

Chapitre 1

Distribution et répartition

Une VARIABLE décrit un *caractère* particulier de chaque INDIVIDU de la POPULATION à étudier : elle peut être de NATURE

- NOMINALE si elle représente les différents caractères par des *mots* ou
- NUMERIQUE si les caractères sont des *nombres* : elle est alors DISCRETE ou CONTINUE.

Elle est dite ORDINALE si ses *modalités* (valeurs) peuvent être ordonnées objectivement et CATEGORISEE si ses modalités sont mises en classes.

Une variable est en plus décrite par un TYPE

- QUALITATIF si elle représente un état ;
 - QUANTITATIF si sa valeur est (ou dépend d')une *quantité* (celle-ci pouvant être la somme de valeurs plus petites qui ont le même sens physique)
- et elle est dite
- SEMI-QUANTITATIVE si elle est qualitative ordinale avec des modalités numériques.

Exemples : Temps, longueur et masse (donc aire, volume et énergie) sont des quantités.

Une concentration n'est pas une quantité (car deux litres d'alcool à 12 degrés donnent de l'alcool à 12 degrés) mais la variable correspondante est quantitative parce qu'elle dépend d'une quantité (dans un volume de liquide).

La température n'est pas une quantité (car on n'obtient pas 2 °C en assemblant deux objets à 1 °C). De même un degré de douleur, défini par un point sur un segment de 10 cm allant de « Douleur absente » à « Douleur insupportable », n'est pas une quantité. Dans les deux cas la variable correspondante n'est pas quantitative.

DESCRIPTION PAR LES FREQUENCES

Une variable se décrit par une DISTRIBUTION (valeurs et fréquences simples) et aussi, dans le cas *ordinal*, par une REPARTITION (valeurs et fréquences cumulées).

Si la distribution est résumable à quelques fréquences on peut présenter graphiquement celles-ci par un diagramme en barres (ou un diagramme circulaire) et, si la variable est ordinale, on peut aussi représenter les fréquences cumulées par un diagramme en barres.

Si la variable est *numérique* la présentation la plus précise se fait avec un diagramme particulier nommé HISTOGRAMME, soit si la variable est

- *discrète* par un diagramme en bâtons, la répartition étant représentée par un diagramme en escalier ;

- *continue* par un diagramme en rectangles (chaque rectangle ayant pour base un intervalle de valeurs et pour hauteur la densité, c'est-à-dire la fréquence des valeurs situées dans l'intervalle rapportée à la largeur de l'intervalle), la répartition étant représentée par un contour polygonal (qui est une ligne brisée).

Exemple : Recensement des N = 259 médecins généralistes ayant soutenu leur thèse à Dijon entre 1998 et 2002 (thèse d'Arnaud Gouget, Dijon-2008) : on a recueilli de l'information sur

- le sexe (homme ou femme),
- l'âge à la soutenance de thèse (de 26 à 48 ans),
- le nombre de diplômes supplémentaires (de 0 à 7),
- le mode d'exercice libéral (1) ou salarié (0),
- le département d'exercice (Yonne, Côte d'Or, Haute-Saône, Nièvre ou autre),
- la présence d'enfants (aucun, un seul, deux, ou au moins 3),

et

- la satisfaction de sa situation (nulle, moyenne ou parfaite).

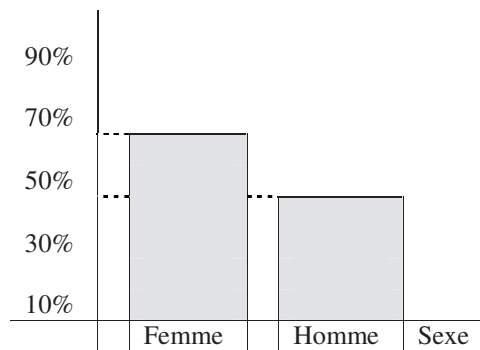
*Fréquences simples***[variable nominale] Diagramme en barres**

Exemple 1-1 : Soit à représenter la variable « Sexe » dont on donne ci-dessous les fréquences en privilégiant les relatives qui permettent une comparaison directe avec d'autres études (faites sur des effectifs différents)

Sexe		(n = 257)
Femme	61,48%	(158)
Homme	38,52%	(99)

(sexe non précisé sur 2 questionnaires)

Cette variable est *nominale* à 2 modalités, et *qualitative*, donc on décrit graphiquement la distribution des sexes avec un diagramme *en barres*



