graphique vous utiliserez pour présenter les fréquences de chaque variable.

Variable	Diagramme circulaire	Diagramme en bâton	Histogramme
Nombre de Couchages			
Département			
Surface			
Type de vacances			
Standing			

Solution. a.

Variable	Qualitative	Quantitative	Ordonnée	Non-ordonnée	Continue	Discrète
Nombre de Couchages		X	X			X
Département	X			X		X
Surface		X	X		X	
Type de vacances	X			X		X
Standing	X		X			X

Quelques commentaires :

- Une variable quantitative est une variable à laquelle on associe une quantité (un nombre). Ce n'est le cas que pour le nombre de couchage et la surface (nombre de mètres carrés). Attention au département : si l'on peut associer à chaque département un *numéro*, ce numéro ne représente pas une quantité. Par exemple, il n'y a aucun sens à faire une moyenne de départements ou de numéros de téléphones.
- Les variables quantitatives (nombre de couchages, surface) sont ordonnées. Le standing aussi (un appartement a un meilleur stading qu'un autre). Le département ou le type de vacances ne sont pas ordonnés (des vacances à la montagne ne sont pas objectivement meilleures que des vacances rurales).
- Une variable discrète ne peut prendre qu'un nombre fini de valeur, alors qu'une variable continue peut prendre toutes les valeurs dans un intervalles (en particulier des valeurs décimales).

b. Suivant les recommandations du cours :

Variable	Diagramme circulaire	Diagramme en bâton	Histogramme
Nombre de Couchages		X	
Département		X	
Surface			X
Type de vacances	X		
Standing		X	

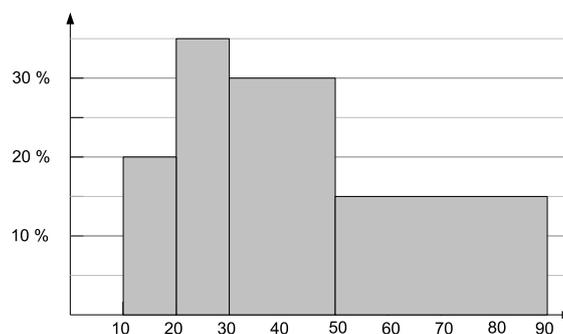
Pour le département : même s'il s'agit d'une variable discrète non-ordonnée, le nombre de modalités est bien trop élevé pour utiliser un diagramme circulaire, et même un diagramme en bâtons est discutable. Un histogramme n'est pas adapté (en l'absence d'ordre). Il faudrait réfléchir à d'autres représentations (réunir les départements par région pour réduire le nombre de modalités ? Faire une carte ?).

SECTION 3. TRI À PLAT PAR CLASSES

Ex 2. Le tableau des fréquences de l'âge des abonnés à la revue "Tattoo & Piercing " est :

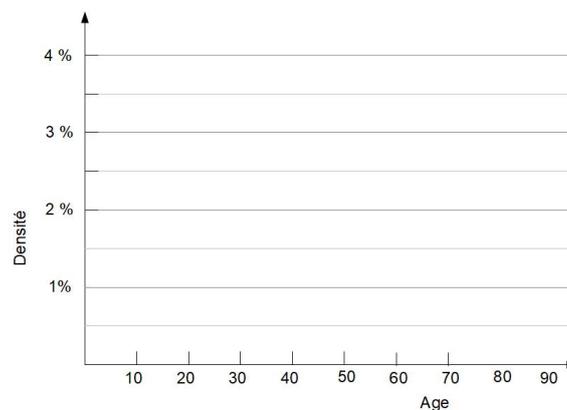
Âge	[10,20[[20,30[[30,50[[50,90[
Fréquence	20%	35%	30%	15%

a. En regardant le graphique, quelle est la tranche d'âge qui semble la plus représentée ? Pourquoi a-t-on cette impression ? Ce graphique est-il une bonne représentation des données ?



b. Compléter le tableau ci-dessous et dessiner l'histogramme associé.

Âge	[10,20[[20,30[[30,50[[50,90[
Fréquence	20%	35%	30%	15%
Amplitude				
Densité				



c. Quelle est l'unité de la densité (axe des ordonnées) ?

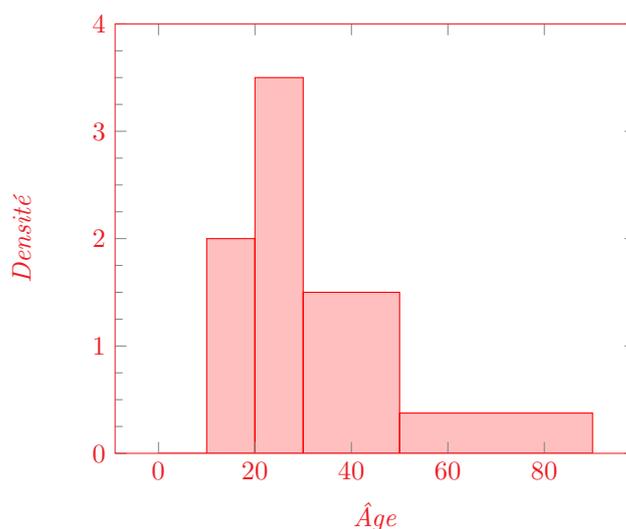
d. Selon les données dont on dispose, environ quelle proportion de lecteurs ont 15 ans ? Et 41 ans ?

