VARIABLES STATISTIQUES CONTINUES

1. Présentation

1.1 Qu'est-ce ?

Voici 2 exemples de variables statistiques :

- **La surface habitable**, pour des logements, peut prendre n'importe quelle valeur dans un intervalle de $\mathbb{R}$ — L'intervalle ici est]0; +∞ [—.: c'est une **variable statistique continue**.
  On ne peut la traiter comme une **variable statistique discrète** (cf leçon 2) : on n'essaiera pas de compter les appartements qui ont exactement la même surface habitable. En revanche, on pourra dénombrer les appartements dont la surface habitable est entre 0 et 15 m^2, ceux dont la surface habitable est supérieure à 30 m^2 ...

.

- **Le nombre d'employés** d'une entreprise n'est pas exactement une variable continue : elle peut prendre les valeurs 236 ou 237, mais pas 236,4. Pourtant, on la traitera comme telle : il serait inutile d'essayer de regrouper les entreprises qui ont exactement le même nombre d'employés. En revanche, il peut être intéressant de regrouper celles qui ont entre 0 et 49 employés, celles qui ont entre 50 et 200 employés ...

Une variable continue est un caractère statistique quantitatif :
dont les modalités décrivent un ou des intervalles entiers de $\mathbb{R}$.
ou dont les modalités sont des valeurs isolées — comme pour une variable discrète — mais très nombreuses.

Les modalités ne peuvent donc être considérées une par une :
on doit les **regrouper.**

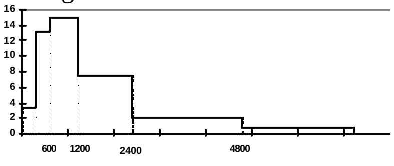
1.2 Variables statistiques continues : qu'en faire ?

Dans cette leçon, nous allons ensemble :

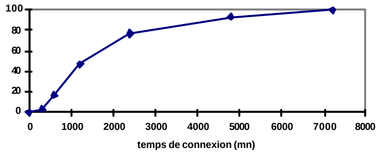
- **regrouper** les valeurs possibles, et les individus, **en classes** homogènes : 2 Regroupements en classes p 2
- exprimer les rapports entre ces classes par **des calculs** : effectifs, fréquences, fréquences par amplitude, fréquences cumulées : 3 Traitements numériques : effectifs et fréquences p 3
- **représenter graphiquement** les rapports entre les classes, par des histogrammes et des courbes cumulatives: 4.1Histogramme, p 7 et 4.2Courbe cumulative ou diagramme intégral : représentation des fréquences cumulées p9

longueur, en mètres	n_i :	f_i	fréquence par amplitude f_i
moins de 12 m	4 631	0,7404	0.0617
[12, 16[	587	0,0938	0.0235
[16, 25 [	875	0,1400	0.0156
[25, 38[	89	0,0142	0.0011
38 et plus	72	0,0115	0.0009
Total	6 255	1	

Histogramme



Courbe cumulative



2. Regroupements en classes

2.1 Regroupements

Pour étudier la surface habitable d'appartements, on regroupe les appartements qui ont des surfaces habitables similaires, à défaut d'avoir la même surface habitable exactement.

surface habitable	nombre d'appartements
[0 ;20 [	123
[20, 70[	130
[70 ; +∞[	96
Total	349

Les classes sont obtenues en choisissant pour la variable étudiée des intervalles **disjoints** , qui recouvrent tout l'ensemble des modalités.

Chaque individu de la population — ici chaque appartement — appartient alors à 1, et 1 seule, classe.

On peut **définir des classes** :

- par des intervalles comme dans l'exemple : [0,20[, [20,70[, [70, +∞[
- ou bien ainsi : moins de 20, 20 à moins de 70, plus de 70
- pour des variables discrètes regroupées en classe, par exemple des âges, on rencontre souvent des expressions telles que
  moins de 20, 20 à 59, 60 à 74, 75 et plus,, ce qui signifie bien sûr [0,20[, [20, 60[, [60, 75[, [75, +∞[.

Mais on **ne peut pas** utiliser: moins de 20, 20 à 60, 60 à 75, 75 et plus:
car les individus de 60 ans seraient dans 2 classes à la fois.

2.2 Amplitude:

C'est la différence entre les 2 bornes d'une classe. L'amplitude de la classe [20; 70 [est 50.
La dernière classe est particulière : elle a souvent une borne infinie. Si on a besoin de lui affecter une amplitude, et si le sens des données n'en impose pas une, on pourra prendre l'amplitude de l'avant dernière classe .
Exemple : nombre de navires de pêche suivant la longueur hors tout en 1997 en France (INSEE, annuaire statistique de France 2000)

longueur, en mètres	nombre de navires
moins de 12 mètres	4 631
[12, 16[	587
[16, 25 [	875
[25, 38[	89
38 et plus	72
Total	6 255

On peut calculer les amplitudes. Pour la dernière classe on prend pour amplitude celle de la classe [25,38[: 13.

longueur, en mètres	amplitude
moins de 12 mètres	12
[12, 16[	4
[16, 25 [	9
[25, 38[	13
38 et plus	13

2.3 Pourquoi des amplitudes différentes:

Il pourrait paraître plus simple de choisir des classes de même amplitude.
Un magazine d'information générale pourrait, pour ses annonceurs, classer ses abonnés par âge.

