

Statistiques descriptives et inférentielles

Abdallah Benaïssa ¹

14 Avril 2020

¹Professeur en mathématiques, chargé du module de biostatistiques à la faculté de médecine de l'université de Batna - Algérie

Table des matières

Preface	ix
---------	----

I Rappels mathématiques. 1

1 Théorie des ensembles.	1
1.1 Ensemble, Elément.	1
1.1.1 Définitions	1
1.1.2 Notation	1
1.1.3 Opération sur les ensembles	1
1.1.4 Propriétés des opérations sur les ensembles	3
1.2 Ensembles particuliers	4
1.2.1 Ensemble fini	4
1.2.2 Ensemble dénombrable	5
1.2.3 Ensemble infini non dénombrable.	5
1.2.4 Ensembles produits	5
1.2.5 Famille d'ensembles	6
1.3 Partition d'un ensemble	7
2 Sommation de nombres et de fonctions.	9
2.1 Le symbole \sum	9
2.1.1 Cas fini	9
2.1.2 Cas infini	10
2.2 Le symbole \int	10
2.2.1 Cas de fonction à valeurs positives.	10
2.2.2 Exemple	10
2.2.3 Cas de fonction à valeurs non positives.	10

2.2.4	Exemple	11
2.2.5	Cas général.	11
2.2.6	Exemple	11
2.2.7	Intégrale sur un intervalle non borné.	11
2.2.8	Exemple	12
2.3	Primitive d'une fonction.	12
2.3.1	Notation	13
2.3.2	Remarque	13
2.3.3	Exemples.	13
2.3.4	Propriétés des primitives	14
2.3.5	Primitive et calcul d'intégrales	14
2.3.6	Exemples	14
2.3.7	propriétés de l'intégrale.	15
 II Statistiques descriptives		 17
 3 Statistique unidimensionnelle		 21
3.1	Définition	21
3.1.1	Population, unité et caractère statistiques.	21
3.1.2	Recensement - Echantillonnage	21
3.1.3	Série statistique simple	22
3.2	Natures des variables statistiques et échelles de mesures.	22
3.2.1	Exemples.	23
3.2.2	Echelles de mesures : Recapitulation par des exemples	23
3.3	La variabilité et l'incertain en biologie	24
3.3.1	Variabilité biologique et variabilité métrologique	24
3.3.2	La décision dans l'incertain	25
3.4	Représentation des données	25
3.4.1	Caractère qualitatif	26
3.4.2	Caractère quantitatif discret	26
3.4.3	Caractère quantitatif représenté en classes de valeurs.	28
3.5	Les différents types de tableaux statistiques	28
3.5.1	Tableaux de données.	28
3.5.2	Tableaux de distributions statistiques.	29

TABLE DES MATIÈRES

vii

3.5.3	Tableau de contingence.	30
3.5.4	Utilisation des effectifs cummulés.	30
3.6	Caractéristiques de tendance centrale	32
3.6.1	La moyenne arithmétique.	32
3.6.2	Exemple	33
3.6.3	La médiane	34
3.6.4	Exemple	35
3.6.5	Le mode	36
3.6.6	Exemple.	36
3.7	Caractéristiques de dispersion	37
3.7.1	L'étendue	37
3.7.2	Quantile	37
3.7.3	Calcul des quartiles : Données discrètes	38
3.7.4	Exemple 1. (données en vrac).	39
3.7.5	Exemple 2. (données groupées en valeurs).	39
3.7.6	Calcul des quartiles : Données groupées en classes de valeurs.	40
3.7.7	Exemple 3.	41
3.7.8	Boite à moustache	42
3.7.9	Variance et Ecart-type	43
3.7.10	Théorème de Konig-Huygens	44
3.7.11	Exemples	45
4	Séries statistiques Doubles	49
4.1	Vocabulaire	49
4.1.1	Définition	49
4.1.2	Notation	49
4.1.3	Exemple	50
4.1.4	Exemple	50
4.2	Variable indépendante Versus Variable dépendante	50
4.2.1	Définition.	50
4.2.2	Exemple.	51
4.2.3	Remarque.	51
4.3	Représentations	51
4.3.1	Représentation brute	51
4.3.2	Représentation dans un tableau à deux lignes	52