Solution. a. A priori, il semble que la tranche d'âge la plus représentée est celle de 50 à 90 ans, car

le rectangle correspondant “domine” le graphique (aire la plus grande). Ce n’est pas ce qu’indique le tableau des fréquences (il y a seulement 15% de la population dans cette tranche, contre plus de 30% dans les deux tranches précédentes). Ce graphique est donc trompeur.

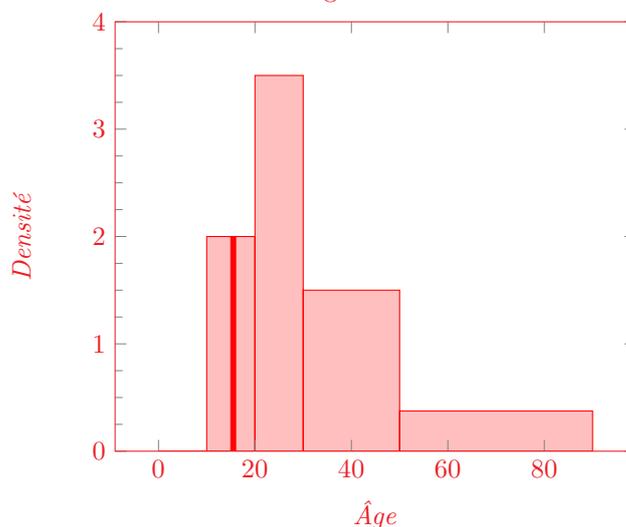
b. L’amplitude est la longueur de la classe d’âge, et la densité est la fréquence divisée par l’amplitude.

Âge	[10,20[[20,30[[30,50[[50,90[
Fréquence	20%	35%	30%	15%
Amplitude	10	10	20	40
Densité	2	3,5	1,5	0,375



c. La fréquence est en % et l’amplitude en années, donc la densité (fréquence divisée par l’amplitude) est en % par an.

d. Un lecteur a 15 ans s’il a entre 15 et 16 ans. On fait l’hypothèse que les lecteurs sont uniformément répartis dans chaque classe (car nous n’avons pas plus d’information). Le nombre de lecteurs ayant entre 15 et 16 ans est alors l’aire du rectangle sombre ci-dessous :

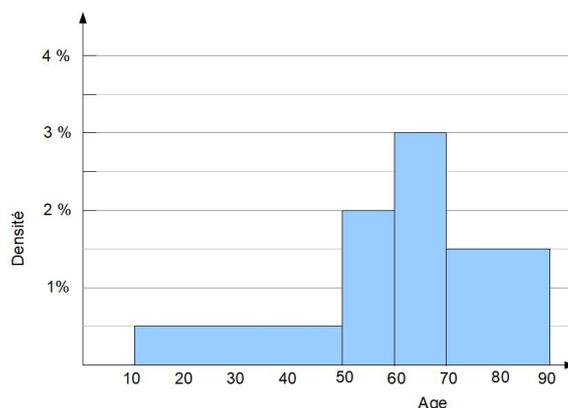


Sa largeur est de 1 an et sa hauteur de 2 % par an. Son aire est donc de

$$(1 \text{ an}) \times (2 \text{ \% par an}) = 2 \text{ \%}.$$

Environ 2% du lectorat a 15 ans. De même, environ 1,5% du lectorat a 41 ans. Cela donne une interprétation de la densité.

Ex 3. L’éditeur du magazine "Tricot & Dentelle " a fait une étude auprès de ses abonnés. Pour ce qui concerne l’âge, on a obtenu l’histogramme suivant :



a. Compléter les affirmations suivantes à l'aide du graphique.

- (1) Le pourcentage d'abonnés dans la tranche d'âge 50-60 est que le pourcentage dans la tranche 70-90.
- (2) Le pourcentage d'abonnés dans la tranche d'âge 60-70 est que le pourcentage dans la tranche 70-90.
- (3) On s'attend à avoir d'abonnés de 55 ans que d'abonnés de 81 ans.
- (4) Le pourcentage d'abonnés qui ont 55 ans est d'environ %.

b. Compléter le tableau ci-dessous et calculer les fréquences de chaque classe.

Âge	[10,50[[50,60[[60,70[[70,90[
Fréquence				
Amplitude				
Densité				

Solution. a.