âge (en années entières)	effectifs
moins de 20	8 000
20 à 39	36 000
40 à 59	7 000
60 et plus	4 000

En fait, des **amplitudes différentes** s'imposent souvent

- par **le sens des données** : un annonceur peut considérer la population de 40 à 59 ans homogène, mais les consommateurs potentiels de 20 à 25 ans ne seront pas abordés de la même façon que ceux de 30 à 39.
- par **l'hétérogénéité des effectifs**:: la classe de 20 à 39 ans a des effectifs supérieurs à ceux des autres classes. La répartition à l'intérieur de cette classe mérite d'être étudiée plus finement.

On essaie donc de déterminer des **classes homogènes et d'effectifs comparables**.

âge (en années entières)	effectifs
moins de 20	8 000
20 à 24	18 000
24 à 30	10 000
30 à 39	9 000
40 à 59	7 000
60 et plus	4 000

2.4 Centre :

Une fois que les individus sont regroupés dans une classe [a; b[, ils sont indiscernables. Si on a besoin de leur affecter une valeur, par exemple pour calculer une moyenne, que faire ?
Dans ce cas on définit une valeur représentative de la classe : le centre . Et on effectuera les calculs, à défaut d'une information plus fine, comme si ce centre était la valeur pour tous les individus de la classe.
De façon générale, pour une variable continue, le centre d'une classe est la moyenne arithmétique entre la borne inférieure de la classe et la borne inférieure de la classe suivante

La ième classe est $[x_i , x_{i+1}[$, son centre est $c_i = \dfrac{x_i + x_{i+1}}{2}$

La dernière classe, est souvent de la forme $[x_k , +\infty[$
Dans ce cas, on lui affecte une amplitude — en général la même que celle de la classe précédente — et on en déduit le centre.
Exemple :
Pour étudier les poids d'enfants d'un service de pédiatrie, on utilise les classes suivantes :
[0, 5[, [5, 10[, [10, 20[, [20, 40[, 40 et plus.
On affecte à la dernière classe la même amplitude que la précédente : 20, et on peut calculer les centres :

poids (kg)	[0, 5[	[5, 10 [	[10, 20[	[20, 40[	40 et plus
centre	$\frac{0+5}{2}$ =2,5	7,5	15	30	50

Attention : **<u>cas des variables discrètes regroupées en classes</u>**.
Pour étudier les âges en années de ces enfants, on désigne souvent les classes de la façon suivante:
0 à 4 ans, 5 à 9 ans, 10 à 15 ans , mais il s'agit des classes [0, 5[, [5, 10[, [10, 16[:dans la première classe, les âges des enfants vont de 0 à 5 ans moins un jour. Le centre ne sera pas $\frac{0+4}{2}$ mais $\frac{0+5}{2}$ = 2,5.

classe (présentation usuelle)	0 à 4 ans	5 à 9 ans	10 à 15 ans
signification	[0, 5[	[5, 10[	[10, 16[
centre	2,5	7,5	12,5

3. Traitements numériques : effectifs et fréquences

Nous allons ensemble calculer

- des effectifs 3.1Effectifs, p 4, des fréquences 3.2 Fréquences p 4
- des densités, ou fréquences par amplitude 3.3 Fréquences par amplitudes , fréquences corrigées des amplitudes , densités: p4
- des fréquences cumulées 3.4Fréquences cumulées :p 5

Ensuite, nous pourrons envisager les graphiques correspondants : 4.1Histogramme, p7

3.1 Effectifs

Pour une classe définie par l'intervalle [x_i , x_{i+1}[, l'effectif , noté n_i , est le nombre d'individus pour lesquels la variable étudiée est dans cet intervalle . la somme des effectifs des classes est bien sûr égale à l'effectif total .
nombre de navires de pêche suivant la longueur hors tout en 1997 en France (INSEE, annuaire statistique de France 2000)

longueur, en mètres	n_i : nombre de navires
moins de 12 mètres	4 631
[12, 16[	587
[16, 25 [	875
[25, 38[	89
38 et plus	72
Total	6 255

3.2 Fréquences

La fréquence d'une classe est le rapport entre l'effectif de cette classe et l'effectif total:

$f_i = \dfrac{n_i}{n}$ pour $1 \leq i \leq k$, k est le nombre de classes, n_i l'effectif de la classe , n l'effectif total.

