Chapitre 3 : **Statistique Descriptive**

Plan du Cours

1) Introduction (cadre général)
2) Notion de Population, d'échantillon, d'individus, de caractère.
3) Tableaux statistiques
4) Représentations graphiques
5) Indicateurs de tendance centrale
6) Indicateurs de dispersion

L'origine du mot "*Statistique*" est "*status*" qui signifie état qui, au début, cette discipline concernée exclusivement les affaires de l'État, en général par des études méthodiques des faits sociaux définissant cet État, par des procédés numériques (dénombrements, inventaires, recensements,...). La statistique est un ensemble de techniques d'interprétation mathématique appliquées à des phénomènes pour lesquels une étude exhaustive de tous les facteurs est impossible à cause de leur grand nombre ou de leur complexité. Pour certain elle est une branche des mathématiques(les anglo-saxons) pour d'autres une discipline à part entière hors des mathématiques.

La statistique descriptive

La statistique descriptive est la branche des statistiques qui regroupe les nombreuses techniques utilisées pour décrire un ensemble relativement important de données. Cette démarche a pour but :

- Résumer et synthétiser l'information contenue dans la série statistique ;
- Mettre en évidence ses propriétés ;
- Suggérer des hypothèses relatives à la population dont est issu l'échantillon.

Les Tableaux

Sont des moyens organisés pour présenter des données numériques afin de faciliter leurs analyses, leurs interprétatiosn et leurs comparaisons. Ces tableaux sont couramment utilisés dans divers domaines tels que les sciences sociales, les sciences économiques, les sciences de la santé, etc.

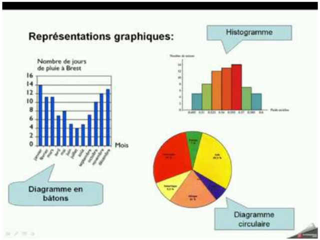
4 L'évolution de la population mondiale de 400 avant J.-C. à 1900 (par grandes régions)

	– 400	J.-C.	500	1000	1300	1400	1500	1700	1800	1900
Europe (avec la Russie)	32	43	41	43	86	65	84	125	195	422
Chine (avec la Corée)	19	70	32	56	83	70	84	150	330	415
Inde (avec le Pakistan et le Bangladesh)	30	46	33	40	100	74	95	175	190	290
Sud-Ouest asiatique	42	47	45	33	21	19	23	30	28	38
Japon	0,1	0,3	2	7	7	8	8	28	30	44
Reste de l'Asie	3	5	8	19	29	29	33	53	68	115
Afrique du Nord	10	13	12	10	9	8	8	9	9	23
Reste de l'Afrique	7	12	20	30	60	60	78	97	92	95
Amérique du Nord	1	2	2	2	3	3	3	2	5	90
Amérique centrale et du Sud	7	10	13	16	29	36	39	10	19	75
Océanie	1	1	1	1	2	2	3	3	2	6
TOTAL MONDIAL	152	250	205	257	429	374	458	682	968	1613

Les graphiques

Les graphiques statistiques sont des outils visuels utilisés pour représenter des données statistiques de manière à en faciliter la compréhension, l'analyse et la communication.

- Histogrammes : Utilisés pour représenter la distribution des données continues.
- Diagrammes en barres : similaires aux histogrammes, mais utilisés pour représenter des données catégoriques ou discrètes.
- Diagrammes circulaires (camemberts) : utilisés pour représenter la composition d'un ensemble sous forme de secteurs d'un cercle.



Les caractéristiques numériques

Sont des grandeurs mathématiques qui servent à résumer dans un sens bien déterminé, toute l'information représentée par l'ensemble de toutes les mesures disponibles.

Les caractéristiques les plus utilisées sont :

- La moyenne arithmétiques
- La médiane
- le mode
- la variance
- l'écartype
- le maximum ou le minimum d'une série de mesures.

