

Année universitaire
2020-2021

Cours de Bio-statistiques
par
Pr. Abdallah Benaïssa

Programme:

Première partie

- 1) Statistiques descriptives à une dimension
- 2) Statistiques descriptives à deux dimensions et régression
- 3) Analyse combinatoire
- 4) notions de probabilités
- 5) Variables aléatoires
- 6) Quelques lois de probabilités.

Deuxième partie: Statistiques inférencielles

- 1- Variable aléatoire "moyenne arithmétique".
- 2- Théorie de l'estimation
- 3- Tests d'hypothèses de conformité
- 4- Tests d'hypothèses d'homogénéité
- 5- Tests de Khi deux
- 6- Analyse de la variance (ANOVA)
- 7- Tests sur les corrélations linéaires

Documentation

Quelques livres dans la
bibliothèque de la faculté:

- 1)- E.A Azoulay et D. Cohen, Cours et exercices de statistique, Réf. 519\11.
- 2)- O. Admane et al. , Statistique, cours et exercices corrigés, Réf. 519\27.

- 3)- R. Atlani et J.M. M\ouren, Probabilités et Statistiques, Réf. 519\08
- 4)- G. Bresson, Probabilités et Statistiques, Réf. 519\07.
- 5)- J.P. Georgin, Probabilités et Statistiques, Réf. 519\10

Documentation via Internet

- Pour la première partie utiliser les mots
- clefs suivants.
- Statistique descriptive, séries statistiques simples, paramètres de position, paramètres de dispersion, représentation graphique des données statistiques, séries statistiques doubles, ajustement et régression linéaire, Bio-statistiques, lois de probabilités, loi normale,....

Statistiques descriptives

Démarches statistiques

La statistique est la science dont l'objet est de recueillir, de traiter et d'analyser des données issues de l'observation de phénomènes aléatoires, c'est-à-dire dans lesquels le hasard intervient. Une étude statistique passe généralement par trois étapes.

1)-

Elle commence par l'identification du phénomène ou problème à étudier, par exemple pour étudier le problème du diabète en Algérie, on doit évaluer le taux de la glycémie, l'effet de ce fléau sur le fonctionnement du corps humain, la relation entre ce fléau et le rendement économique,...

2)-

La deuxième étape concerne la collecte des données, généralement, sur un échantillon représentatif de la population à étudier. La collecte des données se fait, soit par des questionnaires (par exemple, le nombre d'enfants de familles, le niveau d'instruction,...), soit par des manipulations (la mesure de la glycémie, la mesure de tension artérielle,...) .

3)-

La troisième étape concerne l'analyse et l'interprétation des résultats. L'analyse des données est utilisée pour décrire les phénomènes étudiés, faire des prévisions et prendre des décisions à leur sujet.

Les trois démarches qu'on vient de décrire concernent les statistiques descriptives.

Il est important de noter que les statistiques descriptives utilisées il y a quelques siècles, se sont beaucoup développées aujourd'hui en donnant naissance à ce qu'on appelle les statistiques inférentielles.

En effet, on peut répartir la théorie des statistiques en deux grandes classes de statistiques, Statistique descriptive et Statistique inférentielle, la deuxième étant un prolongement de la première

Statistique inférentielle

Le rôle de la statistique inférencielle est de fournir des outils mathématiques basés sur la théorie des probabilités, servant à généraliser les propriétés et les caractéristiques calculées sur l'échantillon à la population toute entière.

Statistique univariée

Population, unité et caractère statistiques.

Une population statistique est un ensemble d'êtres dont on s'intéresse à un de ses caractères, par exemple la population de la ville de Batna, l'ensemble des tremblements de terre en Algérie durant le 20^{ème} siècle.

L'unité statistique est un élément quelconque de la population statistique, par exemple l'unité statistique dans le cas de la population statistique des habitants de la ville de Batna est un habitant de Batna.

