Caractères qualitatifs et discrets

1 - PRESENTATION 1

2 - QUELQUES DEFINITIONS 3

3 - DES CARACTERES DIFFERENTS ... 5

4 - CARACTERES QUALITATIFS 6

4 - 1 Traitement numérique des caractères qualitatifs 6
4 - 1- 1 Effectifs : 6
4 - 1- 2 Fréquences : 7

4 - 2 Traitement graphique des caractères qualitatifs 8
4-2-1 Représentation par secteurs: 9
4-2-2Représentation par tuyaux d'orgues 9
4-2-3Représentation par pictogrammes 10

5- VARIABLES STATISTIQUES DISCRETES 12

5 - 1 Traitement numérique des variables statistiques discrètes 13
5-1-1Effectifs 13
5-1-2-Fréquences : 13
5-1-3Fréquences cumulées : 13
5-1-4 Fonction de répartition : 14

5-2 Traitement graphique des variables statistiques discrètes 15
5-2-1 Représenter les fréquences : le diagramme en bâtons 15
5-2-2 Représenter les fréquences cumulées : la courbe cumulative 15

6- RESUME 16

1 - Présentation

On peut étudier sur les ménages d'une ville,
le nombre de personnes, le nombre d'enfants, le nombre de pièces, le revenu du foyer, le type d'habitation : appartement ou maison individuelle, l'énergie de chauffage : gaz, fuel, électricité, ou autre (charbon, bois...) en voici quelques observations:

n° du ménage	nombre d'enfants	pièces	surface habitable en m^2	Type d'habitation	énergie de chauffage	revenu annuel du ménage, en Euros
1	3	4	77,85	appartement	gaz	32 213,74
2	0	2	37,23	appartement	gaz	22 900,76
3	0	5	92,45	maison individuelle	fuel	48 110,38
4	2	4	89,45	appartement	électricité	55 725 , 46
5	1	3	58,23	appartement	électricité	29 312,52
6	4	5	153,45	maison individuelle	gaz	59 541,98
7	2	3	59,73	appartement	électricité	25 190, 84
8	0	1	35,52	appartement	électricité	29 450,63

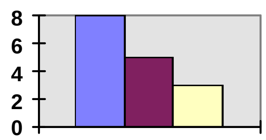
- Après quelques **définitions** :
  paragraphe. 2 - quelques définitions p 3
- vous ferez la distinction entre **caractères qualitatifs** et **quantitatifs**
  paragraphe . 3 - Des caractères **différents...** p5)

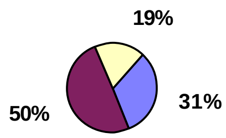
vous établirez des **tableaux d'effectifs** et de **fréquence**

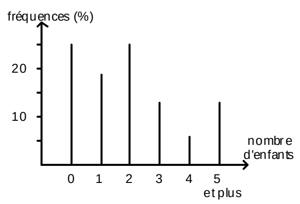
énergie de chauffage	effectifs
gaz	5
électricité	8
fuel	3
Total	16

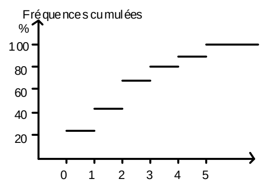
paragraphe 4 - 1 Traitement numérique ***des caractères qualitatifs*** p 6
paragraphe 5 - 1 Traitement numérique des variables statistiques discrètes, p13

- vous ferez connaissance avec quelques représentations graphiques
  4 - 2 Traitement graphique ***des caractères qualitatifs*** p8









Traitement graphique des **variables discrètes**, p 15

2 - quelques définitions

n° du ménage	nombre d’enfants	pièces	surface habitable en m^2	Type d’habitation	énergie de chauffage	revenu annuel du ménage, en Euros
1	3	4	77,85	appartement	gaz	32 213,74
2	0	2	37,23	appartement	gaz	22 900,76
3	0	5	92,45	maison individuelle	fuel	48 110,38
4	2	4	89,45	appartement	électricité	55 725 , 46
5	1	3	58,23	appartement	électricité	29 312,52
6	4	5	153,45	maison individuelle	gaz	59 541,98
7	2	3	59,73	appartement	électricité	25 190, 84
8	0	1	35,52	appartement	électricité	29 450,63