- (1) Le rectangle correspondant à la tranche d'âge 70-90 ans a une aire plus grande que le rectangle correspondant à la tranche d'âge 50-60 ans. Il y a donc moins d'abonnés dans la tranche d'âge 50-60 que dans la tranche 70-90.
- (2) Par un même raisonnement, le nombre d'abonnés dans la tranche d'âge 60-70 est le même que le pourcentage dans la tranche 70-90.
- (3) La densité est plus élevée à 55 ans qu'à 81, donc on s'attend à avoir plus d'abonnés de 55 ans que d'abonnés de 81 ans (voir aussi la dernière question de l'exercice précédent).
- (4) Le pourcentage d'abonnés qui ont 55 ans est d'environ 2% (voir aussi la dernière question de l'exercice précédent).

b. L'amplitude se calcule directement à partir des classes d'âges (par exemple, l'amplitude de la classe [10,50[est de $50 - 10 = 40$ ans, et la densité se lit en ordonnée sur l'histogramme. De plus,

$$\text{Densité} = \frac{\text{Fréquence}}{\text{Amplitude}},$$

donc

$$\text{Fréquence} = \text{Amplitude} \times \text{Densité}.$$

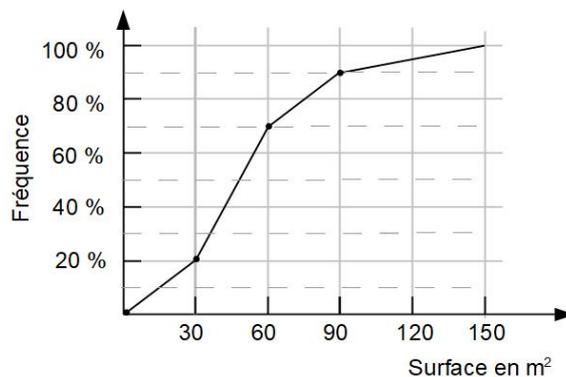
On retrouve au passage que l'aire d'un rectangle est sa base fois sa hauteur. On trouve finalement :

Âge	[10,50[[50,60[[60,70[[70,90[
Fréquence	20%	20%	30%	30%
Amplitude	40	10	10	20
Densité	0,5	2	3	1,5

SECTION 4. EFFECTIFS ET FRÉQUENCES CUMULÉS

Ex 4. Une ville a mené plusieurs études sur les appartements de ses HLM.

a. La première étude porte sur la surface des appartements. La ville a obtenu les fréquences cumulées qui sont représentées dans le graphique ci-dessous.



Compléter les affirmations suivantes à l'aide du graphique.

- (1)% des appartements ont une surface inférieure à $60 m^2$.
- (2)% des appartements ont une surface supérieure à $90 m^2$.
- (3)% des appartements ont une surface comprise entre 30 et $60 m^2$.
- (4) Environ 80% des appartements ont une surface inférieure à m^2 .
- (5) Environ 70% des appartements ont une surface supérieure à m^2 .

b. Compléter le tableau des fréquences

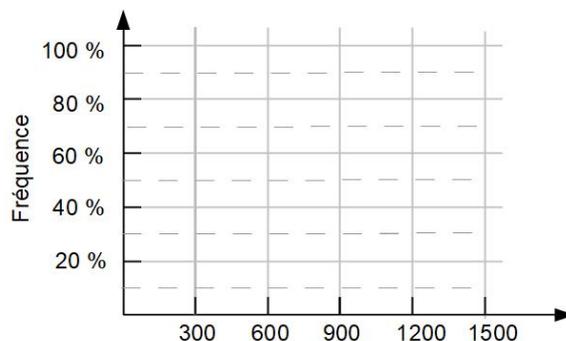
Surface	[0,30[[30,60[[60,90[[90,150[
Fréquence				

c. La seconde étude porte sur le loyer mensuel des appartements. La ville a obtenu le tableau des fréquences suivant.

Loyer (€)	[0,600[[600,900[[900,1200[[1200,1500[
Fréquence	10%	40%	30%	20%

Calculer les fréquences cumulées et les représenter graphiquement.

Loyer (€)					
Fréquence cumulée					



Solution. a.

- (1) 70% des appartements ont une surface inférieure à $60 m^2$ (on regarde l'ordonnée de la courbe à l'abscisse de $60 m^2$).

- (2) 10% des appartements ont une surface supérieure à 90 m^2 (car 90% des appartements ont une surface inférieure à 90 m^2).
- (3) 50% des appartements ont une surface comprise entre 30 et 60 m^2 (car 70% des appartements ont une surface inférieure à 60 m^2 , et 20% des appartements ont une surface inférieure à 30 m^2 . On fait la différence).
- (4) Environ 80% des appartements ont une surface inférieure à 75 m^2 (on se place à une ordonnée de 80%, et on cherche l'abscisse correspondante).
- (5) Environ 70% des appartements ont une surface supérieure à 36 m^2 (on se place à une ordonnée de $100 - 70 = 30\%$, et on cherche l'abscisse correspondante).

b. Il s'agit de retrouver les fréquences à partir des fréquences cumulées. La fréquence d'un intervalle, c'est la différence des fréquences cumulées à ses extrémités.

Surface	$[0,30[$	$[30,60[$	$[60,90[$	$[90,150[$
Fréquence	$20-0 = 20\%$	$70-20 = 50\%$	$90-70 = 20\%$	$100-90 = 10\%$

c.

Loyer (€)	0	600	900	1200	1500
Fréquence cumulée	0%	10%	$10+40=50\%$	$50+30=80\%$	$80+20=100\%$

On place les points de coordonnées (0 euros, 0 %), (600 euros, 10 %), etc. sur le graphique, puis on les relie.