Les outils varient en fonction de :

De la nature de la série

- Si les données ne sont relatives qu'à une seule variable, on parle de statistique descriptive univariée.
  Ex : taille de 100 personnes : 1.65cm, 1.72cm,. . .,1.49cm.
- Dans le cas où l'on s'intéresse à deux variables simultanément, on met en oeuvre la statistique descriptive bivariée.
  Ex : tailles et le poids de 100 personnes : (1.65cm, 55kg), (1.72cm, 80kg),. . ., (1.49cm, 50kg).
- Si l'ensemble des données provient de l'observation de plusieurs variables, on doit faire appel aux méthodes de la statistique descriptive multivariée.
  Ex : taille, poids, âge, couleurs des yeux, catégorie socio-professionnelle, etc.

De la nature des variables (quantitatives ou qualitatives)

.

- Données numériques : taille, poids, age, salaires, ...
- Données non numériques (qualitatives) : nationalité, catégorie socio-professionnelle, couleur des yeux, civilité, etc.

Élément ou unité statistique

qui peut être

- Un individu (êtres vivants) : humain, animal, végétal ... ;
- Un sujet : modules enseignés en biologie, les nationalités, les métiers ou professions ;
- Un objet : Table, chaise, verrerie de laboratoire ;
- Une association (dans les études écologiques en général) : une parcelle d'herbe, une association d'arbustes

Population

C'est un ensemble d'éléments possédant au moins une caractéristique commune et exclusive permettant de l'identifiée et la distinguée sans ambiguïté de toutes les autres.

Exemples

- *Une population algérienne (grande mais finie)*
- *Les étoiles dans l'univers (infinie) ;*
- *Une population de plantes médicinales (très grande mais finie ;*
- *Une population de poissons d'eau douce (finie mais très grande.*
- *Une chaîne de fabrication de voiture (petite mais non limitée)*

Échantillon

Il n'est généralement pas possible de collecter des données sur tous les éléments de la population. En outre, si cette opération est possible il est rarement utile de la faire, car l'analyse d'un groupe restreint d'éléments extraits de la population fournit généralement des résultats de précision satisfaisante. Cette petite partie de la population qu'on va examiner s'appelle échantillon

Exemples

- *Etude de 20 étudiants pris à partir d'une population de 57.*
- *Etude de 5 régions prises à partir d'une population de 25.*
- *Etude de 5 modules pris à partir d'une population de 13.*
- *Etude de 200 patients pris à partir d'une population de 660.*

C 'est des méchanismes scientifiques permettant de prélever une partie représentative de la population (échantillon), c'est-à-dire qu'elle doit refléter fidèlement sa composition et sa structure. Il existe plusieurs méthodes d'échantillonnage qui varient en fonction de la nature de l'étude envisagée, on peut citer l'échantillonnage stratifié, par degré, systématique et le plus utilisé est l'échantillonnage aléatoire simple qui est basé sur le principe que tous les éléments de la population ont une probabilité égale (non nulle) de faire partie de l'échantillon.

Caractère (variable ou mesure)

C'est une propriété possédée par les unités statistiques permettant de les décrire et de les distinguer les unes des autres. Toute unité statistique peut être étudiée selon un ou plusieurs caractères :

Exemples

- *Couleur des yeux ;*
- *Poids des souris ;*
- *Superficie d'une pièce ;*
- *La température de l'air.*
- *Le nombre de particule dans un volume d'air.*
- *....*

Remarque

Le caractère est généralement noté par des lettres capitales : $X, Y, Z, \ldots$

Modalités

Ce sont les diverses situations (cas, état, valeur) susceptibles d'être prises par le caractère. Un caractère peut posséder une ou plusieurs modalités.

