BIOLOGIE MAROC
www.biologie-maroc.com

Statistiques

SCIENCES DE LA VIE

Chap. IV : Introduction à l’analyse statistique des données expérimentales

Pr. Mostafa El Yassa

Introduction

- Lorsque on effectue plusieurs mesures d’une grandeur physique on constate qu’elles ne sont presque jamais constantes : la valeur d’un paramètre physique varie toujours
- Une grandeur physique n’est pas caractérisée par une valeur, mais plutôt par la probabilité de trouver dans une expérience telle ou telle valeur.
- Pour cela on introduit une fonction la *distribution* d'une valeur physique, qui montre quelles sont les valeurs les plus fréquentes ou les plus rares

Introduction

À une grandeur physique on doit associé :

1. La valeur la plus probable : la valeur estimée
2. L'intervalle de confiance : l'incertitude
3. La probabilité que la vraie valeur se trouve dans l'intervalle de confiance : la confiance $(1-\propto)$ ou bien le risque $\propto$

Exemple : mesure de la capacité thermique de l'eau :

$$c = \underbrace{4320}_{\text{Valeur estimée}} \pm \underbrace{660}_{\text{Incertitude}}\ \underbrace{J/K/kg}_{\text{Unitée}} \text{ à } \underbrace{95\%\ (\propto = 5\%)}_{\text{Confiance et risque}} .$$

Introduction

On mesure une même intensité électrique avec 20 multimètres identiques. Les résultats obtenus en mA sont :

119,5	119,9	119,9	119,5	119,2	120,3	119,9	119,2	119,2	119,4
120,0	119,0	120,1	119,8	119,4	120,1	120,1	119,4	119,4	119,5

- Pour exploiter ces valeurs nous devons :
  1. Décrire la distribution de cette série
  2. Caractériser cette intensité électrique :
     - Estimer sa valeur
     - Calculer l'incertitude à un niveau de confiance donné

On applique les méthodes de la statistique descriptive

On utilise les méthodes de la statistique inférentielle

Introduction :
approches et objectifs

- Statistique descriptive ou analyse des données : est l'ensemble des techniques statistiques permettant de synthétiser l'information recueillie au cours d'une enquête ou d'une étude.

Objectifs
  - Ressortir des propriétés de la distribution des données étudiées
  - Suggérer des hypothèses

- Statistique inférentielle : regroupe les méthodes d'estimation statistique et de tests d'hypothèse.

Objectifs
  - Étendre les propriétés constatées sur un échantillon à toute la population
  - Vérifier l'adéquation des hypothèses a priori ou issues d'une phase exploratoire
  - Prévoir en présence de l'aléatoire

Statistique descriptive

Cas unidimensionnel

- étude de chaque variable séparément

Cas bidimensionnel

étude des variables 2 à 2 : corrélation entre les variables

Cas multidimensionnel

étude de plus de deux variables à la fois : analyse des données

Statistique descriptive

Cas unidimensionnel

Terminologie de base

- **Population** ou **champ d'étude Ω** : c'est l'ensemble concerné par l'étude statistique
- **Unité statistique ou individu ω** : c'est un élément ou membre de la population
- **Échantillon** : c'est un sous ensemble de la population sur lequel sont réalisées les observations ou les mesures
- **Taille de l'échantillon** : est le nombre d'individus de l'échantillon; c'est la cardinalité de l'échantillon

Terminologie de base

- **Enquête :** opération qui consiste à observer tous les individus d'un échantillon
  - **Recensement :** est une enquête exhaustive; l'échantillon observé est la population entière
  - **Sondage :** est une enquête sur une partie de la population
- **Variable** ou **Caractère :** est une caractéristique observée ou mesurée sur les individus d'une population ou un échantillon
  - Modalité : est une valeur possible prise par une variable.
  - Domaine d'une variable : ensemble des valeurs possibles ou des modalités

Les différents types de variables

Une variable peut être **quantitative** ou **qualitative :**

- Une variable est dite **quantitative** quand elle est mesurable. Ses modalités sont des valeurs numériques sur lesquelles on peut effectuer des opérations algébriques
- Une variable est dite **qualitative** lorsque ses modalités expriment une qualité et non pas une quantité. Une variable qualitative est définie par un ensemble fini de modalités permettant de caractériser les observations.

Variables qualitatives

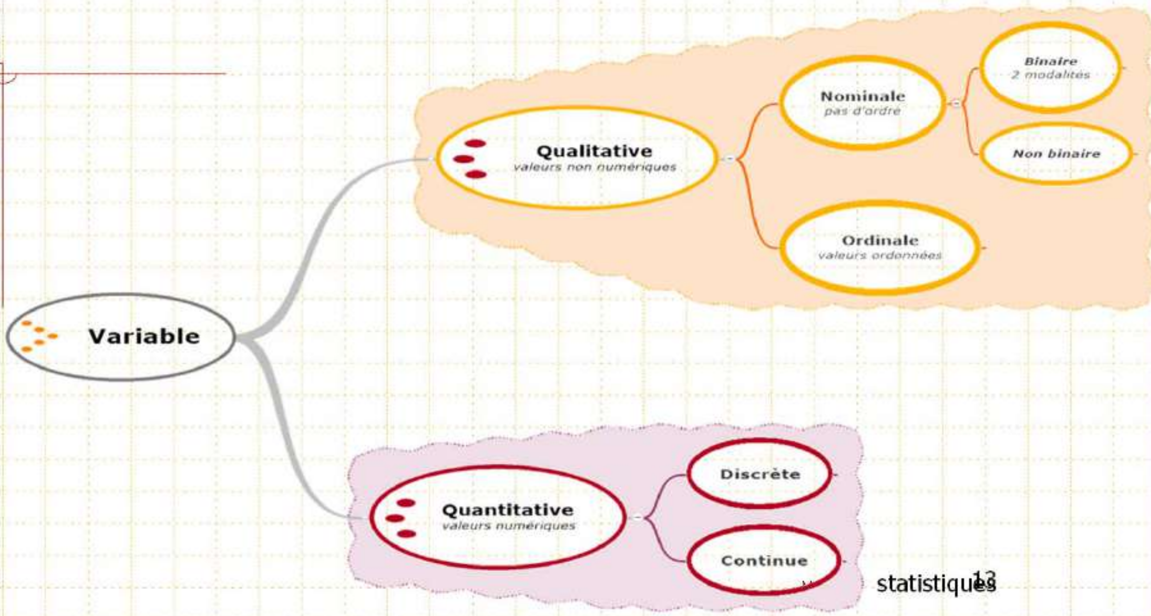
- Variable qualitative nominale **:** les modalités de la variable ne sont pas hiérarchisées ou ordonnées(*exemple: groupe sanguin*).
- Variable qualitative binaire : c'est une variable à deux modalités (*exemple: le succès ou l'échec d'un traitement*). Les deux modalités sont souvent codées 0 et 1, le 1 étant utilisé par convention pour la modalité d'intérêt.
- Variable qualitative ordinale : est une variable dont les modalités sont reliées par une relation d'ordre (*exemple : Niveau d'étude*)

Variables quantitatives :