- **Individus, entités statistiques**: du point de vue statistique, chacun des ménages est appelé une **unité statistique** ou encore un **individu.** (ce terme n’est pas réservé aux individus humains ...). Les individus étudiés ont tous les mêmes **caractères**, mais avec des valeurs éventuellement différentes.

	n° du ménage	nombre d’enfants	pièces	surface habitable en m^2	Type d’habitation	énergie de chauffage	revenu annuel du ménage, en Euros
	1	3	4	77,85	appartement	gaz	32 213,74
	2	0	2	37,23	appartement	gaz	22 900,76
®	3	0	5	92,45	maison individuelle	fuel	48 110,38
	4	2	4	89,45	appartement	électricité	55 725 , 46
	5	1	3	58,23	appartement	électricité	29 312,52
	6	4	5	153,45	maison individuelle	gaz	59 541,98
	7	2	3	59,73	appartement	électricité	25 190, 84

- **Caractère** : le nombre d’enfants, le nombre de pièces, le type d’habitations sont appelés des caractères.. Chaque caractère prend pour chaque individu une valeur. Les caractères sont aussi appelés **variables statistiques** lorsqu’ils sont numériques et traduisent une mesure.

n° du ménage	↓ nombre d’enfants	pièces	surface habitable en m^2	Type d’habitation	énergie de chauffage	revenu annuel du ménage, en Euros
1	3	4	77,85	appartement	gaz	32 213,74
2	0	2	37,23	appartement	gaz	22 900,76
3	0	5	92,45	maison individuelle	fuel	48 110,38
4	2	4	89,45	appartement	électricité	55 725 , 46
5	1	3	58,23	appartement	électricité	29 312,52
6	4	5	153,45	maison individuelle	gaz	59 541,98

- **Modalités** : Les modalités d’un caractère sont les valeurs qu’il peut prendre.
  Leur ensemble doit être précisément défini.
  Exemple le type d’habitation peut être ‘appartement’ ou ‘maison individuelle’ et rien d’autre. Le nombre d’enfants est un entier positif ou nul.
- **Classes** : Pour un caractère donné, on regroupera naturellement dans une classe les individus qui présentent la même modalité pour ce caractère (exemple : la classe des ménages qui disposent d’un appartement de 3 pièces). Chaque individu ayant une, et une seule valeur pour ce caractère, on réalise ainsi une partition de la population : chaque individu appartient à une classe, et une seule. On pourra regrouper ensuite dans une même classe les individus pour lesquels le caractère étudié est dans un ensemble précis de modalités (les ménages disposant de 1 à 3 pièces ..) , mais cela concernera plutôt la leçon sur les variables statistiques continues.

- **Effectifs** : pour chaque modalité d’un caractère, l’effectif correspondant sera **le nombre d’entités** statistiques qui ont cette valeur pour ce caractère**.** On comparera souvent cet effectif à l’effectif total, le nombre total d’entités étudiées.
  Attention ! Ne pas confondre l’effectif et le nombre de classes. Pour le caractère type d’habitation, il y a 2 modalités, donc 2 classes, alors que l’effectif total est le nombre de foyers étudiés.

- **Notations** :
  On note en général **n** le nombre total d’individus de la population, **k** le nombre de modalités,

$\mathbf{x_i}$ la modalité n° i du caractère étudié, et $\mathbf{n_i}$ l'effectif correspondant à cette modalité.

3 - Des caractères différents...

Les différents caractères (nombre de personnes au foyer, nombre de pièces, surface habitable, type d'habitation, énergie de chauffage, revenu annuel) ne peuvent pas être étudiés exactement de la même façon :
On peut se poser les questions :

combien de personnes vivent dans un 4 pièces ?
Quelle est la moyenne des revenus annuels des foyers ?

