

Tableaux de distribution et représentation graphiques

Afin de faciliter l'analyse des données statistiques recueillis lors d'une étude, on les groupe par valeur ou catégories, puis on les présente sous forme de tableau ou e graphique

1 Tableaux de distribution

On obtient un tableau de distribution, en énumérant les valeurs (ou catégories) de la variable ; puis en inscrivant le nombre (effectif ou fréquence absolue) ou le pourcentage de données (ou fréquence relative) de la série statistique qui correspondent à chaque valeur (ou catégories). C'est ce qu'on appelle le tableau de distribution.

1.1 Effectif ou fréquence absolue

On appelle effectif ou fréquence absolue d'une valeur (ou d'une catégorie) le nombre de données égale a cette valeur ou à cette catégorie.

1.2 Fréquence relative

On appelle fréquence relative d'une valeur (ou d'une catégorie) la proportion de données égales à cette valeur (ou à cette catégorie). Cette mesure, généralement exprimée en pourcentage, est particulièrement utile pour comparer deux séries statistiques n'ayant pas le même effectif total.

Répartition des accidents selon le nombre de victimes

Nombre de victimes (x_i)	Nombre d'accidents (n_i)	Pourcentage (%)
1	32	57.1%
2	13	23.2%
3	5	8.9%
4	3	5.4%
5	3	5.4%
TOTAL	56	100%

Le nom de la variable : nombre d'accidents

Les valeurs de la variable : 1 accident, 2 accidents,3 accidents, 4 accidents, 5 accidents

Type de la variable : quantitative discrète

Echelle de mesure : échelle de rapport

Répartition des accidents selon le genre d'accidents

Genre d'accidents	Nombre d'accidents	Pourcentage (%)
Collision avec un véhicule	23	41.1%
Sans collision	24	42.9%
Collision avec un animal	4	7.1%
Collision avec objet fixe	5	8.9%
TOTAL	56	100%

Le nom de la variable : le genre d'accidents

Les catégories de la variable : Collision avec un véhicule, Sans collision, Collision avec un animal, Collision avec objet fixe

Type de la variable : qualitative nominale

Echelle de mesure : échelle nominale

2 Représentations graphiques

Alors que la présentation des données sous la forme d'un tableau de distribution en facilite l'analyse, la représentation graphique de la distribution permet de repérer en un coup d'œil les faits saillants de la distribution.

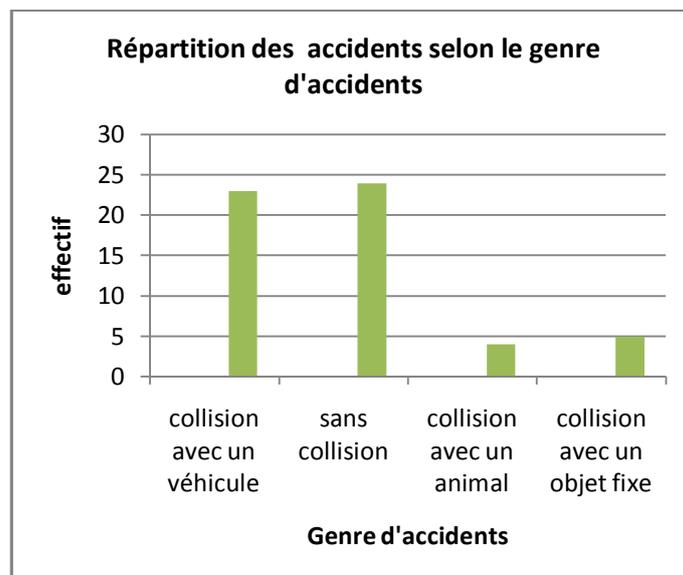
On représente

- Une variable qualitative par un diagramme à rectangles verticaux ou horizontaux
- Une variable quantitative discrète par un diagramme en bâtons
- Une variable quantitative continue par un histogramme