4.3.3	Représentation dans un tableau à trois lignes	52
4.3.4	Représentation dans un tableau de contingence	52
4.3.5	Notation et complétion d'un tableau de contingence	54
4.4	Fréquences associées à une série double	54
4.4.1	Définition	54
4.4.2	Remarques.	55
4.4.3	Tableau de contingence des fréquences.	55
4.4.4	Exemple	56
4.5	Différentes distributions	57
4.5.1	Distribution jointe des effectifs de X et Y	57
4.5.2	Distributions marginales	57
4.5.3	Exemple.	58
4.5.4	Distributions conditionnelles	59
4.5.5	Indépendance de variables statistiques	61
4.6	Séries doubles quantitatives	63
4.6.1	Passage d'une représentation à une autre	63
4.6.2	Exemple	65
4.6.3	Moyennes des distributions marginales :	66
4.6.4	Variances des distributions marginales :	66
4.6.5	Ajustement linéaire d'une série quantitative double	69

III Analyse combinatoire et probabilités 81

5	Analyse combinatoire	83
5.1	Arrangements et permutatations	83
5.1.1	Arrangements.	83
5.1.2	Arrangement avec répétition.	84
5.1.3	Permutation.	85
5.2	Combinaisons.	86
5.2.1	Combinaison (sans répétition)	86
5.2.2	Combinaisons et arrangements.	86
5.2.3	Combinaison avec répétition.	88

TABLE DES MATIÈRES

ix

5.2.4	Formule.	88
5.2.5	Permutation avec répétition.	89
6	Notions de probabilités.	93
6.1	Introduction	93
6.1.1	Exemple	94
6.2	Définitions	94
6.2.1	Ensemble fondamental, événement.	95
6.3	Opérations sur les événements.	95
6.3.1	Négation d'un événement.	95
6.3.2	Intersection de deux événements.	95
6.3.3	Union de deux événements.	96
6.3.4	Exemple.	96
6.4	Probabilité sur un ensemble.	96
6.4.1	Ensemble probabilisable.	97
6.4.2	Ensemble probabilisé.	97
6.4.3	Propriétés d'une probabilité.	98
6.4.4	Cas d'ensemble finis.	99
6.4.5	Ensemble Equiprobable.	100
6.4.6	Incompatibilité- indépendance d'événements.	100
6.5	Exemples.	100
6.5.1	Exemple.	100
6.5.2	Exemple.	102
6.5.3	Exemple.	103
6.5.4	Exemple.	105
6.5.5	Exemple.	106
6.6	Probabilité conditionnelle	108
6.6.1	Définition	108
6.6.2	Remarque	108
6.6.3	Cas particulier : ensemble équiprobable	109
6.6.4	Exemple	109
6.7	Probabilités composées.	110
6.7.1	Cas de deux événements.	110
6.7.2	Cas de plusieurs événements.	111
6.7.3	Exemple.	111

6.8	Système complet d'événements.	113
6.8.1	Définition	113
6.8.2	Exemple.	113
6.8.3	Formule des probabilités totales	114
6.9	Formule de Bayes.	114
6.9.1	Forme simple	114
6.9.2	Formule de Bayes.	114
6.9.3	Remarque.	115
6.9.4	Exemple	115
7	Variables aléatoires	119
7.1	Définitions	119
7.1.1	Exemple.	120
7.2	Variable aléatoire finie	121
7.2.1	Définition	121
7.2.2	Variable aléatoire dénombrable	128
7.2.3	Variable aléatoire définie par une densité	130
7.2.4	Espérance, variance et écart type d'une variable aléatoire définie par une densité	133
7.2.5	Fonction de répartition d'une variable aléatoire	136
7.3	Exemples de lois discrètes	147
7.3.1	Loi binomiale	147
7.3.2	Loi de Poisson	153
7.4	Loi normale $\mathcal{N}(\mu, \sigma)$	156
7.4.1	Définitions	156
7.4.2	Espérance $E(X)$, Variance $V(X)$ et Ecart type $\sigma(X)$	157
7.4.3	Calcul de probabilité pour une loi normale.	157
7.4.4	Approximation de lois discrètes par la loi normale	164
IV	Statistiques inférentielles	167
8	Variable aléatoire "moyenne arithmétique".	169