Exemples

- *Couleur des yeux : vert, bleu, noir.*
- *Poids des souris (en grammes) : 15, 18, 20, 39.*
- *Superficie d'une pièce (en m^2) : 3, 5, 6.*
- *La température de l'air (en $^\circ$) : 8°, 16°, 27°, 30°, 38°.*

Remarque

Si un caractère X possède k modalités, on note alors ces modalités par :

$$\boxed{x_1,\ x_2, \ldots, x_k.}$$

Effectifs

- L'effectif d'une modalité x_i d'un caractère X est le nombre d'individus de la population ayant cette valeur, elle est notée n_i.
- Si X possède k modalités, alors l'effectif total n est la somme de tous les effectifs des k modalités. On écrit alors :

$$\boxed{n = n_1 + n_2 + n_3 + \ldots + n_i + \ldots + n_k}$$

- En rangeant les valeurs du caractère dans l'ordre croissant, on peut calculer l'effectif cumulé croissant en faisant la somme des effectifs de cette valeur et de tous ceux qui la précèdent.

$$\boxed{N_i = n_1 + n_2 + \ldots + n_i}$$

$$n_1 = N_1 \leq N_2 \leq N_3 \leq \ldots \leq N_i \leq \ldots N_k = n$$

Exemple

Dans une promotion de 20 étudiants en L1 informatique, voici les notes obtenues au dernier examen de statistique : 10, 14, 12, 15, 7, 8, 10, 11, 12, 18, 2, 4, 12, 13, 14, 15, 19, 11, 9, 0. On va calculer les effectifs et les effectifs cumulés.
Pour les effectifs on cherche le nombre d'étudiants ayant la note x_i.
Pour les effectifs cumulés on fait la somme des effectifs de la note + la somme des effectifs de toutes les notes qui la précédent.
On obtient alors le tableau suivant :

x_i	0	2	4	7	8	9	10	11	12	13	14	15	18	19
n_i	1	1	1	1	1	1	2	2	3	1	2	2	1	1
N_i	1	2	3	4	5	6	8	10	13	14	16	18	19	20

Table – *Calcul de l'effectif cumulé croisssant*

Remarque

Afin de s'assurer que le calcul des effectifs cumulés est bien correcte, la dernière valeur de l'effectif cumulée doit correspondre aux nombre total d'individus, dans cet exemple égal à 20.

Fréquence d'une modalité ou d'une classe

La fréquence d'une valeur x_i du caractère notée f_i est le quotient de l'effectif de la valeur par l'effectif total c 'est-à-dire :

$$\boxed{f_i = \frac{n_i}{N}.}$$

En rangeant les valeurs du caractère dans l'ordre croissant, on peut calculer les fréquences cumulées croissantes (notée F_i) en faisant la somme des fréquences de cette valeur et de tous ceux qui la précèdent,

$$\boxed{F_i = f_1 + f_2 \ldots + f_i.}$$

Remarque

$$F_k = f_1 + f_2 + \ldots f_k = \frac{n_1 + n_2 + n_3 + \ldots + n_i + \ldots + n_k}{n} = \frac{n}{n} = 1$$

Exemple

Dans l'exemple (2.1) on calcul les effectifs et les effectifs cumulés, les résultats sont résumés dans le tableau suivant.

x_i	0	2	4	7	8	9	10	11	12	13	14	15	18	19
f_i	.05	.05	.05	.05	.05	.05	.1	.1	.15	.05	.1	.1	.05	.05
F_i	.05	.1	.15	.2	.25	.3	.4	.5	.65	.7	.8	.9	.95	1

$$\sum_{i=1}^{14} f_i = 1$$

Nature des caractères

On distingue deux grandes familles de caractères : les caractères qualitatifs et les caractères quantitatifs

Caractère qualitatif

Un caractère est dit qualitatif lorsque ses modalités ne sont pas mesurables, le nombre de valeurs que peut prendre la variable dans ce cas est limité.

On distingue deux types d'échelles pour cette variable : nominale et ordinale.

a) Echelle nominale

Chaque modalité est exprimée par un nom ou un code.

Exemples

Cas des noms

- *Nationalité : Algérienne, Tunisienne ;*
- *Les différentes séquences nucléotidiques (ADN ou ARN) ;*
- *Les hormones : oetradiol, progestérone*
- *État matrimoniale : marié, célibataire, veuf, divorcé ;*
- *Sexe : féminin, masculin ;*
- *Profession : enseignant, médecin ;*

Exemples

Cas des codes

- *État matrimoniale : marié (1), célibataire (2), veuf (3), divorcé (4) ;*
- *Sexe : féminin (1), masculin (2) ;*
- *Profession : enseignant (1), médecin (2) ;*
- *Nationalité : Algérienne (1), Tunisienne (2)*

b) Echelle ordinale

Chaque modalité est explicitement significative du rang pris par chaque individu pour le caractère considéré.