- Variable quantitative continue : on parle de variable continue quand celle-ci peut prendre une quantité indénombrable de valeurs dans un intervalle.
  (exemple : la taille)
- Variable quantitative discrète : sont des variables qui ne peuvent prendre que des valeurs isolées.
  (Exemple : le nombre de frères et sœurs)

SMC4 : 2016-2017 Support de cours : Probabilité & statistiques; chap 4 Pr. ELYASSA

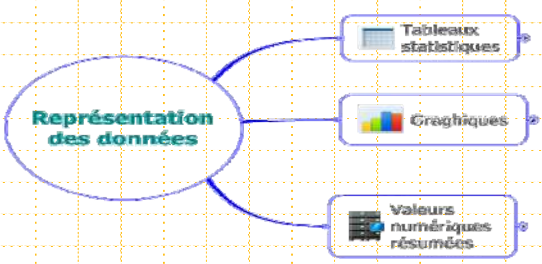
Typologie des variables



Statistique descriptive

Etude d'une variable quantitative

- On dispose d'un échantillon de n individus notés ω_i ; $i=1,\ldots,n$; X est une variable statistique et $X(\omega_i)$ est la valeur prise par cette variable sur l'individu ω_i ; $\{X(\omega_i)\ ,\ i=1,\ldots,n\}$ est la série statistique brute
- Objectif : représenter et résumer les valeurs d'une série statistique



Tableaux statistiques : variables quantitatives

La première étape consiste à organiser les valeurs dans un tableau :

- dans la première colonne du tableau, on liste les valeurs ou les classes distinctes de la série en ordre croissant
- dans les autres colonnes, on présentent les effectifs (ou les fréquences) et les effectifs cumulés (ou les fréquences cumulées) :
  - Effectif d'une modalité : le nombre de répétitions de la modalité, noté $\boldsymbol{n_i}$
  - fréquence d'une modalité : la proportion de l'effectif de la modalité par rapport à l'effectif total, notée $\boldsymbol{f_i}$
  - Effectif cumulé d'une modalité : est le cumul des effectifs des modalités qui lui sont inférieures ou égales, noté $\boldsymbol{N_i}$
  - Fréquence cumulée d'une modalité : est le cumul des fréquences des modalités qui lui sont inférieures ou égales, noté $\boldsymbol{F_i}$

Tableau statistique : Variable quantitative discrète

Ligne des titres

Modalités rangées par ordre croissant

Modalité	Effectif	Effectif cumulé	Fréquence	Fréquence cumulée
$\boldsymbol{x_1}$	n_1	n_1	f_1=n_1/n	f_1
$\boldsymbol{x_2}$	n_2	n_1+n_2	f_2=n_2/n	f_1+f_2
$\boldsymbol{x_3}$	n_3	$n_1+n_2+n_3$	f_3=n_3/n	$f_1+f_2+f_3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\boldsymbol{x_{k-1}}$	n_{k-1}	$n_1+\cdots+n_{k-1}$	f_{k-1}=n_{k-1}/n	$f_1+\cdots+f_{k-1}$
$\boldsymbol{x_k}$	n_k	n	f_k=n_k/n	1
Total	$\sum_{i=1} n_i = n$		$\sum_i f_i = 1$	

Ligne des totaux

SMC4 : 2016-2017 Support de cours : Probabilité & statistiques; chap 4 Pr. ELYASSA

Tableau statistique : Variable quantitative discrète

119,5	119,9	119,9	119,5
120,0	119,0	120,1	119,8
119,2	119,4	120,3	119,9
120,1	120,1	119,4	119,4
119,2	119,2	119,4	119,5

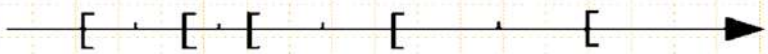
Série statistique

x_i	n_i	N_i	f_i	F_i
119,0	1	1	0,05	0,05
119,2	3	4	0,15	0,20
119,4	4	8	0,20	0,40
119,5	3	11	0,15	0,55
119,8	1	12	0,05	0,60
119,9	3	15	0,15	0,75
120,0	1	16	0,05	0,80
120,1	3	19	0,15	0,95
120,3	1	20	0,05	1
Total	20		1	

Tableau statistique

Tableau statistique : variable quantitative continue

- ✓ On discrétise une variable quantitative continue en découpant son domaine de variation en classes définies par des intervalles semi ouverts à droite sans discontinuité ni chevauchement.



- ✓ On regroupe les valeurs observées par classes pour cela il faut définir :
  - le nombre de classes k
  - les limites des classes (bornes des intervalles) : $a_1, a_2, \ldots, a_{k+1}$
  - les centres de classes $x_1, \ldots, x_k$: $x_i = \frac{a_i + a_{i+1}}{2}$
  - On appelle amplitude de la classe i la longueur de la classe i :

$$l_i = (a_{i+1} - a_i)$$

- ✓ On calcule l'effectif de chaque classe : le nombre d'observations appartenant à la classe

Tableau statistique : variable quantitative continue

Classe	Centre	Effectif	Effectif cumulé	Fréquence	Fréquence cumulée	Densité des effectifs (ou des fréquences)
$[a_1, a_2[$	x_1	n_1	n_1	$f_1=n_1/n$	f_1	$d_1=n_1/(a_2-a_1)$
$[a_2, a_3[$	x_2	n_2	n_1+n_2	$f_2=n_2/n$	f_1+f_2	$d_2=n_2/(a_3-a_2)$
$[a_3, a_4[$	x_3	n_3	$n_1+n_2+n_3$	$f_3=n_3/n$	$f_1+f_2+f_3$	$d_3=n_3/(a_4-a_3)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[a_{k-1}, a_k[$	x_{k-1}	n_{k-1}	$n_1+\cdots+n_{k-1}$	$f_{k-1}=n_{k-1}/n$	$f_1+\cdots+f_{k-1}$	$\vdots$
$[a_k, a_{k+1}[$	x_k	n_k	n	$f_k=n_k/n$	1	$d_k=n_k/(a_k-a_{k-1})$
Total		$\sum_{i=1} n_i = n$		$\sum_i f_i = 1$		

Tableau statistique : variable quantitative continue

SMC4 : 2016-2017 Support de cours : Probabilité & statistiques; chap 4 Pr. ELYASSA

Exemple : On mesure la taille en centimètres de 50 élèves d'une classe :

152 152 152 153 153 154 154 154 155 155 156 156 156 156 156 157 157 157 158 158 159 159 160 160 160 161 160 160 161 162 162 162 163 164 164 164 164 165 166 167 168 168 168 169 169 170 171 171 171 171.