Le caractère statistique est la propriété des unités statistiques à étudier, par exemple la couleur des yeux, la taille pour la population des habitants de Batna. Le caractère statistique peut avoir plusieurs modalités (ou valeurs).

Le caractère étudié est quantitatif si ses modalités (ou ses valeurs possibles) sont mesurables, c'est-à-dire exprimées par des valeurs numériques, et il est qualitatif si ses modalités ne sont pas mesurables.

Un caractère quantitatif est discret si ses modalités sont en nombre fini (ou dénombrables) et il est continu si ses valeurs possibles forment un ou plusieurs intervalles.

Un caractère statistique qualitatif est ordinal si ses modalités peuvent être ordonnées (par exemple le stade d'une maladie), sinon il est nominal.

Recensement - Echantillonnage

Pour étudier un caractère d'une population statistique, on ne peut pas en général, relever la valeur (ou la modalité) de ce caractère sur chaque individu de cette population. Alors, on considère une partie de taille raisonnable de cette population. Cette partie, sélectionnée à priori, s'appelle échantillon de la population et elle doit être représentative de la population.

Mais, parfois on doit considérer tous les éléments de la population statistique, on dit dans ce cas qu'il s'agit d'un recensement, par exemple le recensement de la population de l'Algérie qui se fait tous les dix ans (nombre d'habitants, répartition selon le sexe, selon les tranches d'âges, selon le niveau d'instruction...)

Série statistique simple

Une série statistique simple est une suite de nombres représentant les mesures d'un caractère quantitatif prises sur un échantillon d'une population statistique

La variabilité et l'incertain en biologie

**Variabilité biologique et
variabilité métrologique**

La statistique constitue, en médecine, l'outil permettant de répondre à de nombreuses questions qui se posent en permanence au médecin:

- Quelle est la valeur normale d'une grandeur biologique, taille, poids, glycémie?
 - Quelle est la fiabilité d'un examen complémentaire?
 - Quel est le risque d'un traitement?
 - Le traitement A est-il plus efficace que le traitement B?
-

Toutes ces questions, proprement médicales, reflètent une propriété fondamentale des systèmes biologiques qui est leur variabilité. Cette variabilité est la somme d'une variabilité expérimentale (liée au protocole de mesure et aussi aux appareils utilisés) et d'une variabilité proprement biologique. On peut ainsi décomposer la variabilité d'une grandeur mesurée en deux grandes composantes:

variabilité totale =

variabilité biologique

+ variabilité métrologique.

La variabilité biologique peut être elle-même décomposée en deux termes : d'une part la variabilité intra-individuelle, qui fait que la même grandeur mesurée chez un sujet donné peut être soumise à des variations aléatoires ; et d'autre part la variabilité inter-individuelle qui fait que cette même grandeur varie d'un individu à l'autre.

variabilité biologique =

variabilité intra-individuelle

+variabilité inter-individuelle.

La variabilité métrologique peut être elle aussi décomposée en deux termes : d'une part les conditions expérimentales et d'autre part les erreurs induites par l'appareil de mesure utilisé.

$$\begin{aligned} & \text{variabilité métrologique} = \\ & \text{variabilité expérimentale} + \\ & \text{variabilité appareil de mesure} \end{aligned}$$

La décision dans l'incertain

Pour prendre une décision diagnostique ou thérapeutique, le médecin ne peut pas être sûr, cent pour cent d'avoir fait le bon choix, mais il y a des méthodes statistiques pour minimiser le risque de faire le mauvais choix.

Représentation des données

Il y a trois principales façons de représenter les données concernant un caractère statistique:

- 1)- représentation brute (ou en vrac),
- 2)- dans un tableau statistique et
- 3)- la représentation graphique.

Représentation brute

La représentation brute des données consiste à écrire les valeurs des observations une après une sans aucune manipulation.

Exemple.

Le recueil d'informations concernant le nombre d'enfants pour vingt familles a donné la suite de nombres suivante

3 2 1 0 2 3 4 0 2 3 1 0 3 5 6 6 5 6 5 4.