Exemples

- *Degré d'intelligence : pas intelligent (0), peu intelligent (1), moyennement intelligent(2), très intelligent (3) ;*
- *Forme des fruits : petite (1), moyenne (2), grosse (3) ;*
- *Abondance/Dominance : peu abondant (1), abondant (2), très abondant (3).*

Caractère quantitatif

Un caractère est dit quantitatif si ses modalités s'expriment par des nombres dont les opérations de types somme et produit sont possibles sur les valeurs des modalités. Le nombre de valeurs que peut prendre la variable est illimité. On distingue deux catégories de caractèred quantitatifs :

Les caractères quantitatifs discrets

Sont des caractères dont les modalités sont des nombres isolés, pas nécessairement entiers.

Exemples

- *Nombre de pièces d'un immeuble ;*
- *Nombre d'enfants d'une famille ;*
- *Nombre de personnes atteintes par une maladie.*

Les caractères quantitatifs continus

sont des caractères dont les modalités sont définies sur un intervalle (continu) de valeur donné appelé domaine de variation et défini par les valeurs minimales et maximales. Ses valeurs sont regroupés dans des classes (petit intervalles) qu'on note $e_i =]\min(e_i), \max(e_i)]$

Exemples

la taille, le poids, l'âge, la durée de vie de composantes electroniques,...

Nombre de classes :

Dans le cas continu, les résultats sont regroupés en classes à cause de leur grande masse. Discrétiser une variable quantitative c'est, mathématiquement, transformer un vecteur de nombres réels en un vecteur de nombres entiers nommés " indices de classe ". En statistiques, discrétiser c'est à la fois réaliser cette transformation mathématique, nommer et justifier les classes. Un bon découpage correspond à des classes homogènes et séparées, ce qui correspond respectivement aux notions statistiques de faible variance intra-classe et de forte variance interclasse. Mais d'autres critères sont possibles, comme l'équirépartition, le respect d'un nombre minimal de données par classe, etc. Voici ci-après (tableau suivant) le nombre de classe en fonction du nombre d'élément.

Nombre d'éléments	Nombre de classes
10	2 à 3
20	3 à 4
50	4 à 6
100	5 à 8

Table – Nombre de classes en fonction du nombre d'éléments d'une série

Règle de STURGE

Soit un échantillon de N valeurs observées. Le mathématicien Herbert Sturges (18821958) a proposé une valeur approximative pour le nombre de classe k en fonction de la taille N de l'échantillon :

$$k = 1 + \log_2 N = 1 + \frac{\ln N}{\ln 10}$$

où $\log_2$ est le logarithme en base 2. Le résultat ne sera pas, en général, entier. Il donne une appréciation de ce qui ferait un bon découpage.

Amplitude d'une classes

L'amplitude d'une classe e_i est la longueur de l'intervalle définissant cette classe (noté a_i), c 'est-à-dire :

$$a_i = |e_i| = \max(e_i) - \min(e_i)$$

Centre d'une classes

Le centre d'une classe e_i est égale à :

$$c_i = c(e_i) = \frac{\max(e_i) + \min(e_i)}{2}$$

Étendue d'une série statistique

Dans le cas continu, on s'interesse en particulier à l'étendue des valeurs prise par la variable X qui vaut :

$$E(X) = \max\{x_1, x_2, \ldots, x_n\} - \min\{x_1, x_2, \ldots, x_n\}$$

Un tableau statistique constitu un résumé ou une synthèse numérique des résultats d'une distribution statistique, on distingue trois formes de tableaux statistiques qui sont fonction de l'objectif envisagé et de la nature du caractère étudié.

Tableau brut

Après la collecte des données, celles-ci apparaissent de façon brute. Sous cette forme, elles sont peu informatives. Il nous faut donc des moyens pour en extraire un maximum d'informations.