Mais on ne se pose pas les questions :

quelle est la moyenne de l'énergie de chauffage ?
Quel est le nombre de foyers dont le revenu est 29312,52 Euros? (à la place, on cherchera peut-être quel est le nombre de foyers dont le revenu est entre 20 000 et 30 000 Euros)

- **Caractère qualitatif**.
  Dans le tableau précédent (tableau 1) le type d'habitation, comme l'énergie de chauffage, est un caractère qualitatif...
  C'est un caractère qui ne **traduit pas une mesure**. Il ne s'exprime pas, en général, par un nombre, et ses modalités ne sont **en général pas ordonnées**.
  La situation de famille, le sexe, la commune sont des caractères qualitatifs...
  On répertorie toutes les modalités possibles dans une **nomenclature**:
  par exemple la situation de famille peut être 'célibataire', 'marié', 'divorcé',
  dans le tableau 1 l'énergie de chauffage peut être 'fuel', 'électricité', 'gaz', 'autre'.
- **Variables statistiques discrètes** ou **caractères quantitatifs discrets:** Les caractères nombre de pièces, nombre de personnes du tableau1 en sont. Ils traduisent une mesure, ou plus précisément un comptage, leurs modalités sont des **nombres isolés** (il peut y avoir 1 pièce, ou 2, mais pas de valeur entre 1 et 2) .Ce sont en général des nombres entiers.

  S'il y a un très grand nombre de modalités, la variable sera assimilée à une variable continue : par exemple pour des appartements, le nombre d'habitants est une variable statistique discrète, mais pour des villes, ce sera considéré comme une variable statistique continue.

- **Variables statistiques continues** ou **caractères quantitatifs continus** : il s'agit par exemple des caractères surface habitable et revenu dans le tableau 1. Ils traduisent une mesure et s'expriment par un nombre dont les valeurs possibles forment:
  un ensemble continu (la surface habitable peut être n'importe quel nombre réel positif), ou
  ou **un très grand nombre de valeurs isolées** (l'âge en années sur une population allant de 0 à 100 ans).
  Les variables statistiques continues font l'objet de la leçon 3

4 - Caractères qualitatifs

Dans l'exemple précédent, intéressons-nous à un caractère qualitatif : **l'énergie de chauffage** :

n° du foyer	énergie de chauffage		n° du foyer	énergie de chauffage
1	gaz		9	gaz
2	gaz		10	fuel
3	fuel		11	fuel
4	électricité		12	électricité
5	électricité		13	gaz
6	gaz		14	électricité
7	électricité		15	électricité
8	électricité		16	électricité

Ensemble, nous allons faire :

quelques calculs : traitement numérique des caractères qualitatifs, sur cette page

puis des représentations graphiques : 4 - 2 Traitement graphique ***des caractères qualitatifs***, p 8

4 - 1 Traitement numérique des caractères qualitatifs

On peut se poser les questions ***:***

→ Combien y -a-t-il d'individus dans chaque classe ?

La réponse est le calcul d'**effectifs**

→ Quelles sont les proportions entre les classes ? Comment **comparer** 2 populations ?

La réponse est le calcul de **fréquences**.

4 - 1- 1 Effectifs :

il s'agit simplement, pour chaque modalité de la variable statistique, du nombre d'individus pour lesquels la variable a cette valeur.

Exemple : l'effectif de la classe 'gaz' est 5

On peut dresser le tableau des effectifs:

énergie de chauffage	effectifs
gaz	5
électricité	8
fuel	3
Total	16

Notation : si k est le nombre de classes , ici k = 3, on note :

x_i la modalité n° i

n_i l'effectif correspondant à cette modalité

x_1 : gaz, x_2: électricité, x_3 : fuel.

$n_1 = 5$, $n_2 = 8$, $n_3 = 3$

Important : Le total des effectifs doit être égal au nombre d'individus

ce qui s'écrit $\sum_{1 \le i \le k} n_i = n$

4 - 1- 2 Fréquences :

Pour comparer les parts d'effectifs des différentes modalités, on construit le tableau des fréquences .
L'avantage principal par rapport aux effectifs est de pouvoir **comparer des populations de tailles différentes** :

Exemple : Marins embarqués pour la pêche, par type de pêche, en pourcentages:
(source : Annuaire statistique de France , INSEE)
La population de 1997 (13 353 marins) est plus restreinte que celle de 1985 (18 396 marins)

