

COURS 9 : HISTOGRAMMES ET FRÉQUENCES CUMULÉES

On étudie une variable continue dans une population donnée, par exemple la taille en centimètres.

HISTOGRAMMES

Afin de représenter graphiquement la distribution de cette variable, on va utiliser un **histogramme**. La première étape consiste à trier les modalités dans des **classes** ; idéalement, ces classes ont des extrémités qui sont des nombres “ronds”, et il faut choisir un juste équilibre (trop peu de classes et on perd beaucoup d’informations ; trop de classes et il y aura peu d’individus dans chaque classe). Par exemple, disons que l’on obtient :

Taille (en cm)	[120,150[[150,160[[160,170[[170,180[[180,190[[190, 210[
Effectif	3	22	65	62	16	4

L’**amplitude** d’une classe est la longueur de l’intervalle correspondant. Par exemple, la classe [120, 150[a une amplitude de $150 - 120 = 30$ centimètres.

La première étape consiste à calculer la fréquence de chaque classe. Pour cela, on calcule l’effectif total $N = 3 + 22 + 65 + 62 + 16 + 4 = 172$, et on divise chaque effectif par l’effectif total. Par exemple, la fréquence de la classe [120, 150[est $3/172 \simeq 1,7\%$. On obtient :

Taille (en cm)	[120,150[[150,160[[160,170[[170,180[[180,190[[190, 210[
Fréquence (en %)	1,7	12,8	37,8	36,0	9,3	2,3

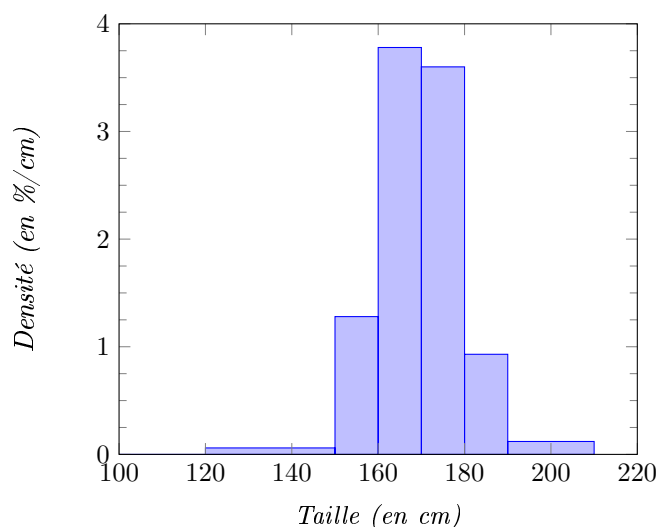
Quand on trace un histogramme, on met en abscisse la variable observée (ici, la taille en centimètres), que l’on divise selon les classes. La population de chaque classe est représentée par un rectangle. Cependant, pour le graphe obtenu ne soit pas trompeur, on ne met pas en ordonnée la fréquence, mais la **densité** :

$$\text{densité} = \frac{\text{fréquence}}{\text{amplitude}}.$$

Le tableau suivant résume cela :

Taille (en cm)	[120,150[[150,160[[160,170[[170,180[[180,190[[190, 210[
Fréquence (en %)	1,7	12,8	37,8	36,0	9,3	2,3
Amplitude (en cm)	30	10	10	10	10	20
Densité (en %/cm)	0,06	1,28	3,78	3,60	0,93	0,12

Nous pouvons maintenant tracer l'histogramme :



Dans un histogramme, la fréquence d'une classe est égale à l'aire du rectangle correspondant. Par exemple, le rectangle correspondant à la classe $[160, 170[$ a une largeur de 10 centimètres (c'est l'amplitude de la classe) et une hauteur de 3,78% par centimètre (c'est la densité). Son aire est donc de :

$$\text{amplitude} \times \text{densité} = (10 \text{ cm}) \times (3,78\%/ \text{cm}) = 37,8\%,$$

ce qui est bien la fréquence de cette classe.

Attention : La densité a une unité étrange (ici en % par centimètre; elle serait, par exemple, en % par an si on mesurait des âges en années, ou en % par M€ si on mesurait des chiffres d'affaires en M€).

FRÉQUENCES CUMULÉES

Étant donnée une variable quantitative, la fréquence cumulée à la valeur x est la fréquence des modalités inférieures ou égale à x . On la notera F_x . Afin de montrer cette construction dans le cas de variables continues triées par classes, reprenons l'exemple ci-dessus :

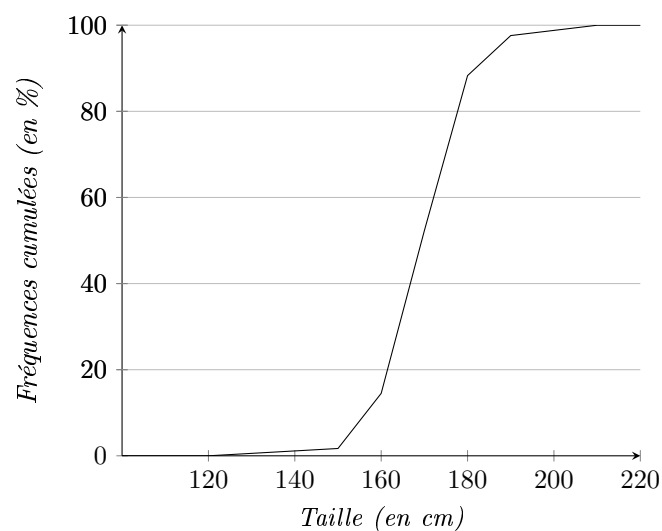
- Personne n'a une taille inférieure ou égale à 120 cm, donc $F_{120} = 0$.
- 1,7% de la population a une taille inférieure ou égale à 150 cm, donc $F_{150} = 1,7\%$.
- $1,7 + 12,8 = 14,5\%$ de la population a une taille inférieure ou égale à 160 cm, donc $F_{160} = 14,5\%$.
- $14,5 + 37,8 = 52,3\%$ de la population a une taille inférieure ou égale à 170 cm, donc $F_{170} = 52,3\%$.

On peut ainsi calculer de proche en proche les fréquences cumulées pour les valeurs de x qui sont les extrémités des classes.

Taille (en cm)	120	150	160	170	180	190	210
Fréquences cumulées (en %)	0	1,7	14,5	52,3	88,3	97,6	100

Attention : La fréquence cumulée au plus bas échelon doit valoir 0% (personne n'est plus petit), et à l'échelon le plus élevée doit valoir 100% (tout le monde est plus petit), éventuellement avec une erreur d'arrondi.

Pour tracer la courbe des fréquences cumulées, on place en abscisse la taille en centimètres, et en ordonnées les fréquences cumulées. On place d'abord les points dont les coordonnées sont données par le tableau ci-dessus (par exemple, le point d'abscisse 150 cm et d'ordonnée 1,7%), puis on relie les points par des segments.



Remarque : La pente d'un segment est donnée par la densité de la classe, donc par la hauteur de l'histogramme : plus un rectangle est haut dans l'histogramme, plus la pente est raide dans la courbe des fréquences cumulées.