Element i	modalité x_i
1	x_1
2	x_2
.	.
.	.
.	.

Table – tableau brut

Masses limites de rupture en charge pour 96 fils de même diamètre (en grammes)

712	772	702	847	910	895	946	764	926	922	893	798
792	865	909	792	821	890	882	706	853	794	790	795
931	925	711	701	907	903	724	868	935	895	851	917
786	892	903	789	784	888	941	758	759	805	704	853
850	922	842	892	862	763	768	925	886	888	933	912
796	887	915	914	892	825	784	801	791	712	710	792
869	914	887	713	912	942	925	935	935	903	899	935
844	890	931	881	864	753	859	857	791	909	758	789

Table – Données à l'état brut peu informatives

Tableau de dénombrement

Il est de la forme suivante :

x_i	n_i	N_i	f_i	F_i
x_1	n_1	$N_1 = n1$	f_1	$F1 = f_1$
x_2	n_2	N_2	f_2	F_2
x_3	n_3	N_3	f_3	F_3
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
x_k	n_k	$N_k = n$	f_k	$F_k = 1$
Total	N		1	

Table – Tableau de dénombrement (cas qualitatif et discret)

Tableau de distribution des fréquences

il est de la forme

Classes	$c_i = \frac{a_{i-1}+a_i}{2}$	n_i	N_i	f_i	F_i
$e_1 = [a_0, \ a_1[$	$c_1 = \frac{a_0+a_1}{2}$	n_1	$N_1 = n_1$	f_1	$F_1 = f_1$
$e_2 = [a_1, \ a_2[$	$c_2 = \frac{a_1+a_2}{2}$	n_2	N_2	f_2	F_2
$e_3 = [a_2, \ a_3[$	$c_3 = \frac{a_2+a_3}{2}$	n_3	N_3	f_3	F_3
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
$e_k = [a_{k-1}, \ a_k[$	$c_k = \frac{a_{k-1}+a_k}{2}$	n_k	$N_k = n$	f_k	$F_k = 1$
Total	$-$	N		1	

Table – Tableau de distribution pour un caractère continu

Exemple

(Tableau statistique pour un caractère qualitatif) :
Les groupes sanguins de 100 étudiants sont resumés dans le tableau suivant :

Groupe sanguin de x_i	n_i	N_i	f_i	F_i
A	40	40	0.40	0.40
B	43	83	0.43	0.83
AB	12	95	0.12	0.95
O	05	100	0.05	1.00
Total	100		1	

Table – Répartition des étudiants en fonction de leurs groupes sanguins

Exemple

(Cas d'un caractère quantitatif discret) Un laboratoire dispose de 20 lots d'animaux, on s'interesse alors au nombre de souris dans chacun des lots. Les résultats obtenus sont résumés dans le tableau suivant :

x_i	0	1	5	10	12	15	18	20	Total
n_i	4	1	2	5	4	1	1	2	20
N_i	4	5	7	12	16	17	18	20	
f_i	0.20	0.05	0.10	0.25	0.20	0.05	0.05	0.1	1
F_i	0.20	0.25	0.35	0.60	0.80	0.85	0.90	1	

Table – répartition des souris sur les 20 lots.

Exemple

(cas continu)
On s'intéresse à la taille (cm) de 20 étudiants, les résultats obtenus sont :

140, 144, 150, 156, 142, 146, 152, 157, 143, 147, 153, 158, 143, 148, 154, 159, 144, 150, 155, 163.