	1985 (en pourcentage, population 18 396 marins)		1997 (en pourcentage, population 13 353 marins)
Grande pêche	2,8		2,6
Pêche au large	18,0		13,5
Pêche côtière	11,2		14,3
Petite pêche, y compris conchyliculture	68,0		69,6

Calcul : pour chaque modalité, la fréquence est le rapport entre l'effectif correspondant à cette modalité et l'effectif total.
S'il s'agit de la modalité n° i,

et que l'effectif correspondant est n_i, la fréquence de cette classe est $f_i = \frac{n_i}{n}$

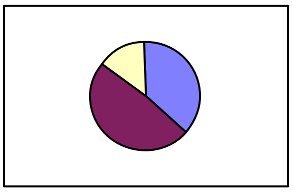
La fréquence est obligatoirement entre 0 et 1.
On peut l'exprimer en pourcentages (dans ce cas elle est entre 0 % et 100 %)
Dans l'exemple des foyers, **la fréquence de la modalité électricité est Erreur != 0,5**
on peut dresser le tableau des fréquences :

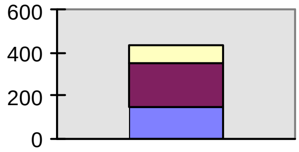
énergie de chauffage	effectifs	fréquences
gaz	5	0,3125
électricité	8	0,5000
fuel	3	0,1875
Total	16	1

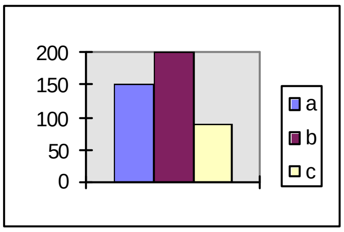
Important : Chaque fréquence est entre 0 et 1 (ou 0 % et 100 %)
le total des fréquences est de 1 , ou de 100%

4 - 2 Traitement graphique des caractères qualitatifs

Dans le cas d'un caractère qualitatif, et contrairement aux caractères quantitatifs, **les modalités ne sont pas ordonnées**. La représentation ne doit donc pas suggérer un ordre. On n'utilise donc pas un axe gradué pour les abscisse comme on le faisait pour un caractère quantitatif..







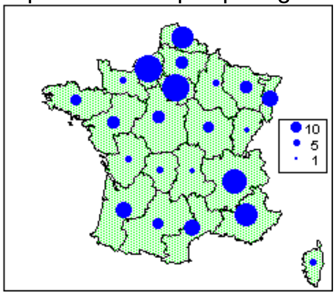
Secteurs

4-2-1 Représentation par secteurs:p8

Tuyaux d'orgues

4-2-2 - Représentation par tuyaux d'orgues p9

Représentation par pictogrammes



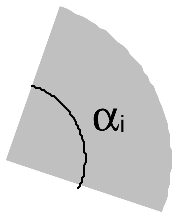
4-2-3 - Représentation par pictogrammes p 10

4-2-1 Représentation par secteurs:

On découpe un disque en secteurs dont les aires sont proportionnelles aux fréquences.
Les angles sont aussi proportionnels aux fréquences

Pour la modalité n° i:

L'angle est $a_i = 360° x \frac{n_i}{n}$

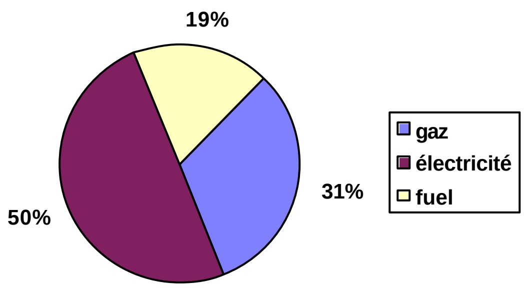


énergie de chauffage	effectifs	fréquences
gaz	5	0,3125
électricité	8	0,5000
fuel	3	0,1875
Total	16	1