Représentation dans un tableau

A- La représentation des données groupées par valeurs de la série précédente du nombre d'enfants des vingt familles est donnée dans le tableau suivant

valeurs (x_i)	0	1	2	3	4	5	6	Total
effectifs (n_i)	3	2	3	4	2	3	3	20

On peut ajouter à ce tableau d'autres lignes, par exemple, une ligne pour les effectifs cummuls (croissants) et une ligne pour les fréquences. On obtient alors le tableau suivant:

(x_i)	0	1	2	3	4	5	6	Total
(n_i)	3	2	3	4	2	3	3	20
$(n_i cum)$	3	5	8	12	14	17	20	□
(f_i)	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{4}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{3}{20}$	1

B- On utilise aussi la représentation des données groupées en classes de valeurs

Exemple

Le tableau suivant représente les données de la série des notes de 20 élèves en classes de valeurs d'amplitude 3.

classes (C_i)	[6;9[[9;12[[12;15[[15;18[total
effectifs (n_i)	4	7	6	3	20

Représentation graphique

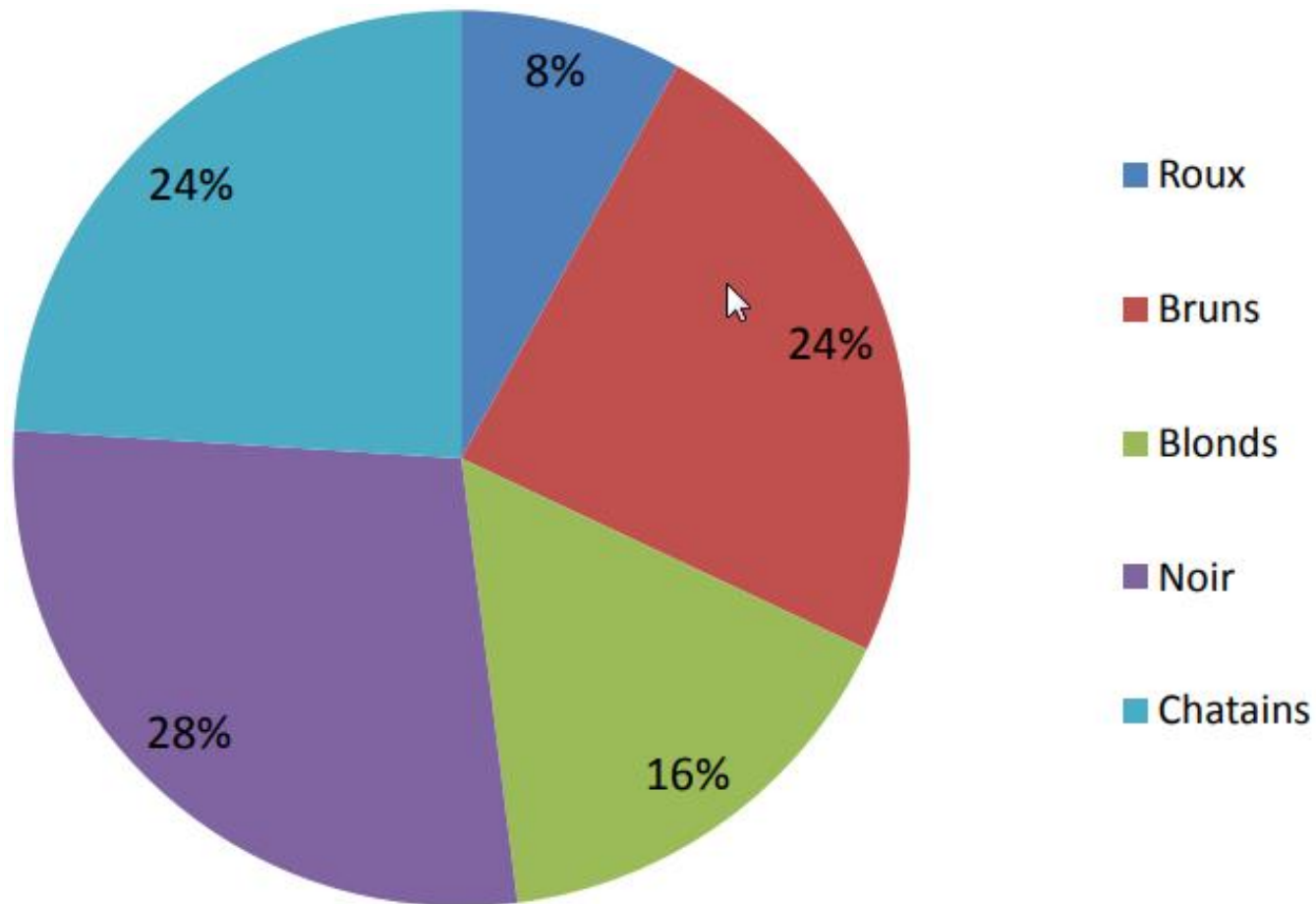
La façon de représenter une variable statistique dépend du genre de cette variable.