Attention à une erreur fréquente: il s'agit bien de diviser par l'effectif total (dans l'exemple : 6 255) et non par le nombre de classes (dans l'exemple : 5) .

La fréquence est toujours un nombre entre 0 et 1 — ou 0 et 100 % — et la somme des fréquences est toujours égale à 1 — ou 100 % —.: $\mathbf{0 \leq f_i \leq 1}$ **pour tout i , et** $\mathbf{\sum_{i=1}^{i=k} f_i = 1}$**, où k est le nombre de classes**

longueur, en mètres	n_i : nombre de navires	f_i
moins de 12 mètres	4 631	0,7404
[12, 16[	587	0,0938
[16, 25 [	875	0,1400
[25, 38[	89	0,0142
38 et plus	72	0,0115
Total	6 255	1

3.3 Fréquences par amplitudes , fréquences corrigées des amplitudes , densités:

Ces trois expressions sont synonymes.
La nécessité de cette notion apparaîtra plus clairement pour le tracé de l'histogramme 4.1 Histogramme , p 7

longueur, en mètres	n_i : nombre de navires	f_i
moins de 12 mètres	4 631	0,7404
[12, 16[	587	0,0938
[16, 25 [	875	0,1400
[25, 38[	89	0,0142
38 et plus	72	0,0115
Total	6 255	1

D'après le tableau précédent, les bateaux de la classe [16, 25[sont plus fréquents que ceux de la classe [12, 16[.
Mais dans la classe [16, 25[se trouvent des navires de longueurs bien plus disparates (9 m d'écart) que dans la classe [12, 16[(4 m d'écart au maximum). Ainsi , pour savoir vers quelle valeur se concentrent le plus d'individus , on calcule pour chaque classe

la densité, ou **fréquence par amplitude :** $\dfrac{\textbf{fréquence de la classe}}{\textbf{amplitude de la classe}}$

Pour la classe [16,25[elle est d'environ 0,016 et pour la classe [12, 16[d'environ 0,23.
Cela montre que vers les tailles 12 à 16 m on a une plus grande densité de navires que vers les tailles 16 à 25m.

longueur, en mètres	n_i : nombre de navires	f_i	amplitude	fréquence par amplitude f_i / a_i
moins de 12 mètres	4 631	0,7404	12	0.0617
[12, 16[	587	0,0938	4	0.0235
[16, 25 [	875	0,1400	9	0.0156
[25, 38[	89	0,0142	13	0.0011
38 et plus	72	0,0115	13	0.0009
Total	6 255	1		

Remarques :

- L'amplitude de la dernière classe a été arbitrairement choisie égale à celle de la classe précédente.
- Alors qu'**une fréquence**, prise **isolément**, a une signification (74 % des bateaux ont moins de 12 m) , la signification des **fréquences par amplitudes** ne réside que dans les **rapports entre elles**:
  dire que la fréquence par amplitude de la classe [12,16[est de 0,0617 ne donne aucune information, tant qu'on ne la compare pas à celle d'une autre classe..

L'exemple suivant illustre de façon plus concrète cette notion de densité.

Une autoroute relie 2 villes A et B sur une distance de 110 km. Pour évaluer le tarif d'affichages publicitaires, on recense les véhicules entrant sur l'autoroute, le point d'entrée étant repéré par sa distance à la ville A.
Voilà un résultat sur une semaine :

distance à la ville A	fréquence (%)	amplitude (unité : 1 km)	fréquence/amplitude
[0, 5[	25,12	5	5,02
[5,10[	19,45	5	3,89
[10, 50[	12,16	40	0,30
[50 ,90[	10,50	40	0,26
[90,105[	24,45	15	0 ,96
[105,110[	18,32	5	3,67

La densité de véhicules entrant est beaucoup plus forte près des 2 villes.
Les fréquences ne rendent pas bien compte de ce phénomène, puisqu'elles portent sur des zones de tailles différentes.
La dernière colonne, elle, est représentative du **nombre de voitures entrant par kilomètre**, c'est-à-dire de la **densité** du trafic.
Elle montre qu'un affichage placé à un péage entre 105 et 110 km de A a plus de valeur qu'à un péage situé entre 90 et 105 km de A.

3.4 Fréquences cumulées :

Alors que les fréquences permettent de répondre aux questions : " quelle proportion d'individu est dans cette classe ? ", les fréquences cumulées permettent de répondre à
Quelle proportion de la population est en dessous d'un niveau donné ?
Quelle est la valeur minimum atteinte par 15 % des individus ?