Dans ce cas, on doit regrouper cette série en classes. Par la règle de Sturges le nombre de classe est

$$k = 1 + \log_2 N = 1 + \frac{\ln 20}{\ln 2} = 5.32 \simeq 5.$$

On calcul ensuite l'amplitude de chaque classe, qui est égal

$$a_i = \frac{E(X)}{k} = \frac{163 - 140}{5} = 4.6 \simeq 5.$$

On obtient alors le tableau suivant

Classes	c_i	n_i	f_i	N_i	F_i
$[140 - 145[$	142.5	6	0.30	6	0.30
$[145 - 150[$	147.5	3	0.15	9	0.45
$[150 - 155[$	152.5	5	0.25	14	0.70
$[155 - 160[$	157.5	5	0.25	19	0.95
$[160 - 165[$	162.5	1	0.05	20	1
Total	-	20	1		

Diagramme en bâton

Le diagramme en bâton (ou tuyaux d'orgue) est une représentation graphique de la distribution des fréquences d'une variable qualitative. Les "bâtons" sont bien séparés pour indiquer les différentes catégories (modalités). La hauteur d'un bâton est proportionnelle à la fréquence de la catégorie correspondante.

Dans le cas des groupes sanguins des 100 étudiants, on a obtenu le diagrame suivant (le groupe sanguin B est le plus répondu dans la population des étudiants)

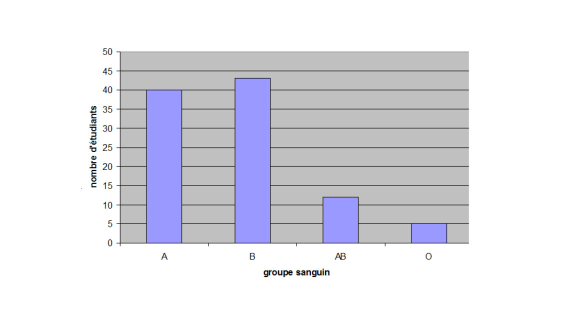


Figure – diagramme en bâton des groupes sanguins de 100 étudiants

Le camembert (diagramme circulaire)

Dans le diagramme circulaire, chaque secteur a une surface proportionnelle à la fréquence de chaque modalité, pour calculer cette surface, on utilise la règle de trois. Par exemple la surface occupée par la modalité B se calcul comme suit :

360° correspond à une fréquence de 1
x correspond à la fréquence de la modalité B, c 'est-à-dire 0.43.
On déduit que $x = \frac{0.43 \times 360}{1} = 154.8°$.

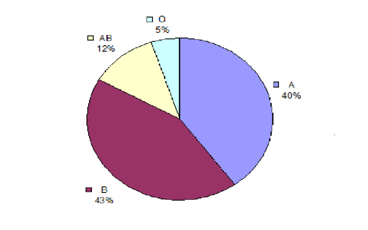


Figure – Diagramme en circulaire des groupes sanguins de 100 étudiants

Cas d'un caractère quantitatif

Il existe deux types de représentation graphique d'une distribution statistique à caractère quantitatif :

- Le diagramme différentiel correspond à une représentation des effectifs ou des fréquences.
- Le diagramme intégral correspond à une représentation des effectifs cumulés, ou des fréquences cumulées.

Le diagramme différentiel en bâtons (cas discret)

Il est réalisé en fonction des effectifs ou des fréquences, à la différence du cas qualitatif, les abscisses sont des valeurs de la variable.

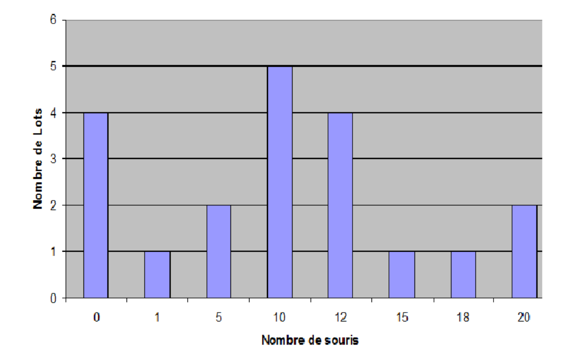
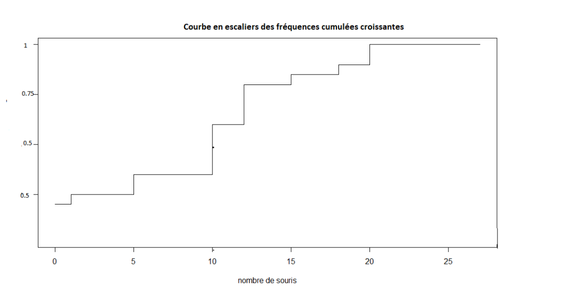