8.1	Théorie d'échantillonnage	169
8.1.1	Les différentes méthodes d'échantillonnage aléatoire. . .	169
8.1.2	Définitions	170
8.1.3	Exemple.	171
8.2	La loi de probabilité de la moyenne arithmétique	177
8.2.1	Proposition	177
8.2.2	Théorème central limite.	177
8.2.3	Remarque	178
8.2.4	Exemple 1	178
8.2.5	Exemple 2	179
9	Théorie de l'estimation	181
9.1	Estimation ponctuelle	181
9.1.1	Biais	182
9.1.2	Exemple	182
9.2	Estimation par intervalle de confiance	182
9.2.1	Signification de l'intervalle de confiance	183
9.2.2	Taille de la population	183
9.2.3	Taille de l'échantillon	183
9.2.4	Intervalle de pari (ou de fluctuation).	184
9.2.5	Intervalle de pari pour \bar{X}_n	184
9.3	Intervalles de confiance pour μ	185
9.3.1	Cas de grands échantillons	185
9.3.2	Cas de petits échantillons, quand X suit la loi normale et σ est inconnu	187
9.4	Intervalle de confiance pour la différence $\mu_1 - \mu_2$	188
9.4.1	Cas de grands échantillons ($m > 30$ et $n > 30$)	189
9.4.2	Cas où les lois mères sont normales, de variances σ_1^2 et σ_2^2 connues	190
9.4.3	Cas où les lois mères sont normales, mais de variances σ_1^2 et σ_2^2 inconnues	190
9.5	Intervalle de confiance pour la fréquence vraie P	191
9.5.1	Introduction	191
9.5.2	Exemple.	192
9.5.3	Intervalle de pari pour la fréquence \tilde{P} d'un échantillon	193

10 Tests d'hypothèses de conformité	195
10.1 Conformité de m à μ	195
10.1.1 Principe	195
10.1.2 Cas de grand échantillons ($n > 30$)	195
10.1.3 P- valeur.	196
10.1.4 Exemple	197
10.1.5 Cas où la loi de X est normale et σ est connu	198
10.1.6 Test de conformité unilatéral	198
10.1.7 Exemple.	199
10.1.8 Cas où la loi de X est normale, $n < 30$ et σ est inconnu	199
10.1.9 Exemple.	200
10.2 Comparaison \tilde{p} calculée et P vraie	201
10.2.1 Test bilatéral de l'hypothèse ($\tilde{p} = P$)	201
10.2.2 Test unilatéral de l'hypothèse ($\tilde{p} = P$)	202
10.2.3 Exemple.	202
11 Tests d'hypothèses d'homogénéité	205
11.1 Comparaison de deux moyennes μ_1 et μ_2	205
11.1.1 Cas de grands échantillons ($m > 30$ et $n > 30$)	205
11.1.2 Test bilatéral de l'hypothèse $\mu_1 - \mu_2 = 0$	206
11.1.3 Test unilatéral de l'hypothèse $\mu_1 - \mu_2 = 0$	206
11.1.4 Exemple.	207
11.1.5 Cas de lois normales et variances σ_X^2 et σ_Y^2 connues	208
11.1.6 Cas de lois normales, et variances σ_X^2 et σ_Y^2 inconnues	209
11.1.7 Exemple	209
11.2 Comparaison de P_1 et P_2	211
11.2.1 Introduction	211
11.2.2 Test bilatéral de l'hypothèse $P_1 = P_2$	212
11.2.3 Test unilatéral de l'hypothèse $P_1 = P_2$	213
11.2.4 Exemple.	213
12 Tests de Khi deux	215
12.1 Test de Khi deux de conformité	215
12.1.1 Introduction	215
12.1.2 Les effectifs observés et effectifs théoriques.	216
12.1.3 Les étapes du test.	216

TABLE DES MATIÈRES

xiii

12.1.4	Validité de l'application du test.	217
12.1.5	Exemple	217
12.1.6	Exemple	218
12.1.7	Exemple (Cas particulier)	220
12.2	Test d'homogénéité	221
12.2.1	Introduction	221
12.2.2	Exemple	222
12.2.3	Exemple	223
12.3	Test d'indépendance	225
12.3.1	Introduction	225
12.3.2	Exemple	226
13	Analyse de la variance (ANOVA)	229
13.1	Comparaison de plusieurs moyennes.	229
13.1.1	Conditions de validité du test.	229
13.1.2	Exemple	230
13.1.3	Les étapes du test.	232
13.2	Comparaison de deux variances	233
13.2.1	Exemple	233
13.2.2	Exemple	234
14	Corrélations linéaires	237
14.1	Introduction	237
14.2	Notations	238
14.3	Regression linéaire	239
14.3.1	Rappel	239
14.3.2	Ajustement linéaire.	239
14.3.3	Intervalle de confiance pour $E(Y x_0)$	240
14.4	Test sur le coefficient de corrélation ρ	240
14.4.1	Test d'indépendance : $\rho = 0$	240
14.4.2	Exemple	241
14.4.3	Exemple	242
14.4.4	Comparaison de ρ à une valeur donnée ρ_0	243
14.5	Tests sur la droite d'ajustement	244
14.5.1	Test de l'hypothèse $\beta = \beta_0$	245
14.5.2	Test de l'hypothèse $\alpha = \alpha_0$	246
14.5.3	Exemples	247

Preface

Ce polycopié est adressé aux étudiants de la première année en médecine. Il est conçu selon le nouveau programme officielle élaboré en Juin 2017.