Pour y répondre, **on cumule les fréquences** (**et non les fréquences par amplitudes** !), en partant des valeurs les moins élevées de la variables statistique;
$F_1 = f_1$ proportion de navires de moins de 12 m
$F_2 = f_1 + f_2$ proportion de navires de 16 à moins de 25 m
$F_3 = f_1 + f_2 + f_3$ proportion de navires de 25 à moins de 38 m. ...
...
F_i s'appelle la ***fréquence cumulée*** de la classe i

$\boxed{F_i = f_1 + f_2 + ... + f_i = \sum_{p=1}^{i} f_p}$ *est la proportion d'individus pour lesquels la valeur de la variable étudiée est inférieure ou égale à la borne supérieure de la classe n° i*

Longueur, en mètres	n_i : nombre de navires	f_i : fréquence	F_i fréquences cumulées
moins de 12	4 631	0,7404	0,7404
[12, 16[	587	0,0938	0.7404 + 0.0938 = 0,8342
[16, 25 [	875	0,1400	0.7404 + 0.0938 +0.1400 = 0,9742
[25, 38[	89	0,0142	0,9884
38 et plus	72	0,0115	1
Total	6 255	1	

Remarque : si l'on ne connaît que les fréquences cumulées F_i , on peut calculer les fréquences f_i par :
$F_i - F_{i-1} = f_i$., par exemple $f_2 = 0,8342 - 0,7404 = 0,0938$

3.5 Fonction de répartition :

pour **tout nombre** réel x,

F(x) est la proportion d'individus pour lesquels la variable étudiée est inférieure à x.

Le tableau indique F(12) = 0, 7404 , F(16) = 0,8342 : on connaît les valeurs de F aux bornes des classes.
Mais pour 12 < x < 16, du fait du regroupement en classes, on ne connaît pas F(x). Tout ce qu'on sait est que 0,7404 < x < 0,8342.
On fait donc l'hypothèse qu'à l'intérieur d'une classe, la répartition est uniforme, et que F y coïncide avec une fonction affine.

F est donc considérée comme une fonction affine par morceaux, dont le graphe est construit au paragraphe 4.2 Courbe cumulative ou diagramme intégral : représentation des fréquences cumulées p 9

3.6 Fréquences cumulées décroissantes :

On cumule les fréquences des classes pour lesquelles la variable est **supérieure** à une valeur donnée.
En pratique, on cumule en partant des valeurs les plus élevées de la variable statistique, pour finir par les valeurs les moins élevées.

Voici un exemple : on s'intéresse, chaque année, à la proportion des entreprises créées ou reprises en 1990, en Bretagne, qui ont survécu (source : Tableaux Economiques de Bretagne 2000-2001). Naturellement, cette proportion décroît avec les années.
Voici le tableau des fréquences en fonction de la durée de vie.

durée (années)	fréquence (%)
[0, 1[	13,7
[1,2[	10,7
[2,3[	8,5
[3,4[	6,5
[4,5[	5,3
[5,6[	4,4
[6,7[	4,1
7 et plus	46,8

On cumule pour chaque année les proportions d'entreprises dont la durée est supérieure. On cumule donc de la plus grande valeur de la variable (7 et plus) à la plus petite ([0,1[)

durée (années)	fréquence (%)	fréquences cumulées décroissantes (%)
[0, 1[	13,7	86,3 + 13,7 = 100
[1, 2[	10,7	75,6 + 10,7 = 86,3
[2, 3[	8,5	67,1 + 8,5 = 75,6
[3, 4[	6,5	46,8 + 4,1+ 4,4 + 5,3 + 6,5 = 67,1
[4, 5[	5,3	46,8 + 4,1 + 4,4 + 5,3 = 60,6
[5, 6[	4,4	46,8 + 4,1 + 4,4 = 55,3
[6, 7[	4,1	46,8 + 4,1 = 50,9
7 et plus	46,8	46,8

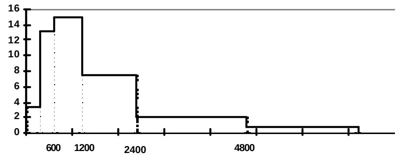
La dernière colonne est le taux de survie..

date	Taux de survie (%)
1er anniversaire	86,3
2ème anniversaire	75,6
3ème anniversaire	67,1
4ème anniversaire	60,6
5ème anniversaire	55,3
6ème anniversaire	50,9
7ème anniversaire	46,8

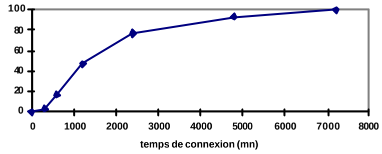
4. Traitements graphiques

2 graphiques sont à votre disposition :

- **L'histogramme** , ou **diagramme différentiel**, représente les importances relatives des différentes classes Histogramme p 7



- **La courbe cumulative** , ou **diagramme intégral** , représente la fonction de répartition, c'est-à-dire, pour toute valeur, la proportion de population en dessous de cette valeur Courbe cumulative ou diagramme intégral : représentation des fréquences cumulées, p 9



4.1 Histogramme

Il s'agit de représenter les rapports entre les classes par des rectangles **d'aires proportionnelles aux fréquences.** Dans le cas de classes d'amplitudes différentes , les fréquences ne seront donc pas proportionnelles aux ordonnées du graphique.
Cette méthode n'apparaît pas naturelle à tous. Elle est jusitifiée en 4.1.1 Histogramme : une justification de la construction p 7, puis décrite en détail en 4.1.2 Construction p 9.