- Nous allons grouper les valeurs en cinq classes de même amplitude =4 :
  [151,5 ; 155,5[, [155,5 ; 159,5[, [159,5 ; 163,5[, [163,5 ; 167,5[, [167,5 ; 171,5[

Exemple de tableau statistique

variable quantitative continue

Tableau Statistique

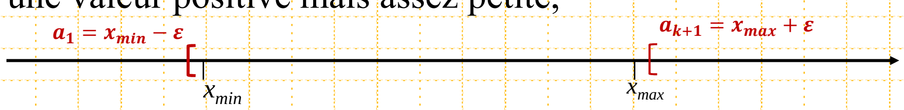
$[a_i ; a_{i+1}[$	x_i	n_i	N_i	f_i	F_i	d_i
[151,5 ; 155,5[	153,5	10	10	0,20	0,20	2,50
[155,5 ; 159,5[	157,5	12	22	0,24	0,44	3
[159,5 ; 163,5[	161,5	11	33	0,22	0,66	2,75
[163,5 ; 167,5[	165,5	7	40	0,14	0,80	1,75
[167,5 ; 171,5[	169,5	10	50	0,20	1,00	2,50
Total		**50**		**1,00**		

Tableau statistique

Variable quantitative continue

En générale les classes sont déterminées selon la nature du caractère étudier. Mais il existent des formules pour calculer de manière théorique le nombre de classes :

- La règle de Sturge : $k = [1 + 3{,}3 \log_{10}(n)]$
- La règle de Yule : $k = [2{,}5\sqrt[4]{n}]$
- L'amplitude commune des classe est obtenue de la manière suivante : $l = \frac{1}{k}\left((x_{max} + \varepsilon) - (x_{min} - \varepsilon)\right)$,
  x_{max} *et* x_{min} désignent la plus grande et la plus petite valeur observées
  ε est une valeur positive mais assez petite,



SMC4 : 2016-2017 Support de cours : Probabilité & statistiques; chap 4 Pr. ELYASSA

Tableau statistique : Variable quantitative Continue

119,5	119,9	119,9	119,5
120,0	119,0	120,1	119,8
119,2	119,4	120,3	119,9
120,1	120,1	119,4	119,4
119,2	119,2	119,4	119,5

Si on applique la régle de Yule on trouve :

$k = \left[2{,}5\sqrt[4]{20}\right] = 5$

a_1 =118,95 et a_6 =120,35

$l = 0,2(120{,}35 - 118{,}95) = 0{,}28$

Classes	x_i	n_i	N_i	f_i	F_i	d_i
[118,95 , 119,23[	119,09	4	4	0,20	0,20	14,29
[119,23 , 119,51[	119,37	7	11	0,35	0,55	25
[119,51 , 119,79[	119,65	0	11	0,00	0,55	0
[119,79 , 120,07[	119,93	5	16	0,25	0,80	17,86
[120,07 , 120,35[	120,21	4	20	0,20	1,00	14,29
Total		20		1		

Tableau statistique

Statistique exploratoire

REPRÉSENTATIONS GRAPHIQUES

SMC4 : 2016-2017 Support de cours : Probabilité & statistiques; chap 4 Pr. ELYASSA

Représentations graphiques de variables quantitatives

- ✓ Le choix de la représentation graphique est fortement lié à la nature des variables étudiées.
- ✓ Les représentations les plus fréquemment utilisées sont :
  - ➢ Représentations des effectifs (ou des fréquences):
    - diagramme en bâtons pour les variables discrètes
    - histogramme pour les variables continues
  - ➢ Représentations des effectifs cumulés (ou des fréquences cumulées)
  - ➢ Représentations de la dispersion des valeurs : boîte à moustaches

Représentation graphique de variables *discrètes*

Diagrammes en bâtons

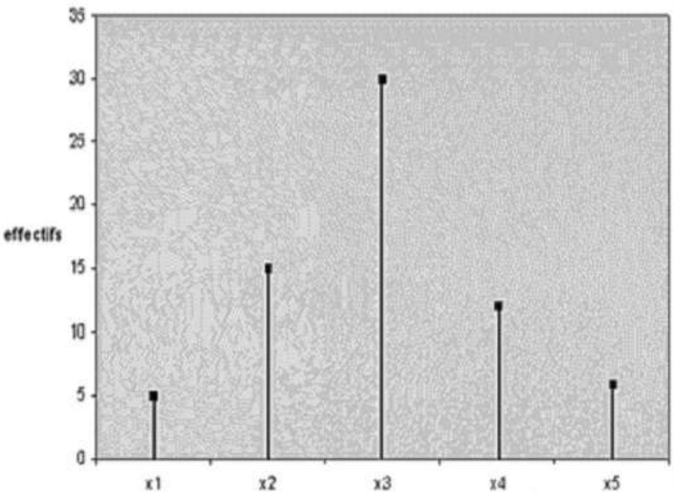


Diagramme en bâtons

Il est constitué de segments de droite verticaux dont les hauteurs sont égales aux effectifs ou aux fréquences de chaque modalité.

- Sur l'axe des abscisses (horizontal) sont reportées les modalités de la série par ordre croissant.
- Sur l'axe des ordonnés sont reportées les effectifs ou les fréquence

Représentation graphique de variables *discrètes*

Diagrammes des fréquences cumulées

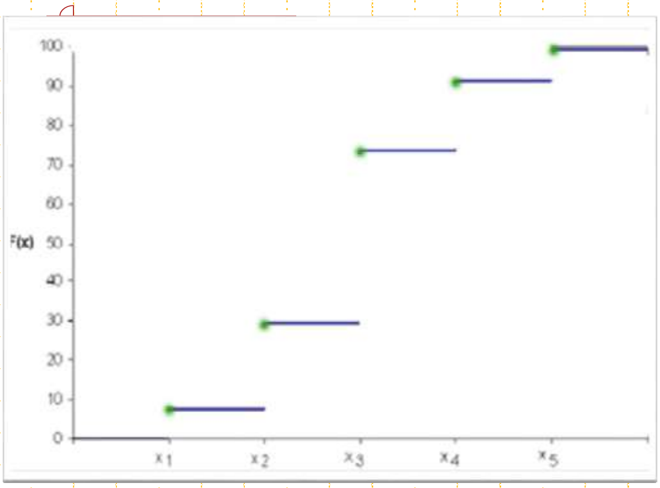


Diagramme des fréquences cumulées

Il représente une fonction en escaliers (fonction de répartition de la distribution) :

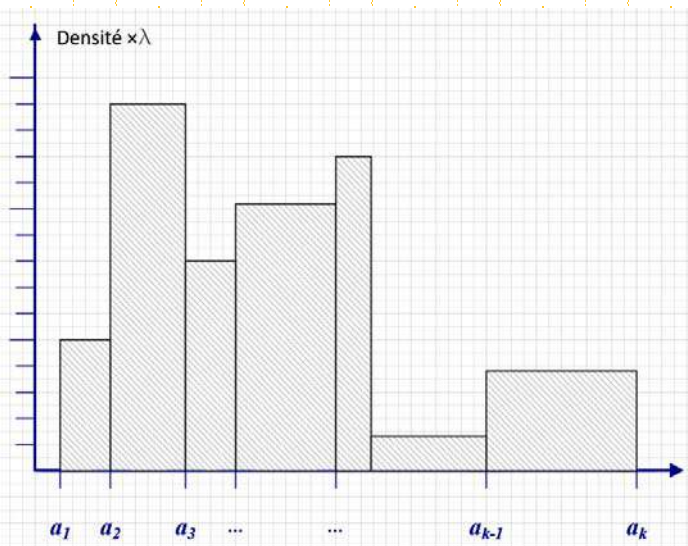
$$F(x) = F_i \quad pour \quad x_i \leq x < x_{i+1}$$

les paliers correspondent aux l'effectifs cumulés (ou aux fréquences cumulées) de la série.