Gaz $360° \times \frac{5}{16} = 112°30'$

Electricité : $360° \times \frac{8}{16} = 180°$

Fuel : $360° \times \frac{3}{16} = 67°30'$

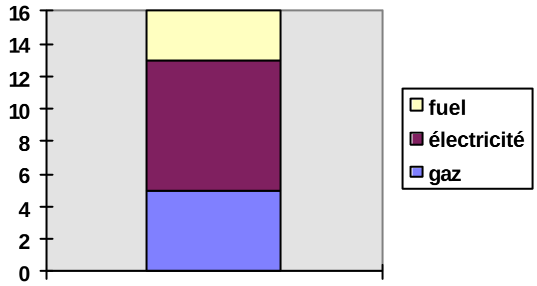


4-2-2 - Représentation par tuyaux d'orgues

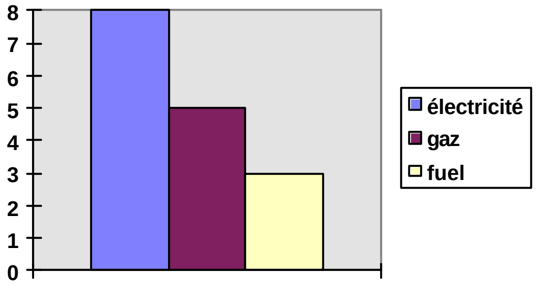
On représente les effectifs ou les fréquences par des rectangles de même largeur, et de longueur proportionnelle aux effectifs, et aux fréquences.

énergie de chauffage	effectifs	fréquences
gaz	5	0,3125 ≈ 0,31
électricité	8	0,50
fuel	3	0,1875 ≈ 0,19
Total	16	1

2 possibilités :



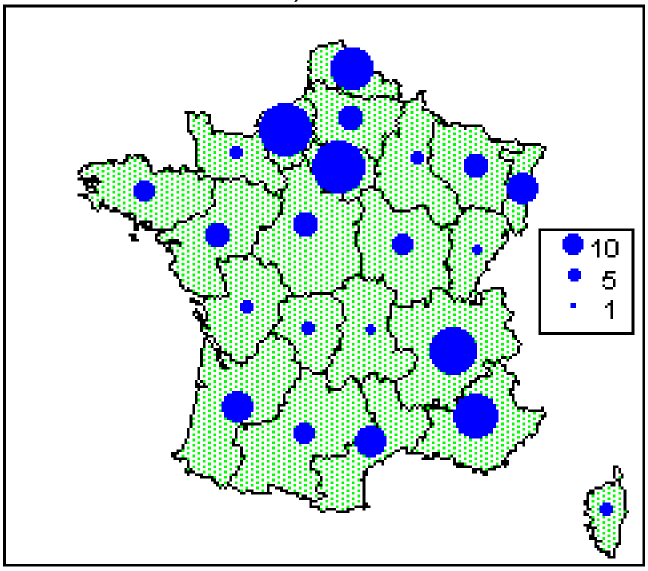
OU



N'oubliez pas que dans le graphique de droite, les modalités ne sont **pas ordonnées** a priori. .
On peut donc les placer dans l'ordre croissant, ou décroissant comme ici, des effectifs, ce qui facilite les comparaisons.

4-2-3 - Représentation par pictogrammes

On représente les données par des figures homothétiques **d'aires proportionnelles aux effectifs**, et donc aux fréquences.
Exemple : Nombre d'installations industrielles à risques (Seveso1) par régions françaises, en 1998 (source: INSEE, TEF 2000/20001)

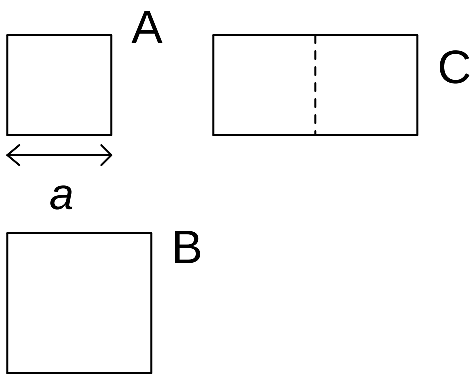


Attention : Ce sont bien **les aires** et **non les dimensions** qui sont proportionnelles aux effectifs.
Détaillons ce point :

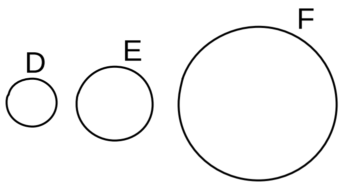
Si le carré A, de côté a, représente 1 unité ,
pour représenter 2 unités :

on pourrait utiliser C, mais ce n’est pas un carré

On utilise le carré B , mais attention son côté est
seulement $a\sqrt{2} \approx 1{,}4$ a et non 2a



:
Un exemple avec des cercles :
D représente une grandeur d’une unité
E de 2 unités et F de 10 unités.