Première partie

Rappels mathématiques.

Chapitre 1

Théorie des ensembles.

1.1 Ensemble, Élément.

1.1.1 Définitions

On appelle ensemble, toute liste où collection d'objets bien définis, explicitement où implicitement ; on appelle éléments ou membres de l'ensemble, les objets appartenant à l'ensemble et on note :

- $p \in A$ si p est un élément de l'ensemble A .
- $B \subset A$ ou $A \supset B$, si $x \in B \Rightarrow x \in A$.
- Si $B \subset A$, on dit que B est une partie de A , ou que B est un sous ensemble de A .

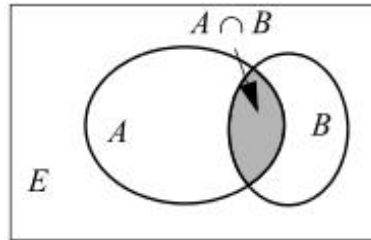
On définit un ensemble, soit en listant ses éléments : $A = \{1, 2, 3\}$, soit en donnant la définition de ses éléments : $X = \{x : x \text{ est un entier pair}\}$.

1.1.2 Notation

- la négation de $x \in A$ est $x \notin A$.
- \emptyset est l'ensemble vide.
- E est l'ensemble universel.

1.1.3 Opération sur les ensembles

Soient A et B deux ensembles quelconques.

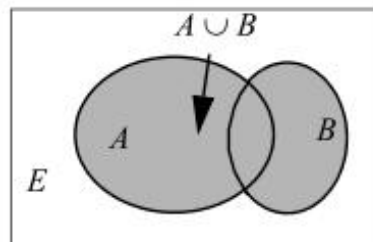
Intersection

L'intersection de A et B , notée $A \cap B$, est l'ensemble des éléments x tels que $x \in A$ et $x \in B$,
soit :

$$A \cap B = \{x : x \in A \text{ et } x \in B\}.$$

Le terme « et » est employé au sens, $x \in A$ et B , si x appartient à la fois à A et à B .

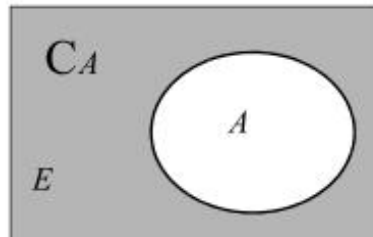
Cas particulier. Si $A \cap B = \emptyset$, on dit que A et B sont disjoints, on dit aussi en langage probabiliste incompatible

Réunion

La réunion de A et B , notée $A \cup B$, est l'ensemble des éléments x tels que $x \in A$ ou $x \in B$, soit:

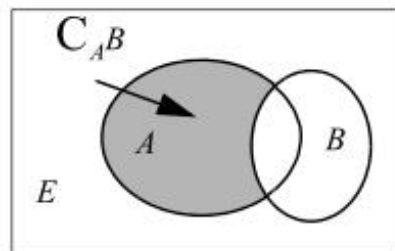
$$A \cup B = \{x : x \in A \text{ ou } x \in B\}.$$

le terme « ou » est employé au sens, $x \in A$ ou $x \in B$, si x appartient à A (uniquement), ou x appartient à B (uniquement), ou à la fois à A et à B .

Complémentaire

Le complémentaire de A , noté CA ou \bar{A} , est l'ensemble des éléments de E qui n'appartiennent pas à A :

$$CA = \{x : x \notin A\}$$

Différence

La différence entre A et B , ou complémentaire de B relatif à A , noté $A - B$, ou C_A^B , est l'ensemble des éléments de A qui n'appartiennent pas à B :

$$A - B = \{x : x \in A \text{ et } x \notin B\}$$

1.1.4 Propriétés des opérations sur les ensembles**Loi idempotente**

$$A \cup A = A; \quad A \cap A = A$$

Loi associative

$$(A \cup B) \cup C = A \cup (B \cup C); \quad (A \cap B) \cap C = A \cap (B \cap C)$$

Loi commutative

$$A \cup B = B \cup A; \quad A \cap B = B \cap A$$

Loi de distributivité

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C); \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

Loi d'identité

$$\begin{aligned} A \cup \emptyset &= A; & A \cap \emptyset &= \emptyset \\ A \cup E &= E; & A \cap E &= A \end{aligned}$$

Loi de complémentarité

$$\begin{aligned} A \cup C^c &= E; & A \cap C^c &= \emptyset \\ C^c C &= A; & C^c E &= \emptyset; & C^c \emptyset &= E \end{aligned}$$

Loi de De Morgan

$$C^c (A \cup B) = C^c A \cap C^c B; \quad C^c (A \cap B) = C^c A \cup C^c B$$

1.2 Ensembles particuliers**1.2.1 Ensemble fini**

Un ensemble E est fini s'il est vide ou s'il contient un nombre fini d'éléments, sinon il est infini.