- Sur l'axe des abscisses (horizontal) sont reportées les modalités de la série par ordre croissant.
- Sur l'axe des ordonnées sont reportées les effectifs cumulés ou les fréquences cumulées

Représentation graphique de v*ariables continues :*

Histogrammes



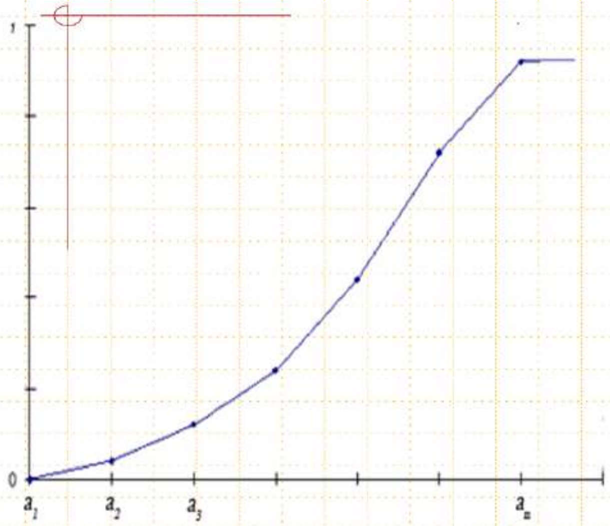
Un histogramme est constitué de rectangles contigus dont les aires sont proportionnelles aux effectifs (ou bien aux fréquences) de chaque classe :

- Sur l'axe des abscisses sont reportées les bornes des classes de la série.
- Chaque rectangle correspond à une classe et sa hauteur est $\boldsymbol{h_i = \lambda \times di}$ où λ est une constante de mise à l'échelle.
- ✓ Remarque : si les classes ont tous la même amplitude, alors on peut prendre $\boldsymbol{h_i = n_i}$ ou $\boldsymbol{h_i = f_i}$

SMC4 : 2016-2017 Support de cours : Probabilité & statistiques; chap 4 Pr. ELYASSA

Représentation graphique de v*ariables continues :*

La courbe cumulative des fréquences



Elle représente la fonction de répartition de la distribution qui doit être continue :

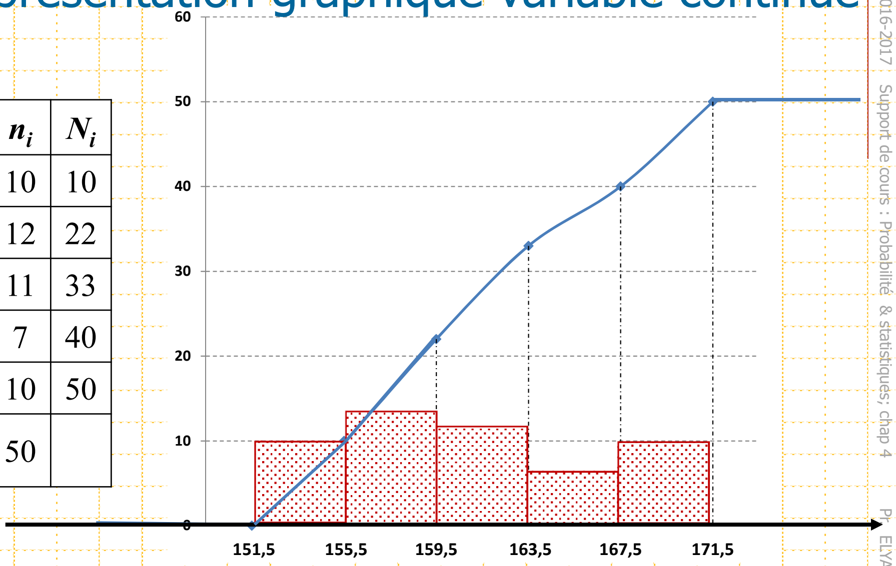
- Sur l'axe des abscisses sont reportées les bornes des classes de la série.
- Sur l'axe des ordonnées sont reportées les effectifs cumulés ou les fréquences cumulées

A la borne supérieure de chaque classe on fait correspondre en ordonnée la fréquence cumulée (ou l'effectif cumulé) de la classe puis on relie les points obtenus par des segments.

Exemple de représentation graphique variable continue

$[a_i\,;\,a_{i+1}[$	x_i	n_i	N_i
[151,5 , 155,5[	153,5	10	10
[155,5 , 159,5[	157,5	12	22
[159,5 , 163,5[	161,5	11	33
[163,5 , 167,5[	165,5	7	40
[167,5 , 171,5[	169,5	10	50
Total		50	

Tableau Statistique



Histogramme et courbe cumulative des effectifs

Statistique descriptive

VALEURS CARACTÉRISTIQUES

Valeurs caractéristiques
(nombres résumés)

Trois types de valeurs caractéristiques


- Indicateurs de position

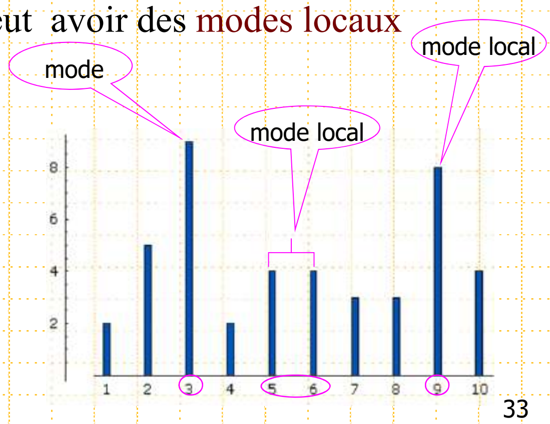

- Indicateurs de dispersion


- Indicateurs de forme

SMC4 : 2016-2017 Support de cours : Probabilité & statistiques; chap 4 Pr. ELYASSA

Indicateurs de position
Le mode

- Est la modalité qui admet le plus grand effectif la plus grande fréquence
- Une distribution peut plusieurs modes : plurimodale
- Une distribution plurimodale peut avoir des modes locaux



SMC4 : 2016-2017 Support de cours : Probabilité & statistiques; chap 4 Pr. ELYASSA

Indicateurs de position
Le mode

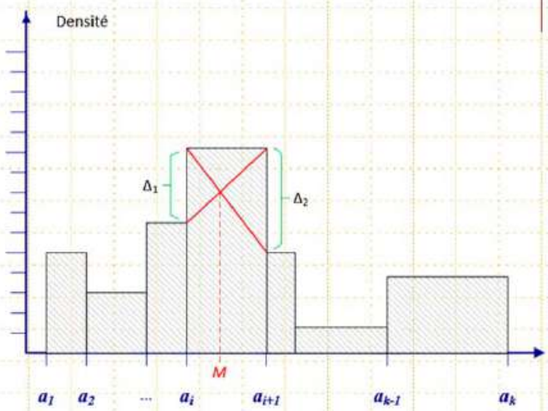
- Le mode est parfaitement défini pour une variable qualitative ou une variable quantitative discrète.
- Lorsque cette valeur est unique, on dit que la distribution est ***unimodale***, dans le cas contraire on dit que la distribution est **plurimodale.**
- Une variable des modes locaux (modalités dont la fréquence est supérieure ou égale aux fréquences adjacentes). Cette situation est intéressante : elle met en évidence l'existence de plusieurs sous-populations, donc l'hétérogénéité de la population étudiée.