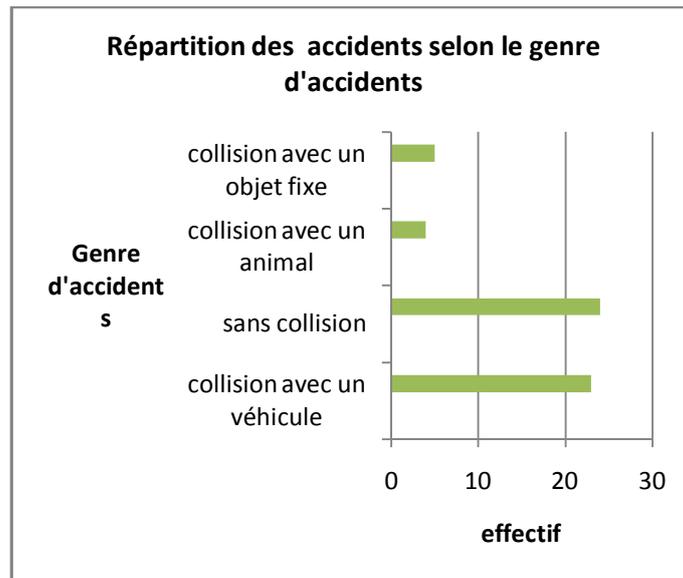
2.1 Cas d'une variable qualitative « diagramme à rectangles verticaux ou horizontaux »

Pour construire un diagramme à rectangles verticaux, on inscrit les catégories de la variable sous l'axe horizontal d'un système d'axes, puis on érige des rectangles, non adjacents, de hauteur égale à l'effectif ou à la fréquence relative (en pourcentage) au-dessous de chaque catégorie. Enfin, on nomme les axes et on donne un titre au diagramme

« Diagramme à rectangles verticaux »

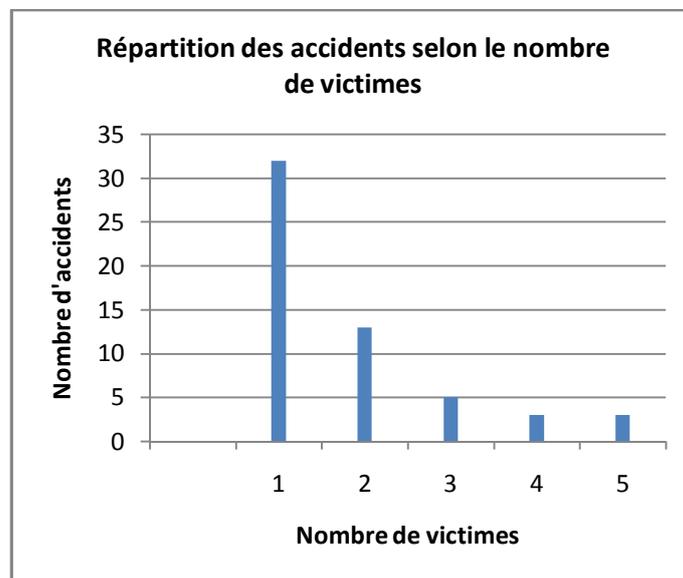


« Diagramme à rectangles horizontaux »



2.2 Cas d'une variable quantitative discrète « diagramme en bâtons »

Pour construire un diagramme en bâtons, on porte sur l'axe horizontal les différentes valeurs de la variable selon une échelle choisie arbitrairement, puis on élève sur chaque valeur un bâton de longueur proportionnelle à l'effectif ou à la fréquence relative. Par la suite, on donne un titre au diagramme et on nomme les axes



2.3 Cas d'une variable quantitative continue « HISTOGRAMME »

2.3.1 Histogramme à classes égales

Un histogramme est formé de rectangles adjacents ; ceux-ci sont accolés afin d'indiquer que l'étude porte sur une variable continue. La base des rectangles correspond aux classes de la distribution, et la surface de chacun est proportionnelle à l'effectif ou à la fréquence de la classe à laquelle il est

associé. Quand toutes les classes ont la même amplitude, on respecte le principe de proportionnalité entre la surface des rectangles et les fréquences en érigeant, sur chaque classe préalablement indiquée sur l'axe horizontal, un rectangle dont la hauteur est égale à la fréquence de la classe