Exemple. L'ensemble E des cinq lettres a, b, c, d, e est fini

Exemple. l'ensemble des habitants de l'algérie est un ensemble fini (même s'il est relativement grand plus de trente millions).

1.2.2 Ensemble dénombrable

– Un ensemble est dénombrable s'il existe une bijection entre une partie de l'ensemble des entiers naturels \mathbb{N} et cet ensemble. On a trois genres d'ensemble dénombrable :

- a- L'ensemble vide \emptyset est dénombrable par convention.
- b- Un ensemble fini A , s'écrit évidemment sous la forme $\{a_1, a_2, \dots, a_n\}$, n étant un élément de \mathbb{N} .
- c- Un ensemble dénombrable non fini A est un ensemble qui s'écrit sous la forme $\{a_1, a_2, \dots\}$.

L'ensemble des entiers naturels \mathbb{N} est évidemment un ensemble dénombrable non fini.

1.2.3 Ensemble infini non dénombrable.

Un ensemble A est non dénombrable s'il n'existe pas de bijection entre une partie de \mathbb{N} et cet ensemble. Intuitivement çà signifie qu'il est plus grand que \mathbb{N} .

Exemple. Les ensembles infinis non dénombrables qu'on rencontrera dans notre contexte sont surtout, les intervalles de \mathbb{R} .

1.2.4 Ensembles produits

L'ensemble produit de deux ensembles A et B est l'ensemble noté $A \times B$ formé des couples ordonnés (a, b) où a est un élément de A et b est un élément de B :

$$A \times B = \{(a, b) : a \in A, b \in B\}$$

Exemple 1. Si $A = \{1, 2, d\}$ et $B = \{a, d\}$, alors

$$A \times B = \{(1, a), (1, d), (2, a), (2, d), (d, a), (d, d)\}.$$

Exemple 2. Si $A = B = \mathbb{R}$, alors

$$A \times B = \mathbb{R}^2 = \{(x, y) : x \in \mathbb{R} \text{ et } y \in \mathbb{R}\}$$

1.2.5 Famille d'ensembles

Il peut arriver qu'on considère un ensemble dont les éléments sont eux mêmes des ensembles. Un tel ensemble est appelé une famille d'ensembles. On représente généralement une famille d'ensembles en fixant un ensemble I appelé ensemble d'indices et on associe à chaque indice $i \in I$, un élément de la famille qu'on note A_i , la famille d'ensembles est alors représentée par $A_i, i \in I$. Parties d'un ensemble

Soit A un ensemble quelconque. On note l'ensemble des parties de l'ensemble A , par $P(A)$.

Exemple 1. Si $A = \{a, b, d\}$ alors

$$P(A) = \{\emptyset, \{a\}, \{b\}, \{d\}, \{a, b\}, \{a, d\}, \{b, d\}, \{a, b, d\}\}$$

Union et intersection infinies

Soient $A_i, i \in I$ une famille infinie d'ensembles.

a- Union L'union des éléments de la famille $A_i, i \in I$, noté $\bigcup_{i \in I} A_i$, est l'ensemble des éléments qui appartiennent au moins à un certain A_i :

$$\bigcup_{i \in I} A_i = \{x : \exists i \in I, x \in A_i\}.$$

Si l'ensemble des indices est une partie $I = \{n \in \mathbb{N} : n \geq k\}$ de l'ensemble \mathbb{N} des entiers naturels on utilise aussi la notation $\bigcup_{i=k}^{\infty} A_i$ ou $\bigcup_{i \geq k} A_i$.

Exemple 1. Soient pour tout entier naturel n , l'intervalle $A_n = [0, n[$. On a $\bigcup_{n=1}^{\infty} A_n = \mathbb{R}$.