Indicateurs de position
Le mode

Pour une distributions regroupées en classes,

1. On détermine d'abord la classe modale : c'est la classe qui possède la plus forte densité.
2. Puis on calcule la valeur du mode par interpolation linéaire :
   si [a_i, a_{i+1} [est la classe modale alors le mode M vérifie :

$$Mo = a_i + \frac{\Delta_1}{\Delta_1 + \Delta_2}(a_{i+1} - a_i)$$



Indicateurs de position
La médiane

La médiane est la valeur qui divise les observations en deux parties égales : la moitié des données sont inférieures ou égales à la médiane et la moitié lui sont supérieures ou égales.

- Pour une distribution discrète, il suffit d'ordonner les valeurs en une liste croissante et de choisir la valeur qui est au centre de cette liste :
  - si n est impair alors la médiane se trouve à la position $\left(\frac{n+1}{2}\right)$
  - Si n est pair est égale à la moyenne des deux valeurs centrales $\frac{n}{2}$ et $(\frac{n}{2} + 1)$

Indicateurs de position
La médiane

- Pour une distributions regroupées en classes, le calcul de la médiane se fait en deux étapes :

1. On détermine la classe qui contient la médiane : c'est la classe dont la fréquence cumulée vérifie $F_i \leq 0{,}5 \; et \; F_{i+1} > 0{,}5$
2. On calcule la médiane par interpolation linéaire : si $[\, a_i \,, a_{i+1} \,[$ est la classe de la médiane, alors

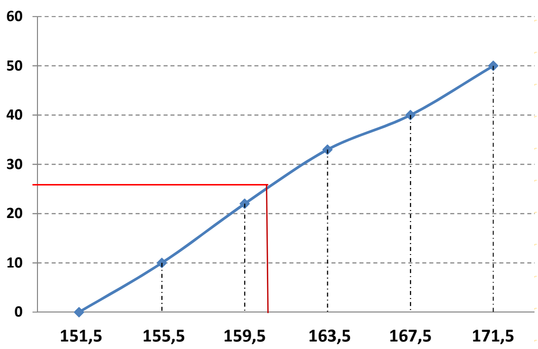
$$Me = a_i + (a_{i+1} - a_i)\frac{0{,}5 - F_i}{F_{i+1} - F_I}$$

Indicateurs de position
La médiane

Tableau Statistique

$[a_i ; a_{i+1}[$	n_i	N_i	f_i	F_i
[151,5 ; 155,5[	10	10	0,20	0,20
[155,5 ; 159,5[	12	22	0,24	0,44
[159,5 ; 163,5[	11	33	0,22	0,66
[163,5 ; 167,5[	7	40	0,14	0,80
[167,5 ; 171,5[	10	50	0,20	1
Total	**50**			

Courbe cumulative des fréquences



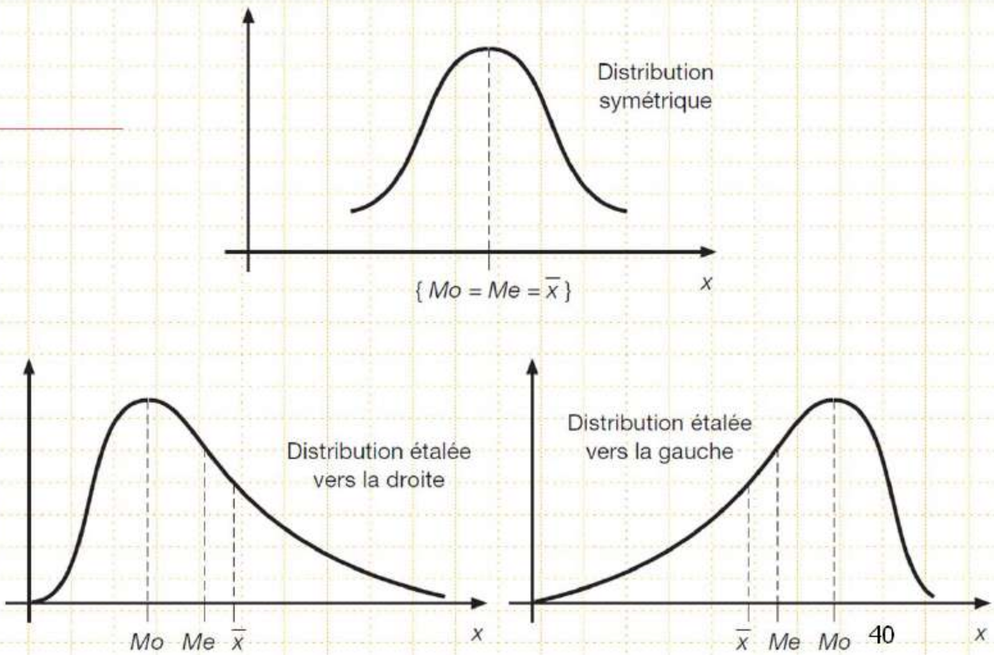
$$Me = a_i + (a_{i+1} - a_i)\frac{0{,}5 - F_i}{F_{i+1} - F_I} = 159{,}5 + 4\,\frac{0{,}5 - 0{,}44}{0{,}66 - 0{,}44} = 160{,}59$$

Indicateurs de position

La Moyenne

- La moyenne est égale à la somme de toutes les valeurs observées divisée par le nombre d'observations
- Pour une distribution groupée en classes, le calcul de la moyenne utilise les centres des classes
- Le calcul de la moyenne s'effectue directement sur les valeurs observées brutes ou bien en utilisant les effectifs ou les fréquences des observations.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \; ; \; \bar{X} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i \; ; \; \bar{X} = \sum_{i=1}^{k} f_i x_i$$



Indicateurs de position
Exemple de calcul de la moyenne

Tableau I

x_i	n_i	f_i	$n_i{*}x_i$	$f_i{*}x_i$
1	5	0,10	5	0,1
2	9	0,18	18	0,36
3	15	0,30	45	0,90
4	11	0,22	44	0,88
5	3	0,06	15	0,3
6	7	0,14	42	0,84
Total	**50**	**1**	**169**	**3,38**

Tableau II

$[a_i ; a_{i+1}[$	x_i	n_i	f_i	$n_i{*}x_i$	$f_i{*}x_i$
[151,5 ; 155,5[	153,5	10	0,20		30,7
[155,5 ; 159,5[	157,5	12	0,24		37,8
[159,5 ; 163,5[	161,5	11	0,22		35,53
[163,5 ; 167,5[	165,5	7	0,14		23,17
[167,5 ; 171,5[	169,5	10	0,20		33,9
Total		**50**	**1**		**161,1**

Colonnes ajoutées pour calculer la moyenne

Indicateurs de position
Les quartiles

Trois valeurs $Q1$, $Q2$, $Q3$ qui partagent la série ordonnée des observations en 4 groupes d'effectifs égaux