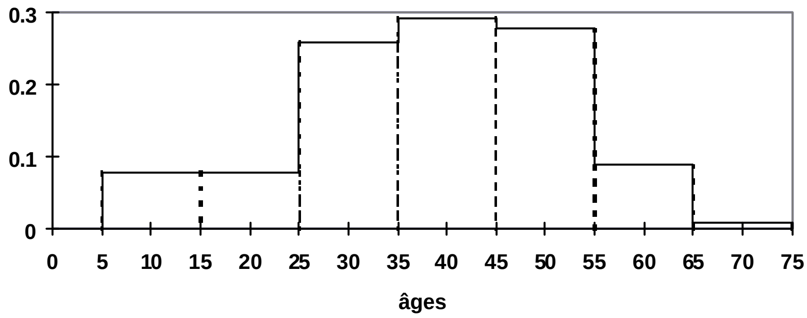
4.1.1 Histogramme : une justification de la construction

On représente chaque classe par un rectangle, dont la largeur , tracée sur l'axe des abscisses, est égale à l'amp litude de la classe.

4.1.1.1 Cas particulier de classes de même amplitude .

De la même façon que pour des **variables discrètes**, les rectangles ont même largeur, et des **aires proportionnelles aux fréquences**. Leurs longueurs sont donc aussi proportionnelles aux fréquences
Exemple :voici la répartition par tranches d'âge de la population active de l'Aveyron au recensement de 1999 (source : recensement)

Age	Effectifs	Fréquence
[15,25[	8811	0.078
[25, 35 [	29085	0.258
[35,45[	33030	0.293
[45, 55[	31220	0.277
[55, 65[	9977	0.088
65 et plus	774	0.007
Total	112897	1.00



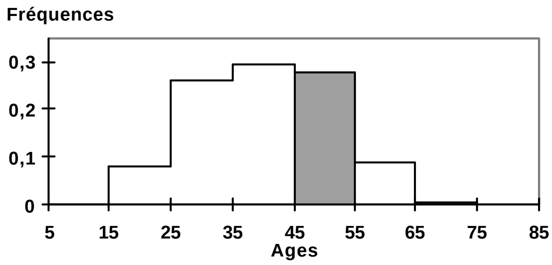
4.1.1.2 Cas général : classes d'amplitudes différentes

Dans le même exemple, scindons la classe [45 ,55 [en 2 classes [45 , 50[et [50,55[

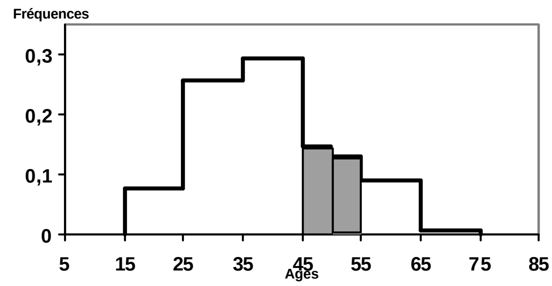
Age	Effectifs	Fréquence
[15,25[	8811	0.078
[25, 35 [	29085	0.258
[35,45[	33030	0.293
[45 ,50[	**16 428**	**0,146**
[50, 55[	**14 792**	**0,131**
[55, 65[	9977	0.088
65 et plus	774	0.007
Total	112897	1.00

Attention !
Si on représentait les classes en abscisses et les fréquences en ordonnée, voici ce qu'il se passerait :

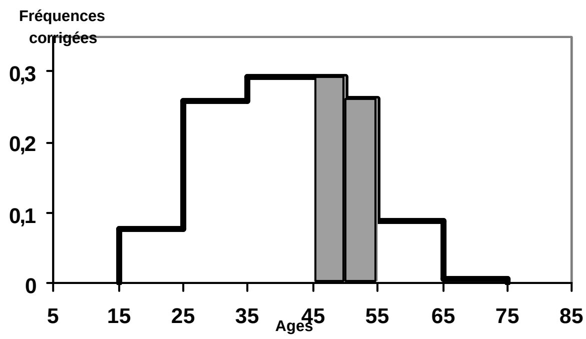
Graphique initial :
Classes de même amplitude



Classes d'amplitudes différentes.
Graphique erroné :
les fréquences sont en ordonnées ,
la classe [45, 55[apparaît sous-évaluée



Ce ne sont pas les ordonnées, mais **les AIRES** qui doivent être **proportionnelles aux fréquences** et aux effectifs.
Les deux rectangles représentant les classes [45,50[et [50,55[ayant des largeurs moitié plus petites que les autres, il faut multiplier leurs longueurs par 2 pour conserver les mêmes aires. Leurs ordonnées sont donc les fréquences multipliées par 2.