Exemple 2. Soient pour tout entier naturel $k \geq 1$, l'intervalle $B_k = [\frac{1}{k}, k[$. on peut vérifier que $\bigcup_{k \geq 1} B_k = \mathbb{R}^* =]0, +\infty[$.

b- Intersection l'intersection des éléments de la famille $A_i, i \in I$, est l'ensemble des éléments qui appartiennent à tout élément A_i de la famille d'ensembles en question.

Exemples. Reprenons les ensembles $A_n = [0, n[$ et $B_k = [\frac{1}{k}, k[$, considérés précédemment.

a- $\bigcap_{n=1}^{\infty} A_n = [0, 1[$, car $A_1 = [0, 1[$ est contenu dans tous les autres ensembles $A_n, n > 1$.

$$b- \bigcap_{n=1}^{\infty} B_n = \emptyset, \text{ car } B_1 = [1, 1[= \emptyset.$$

1.3 Partition d'un ensemble

Une partition d'un ensemble A est une subdivision de l'ensemble A . Rigoureusement parlant, une partition de A est une famille de parties de A telle que l'union des éléments de cette famille est égale à A , et l'intersection de deux éléments (quelconques) différents de cette famille est égale à l'ensemble vide.

Exemple 1. Soient $A = \{a, b, c, 4\}$, $A_1 = \{a\}$, $A_2 = \{a, b\}$, $A_3 = \{c, 4\}$, $A_4 = \{b, c\}$, $A_5 = \{4\}$.

A_1, A_2, A_4 est une famille de parties de A , qui ne forme pas une partition de A , car $A_1 \cap A_2 = \{a\} \neq \emptyset$.

A_2, A_5 est une famille de parties de A , qui n'est pas une partition de A , car $A_2 \cup A_5 = \{a, b, 4\} \neq A$.

A_2, A_3 est une partition de A .

Exemple 2. Soient A l'ensemble des êtres humains, A_1 l'ensemble des êtres humains de sexe masculin, A_2 l'ensemble des êtres humains de sexe féminin.

La famille A_1, A_2 est une partition de A .

Chapitre 2

Sommation de nombres et de fonctions.

2.1 Le symbole \sum .

2.1.1 Cas fini

Si a_1, a_2, \dots, a_n est une suite finie de nombres, on désigne la somme $a_1 + a_2 + \dots + a_n$ par $\sum_{i=1}^n a_i$:

$$\sum_{i=1}^n a_i = a_1 + a_2 + \dots + a_n. \quad (2.1)$$

Exemple 1.

$$\begin{aligned} \sum_{i=1}^5 \frac{1}{2^i} &= \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \frac{1}{2^5} \\ &= \frac{1}{2} \frac{1 - \left(\frac{1}{2}\right)^5}{\frac{1}{2}} \text{ (somme des cinq premiers termes d'une suite géométrique)} \\ &= 1 - \left(\frac{1}{2}\right)^5 \end{aligned}$$

10 CHAPITRE 2 SOMMATION DE NOMBRES ET DE FONCTIONS.

Exemple 2.

$$\sum_{i=1}^n 2^i = 2^1 + 2^2 + \dots + 2^n.$$

2.1.2 Cas infini

Si a_1, a_2, \dots est une suite infinie de nombres, on désigne par $\sum_{i=1}^{\infty} a_i$ la limite de la suite $S_n = \sum_{i=1}^n a_i$:

$$\sum_{i=1}^{+\infty} a_i = \lim_n \sum_{i=1}^n a_i.$$

Exemple .

$$\begin{aligned} \sum_{i=1}^{+\infty} \frac{1}{2^i} &= \lim_n \left[\frac{1}{2^1} + \frac{1}{2^2} + \dots + \frac{1}{2^n} \right] \\ &= \lim_n \frac{1 - \left(\frac{1}{2}\right)^{n+1}}{1 - \frac{1}{2}} \quad (\text{somme des } n \text{ premiers termes d'une suite géométrique}) \\ &= 1. \end{aligned}$$

2.2 Le symbole \int .

2.2.1 Cas de fonction à valeurs positives.

Soit $f(x)$ une fonction à valeurs réelles non négatives ($f(x) \geq 0$) définie sur un intervalle $[a, b]$. On désigne par $\int_a^b f(x) dx$, l'air sous le graphe de la fonction $f(x)$ entre les droites $x = a$ et $x = b$.

2.2.2 Exemple

$\int_1^3 (2x + 1) dx$ est l'air d'un trapèze. Cette aire est égale à $\int_1^3 (2x + 1) dx = \left(\frac{7+3}{2} \times 2\right) = 10$

2.2.3 Cas de fonction à valeurs non positives.

Si $f(x) \leq 0$, $\int_a^b f(x) dx$ désigne l'air au dessus du graphe de la fonction $f(x)$, située entre les deux droites $y = a$ et $y = b$, avec le signe $-$.