§

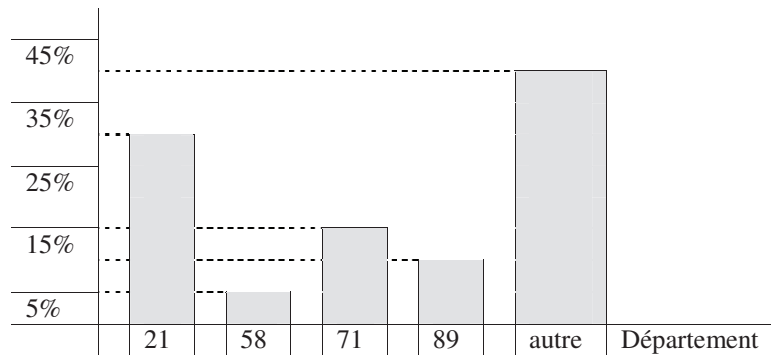
Exemple 1-2 : Soit à représenter la variable « Département d'exercice » dont la distribution est décrite ci-dessous en fréquences simples :

Département		(n = 236)
21-Côte d'Or	30,51%	(72)
58-Nièvre	2,97%	(7)
71-Saône et Loire	16,95%	(40)
89-Yonne	8,05%	(19)
Autre	41,52%	(98)

(département non précisé sur 23 questionnaires)

14 ELEMENTS DE STATISTIQUE DESCRIPTIVE

Cette variable est *nominale* à 5 modalités, et *qualitative pure*, donc on la représente graphiquement par un diagramme *en barres* :



Les départements auraient pu être présentés du nord au sud (Yonne, Côte d'Or, Nièvre, Saône et Loire) ou par nombre d'habitants (Côte d'Or, Saône et Loire, Yonne, Nièvre) : il n'y a pas d'ordre principal, toutes les conventions sont acceptables.

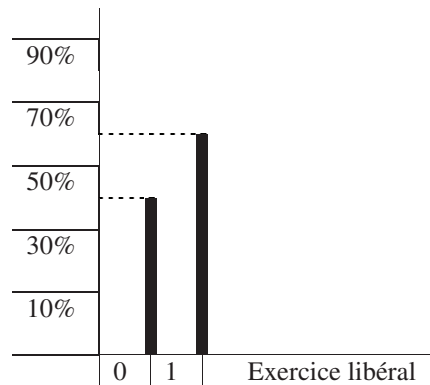
[variable binaire] Histogramme en bâtons

Exemple 1-3 : Soit à représenter, pour le Mode d'exercice (salarié ou libéral), la variable « Exercice libéral » (1 si oui, 0 sinon) dont la distribution des valeurs est la suivante

Exercice libéral		(n = 236)
1	57,20%	(135)
0	42,80%	(101)

(exercice non précisé sur 23 questionnaires)

On présente graphiquement cette variable *numérique* dite *binaire* (en 0 ou 1), qui est de type *qualitatif*, avec un *histogramme en bâtons* (pour rappeler qu'aucun nombre autre que 0 ou 1 ne peut être observé, ce qui fait dire que la variable numérique est *discrète*) :



Remarque : le *bâton* est idéalement une ligne (sans épaisseur) tandis que la *barre* a une épaisseur bien visible, mais faire un graphique clair imposant de donner au bâton une épaisseur, il faut simplement veiller à ce que cette dernière soit nettement plus faible que celle d'une barre.

§

Fréquences simples et cumulées

[variable ordinale] Diagramme en barres

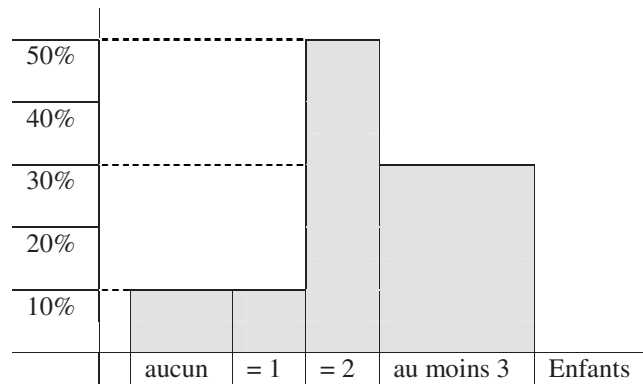
Exemple 1-4 : Soit à représenter la variable « Présence d'enfants » dont la distribution ci-dessous est complétée par la répartition obtenue en cumulant les fréquences à partir de la première :

Enfants		(n = 227)		Pas d'enfant	
Aucun	11,01%	(25)		1 au plus	11,01%
1	11,89%	(27)		2 au plus	22,90%
2	47,14%	(107)		Tous les cas	70,04%
Au moins 3	29,96%	(68)			100,00%

