

Nous disposons des données statistiques suivantes :

Thème : Population Sous-thème : Évolution et structure de la population POP1A - Population par âge regroupé en 2011 Commune de Paris <i>Source : Insee, RP 2011 exploitation principale</i>		Effectif
	Moins de 3 ans	73 946
	3 à 5 ans	65 679
	6 à 10 ans	103 810
	11 à 17 ans	140 886
	18 à 24 ans	244 026
	25 à 39 ans	598 565
	40 à 54 ans	435 966
	55 à 64 ans	257 600
	65 à 79 ans	222 118
	80 ans ou plus	107 378
	Ensemble	2 249 974

1. Décrivons cette série statistique :

- population étudiée : La population de Paris en 2011
- caractère étudié : Age
- caractère qualitatif ou quantitatif : quantitatif
- Caractère discret ou continu : continu (on peut le considérer comme discret mais comme il y a beaucoup de valeurs on regroupe en classe comme pour des variables continues)

2. Pour étudier cette série statistique, nous allons utiliser un tableur (fichier « population Paris » à télécharger)

Classe		Effectif	Fréquence	Amplitude	Densité de fréquence
borne inf	borne sup				
0	3	73 946	0,03	3	0,01
3	6	65 679	0,03	3	0,01
6	11	103 810	0,05	5	0,01
11	18	140 886	0,06	7	0,009
18	25	244 026	0,11	7	0,016
25	40	598 565	0,27	15	0,018
40	55	435 966	0,19	15	0,013
55	65	257 600	0,11	10	0,011
65	80	222 118	0,10	15	0,007
80	105	107 378	0,05	25	0,002
0	105	2 249 974			

- (a) Quelle **formule** faut-il écrire dans la cellule C13 pour obtenir l'effectif total ?
 =SOMME(C3 :C12)
- (b) Quelle formule faut-il écrire dans la cellule D2 pour pouvoir ensuite avoir toutes les valeurs de la colonne D en étendant cette formule vers le bas ?
 =C2/\$C\$13
- (c) Compléter la colonne D (arrondir au centième près)

Les classes [6; 11[et [80; 105[ont la même fréquence mais l'amplitude de la première est inférieure à celle de la seconde.

Faisons une analogie avec la géographie :

D'après les dernières données Monaco avait 37 308 habitants (en 2016) et le Japon 126 477 000 (en 2020) (sources Institut Monégasque de la Statistique et des Études Économiques et Institut nationale d'études démographiques).

Bien sûr, les démographes diront que ces renseignements sont très largement insuffisants pour comparer la démographie des deux pays : il faut au minimum s'intéresser aux superficies de ces deux pays et calculer pour chacun la densité de population, c'est-à-dire le nombre d'habitants au kilomètre carré.

Avec une superficie de $2,02 \text{ km}^2$ pour Monaco et de $378\,000 \text{ km}^2$ pour le Japon, les densités sont respectivement $d_1 = \frac{37\,308}{2,02} \simeq 18\,469 \text{ hab/km}^2$ pour Monaco et $d_2 = \frac{126\,477\,000}{378\,000} \simeq 335 \text{ hab/km}^2$ pour le Japon.

Autrement dit, alors que la population de Monaco est la moins importante en taille, sa densité est plus importante que celle du Japon.

Ainsi, une représentation pertinente des populations de Monaco et du Japon doit rendre visible cette différence de densité.

De même dans notre exemple, il n'est pas pertinent de représenter de la même façon la fréquence égale à 0,05 de la tranche d'âge $[6; 11[$ et de la tranche d'âge $[80; 105[$.

Ces deux classes d'âge ont la même fréquence mais l'amplitude de la première classe est inférieure à celle de la seconde, on pressent que la densité de la première classe est supérieure à celle de la seconde. Par analogie avec la densité de population utilisée en géographie, on calculera le quotient de la fréquence d'une classe d'âge par l'amplitude de cette classe, et on parlera de densité de fréquence.

Dans notre exemple, la densité de fréquence pour la tranche d'âge $[6; 11[$ est $\frac{0,05}{5} = 0,01$ et la densité pour la tranche d'âge $[80; 105[$ est $\frac{0,05}{25} = 0,002$.

Comme perçue, la densité de fréquence de la première tranche d'âge est plus grande que la densité de fréquence de la seconde tranche d'âge.

Utilisons une **représentation graphique** rendant visible ce phénomène.

Définition 1. *Un histogramme de fréquences est un diagramme composé de rectangles collés dont les aires sont proportionnelles aux fréquences et dont les bases sont déterminées par les intervalles des classes.*

Définition 2. *Dans le cas d'une variable quantitative continue, on définit la densité de fréquence d_i d'une classe de fréquence f_i et d'amplitude a_i par : $d_i = \frac{f_i}{a_i}$.*

Rappel : L'amplitude a_i de la classe $[x_i; x_{i+1}[$ est le nombre $a_i = x_{i+1} - x_i$.

3. (Niveau **) Montrer que, dans un histogramme de fréquences, les hauteurs des rectangles sont proportionnelles à la densité.

Considérons une classe $[x_i; x_{i+1}[$.