Les démarches à suivre pour construire le tableau de distribution

Pour construire un tableau de distribution d'une variable quantitative continue, on procède comme suit :

- a- Déterminer le nombre de classes

Le nombre de classes à créer dépend du nombre de données de la série statistique. On se sert de la formule de STURGES pour fixer temporairement le nombre de classes :

$$\text{Nbre de classes} = 1 + 3.32 \log N$$

N est le nombre de données

Nb : il existe d'autres formules comme celle de YULE (Nbre de classes = $2.5 \times \sqrt{N}$)

- b- Calculer l'étendu de la série statistique

L'étendue, que l'on note E, est égale à la différence entre la plus grande valeur de la série statistique et la plus petite valeur

$$E = X_{\max} - X_{\min}$$

- c- Déterminer l'amplitude des classes

Nous voulons savoir quelle sera la largeur de chaque classe, cette réponse est donnée par la formule suivante

$$a = \frac{\text{Etendu}}{\text{Nbr de classe}}$$

Il est préférable de choisir comme amplitude un multiple de 5 ou un nombre pair. Si les données à grouper sont des entiers il conviendra de choisir comme amplitude un entier, et si les données sont précises aux dixièmes près ou aux centièmes près, nous utiliserons la même précision pour l'amplitude choisie

- d- Choisir la limite inférieure de la première classe

Le choix de la limite inférieure de la première classe est déterminant dans la construction des classes. Nous choisirons donc un nombre qui permettra de produire un tableau de distribution agréable à lire. La limite inférieure de la première classe devra être plus petite ou égale à la plus petite donnée de la série statistique.

Exemple

La série suivante représente la longueur d'un échantillon de 51 branches d'arbres. Les mesures sont exprimées en cm

10.8	8.9	11.7	13.2	7.9	15.2
6.6	7.9	6.0	9.8	5.7	11.5
4.2	3.2	4.2	12.0	2.3	4.9
16.0	10.4	6.9	4.0	10.1	8.0
11.3	8.9	10.0	12.0	5.8	11.8
5.8	12.7	6.5	7.8	18.5	4.8
11.1	13.7	6.1	5.6	6.7	7.0
8.7	9.5	5.8	9.4	6.8	8.6
9.6	3.5	4.5			

1- fixer le nombre de classe

Selon la règle de STURGES : Nbre de classes= $1 + 3 \cdot 32 \log N$

$$\text{Nbre de classes} = 1 + 3 \cdot 32 \log 51$$

$$\text{Nbre de classes} = 1 + 3 \cdot 32 (1.71) = 6.64 = 7 \text{ classes}$$

2- calculer l'étendu E :