- Le premier quartile $Q1$ est obtenu lorsqu'on a cumulé 25% de la population
- Le second quartile $Q2$ est obtenu lorsqu'on a cumulé 50% de la population : c'est la médiane
- Le troisième quartile $Q3$ est obtenu lorsqu'on a cumulé 75% de la population

SMC4 : 2016-2017 Support de cours : Probabilité & statistiques; chap 4 Pr. ELYASSA

Indicateurs de position
Les quartiles

- Pour déterminer les quartiles Q_1 et Q_3 d'une série de n valeurs :
- On calcule la quantité **q=n/4.**
  Deux cas sont possibles: soit le résultat est entier (la division tombe juste), soit non.
  - **1ier cas :** q est un entier alors Q_1 est la valeur qui se trouve à la position q et Q_3 est la valeur située à la position 3q
  - **2ème cas :** q n'est pas un entier alors on arrondi q à l'entier supérieur, Q_1 est la valeur qui se trouve à cette position et on arrondi 3q à l'entier supérieur pour trouver la position de la valeur de Q_3 .

Indicateurs de position
Les quartiles

- Pour une distributions regroupées en classes, le calcul des quartiles Q_1 et Q_3 se fait en deux étapes :

1. On détermine la classe qui contient le quartile Q_1 (resp. Q_3) : c'est la classe dont la fréquence cumulée vérifie
   $F_i \leq 0{,}25 \; et \; F_{i+1} > 0{,}25$ (resp. $F_i \leq 0{,}75 \; et \; F_{i+1} > 0{,}75$)
2. On calcule la médiane par interpolation linéaire :
   si $[\, a_i , a_{i+1} [$ est la classe de Q_1 (resp. Q_3) alors

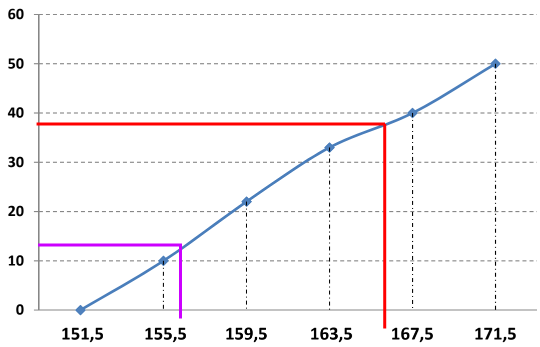
$$Q_1 = a_i + (a_{i+1} - a_i)\frac{0{,}25-F_i}{F_{i+1}-F_I} \text{ (resp. } Q_3 = a_i + (a_{i+1} - a_i)\frac{0{,}75-F_i}{F_{i+1}-F_I})$$

SMC4 : 2016-2017 Support de cours : Probabilité & statistiques; chap 4 Pr. ELYASSA

Indicateurs de position

Les quartiles

$[a_i ; a_{i+1}[$	n_i	N_i	f_i	F_i
[151,5 ; 155,5[	10	10	0,20	0,20
[155,5 ; 159,5[	12	22	0,24	0,44
[159,5 ; 163,5[	11	33	0,22	0,66
[163,5 ; 167,5[	7	40	0,14	0,80
[167,5 ; 171,5[	10	50	0,20	1
Total	**50**			



$$Q1 = a_i + (a_{i+1} - a_i)\frac{0,25 - F_i}{F_{i+1} - F_I} = 155,5 + 4\,\frac{0,25 - 0,20}{0,44 - 0,20} = 156,33$$

$$Q3 = a_i + (a_{i+1} - a_i)\frac{0,75 - F_i}{F_{i+1} - F_I} = 163,5 + 4\,\frac{0,75 - 0,66}{0,80 - 0,66} = 166,07$$

Indicateurs de dispersion

- Variance et écart type
- Coefficient de variation
- Écart interquartiles

SMC4 : 2016-2017 Support de cours : Probabilité & statistiques; chap 4 Pr. ELYASSA

Indicateurs de dispersion
Ecart type expérimental

- L'écart type s mesure la dispersion de l'échantillon autour de la moyenne.
- Pour calculer l'écart type, il faut calculer la variance de l'échantillon :

$$s^2 = \frac{1}{n-1}\sum_i n_i(x_i - \bar{X})^2 = \frac{n}{n-1}\sum_i f_i(x_i - \bar{X})^2$$

$$= \frac{1}{n-1}\sum n_i {x_i}^2 - \frac{n}{n-1}(\bar{X})^2 = \frac{n}{n-1}\left(\sum f_i {x_i}^2 - (\bar{X})^2\right)$$

Indicateurs de dispersion
Coefficient de variation

- Le coefficient de variation (CV) est le rapport de l'écart-type à la moyenne;

$$\boldsymbol{CV} = \frac{\boldsymbol{s}}{|\boldsymbol{\bar{X}}|}$$

- Le coefficient de variation n'a pas d'unité
- Ce coefficient est souvent exprimé sous forme de pourcentage

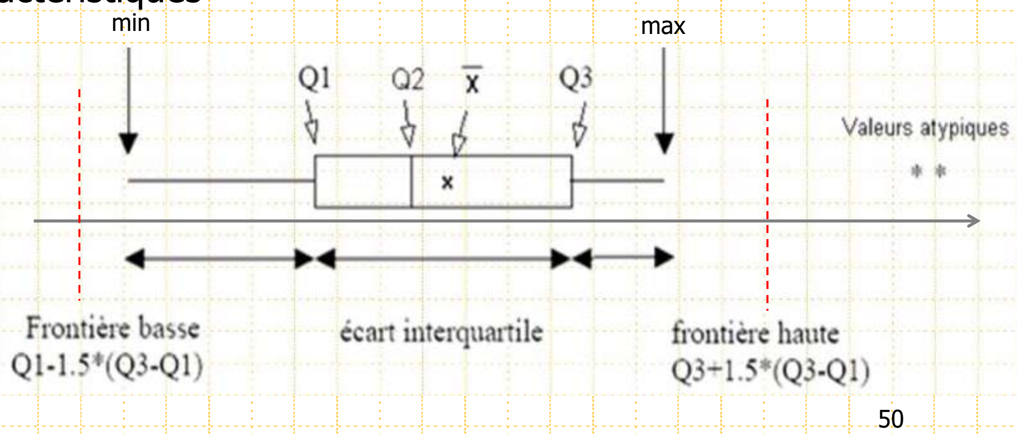
Mostafa El Yassa SMC4 : 2016-2017 Support de cours : Probabilité & statistiques; chap 4 Pr. ELYASSA

Indicateurs de dispersion
Intervalle et écart interquartile

- On appelle intervalle interquartile l'intervalle $[Q1\ ;\ Q3]$;
- L'écart interquartile est l'amplitude de l'intervalle $[Q1\ ;\ Q3]$, c'est-à-dire le nombre $(Q3 - Q1)$;
- L'écart interquartile est utilisé comme indicateur de dispersion. Il correspond à 50% des effectifs situés dans la partie centrale de la distribution
- Toute valeur n'appartenant pas à l'intervalle $[Q1 - 1{,}5(Q3 - Q1)\ ,\ \ Q3 + 1{,}5(Q3 - Q1)]$ est considérée comme aberrante

Représentation graphique : Boite à moustaches

Graphique très pratique qui permet de résumer quelques caractéristiques



SMC4 : 2016-2017 Support de cours : Probabilité & statistiques; chap 4 Pr. ELYASSA

Représentation graphique : Boite à moustaches

Ce diagramme est constitué de la façon suivante:

- On trace un axe des unités horizontalement ou verticalement
- On trace une "boîte" : un rectangle dont la longueur s'étend du premier quartile au troisième quartile, et qui est coupé par un trait vertical à hauteur de la médiane. On matérialise la position de la moyenne par une croix.
- De cette boîte partent deux traits (moustaches) dans le sens de l'axe : l'un va du premier quartile Q1 à la valeur minimale non aberrante de la série, l'autre du troisième quartile Q3 à la valeur maximale non atypique.
- On représente également les valeurs aberrantes de la série.