Figure – Répartition des nombre de souris sur les 20 lots

Diagramme (courbe) intégral (cas discret)

C'est une courbe en escalier réalisée en fonction des effectifs cumulés ou des fréquences cumulées. Dans cette représentation les effectifs ou les fréquences des diverses valeurs de la variable statistique correspondent aux hauteurs des marches de la courbe. Cette courbe est définie comme suit :

$$F(x) = \begin{cases} N_i, & \text{si} \quad x_i \leq x < x_{i+1} \\ n, & \text{si} \quad x \geq x_k \end{cases} \quad \text{ou} \quad F(x) = \begin{cases} F_i, & \text{si} \quad x_i \leq x < x_{i+1} \\ 1, & \text{si} \quad x \geq x_k \end{cases}$$



Histogramme et polygone des effectifs ou des fréquences (cas continu)

L'histogramme est une représentation graphique (en tuyaux d'orgue) de la distribution des effectifs ou des fréquences d'une variable quantitative. Souvent, les "tuyaux" sont accolés pour montrer la continuité de la variable. La hauteur du tuyau est proportionnelle à l'effectif ou la fréquence de la classe correspondante.

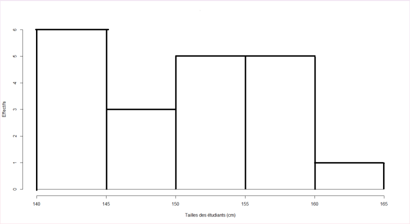


Figure – La distribution de la taille des étudiants, histogramme des effectifs

Important

Histogramme à longueur non égale

Dans le cas où les classes ont des amplitudes différentes, une modification des effectifs (effectifs corrigés) s'impose avant de tracer l'histogramme. On ulilise pour cela les transformation suivante :
Si n_i est a_i sont respectivement l'effectif et l'amplitude de la classe e_i, l'effectif corrigé est défini par

$$\tilde{n}_i = \frac{n_i}{a_i}$$

Exemple

Classes	a_i	n_i	$\tilde{n}_i$
$[100 - 145[$	45	300	6.67
$[145 - 150[$	5	120	24.00
$[150 - 160[$	10	152	15.2
$[160 - 200[$	40	141	3.52
$[200 - 220[$	20	164	8.2

Le polygone des effectifs ou des fréquences

est une autre représentation graphique (en ligne brisée) de la distribution des effectifs ou des fréquences d'une variable quantitative. Pour tracer le polygone, on joint les points milieu du sommet des rectangles adjacents par un segment de droite. Le polygone est fermé aux deux bouts en le prolongeant sur l'axe horizontal

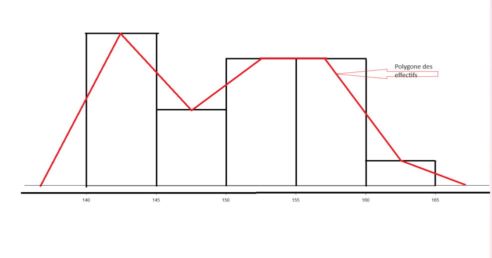


Diagramme (courbe intégral)

La courbe intégrale en fonction des effectifs cumulés ou des fréquences cumulées appelée parfois ogive. Une telle représentation est utile par exemple pour un calcul graphique de la valeur médiane, elle est défnie comme suit :

$$f(x) = \begin{cases} \frac{f_{i+1}}{a_i} x + (x_{i+1} - \frac{f_{i+1}}{a_i} x_i), & \text{si} \quad x \in [x_i,\ x_{i+1}[\\ 1, & \text{si} \quad x \geq 165 \end{cases}$$

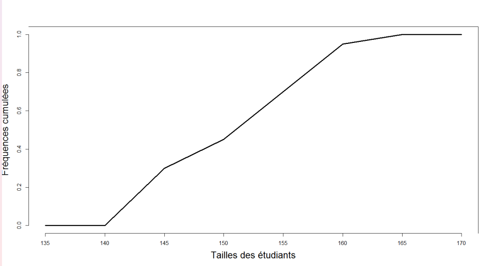


Figure – Polygones des fréquences cumulées.