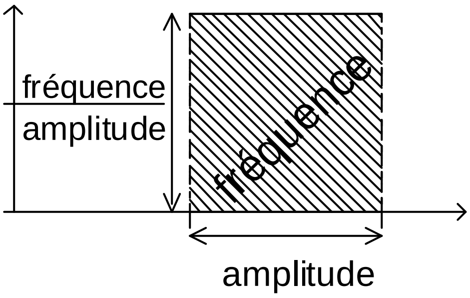


Graphique correct : les aires sont conservées.

Ce ne sont donc pas les fréquences que l'on reporte sur les ordonnées

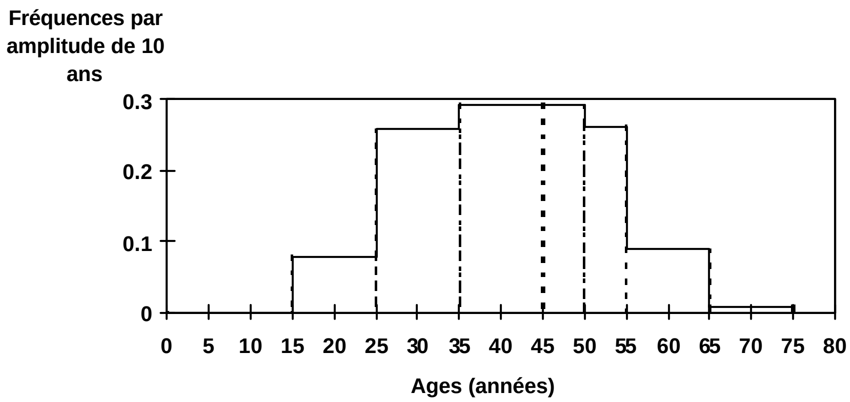
4.1.2 Construction de l'histogramme

Chaque classe est représentée par un rectangle.
L'aire du rectangle doit être proportionnelle à la fréquence de la classe
La largeur du rectangle est l'amplitude de la classe.
C'est donc la **fréquence par amplitude =** $\dfrac{\textbf{fréquence de la classe}}{\textbf{amplitude de la classe}}$ que l'on porte en ordonnées.



Pratiquement, après avoir construit le tableau des fréquences, il faudra calculer les amplitudes des classes, puis les fréquences par amplitude; On peut comme ici, choisir une unité d'amplitude commode, dont toutes les amplitudes sont un multiple simple.

Classes d'âge	Effectif	Fréquence	Amplitude (unité 10 ans)	Fréquence par amplitude de 10 ans		
[15,25[	8811	0,078	1	0,078/1	=	0,078
[25, 35 [	29085	0,258	1	0,258/1	=	0,258
[35,45[	33030	0,293	1	0,293/1	=	0,293
[45 ,50[	16 428	0,146	0,5	0,146/0,5	=	0,292
[50, 55[	14 792	0,131	0,5	0,131/0,5	=	0,262
[55, 65[	9977	0,088	1	0,088/1	=	0,088
65 et plus	774	0,007	1	0,007/1	=	0,007
Total	112897	1,00				



N'oubliez pas que sur ce graphique, ce sont surtout les aires, plutôt que les ordonnées, qui sont significatives.

4.2 Courbe cumulative ou diagramme intégral : représentation des fréquences cumulées

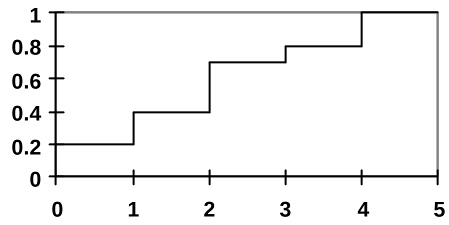
Comme pour les variables discrètes (cf leçon 2) il s'agit de représenter la courbe de la fonction de répartition : $x \in \mathbb{R} \mapsto F(x)$: proportion d'individus pour lesquels la variable est inférieure ou égale à x.

4.2.1 Rappel sur le cas des variables discrètes :

Dans le cas des variables discrètes, la variable a la même valeur pour toute la classe, la fonction de répartition est en escalier.

Tant que x est entre deux valeurs possibles x_i et x_{i+1}, cette proportion est constante , et pour $x = x_{i+1}$, la proportion est subitement augmentée de la fréquence de la classe n°i+1. F a donc une discontinuité en x_{i+1} , et est constante ensuite sur $[x_{i+1}, x_{i+2}[$.

exemple : fonction de répartition du nombre d'enfants dans un ensemble de foyers.
La variable nombre d'enfants est inférieure ou égale à 2 pour 70% de la population.
Elle est inférieure ou égale à 1, mais aussi à toute valeur de [1, 2[pour 40 % de la population.