2.2.4 Exemple

$\int_1^3 (-2x - 1) dx$ est l'air du même trapèze avec le signe $-$. on a donc

$$\int_1^3 (-2x - 1) dx = -10.$$

2.2.5 Cas général.

Si $f(x)$ a des valeurs positives et aussi des valeurs négatives sur l'intervalle $[a, b]$, on divise l'intervalle $[a, b]$ en n sous intervalles $[a, a_2[$, $[a_2, a_3[$, ... $[a_n, b[$ de sorte que le signe de la fonction $f(x)$ est, soit \leq ou \geq sur chaque sous intervalle de la subdivision. L'intégrale de $f(x)$ sur l'intervalle $[a, b]$ est la somme des integrales de $f(x)$ sur chaque sous intervalle de la subdivision.

2.2.6 Exemple

Soit la fonction $f(x)$ définie sur l'intervalle $[-2, 7]$ comme suit

$$f(x) = \begin{cases} \frac{x}{2}; & \text{pour } -2 \leq x < 2 \\ -x, & \text{pour } 2 \leq x < 4 \\ x + b, & \text{pour } 4 \leq x \leq 7 \end{cases}$$

On a

$$\begin{aligned} \int_{-2}^7 f(x) dx &= \int_{-2}^0 f(x) dx + \int_0^3 f(x) dx + \int_3^5 f(x) dx + \int_5^7 f(x) dx \\ &= -\frac{2}{2} + \frac{3 \times 1}{2} - \frac{2 \times 1}{2} + \frac{2 \times 2}{2} = 1,5 \end{aligned}$$

2.2.7 Intégrale sur un intervalle non borné.

Soit une fonction $f(x)$ définie sur l'intervalle $[a, +\infty[$. Si la suite $S_n = \int_a^n f(x) dx$, $n > a$, converge vers une limite $S \in R$, On appelle cette limite l'intégrale indéfinie de la fonction $f(x)$ sur l'intervalle $[a, +\infty[$ et on note

$$\int_a^{+\infty} f(x) dx = \lim_n \int_a^n f(x) dx. \quad (2.2)$$

12CHAPITRE 2 SOMMATION DE NOMBRES ET DE FONCTIONS.

On définit de la même façon les intégrales indéfinies

$$\int_{-\infty}^b f(x) dx = \lim_n \int_{-n}^{+\infty} f(x) dx, (n \in \mathbb{N}, -n < b) \quad (2.3)$$

$$\int_{-\infty}^{+\infty} f(x) dx = \lim_n \int_{-n}^0 f(x) dx + \lim_n \int_0^n f(x) dx, (n \in \mathbb{N}, n \geq 1) \quad (2.4)$$

2.2.8 Exemple

Considérons la fonction $f(x)$ définie sur l'intervalle $[1, +\infty[$ comme suit.

$$f(x) = \begin{cases} \frac{1}{2^1}, & \text{pour } 1 \leq x < 2, \\ \frac{1}{2^2}, & \text{pour } 2 \leq x < 3, \\ \dots & \\ \frac{1}{2^n}, & \text{pour } n \leq x < n+1. \end{cases}$$

$f(x)$ est constante sur tout intervalle $[n, n+1[$, $n = 1, 2, \dots$

$$\begin{aligned} \int_1^{+\infty} f(x) dx &= \lim_n \int_1^n f(x) dx \\ &= \lim_n \sum_{k=1}^n \frac{1}{2^k} \\ &= \lim_n \frac{1}{2} \frac{1 - \left(\frac{1}{2}\right)^n}{1 - \frac{1}{2}} \\ &= 1. \end{aligned}$$

2.3 Primitive d'une fonction.

Soit une fonction $f(x)$ définie un intervalle $I = [a, b]$, ($a < b$, l'un où les deux pouvant être infinis), ou définie sur l'union d'intervalles.

On dit qu'une fonction $F(x)$, définie sur le même ensemble, est une primitive de la fonction $f(x)$, si la dérivée $F'(x)$ de la fonction F au point $x \in I$, est égale à $f(x)$, valeur de f au point x .

2.3.1 Notation

Si une fonction $F(x)$ est une primitive de la fonction $f(x)$, on écrit

$$F(x) = \int f(x)dx.$$

2.3.2 Remarque

Faites attention! Il faut bien distinguer entre l'intégrale définie $\int_a^b f(x)dx$ et l'intégrale indéfinie $\int f(x)dx$. L'intégrale définie $\int_a^b f(x)dx$ est un nombre lié à l'intervalle $[a, b]$. L'intégrale indéfinie est une fonction qui à tout nombre $x \in [a, b]$ associe un nombre noté usuellement par $\int f(x)dx$.