A- Variable qualitative.

Les deux principales façons de représenter graphiquement une variable statistique qualitative sont le diagramme en secteurs et le diagramme en bandes.

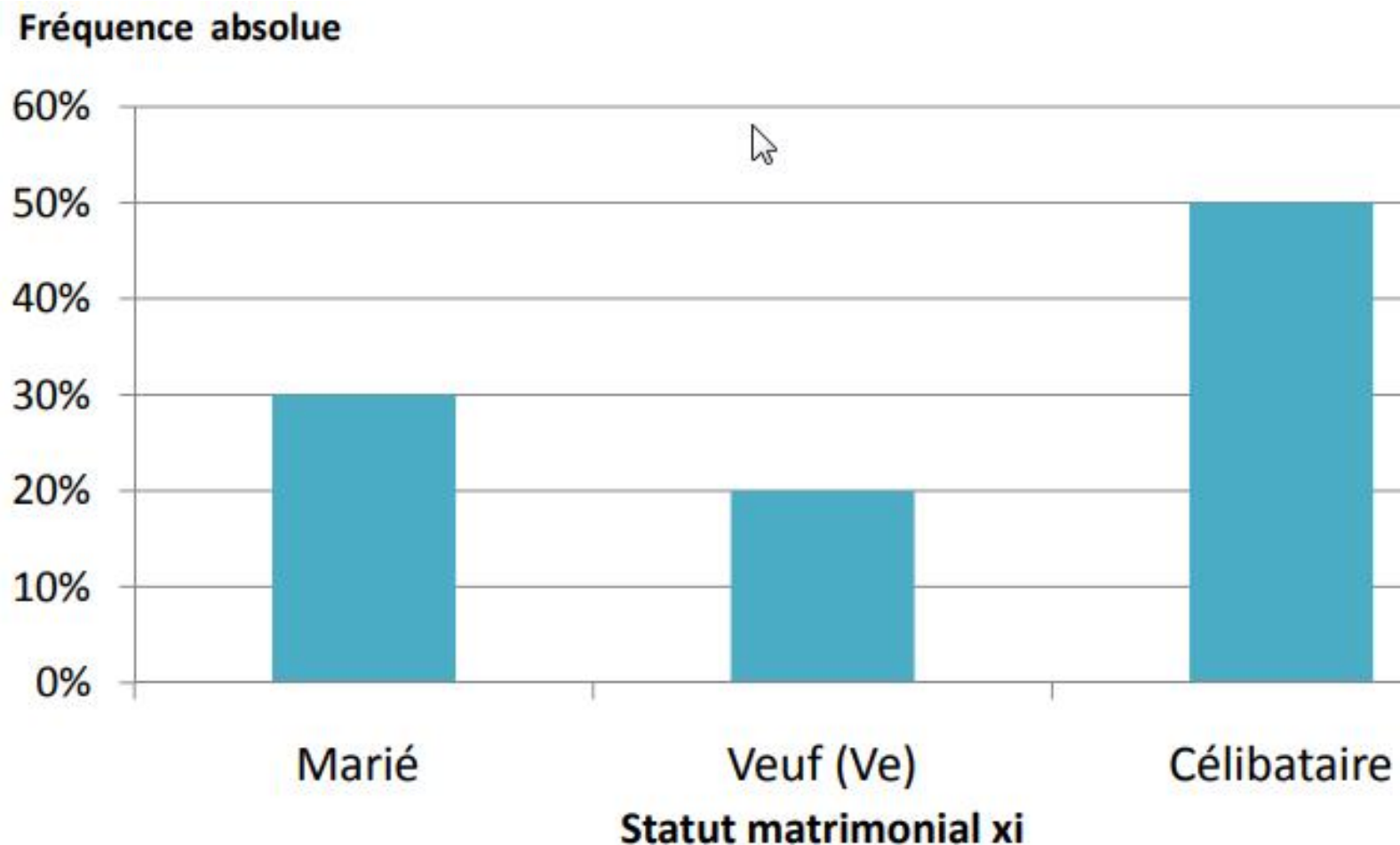
Exemple

Représentation par **un diagramme en secteurs** de