Par rapport à l’aire de D ,
celle de E est multipliée par 2 , celle de F par 10,
par rapport au rayon de D,
celui de E est multiplié par $\sqrt{2} \approx 1{,}4$, celui de F par $\sqrt{10} \approx 3{,}2$

Les pictogrammes ne sont bien adaptés que pour comparer de grandeurs largement différentes : pour comparer avec précision des grandeurs proches, il est plus efficace d’utiliser des diagrammes en tuyaux d’orgues.

5 - Variables statistiques discrètes (ou caractères quantitatifs discrets)

Ce sont des caractères mesurables, mais qui ne peuvent prendre que certaines valeurs isolées.:
nombre d'enfants, nombre de parts entrant dans le calcul de l'impôt sur le revenu...
Dans les données qui servent d'exemple,
il s'agira du nombre d'enfants, et du nombre de pièces

n° du ménage	nombre d'enfants	pièces	surface habitable en m^2	Type d'habitation	énergie de chauffage	revenu annuel du ménage, en Euros
1	3	4	77,85	appartement	gaz	32 213,74
2	0	2	37,23	appartement	gaz	22 900,76
3	0	5	92,45	maison individuelle	fuel	48 110,38
4	2	4	89,45	appartement	électricité	55 725 , 46
5	1	3	58,23	appartement	électricité	29 312,52
6	4	4	153,45	maison individuelle	gaz	59 541,98
7	2	3	59,73	appartement	électricité	25 190, 84
8	0	1	35,52	appartement	électricité	29 450,63

- Comme pour les variables qualitatives, des **classes** apparaissent naturellement: on pourra **regrouper** les logements qui ont le même nombre de pièces
- Mais en plus , on pourra **ordonner** les données : on peut étudier les appartements qui ont plus de 2 pièces.

Il y a , a priori, **une classe par valeur possible** : 0,1 , 2, 3, 4, 5, 6, ...
Mais il arrive souvent que pour la **dernière** classe on prenne un regroupement ; 0,1 , 2, 3, 4, 5 et plus

Dans les écrans suivants, vous allez voir:

quelques calculs : Traitement numérique des variables quantitatives discrètes, p13
des représentations graphiques : - Traitement graphique des variables statistiques discrètes, p15

5 - 1 Traitement numérique des variables statistiques discrètes

- Les **effectifs** **r**épondront à la question : **Combien** y a-t-il d'individus pour chaque valeur de la variable ?

- Les **fréquences** répondront à : Quelle est **la part** de population correspondant à une valeur ? Comment **comparer** deux populations de tailles différentes ?

- Les **fréquences cumulées** et la **fonction de répartition** répondront à : quelle proportion d'individus a moins d'une certaine valeur ? Quelle valeur maximale atteint la première moitié des effectifs ?

5-1-1 - Effectifs

Dans le tableau précédent, pour chaque classe, on compte le nombre de logements.

nombre d'enfants	effectif
0	4
1	3
2	4
3	2
4	1
5 et plus	2
Total	16

Notation : on note **n l'effectif total** : n = 16 , **n_i l'effectif de la classe n° i**. , $n_3 = 5$

5-1-2- Fréquences :

Pour chaque modalité, on calcule la proportion d'individus. $\boxed{f_i = \frac{n_i}{n}}$

Les fréquences sont utiles notamment pour **comparer** des populations de tailles différentes.

nombre d'enfants	effectif : n_i	fréquence . f_i = **Erreur !**
0	4	25 %
1	3	18,75 %
2	4	25 %
3	2	12,5 %
4	1	6,25 %
5 et plus	2	12,5 %
Total	16	100

Important : t la somme des fréquences est égale à 1 , ou à 100 %

5-1-3 - Fréquences cumulées :

On peut se poser les questions :
- Quelle est la proportion de ménages de moins de 2 enfants ?
- A partir de combien d'enfants un ménage en a-t-il plus que la moitié de la population ?