$$E = X_{\max} - X_{\min} = 18.5 - 2.3 = 16.2$$

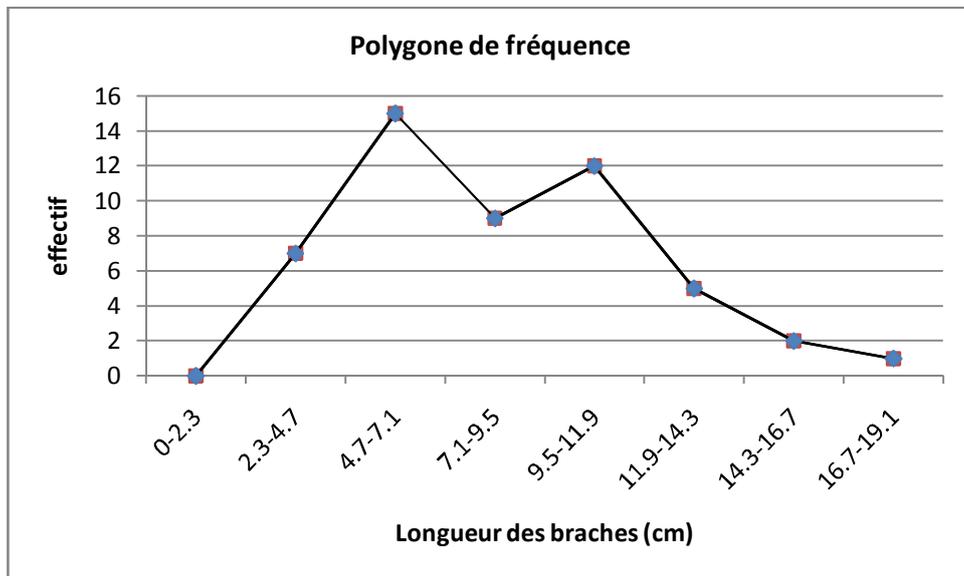
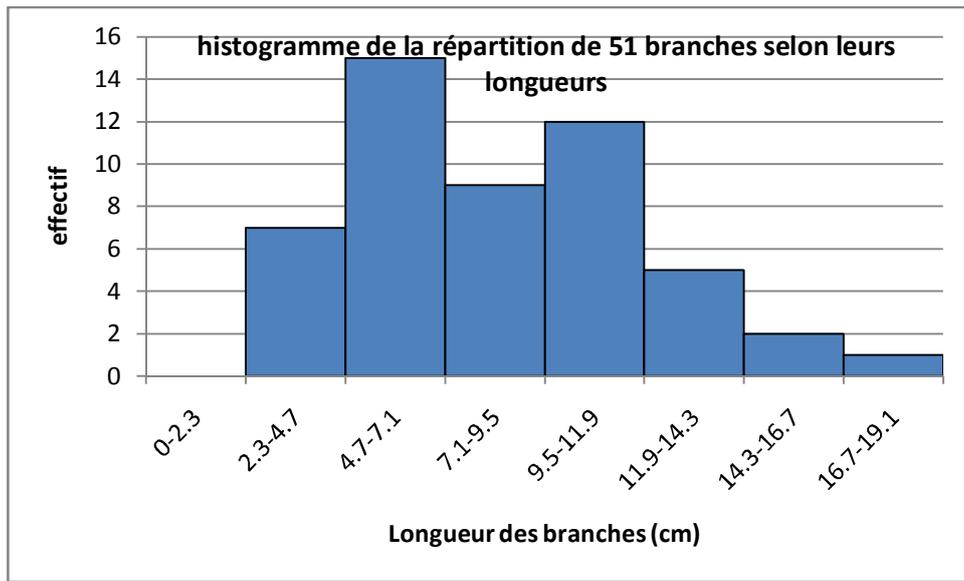
3- déterminer l'amplitude

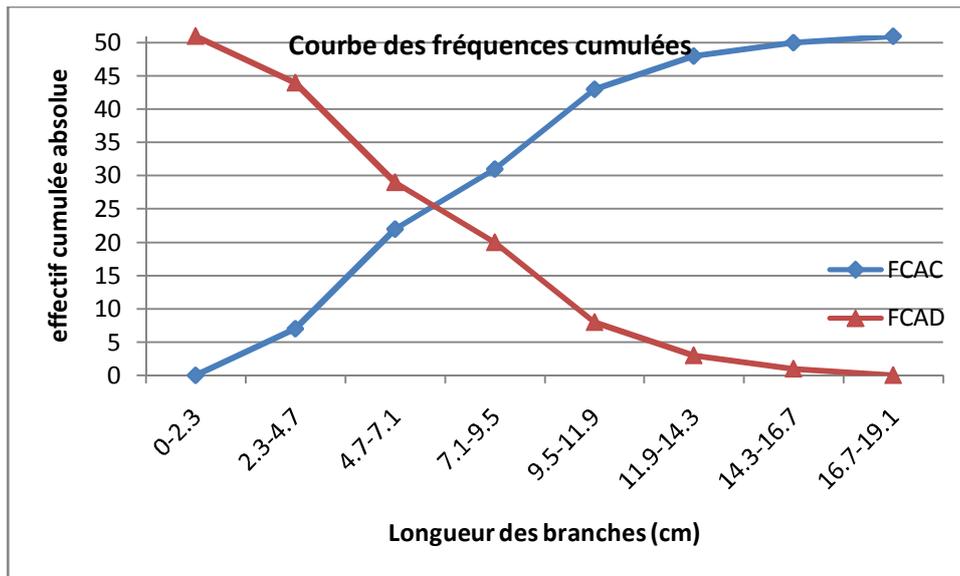
$$a = \frac{\text{étendu}}{\text{nbr classes}}, \quad a = \frac{16.2}{7} = 2.31, \text{ donc l'amplitude choisie est de } 2.4$$