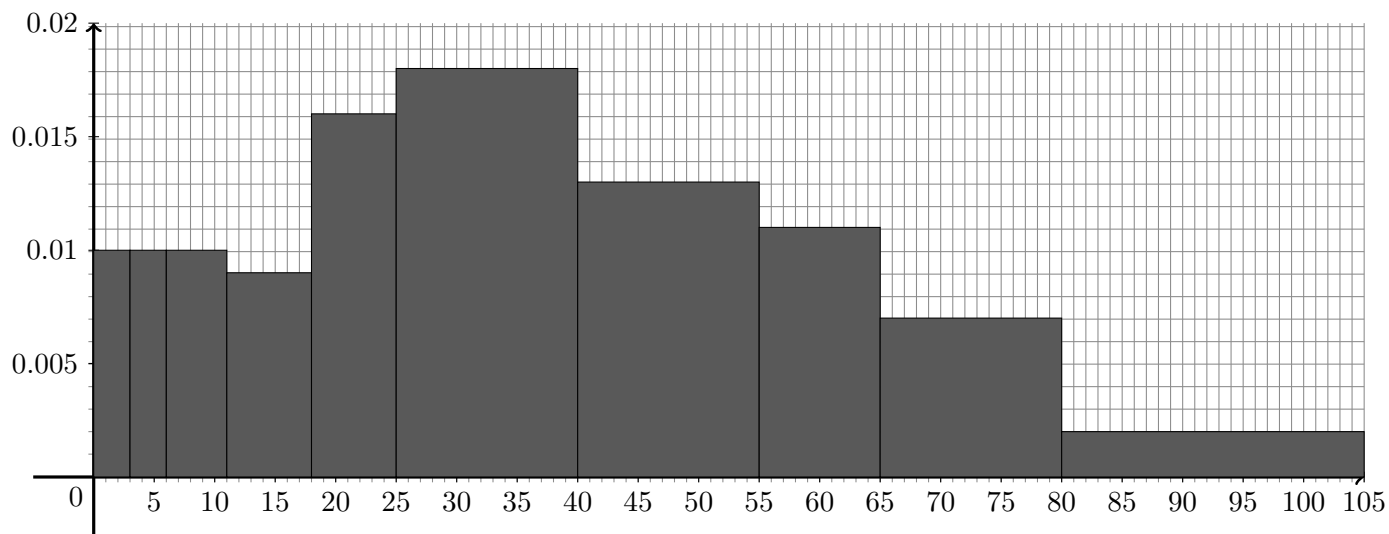
L'aire \mathcal{A}_i du rectangle représentant la classe est proportionnelle à la fréquence f_i donc il existe un réel p tel que $\mathcal{A}_i = p \times f_i$.

Or ce rectangle a pour largeur $x_{i+1} - x_i = a_i$ et pour hauteur h_i donc son aire est égale à $h_i \times a_i$

On a donc $h_i \times a_i = p \times f_i$ donc $h_i = p \times \frac{f_i}{a_i} = p \times d_i$

La hauteur est proportionnelle à la densité.

4. Compléter l'histogramme ci-dessous (vous pourrez ajouter une colonne au tableau pour les calculs) :



5. Que représente l'axe des ordonnées ?

L'axe des ordonnées représente la densité de fréquence.

6. Les rectangles représentant les classes $[6;11[$ et $[11;18[$ ont presque la même hauteur, que peut-on en déduire ?

Ces deux classes ont la même densité de fréquence.

7. Dans quelle tranche d'âge la densité est-elle la plus importante ? justifier à l'aide de vos connaissances en géographie.

La densité est la plus importante pour les 25 - 40 ans. Cette classe d'âge correspond aux jeunes actifs.

8. Dans un histogramme, si un rectangle est plus haut qu'un autre, peut-on affirmer que la fréquence associée est la plus grande ?

Non il faut prendre en compte l'amplitude de la classe. Dans cet exemple le rectangle associé à la classe $[18;25[$ est plus haut que celui associé à la classe $[40;55[$ mais c'est la classe $[40;55[$ qui a la fréquence la plus importante.

9. Tracer les histogrammes représentant la population du Haut-Rhin (*niveau ***) puis celle de Paris en janvier 2021 (*niveau ****) (tableur accessible sur moodle)

Exercice à rendre Tirage aléatoire

On souhaite simuler le tirage aléatoire d'un nombre réel compris entre 0 et 12 ; et représenter les résultats pas un histogramme.

1. Sur tableur la formule =ALEA() génère un nombre compris entre 0 et 1.
Quelle formule va générer un nombre aléatoire compris entre 0 et 12 ?
2. Générer des nombres aléatoires compris entre 0 et 12(12 exclu) sur la plage A1 :A20 d'une feuille de tableur.
3. Copier les valeurs générée sur la plage B1 :B20 pour les figer (collage valeurs).
4. On souhaite construire un histogramme représentant la répartition de ces nombres.

min	max	nb valeurs	fréquence	Amplitude	Densité de fréquence
0	2				
2	6				
6	9				
9	12				

- (a) Expliquer pourquoi la formule : =NB.SI(B1 :B20 ; "<6")-NB.SI(B1 :B20 ; "<2") permet de compléter la case grisée.
- (b) Compléter l'ensemble du tableau.
- (c) Construire l'historgramme correspondant.

Déposer le tableau excel utilisé et une photo de votre histogramme sur moodle