Exemple

x_i	n_i	N_i	f_i	F_i
119,0	1	1	0,05	0,05
119,2	3	4	0,15	0,20
119,4	4	8	0,20	0,40
119,5	3	11	0,15	0,55
119,8	1	12	0,05	0,60
119,9	3	15	0,15	0,75
120,0	1	16	0,05	0,80
120,1	3	19	0,15	0,95
120,3	1	20	0,05	1
Total	20		1	

$\boldsymbol{Mode = 119,40}$

$\boldsymbol{Q_1 = 119,4;\ \ Q_2 = 119,5\ ;\ \ Q_3 = 119,9}$

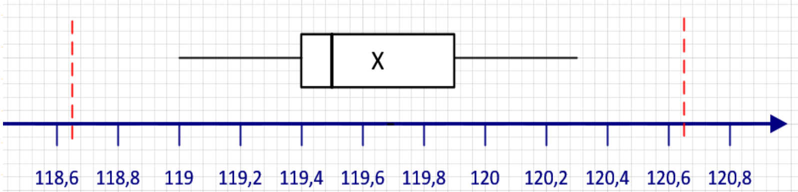
$\boldsymbol{IQ = Q_3 - Q_1 = 0,5}$

$\boldsymbol{Q_1 - 1,5(Q_3 - Q_1) = 118,65}$

$\boldsymbol{Q_3 + 1,5(Q_3 - Q_1) = 120,65}$

$\boldsymbol{\overline{X} = 119,64}$

$\boldsymbol{\sigma = 0,38}$



ESTIMATION
INTERVALLE DE CONFIANCE

Exemple

On mesure une même intensité électrique avec 20 multimètres identiques. Les résultats obtenus en mA sont :

119,5	119,9	119,9	119,5	119,2	120,3	119,9	119,2	119,2	119,4
120,0	119,0	120,1	119,8	119,4	120,1	120,1	119,4	119,4	119,5

- Les indicateurs de position de cette série de mesures sont : mode =119,40 ; médiane=119,50 ; ***moyenne=119,64***
- Les indicateurs de dispersion : Ecart type ***s=0,38*** ***, coefficient de variation = 0,3%***

Alors :

- Quelle est la valeur de cette intensité ?
- Quelle l'incertitude de cette valeur et quel son niveau de confiance ?

Estimateur et intervalle de confiance définitions

- ✓ L'estimation consiste à évaluer un paramètre θ de la population-mère à l'aide d'un paramètre T calculé à partir des mesures effectuées sur un échantillon.
  T est appelé « estimateur du paramètre θ »
- ✓ Un estimateur T doit posséder deux propriétés essentielles :
  1. Son espérance doit tendre vers la vraie valeur : $\lim_{n\to\infty} E(T) = \theta$
  2. Sa variance doit tendre vers 0 : $\lim_{n\to\infty} Var(T) = 0$

Base théorique de l'estimation : Le théorème central limite

Soit un échantillon de taille n issue d'une population mère dont les valeurs d'un paramètre suivent une distribution de moyenne μ *et d'écart type* σ , alors :

- la valeur moyenne des mesures observées sur un grand nombre n *varie autour de* μ *avec un écart-type égal à* $\sigma/\sqrt{n}$
- quand n croît la distribution des mesures observées est de plus en plus concentrée autour de μ et devient de plus en plus proche d'une distribution de la loi normal $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

SMC4 : 2016-2017 Support de cours : Probabilité & statistiques; chap 4 Pr. ELYASSA

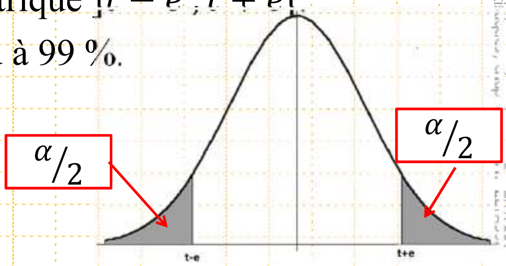
Estimateur et intervalle de confiance

- Dans le cas des mesures physico-chimiques, les paramètres de la population-mère représentent les *vraies valeurs des grandeurs physiques* que l'on veut mesurer et qu'on ne peut atteindre théoriquement qu'au bout d'une infinité de mesures.
- Dans ce contexte les seules paramètres traités sont la moyenne et l'écart-type

Paramètre	Valeur réelle	Estimateur
Moyenne	μ	$\bar{X} = \frac{S_n}{n}$
Ecart- type	σ	$s = \frac{n}{n-1}\left(\sum f_i (x_i)^2 - \bar{X}^2\right)$

Intervalle de confiance définitions

- L'intervalle de confiance est un intervalle dans lequel on a une probabilité connue de trouver la vraie valeur, donc c'est un intervalle t_1, t_2 tel que $P(t_1 < \theta < t_2) = 1 - \alpha$;
  $(1 - \alpha)$est le seuil de confiance, α est le seuil de risque de l'estimation
- Dans le cas de mesures physiques on considère pratiquement toujours :
  - l'intervalle de confiance bilatéral symétrique $]t - e\ ; t + e[$
  - le seuil de confiance est fixé à 95 % ou à 99 %.



SMC4 : 2016-2017 Support de cours : Probabilité & statistiques; chap 4 Pr. ELYASSA

Estimation de la moyenne

Cas 1 : la valeur de σ est connue

Par application du théorème central limite l'estimateur de la moyenne $T = \bar{X}$ suit une loi normale $\mathcal{N}\left(\mu\,,\frac{\sigma^2}{n}\right)$,

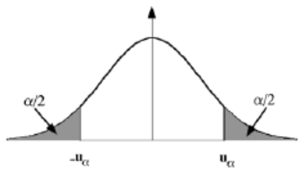
d'où pour un intervalle de confiance au seuil α :

$$P\left(T - a\frac{\sigma}{\sqrt{n}} < \mu < T + a\frac{\sigma}{\sqrt{n}}\right) = P\left(\mu - a\frac{\sigma}{\sqrt{n}} < T < \mu + a\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$\mu = T \pm u_\alpha \frac{\sigma}{\sqrt{n}} \;\; ; P(|Z| < u_\alpha) = 1 - \alpha\,, Z \sim \mathcal{N}(0\,,1)$$

- Pour $\alpha = 5\%$, on aura $u_\alpha = 1{,}960$; $\mu = T \pm 1{,}96\frac{\sigma}{\sqrt{n}}$
- Pour $\alpha = 1\%$, on aura $u_\alpha = 2{,}576$; $\mu = T \pm 2{,}576\frac{\sigma}{\sqrt{n}}$

Table des fractiles symétriques de la loi normale $\mathcal{N}(0\,,1)$



α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,10	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,20	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,30	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,40	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690
0,50	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,60	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399
0,70	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,80	0,253	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,90	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013

Cette table donne la valeur de u_α telle que

$$P(|Z| < u_\alpha) = 1 - \alpha$$

Ainsi on a :

- $u_{0,15} = 1{,}440$;
- $u_{0,05} = 1{,}960$;
- $u_{0,02} = 2{,}326$;
- $u_{0,01} = 2{,}576$.