la répartition de 25 sujets en fonction de la couleur des cheveux



Exemple

Représentation par un **diagramme en bandes** de la distribution d'une population selon la situation matrimoniale



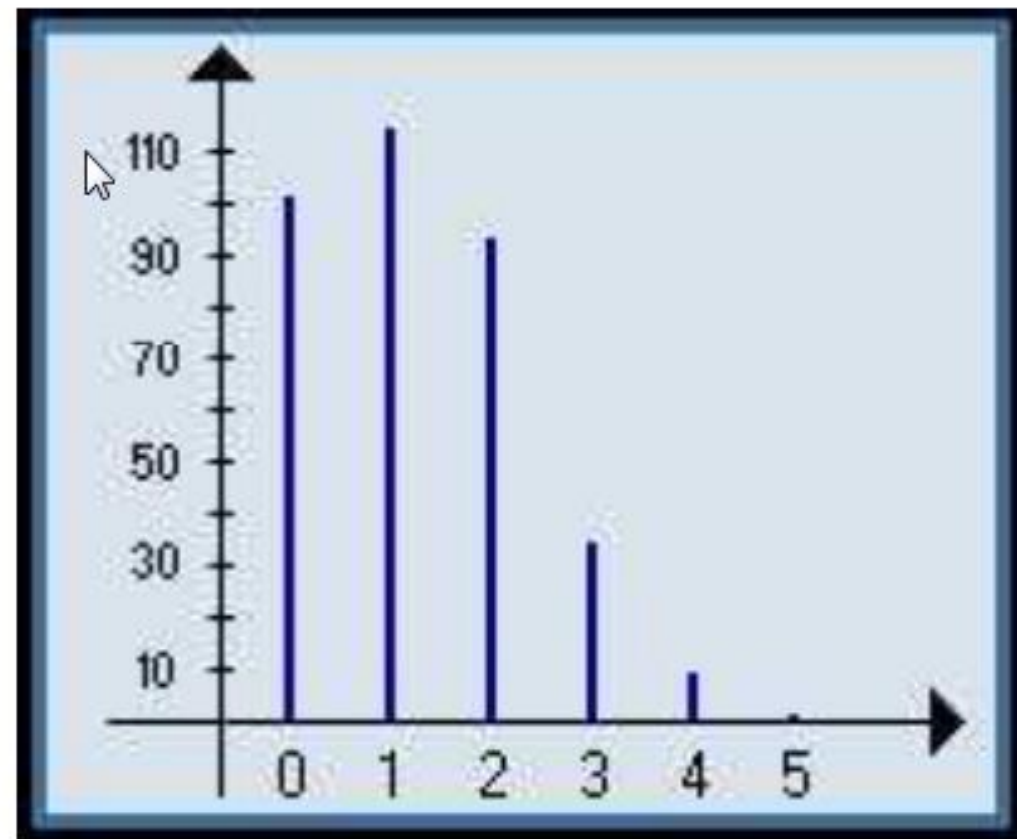
B- Variable quantitative discrète.