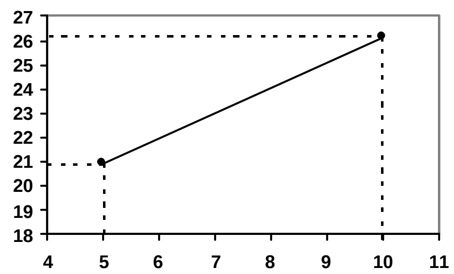
4.2.2 Cas des variables continues :

On cumule les fréquences des classes, en partant des valeurs les plus petites de la variable.
Prenons l'exemple de la répartition par taille des exploitations agricoles des Côtes d'Armor en 1997 (source : Tableaux économiques de Bretagne 2000-2001)

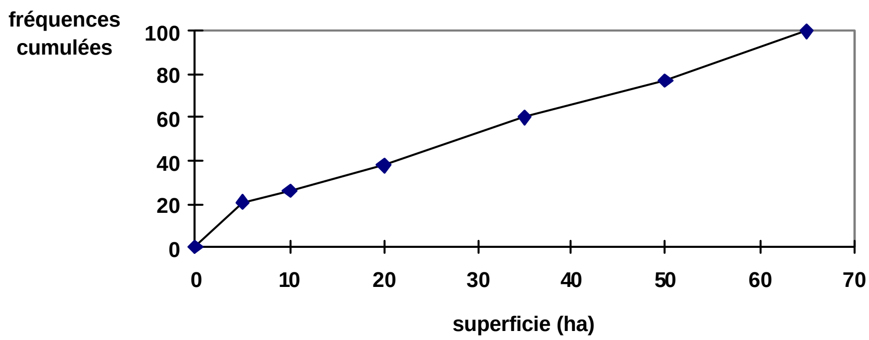
Taille des exploitations agricoles (ha)	fréquence (%)	fréquence cumulée (%)
[0, 5[	20,90	20,90
[5,10[	5,24	26,14
[10, 20[	11,71	37,85
[20,35[	22,13	59,98
[35,50[	17,05	77,03
50 et plus	22,97	100
Total	100	

Le tableau indique qu'il y a 20,90 % d'exploitations de moins de 5 hectares, 26,14% de moins de 10 hectares. Si F est la fonction de répartition, F(5) = 20,90 %, F(10)= 26,14 %, mais on ignore les valeurs de F(x) pour 5< x< 10.

Pour construire le graphique, **on fait l'hypothèse que la distribution est uniforme à l'intérieur d'une classe**, et que la fonction F s'y comporte comme une fonction affine. Sa représentation est un segment de droite



La courbe cumulative est donc obtenue en traçant pour chaque extrémité de classe les points correspondant aux fréquences cumulées, et en les reliant par des segments de droites .
Pour la dernière classe, on utilise la même amplitude que celle de la classe précédente : 15, donc une borne supérieure de 50+15=65. Bien entendu, ceci est tout à fait artificiel et n'exclut pas qu'il y ait des exploitations de plus de 65 ha.

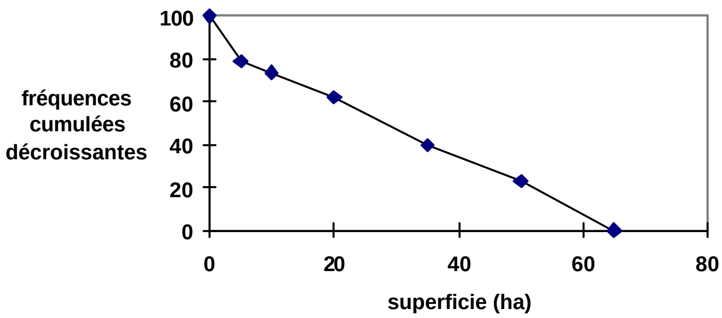


courbe cumulative, ou diagramme intégral

4.2.3 Diagramme des fréquences cumulées décroissantes :

Il est moins utilisé que le précédent. Il représente les fréquences cumulées décroissantes, c'est-à-dire que l'on cumule les fréquences en commençant par les classes de plus grandes valeurs.

Taille des exploitations (ha)	fréquence (%)	fréquences cumulées décroissantes (%)
[0, 5[	20,90	100
[5,10[	5,24	79,10
[10, 20[	11,71	73,86
[20,35[	22,13	22,97 + 17,05 +22,13 = 62,15
[35,50[	17,05	22,97+17,05 = 40,02
50 et plus	22,97	22,97
Total	100	



Encore une fos, la borne supérieure de la dernière classe a été arbitrairement prise à 65 ha

5. Résumé

Une variable statistique continue prend ses valeurs dans tout un intervalle. Une variable discrète est assimilée à une variable continue quand elle prend un très grand nombre de valeurs.