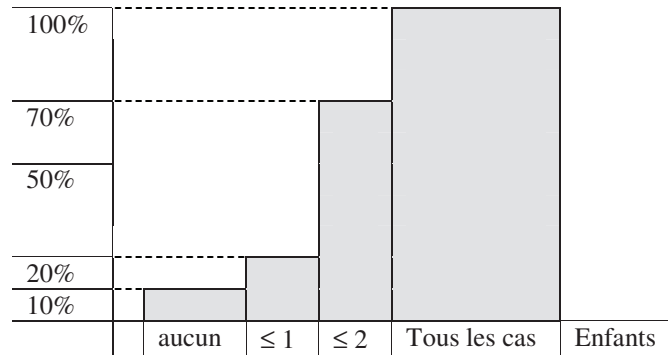
(nombre d'enfants non précisé sur 32 questionnaires)

16 ELEMENTS DE STATISTIQUE DESCRIPTIVE

Pour cette variable *nominale*, qui est *qualitative ordinale* puisque ses 4 modalités peuvent être ordonnées de façon objective, on présente graphiquement la distribution avec un *diagramme en barres*



et on fait de même pour la répartition (utile puisque la variable est ordinale)



ce qui montre que 70% des généralistes ont au plus 2 enfants.

§

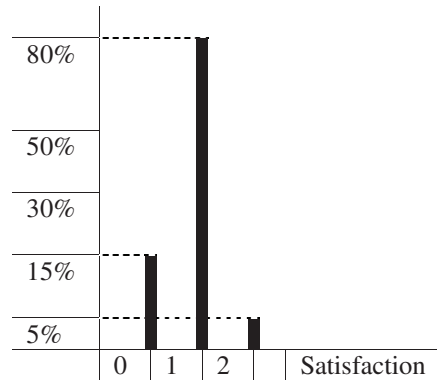
[variable discrète] Histogramme en bâtons et diagramme en escalier

Exemple 1-5 : Soit à représenter la variable « Satisfaction de sa situation » (0 = nulle, 1 = moyenne et 2 = parfaite) dont on a ci-dessous les fréquences simples et les fréquences cumulées à partir de la première :

Satisfaction	(n = 231)			
0	14,72%	(34)	0 au plus	14,72%
1	82,25%	(190)	1 au plus	96,97%
2	3,03%	(7)	2 au plus	100,00%

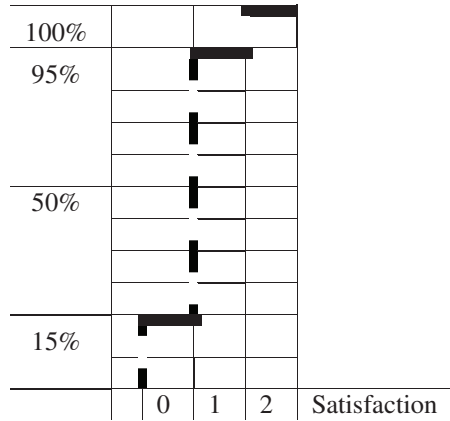
(satisfaction non précisée sur 28 questionnaires)

Pour cette variable *numérique discrète*, qui est *semi-quantitative*, le graphique est un *histogramme en bâtons* pour la fonction de distribution :



18 ELEMENTS DE STATISTIQUE DESCRIPTIVE

et un diagramme *en escalier* pour la fonction de répartition :



Remarque : On peut aussi cumuler les fréquences à partir de la fin :

Satisfaction	Fréquence relative (%)	Effectif (n)	0 au moins	Fréquence relative (%)
0	14,72%	(34)	0 au moins	100,00%
1	82,25%	(190)	1 au moins	85,28%
2	3,03%	(7)	2 au moins	3,03%
			3 au moins	0,00%

§

Exemple 1-6 : Soit à représenter la variable « Nombre de diplômes supplémentaires »

Diplômes en plus	Fréquence relative (%)	Effectif (n)	≤	Fréquence relative (%)
0	38,99%	(85)	≤ 0	38,99%
1	27,06%	(59)	≤ 1	66,06%
2	16,06%	(35)	≤ 2	82,11%
3	8,72%	(19)	≤ 3	90,83%
4	4,59%	(10)	≤ 4	95,41%
5	3,21%	(7)	≤ 5	98,62%
6	1,38%	(3)	≤ 6	100,00%

(nombre de diplômes non précisé sur 41 questionnaires)