On représente généralement une variable statistique discrète par un diagramme en bâtons.

Exemple

Représentation par un diagramme en bâtons de la distribution du nombre d'enfants par famille

Effectifs



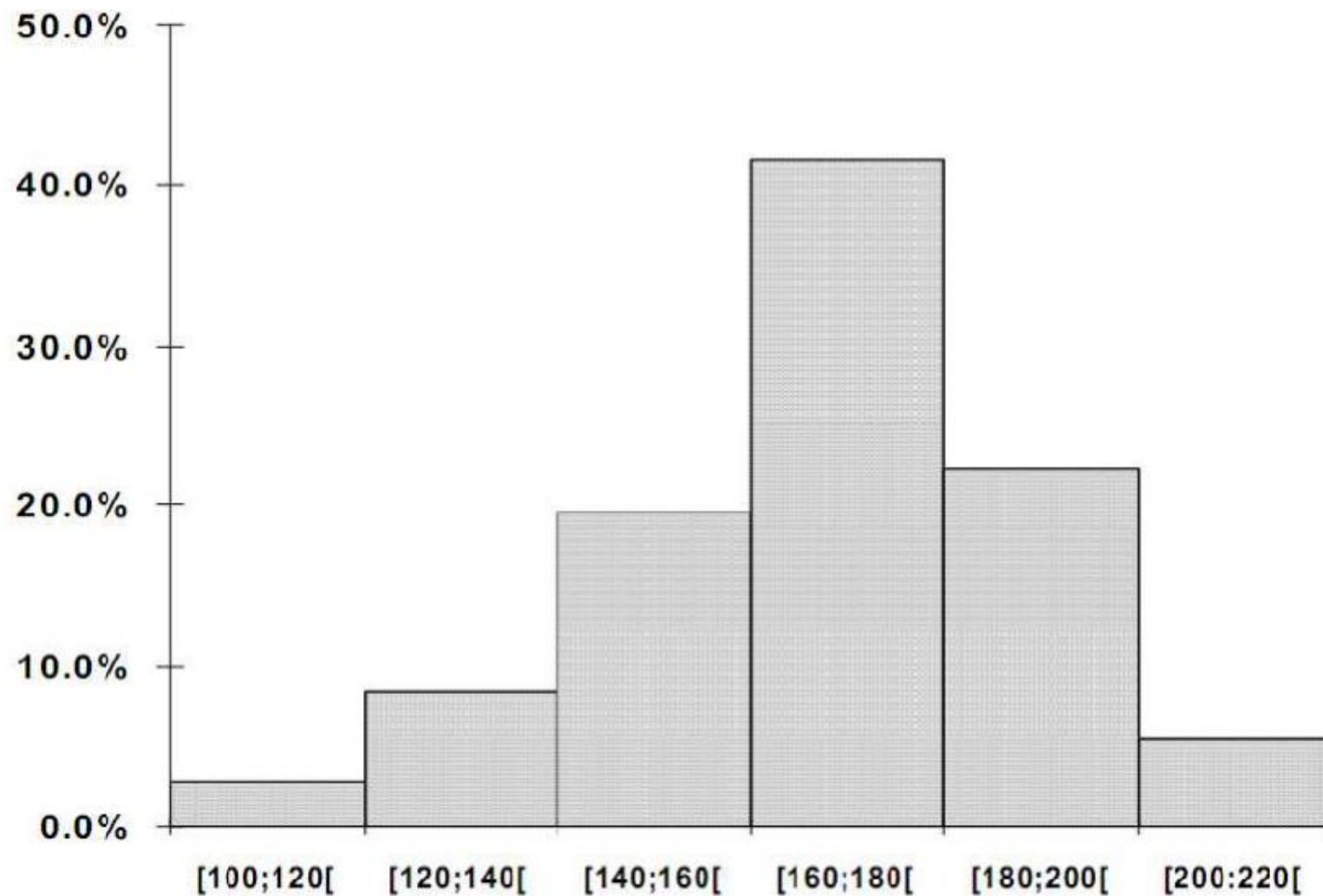
Nbre d'Enfants

C- Données groupées en classes de valeurs

La représentation graphique d'un caractère, dont les valeurs sont représentées en classes de valeurs et non en valeurs discrètes, se fait par un histogramme. Dans un histogramme on représente les classes par des intervalles sur l'axe des abscisses, et on représente l'effectif (ou la fréquence de chaque classe) par un rectangle dont la superficie est proportionnelle à l'effectif (ou à la fréquence).

Exemple

Représentation par un histogramme de la distribution fréquentielle de la taille (en cm) de 36 plantes