Pour y répondre, on cumule les fréquences :
$F_1 = f_1$ proportion de ménages sans enfants
$F_2 = f_1 + f_2$ proportion de ménages avec au plus 1 enfant
$F_3 = f_1 + f_2 + f_3$ proportion de ménages avec au plus 2 enfants...
...
F_i s'appelle la ***fréquence cumulée*** de la classe i

$\boxed{F_i = f_1 + f_2 + \ldots + f_i = \sum_{k=1}^{i} f_k}$ *est la proportion d'individus pour lesquels la valeur de la variable étudiée est inférieure ou égale à la modalité n° i.*

nombre d'enfants	effectif : n_i	fréquence . f_i = (en pourcentage)	fréquences cumulées F_i (en pourcentage)
0	4	25	25
1	3	18,75	25 +18,75 = 43,75
2	4	25	25 + 18,75 +25 = 68,75
3	2	12,5	25+18,75+25+12,5 = 81,25
4	1	6,25	25+18,75+25+12,5 + 6,25 = 87,5
5 et plus	2	12,5	100
Total	16	100 %	

On peut maintenant répondre aux 2 questions posées :

Il y a environ 44 % de ménages qui ont moins de 2 enfants,

et si on regroupe 50 % des ménages en prenant ceux qui ont le moins d'enfants, ils ont au plus 2 enfants

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

5-1-4 -Fonction de répartition :

Cette notion apparaîtra plus naturellement pour les variables statistiques continues.

pour **tout nombre** réel x, et pas seulement pour les valeurs de la variable,

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

Pour une variable statistique discrète, si x est compris entre la i ème modalité et la suivante: $x_i \leq x < x_{i+1}$
alors $F(x) = F_i$

ici,
$x_4 = 3$ $x_5 = 4$.
Donc $F(3) = F_4 = 0{,}8125 = 81{,}25\ \%$,
et F(3,2) = 81,25 % aussi. puisque $3 \leq 3{,}2 < 4$

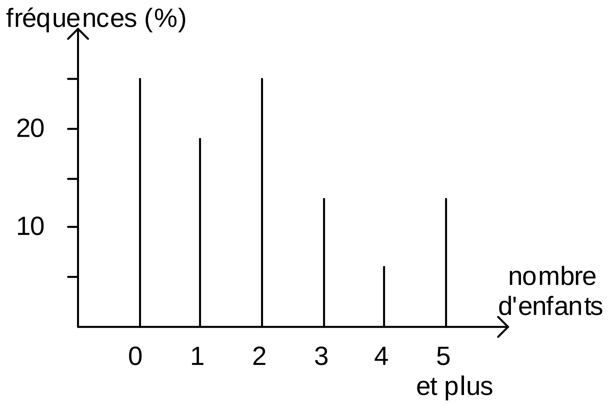
Attention : Malheureusement, pour cette notion, **deux conventions** différentes **existent**: dans certains ouvrages, mais jamais dans ce cours, vous pourrez rencontrer F(x) définie comme la proportion d'individus pour lesquels la valeur du paramètre est strictement inférieure à x . Avec cette définition, F(3) serait de 68,75 %

5 -2 - Traitement graphique des variables statistiques discrètes

- diagramme des fréquences (en bâtons)
- diagramme des fréquences cumulées

5-2-1 - Représenter les fréquences : le diagramme en bâtons

Ce diagramme met bien en évidence le fait que la variable ne prend que des valeurs isolées.
Pour chacune de ces valeurs, on trace un segment de longueur proportionnelle à la fréquence correspondante.
Le calcul des fréquences vient de l'écran précédent