Estimation de la moyenne
Cas 1 : la valeur de σ est connue

Exemple : On mesure une même intensité électrique avec 20 multimètres identiques. On suppose que l'écart type réel est connu : $\sigma = 0{,}36$.

Les mesures obtenus en mA sont :

119,5	119,9	119,9	119,5	119,2	120,3	119,9	119,2	119,2	119,4
120,0	119,0	120,1	119,8	119,4	120,1	120,1	119,4	119,4	119,5

La moyenne de l'échantillon $\bar{X} = 119{,}64$, alors l'estimation au risque α de la valeur de l'intensité est :

$$\mu = 119{,}64 \pm u_\alpha \frac{0{,}36}{\sqrt{20}} = 119{,}64 \pm 0{,}080\, u_\alpha \text{ mA}$$

- $\mu = 119{,}64 \pm 0{,}16$ mA , avec un risque $\alpha = 5\%$
- $\mu = 119{,}64 \pm 0{,}21$ mA , avec un risque de $\alpha = 1\%$

Estimation de la moyenne
Cas 2 : la valeur de σ est inconnue

Dans ce cas on doit estimer la valeur de σ par s, d'où pour un intervalle de confiance au seuil α :

$$1 - \alpha = P\left(T - a\frac{s}{\sqrt{n}} < \mu < T + a\frac{s}{\sqrt{n}}\right) = P\left(\mu - a\frac{s}{\sqrt{n}} < T < \mu + a\frac{s}{\sqrt{n}}\right)$$
$$= P\left(-a < \frac{T-\mu}{\frac{s}{\sqrt{n}}} < a\right)$$

Mais $\left(\frac{T-\mu}{\frac{s}{\sqrt{n}}}\right)$ suit la loi de student à $n-1$ degrés de liberté : $\boldsymbol{t(n-1)}$

$$\mu = T \pm t_{n-1,\alpha} \times \frac{s}{\sqrt{n}} \quad ; P\left(|t| < t_{n-1,\alpha}\right) = 1 - \alpha$$

Table des fractiles symétriques de la loi student $t(n)$

Cette table donne la valeur de $t_{n-1,\,\alpha}$ telle que

$$P\left(|t| < t_{n-1,\,\alpha}\right) = 1 - \alpha$$

Ainsi on a :

- $t_{4\,,\,0{,}05} = 2{,}776$;
- $t_{4\,,\,0{,}01} = 4{,}604$;
- $t_{19\,,0{,}05} = 2{,}093$;
- $t_{19\,,\,0{,}01} = 2{,}861$.

α / ddl	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	1,000	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,816	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,765	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,134	0,741	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,727	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,718	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,711	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,706	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,703	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,700	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,697	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,695	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,694	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,692	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,691	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,690	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,689	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,688	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,688	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,687	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,686	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,686	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,685	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,685	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,684	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,684	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,684	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,683	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,683	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,683	1,055	1,310	1,697	2,042	2,457	2,750	3,646
∞	0,126	0,674	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Estimation de la moyenne
Cas 2 : la valeur de σ est inconnue

Exemple :

Les mesures obtenus en mA sont :

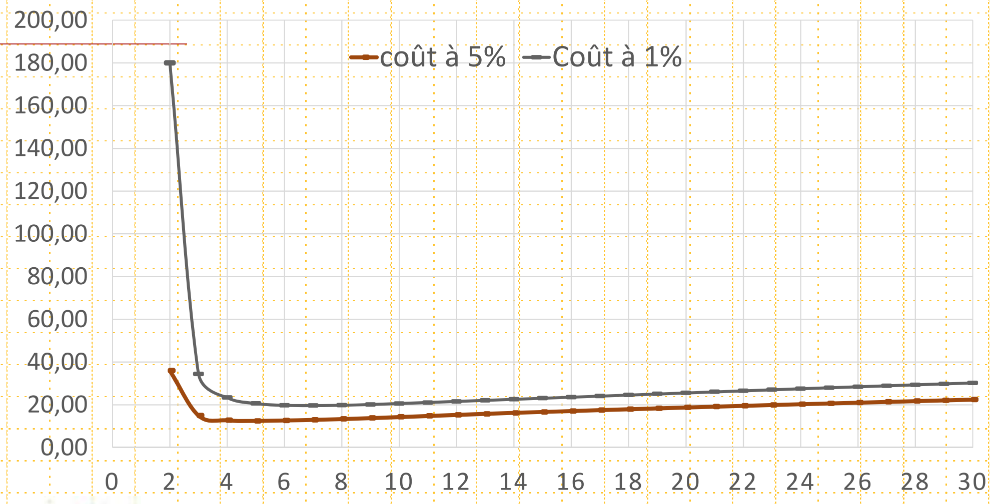
119,5	119,9	119,9	119,5	119,2	120,3	119,9	119,2	119,2	119,4
120,0	119,0	120,1	119,8	119,4	120,1	120,1	119,4	119,4	119,5

La moyenne de l'échantillon $\bar{X} = 119{,}64$, et ***s=0,38*** alors l'estimation au risque α de la valeur de l'intensité est :

$$\mu = 119{,}64 \pm t_{(19\,,\alpha)} \frac{0{,}38}{\sqrt{20}} = 119{,}64 \pm 0{,}084 \times t_{(19\,,\alpha)}\ mA$$

- $\mu = 119{,}64 \pm 0{,}18\ mA$, avec un risque $\alpha = 5\%$
- $\mu = 119{,}64 \pm 0{,}24\ mA$, avec un risque de $\alpha = 1\%$

COUT D'ESTIMATION DE LA MOYENNE EN FONCTION DE ddl



il ne faut jamais travailler en dessous de 4 mesures pour faire des estimations

Bon courage

LIENS UTILES ☝

Visiter :

1. https://biologie-maroc.com
   - Télécharger des cours, TD, TP et examens résolus (PDF Gratuit)

2. https://biologie-maroc.com/shop/
   - Acheter des cahiers personnalisés + Lexiques et notions.
   - Trouver des cadeaux et accessoires pour biologistes et géologues.
   - Trouver des bourses et des écoles privées

3. https://biologie-maroc.com/emploi/
   - Télécharger des exemples des CV, lettres de motivation, demandes de ...
   - Trouver des offres d'emploi et de stage