Dans une description statistique des données ils est intéressant d'avoir la possibilité de résumé un partie de l'information par des paramètres numériques appelés valeurs atypique. Ils existent deux grandes familles de ces valeurs : paramètres de tendance centrale et et paramètres de dispersion.

Le mode

Le mode est la valeur la plus fréquente d'une distribution. Il se calcule toujours à partir d'un dénombrement des modalités du caractère. Comme pour le tableau de dénombrement, il faut distinguer le cas des caractères discrets et des caractères continus.

Caractère discret

pour un caractère quantitatif discret le mode noté *Mod* est la modalité qui a la fréquence la plus élevée (ou l'effectif le plus élevé).

Exemple

- *10, 11, 12, 10, 10, 10, 9, 14 ? $Mod = 10$ (4 fois), la distribution unimodale.*
- *10, 11, 12, 10, 10, 12, 12, 9, 14 ? Deux modes : $Mod_1 = 10$ (3 fois) et $Mod_2 = 12$ (3 fois), la distribution est bimodale.*

Caractère quantitatif continu :

Les modalités étant en nombre infini, il est peu probable que deux éléments aient la même valeur. Dans ce cas, le mode ne peut pas être défini directement, il faut au préalable établir une partition en classes. Le mode est avant tous situer dans une classe appelée "**classe modale**", la deuxième étape consiste à déterminer une valeur approchée du mode en utilisant la formule suivante : Si e_i est la classe modale (la classe contenant le plus grands effectifs ou la plus grande fréquence, alors :

$$Mod = \min(e_i) + a_i \frac{\Delta_1}{\Delta_1 + \Delta_2} \tag{1}$$

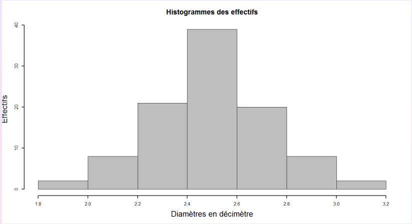
- a_i l'amplitude de la classe modale.
- Δ_1 c'est l'excès de la classe modale par rapport à la classe précédente.
- Δ_2 c'est l'excès de la classe modale par rapport à la classe suivante.

Exemple

Chez un fabricant de tubes de plastiques, on a prélevé un échantillon de 100 tubes dont on a mesuré le diamètre en centimètre. Les résultats sont regroupés dans le tableaux suivant.

1.94	2.20	2.33	2.39	2.45	2.50	2.54	2.61	2.66	2.85
1.96	2.21	2.33	2.40	2.46	2.51	2.54	2.62	2.68	2.87
2.07	2.26	2.34	2.40	2.47	2.52	2.55	2.62	2.68	2.90
2.09	2.26	2.34	2.40	2.47	2.52	2.55	2.62	2.68	2.91
2.09	2.28	2.35	2.40	2.48	2.52	2.56	2.62	2.71	2.94
2.12	2.29	2.36	2.41	2.49	2.52	2.56	2.63	2.73	2.95
2.13	2.30	2.37	2.42	2.49	2.53	2.57	2.63	2.75	2.99
2.14	2.31	2.38	2.42	2.49	2.53	2.57	2.65	2.76	2.99
2.19	2.31	2.38	2.42	2.49	2.53	2.59	2.66	2.77	3.09
2.19	2.31	2.38	2.42	2.50	2.54	2.59	2.66	2.78	3.12

La règle de Sturge permet d'obtenir 7 classes d'amplitude $a_i = 0.2$ décimètre, ce qui permet de construire l'histogramme suivant :



La classe modale est donc la quatrième classe $e_4 = [2.4 - 2.6[$, le mode est égal alors :

$$\boxed{Mod = 2.4 + 0.2\frac{39-21}{(39-21)+(39-20)} = 2.50}$$