4- la limite inférieure de la 1 ère classes

La limite inférieure est égale à la plus petite donnée de la série statistique qui est de 2.3

Classes (x_i)	Effectif (n_i)	Fréquence relative	Fréquence cumulée absolue croissante	Fréquence cumulée absolue décroissante
[2.3 ; 4.7 [7	0.14 (7÷51)	7	51
[4.7 ; 7.1 [15	0.29 (15÷51)	22 (7+15)	44 (51-7)
[7.1 ; 9.5 [9	0.18 (9÷51)	31 (22+9)	29 (44-15)
[9.5 ; 11.9 [12	0.23 (12÷51)	43 (31+12)	20 (29-9)
[11.9 ; 14.3 [5	0.10 (5÷51)	48 (43+5)	8 (20-12)
[14.3 ; 16.7 [2	0.04 (2÷51)	50 (48+2)	3 (8- 5)
[16.7 ; 19.1 [1	0.02 (1÷51)	51 (50+1)	1 (3-2)
TOTAL	51	1		





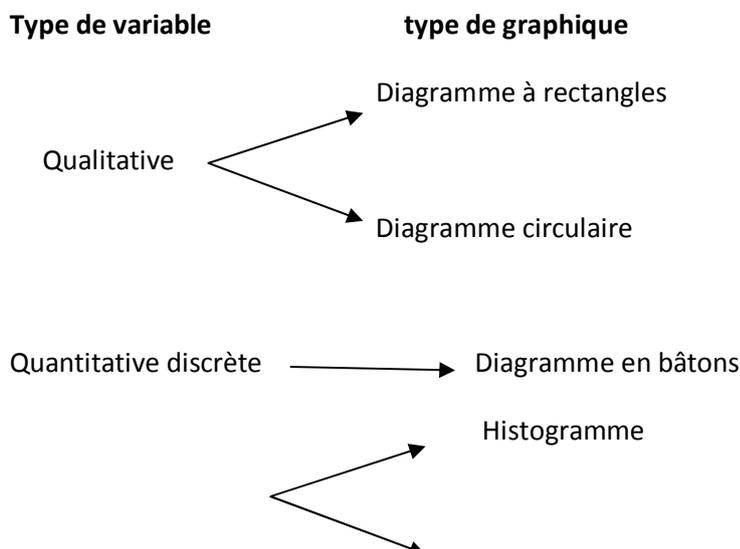
Le principe de proportionnalité nous permet de faire la déduction suivante :

$$\text{Pourcentage de données d'une classe} = \frac{\text{AIRE DU RECTANGLE DE CETTE CLASSE}}{\text{AIRE TOTALE DE L'HISTOGRAMME}}$$

2.3.2 Histogramme à classes inégales

La technique utilisée pour construire un histogramme composé de classes égales ne permet pas de respecter le principe de proportionnalité lorsque les classes sont inégales, il faut donc choisir une méthode de construction d'histogramme dans ce cas. Nous utilisons la méthode suivante :

- On détermine une amplitude de base, généralement celle de la majorité des classes de la distribution
- Si l'amplitude d'une classe est égale au double de l'amplitude de base, on considère qu'elle équivaut au regroupement de deux classes standard et, par conséquent, la hauteur du rectangle de cette classe devra être égale à la moitié de sa fréquence. On applique ce même principe si une classe correspond à plus de deux fois l'amplitude de base. Le processus est appelé rectification de fréquences.



Qualitative continue

Polygone de fréquence