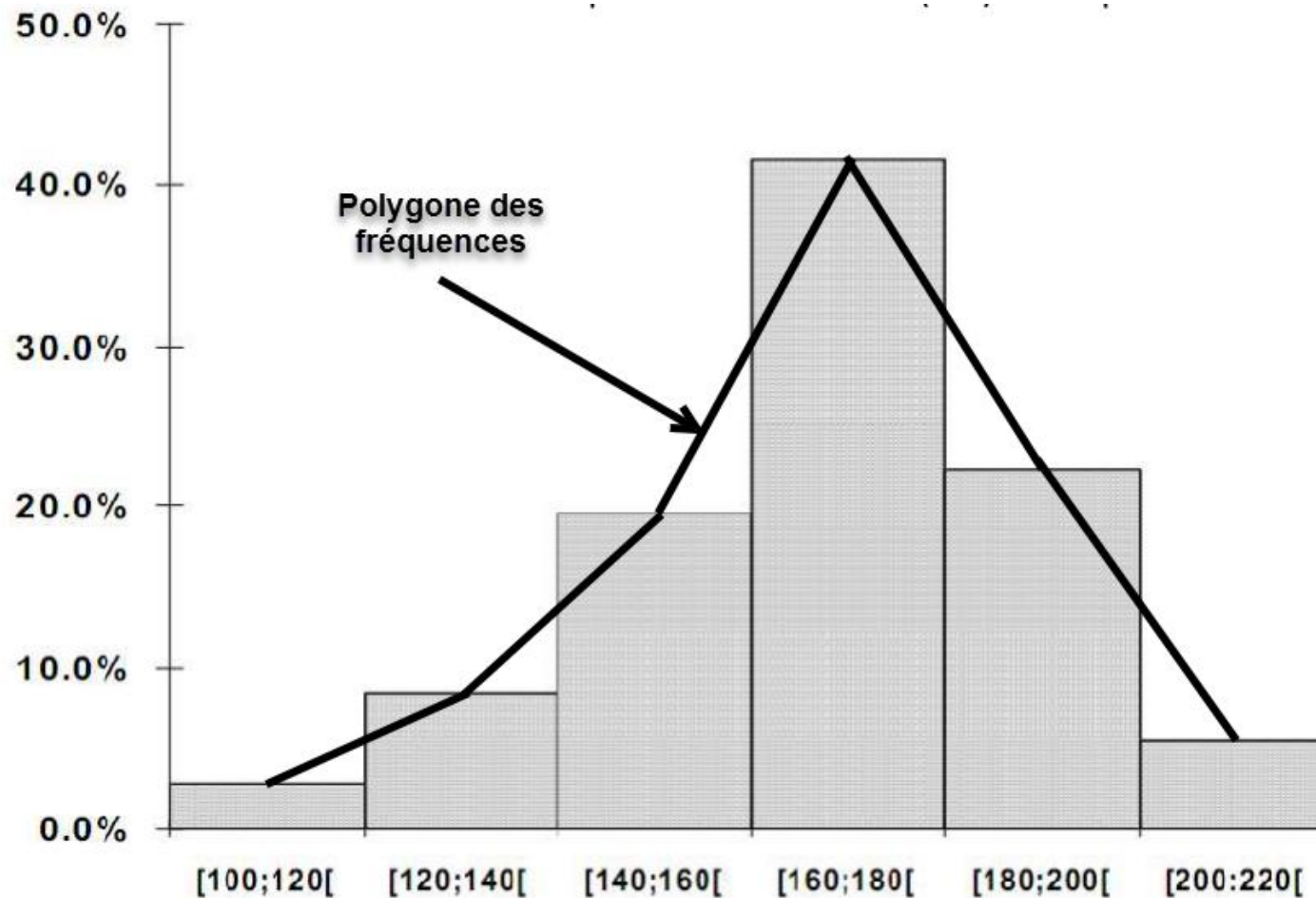
Polygone des fréquences

Le polygone des fréquences est aussi une autre représentation graphique qui peut remplacer un histogramme.

Le polygone est obtenu à partir de l'histogramme en rejoignant le point milieu du sommet de chaque rectangle au milieu du sommet du rectangle adjacent

Exemple

L'histogramme précédent engendre le polygone des fréquences suivant



Les différents types de tableaux statistiques

Il y a trois types de tableaux en statistiques.

1)- Tableaux de données.

Ces tableaux sont confectionnés pour collecter les données d'un ou de plusieurs caractères sur les sujets, d'une population statistique ou d'un échantillon d'une population statistique.

Exemple

Dans une classe de 15 élèves on a relevé sur le tableau suivant, les données concernant le sexe, le niveau de la langue française et la nationalité.

N°	sexe		français			moyenne générale			
	M	F	A	B	C	[7;10[[10;13[[13;16[[16;19[
1	x	<input type="checkbox"/>	<input type="checkbox"/>	x	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	<input type="checkbox"/>	x	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>
3	x	<input type="checkbox"/>	x	<input type="checkbox"/>	x	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>
4	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>	x	<input type="checkbox"/>
5	x	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	x	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>
6	<input type="checkbox"/>	x	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	x	<input type="checkbox"/>
7	<input type="checkbox"/>	x	x	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>
8	x	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>
9	<input type="checkbox"/>	x	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>
10	x	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	x	<input type="checkbox"/>