Pour traiter une variable continue il faut **donc regrouper les valeurs possibles** en un petit nombre d'intervalles disjoints (2 Regroupements en classes, p3) .Pour chacun de ces intervalles, il y a **une classe d'individus** pour lesquels la variable est dans cet intervalle. Ces classes sont disjointes et leur réunion recouvre toute la population.

Chaque classe $[x_i, x_{i+1}[$ a

- 2 **bornes** : x_i et x_{i+1},
- un **centre**. $c_i = \dfrac{x_i + x_{i+1}}{2}$,
- une **amplitude** : $a_i = x_{i+1} - x_i$
- Si la dernière classe n'a pas de borne supérieure, on lui affecte l'amplitude de la classe précédente, ce qui permet de lui associer un centre.

Effectifs : pour une classe, c'est le nombre d'individus dans cette classe. On note n_i l'effectif de la classe n°i, n l'effectif total.

Fréquence : 3.2Fréquences p 4 . Pour une classe, c'est la proportion d'individus de cette classe par rapport à la population totale. On la note f_i

$$\mathbf{f_i} = \frac{\textbf{nombre d'individus dans la classe}}{\textbf{nombre d'individus dans la population entière}} = \frac{n_i}{n}$$

Elle peut être exprimée en valeur brute , entre 0 et 1, ou en pourcentage, entre 0% et 100%.

La somme des fréquences est égale à 1 , ou 100 %.

$0 \leq f_i \leq 1$ et $\sum_{i=1}^{k} f_i = 1 = 100\%$, où k est le nombre de classes.

Fréquences cumulées : 3.4 Fréquences cumulées :p 5. Pour la classe n°i , [x_i , x_{i+1}[la fréquence cumulée est

$$\boxed{F_i = f_1 + f_2 + \ldots + f_i = \sum_{p=1}^{i} f_p}$$

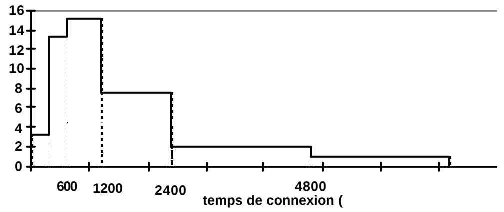
C'est la proportion d'individus dans cette classe ou les précédentes , c'est-à-dire pour lesquels la variable est strictement inférieure à x_{i+1}.
s'il y a k classes, $F_k = 1$

Fréquence corrigée de l'amplitude , fréquence par amplitude, **densité** : 3.3 Fréquences par amplitudes , fréquences corrigées des amplitudes , densités: p4 . C'est le rapport entre la fréquence de la classe et son amplitude.: $\frac{f_i}{a_i}$

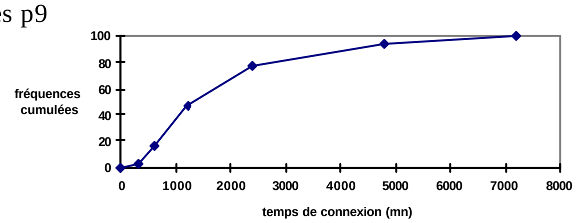
Elle mesure la **densité** d'individus par unité d'amplitude. Elle est utilisées pour construire l'histogramme.

Fonction de répartition : 3.5 Fonction de répartition :p6 Souvent notée F . Pour tout nombre réel x, F(x) est la proportion d'individus pour lesquels la variable est inférieure ou égale à x. Pour une variable discrète, F est une fonction en escalier, mais pour une variable continue, F est une fonction continue. Après un regroupement en classes, on ne connaît les valers de F qu'aux bornes des classes. On assimile donc F à une fonction affine par morceaux.
Lorsque x varie de $-\infty$ à $+\infty$, F(x) croît de 0 à 1.

Histogramme : C'est la représentation graphique des fréquences des classes.
On reporte sur les abscisses les bornes des classes, et sur les ordonnées les fréquences par amplitude. Ainsi les classes sont représentées par des rectangles d'aires proportionnelles aux fréquences de ces classes.
4.1 Histogramme p7



Courbe cumulative , ou **diagramme intégral,** ou **diagramme des fréquences cumulées**: 4.2 Courbe cumulative ou diagramme intégral : représentation des fréquences cumulées p9
Il s'agit de la représentation graphique des fréquences cumulées. Dans le cas d'une variable continue cette courbe est continue, c'est le graphe d'une fonction affine par morceaux, croissante. Le premier point de la courbe a pour ordonnée 0, le dernier a pour ordonnée 1 , ou 100%.



Pour chaque classe [x_i , x_{i+1}[on construit

- le point d'abscisse x_i et d'ordonnée F_{i-1} , fréquence cumulée de la classe n° i -1,
- celui d'abscisse x_{i+1} et d'ordonnée F_i , fréquence cumulée de la classe n°i, $F_i = F_{i-1} + f_i = \sum_{p=1}^{i} f_p$

et on les joint par un segment de droite.