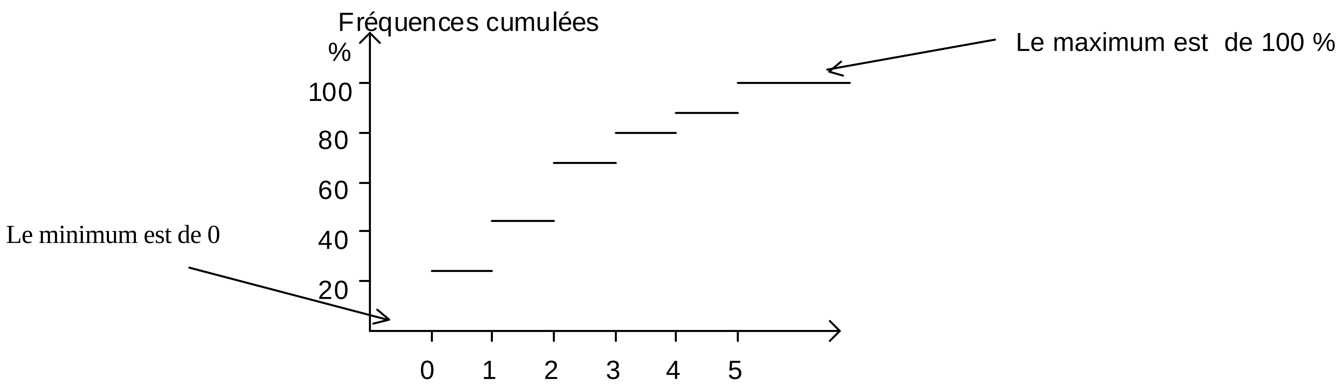


nombre d'enfants	fréquence . f_i = **Erreur !** (en pourcentage)
0	25
1	18,75
2	25
3	12,5
4	6,25
5 et plus	12,5
Total	100 %

5-2-2 - Représenter les fréquences cumulées : la courbe cumulative

Ce graphique représente la fonction de répartition F: pour tout nombre x , F(x) est la proportion d'individus dont la valeur de la variable est inférieure ou égale à x.
Si x_i est la valeur de la variable pour la classe n°i, F (x) = F_i pour le i tel que $x_i \leq x < x_{i+1}$
F(3) = F_4 = 81,25 , F(2, 7) = F(2) = F_3 = 68,75

Dans le cas d'une variable discrète, F évolue par paliers .



Important :

- la fonction de répartition est **croissante**
- elle vaut **0 en -¥** et **1 (ou 100 %) en +¥**
- dans le cas d'une **variable discrète**, elle évolue **par paliers**. A la i ème modalité x_i ,
  le saut est $F_i - F_{i-1} = f_i$.

6- Résumé

Les **caractères statistiques** peuvent être **quantitatifs**, s'ils traduisent une mesure — on parle alors de **variables statistiques —** , **qualitatifs** sinon. Les caractères quantitatifs peuvent être **discrets** : ils ne prennent que certaines valeurs isolées; ou **continus**: ils prennent leurs valeurs dans des intervalles de R.

Les valeurs possibles d'un caractère sont appelées ses **modalités**.

Pour étudier un caractère qualitatif :

- on calcule les **effectifs** : n_i est le nombre d'individus pour lesquels le caractère est égal à la modalité n°i
- on calcule les **fréquences** : **f_i = Erreur !**
- on représente les **effectifs** par une **représentation en tuyaux d'orgues**
- on représente la **structure** (les différentes **proportions**) par un **graphique en secteurs**
- on peut dans certains cas utiliser des **pictogrammes** dont les **aires** (et non les dimensions) sont proportionnelles aux effectifs et aux fréquences.

Pour étudier un caractère quantitatif discret ou variable statistique discrète

- on calcule les **effectifs**, les **fréquences** comme pour un caractère qualitatif, mais aussi les **fréquences cumulées : $F_i = f_1 + f_2 + \dots + f_i$**
- on représente les effectifs ou les fréquences par un **diagramme en bâtons** , et les fréquences cumulées par une **courbe cumulative**.

Quelques égalités : si n est le nombre d'entités, et k est le nombre de modalités :

$\sum_{1 \le i \le k} n_i = n$: la somme des effectifs pour les différentes modalités est égale à l'effectif total.

$\sum_{1 \le i \le k} f_i = 1 = 100\%$: la somme des fréquences est égale à 1 (ou à 100%)

$F_k = 1$ **La fréquence cumulée de la dernière classe est égale à 1.**