11	<input type="checkbox"/>	x	x	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	x
12	<input type="checkbox"/>	x	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>
13	<input type="checkbox"/>	x	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	x	<input type="checkbox"/>
14	<input type="checkbox"/>	x	x	<input type="checkbox"/>	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15	x	<input type="checkbox"/>	<input type="checkbox"/>	x	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	x

2)-Tableaux de distributions statistiques.

Ce type de tableaux est le plus connu. Il utilise généralement un tableau des données, pour exprimer la distribution statistique (c'est-à-dire la distribution des effectifs ou des fréquences) d'un ou de plusieurs caractères statistiques. Les tableaux statistiques déjà rencontrés dans ce cours sont de ce type.

Exemple.

L'utilisation du tableau des données de l'exemple précédent donne le tableau suivant de la distribution statistique des moyennes générales des quinze élèves.

M.G (x_j)	[7; 10[[10; 13[[13; 16[[16; 19[total
(n_j)	2	7	4	2	15

3)- Tableau de contingence.

Ces tableaux sont constitués par croisement de deux variables statistiques. Leur élaboration nécessite de revenir au tableau initial de données et le dénombrement des sujets présentant simultanément deux valeurs des variables considérées. La fonction d'un tableau de contingence est de permettre de tester l'indépendance ou le lien entre deux variables.

Exemple.

Pour étudier la relation entre le sexe et la réussite scolaire des quinze élèves de l'exemple précédent, l'utilisation du tableau des données dans le même exemple donne le tableau de contingence suivant.

sexe \ M.G	[7; 10[[10; 13[[13; 16[[16; 19[totaux
masculin	1	3	1	1	6
feminin	1	4	3	1	9
totaux	2	7	4	2	15