2.3.3 Exemples.

1-

$$f(x) = x^2 + 1,$$

$$F(x) = \frac{x^3}{3} + x$$

est une primitive de $f(x)$. Et, en général si C est un nombre quelconque la fonction $\frac{x^3}{3} + x + C$ est une primitive de la même fonction

$$f(x) = x^2 + 1.$$

2- Si n est un entier différent de -1 , alors

$$\int x^n dx = \frac{1}{n+1} x^{n+1}$$

3-

$$\int \sin x dx = \cos x + C.$$

4- Si $a \neq 0$,

$$\int \exp(ax) dx = \frac{1}{a} \exp(ax) + C.$$

2.3.4 Propriétés des primitives

1- En ajoutant un nombre quelconque à une primitive d'une fonction, on obtient une autre primitive de la même fonction : Si $F(x) = \int f(x)dx$ et C est un nombre, alors $F(x) + C = \int f(x)dx$.

2- Linéarité :

$$\int (f(x) + g(x)) dx = \int f(x)dx + \int g(x)dx, \text{ et} \quad (2.5)$$

$$\int \lambda f(x)dx = \lambda \int f(x)dx \text{ (pour tout nombre } \lambda)$$

2.3.5 Primitive et calcul d'intégrales

a- Si elle existe, alors la fonction $F(x) = \int_{-\infty}^x f(t)dt$ est une primitive de la fonction $f(x)$.

Ici x est fixé, c'est la borne supérieure de l'intervalle $]-\infty, x[$ sur laquelle on calcule l'intégrale, t est une lettre muette, elle n'a pas de sens toute seule.

b- Si une fonction $F(x)$ est une primitive de la fonction $f(x)$, alors on la formule suivante pour calculer l'intégrale définie de $f(x)$ sur l'intervalle $[a, b]$.

$$\int_a^b f(x) dx = F(b) - F(a) \quad (2.6)$$

Notation Dans les calculs, on exprime la quantité $F(b) - F(a)$ par $[F(x)]_a^b$.

2.3.6 Exemples

1-

$$\begin{aligned} \int_0^2 (x^2 + 1) dx &= [x^3 + x]_0^2 = (2^3 + 2) - (0^3 + 0) \\ &= 6 - 0 = 6. \end{aligned}$$

2-

$$\begin{aligned}\int_0^{\frac{\pi}{2}} \sin x dx &= [\cos x]_0^{\frac{\pi}{2}} \\ &= \cos \frac{\pi}{2} - \cos 0 \\ &= 0 - 1 = -1.\end{aligned}$$

3-

$$\begin{aligned}\int_1^5 \text{Exp}(-2x) dx &= \left[-\frac{1}{2} \text{Exp}(-2x) \right]_1^5 \\ &= \left(-\frac{1}{2} \text{Exp}(-2 \times 5) \right) - \left(-\frac{1}{2} \text{Exp}(-2 \times 1) \right) \\ &= -\frac{1}{2} (e^{-10} - e^{-2}).\end{aligned}$$

Remarque La formule (2.6) est valable même dans le cas où l'une ou les deux bornes a et b sont infinies, par exemple

$$\begin{aligned}\int_a^{+\infty} f(x) dx &= F(+\infty) - F(a). \\ &= \left(\lim_n F(n) \right) - F(a).\end{aligned}\tag{2.7}$$

2.3.7 propriétés de l'intégrale.

soit $f(x)$ et $g(x)$ deux fonctions à valeurs réelles, λ un nombre réel.

Linéarité a- $\int_a^b \lambda f(x) dx = \lambda \int_a^b f(x) dx$

b- $\int_a^b (f(x) + g(x)) dx = \int_a^b f(x) dx + \int_a^b g(x) dx.$

Exemple 1. $\int_{-1}^1 (-x^3 + 2x^2 - x^5) dx = -\int_{-1}^1 x^3 dx + 2 \int_{-1}^1 x^2 dx - \int_{-1}^1 x^5 dx =$
 $= -\left[\frac{x^4}{4} \right]_{-1}^{+1} + 2 \left[\frac{x^3}{3} \right]_{-1}^{+1} - \left[\frac{x^6}{6} \right]_{-1}^{+1} =$
 $= -\left(\frac{1^4}{4} - \frac{(-1)^4}{4} \right) + 2 \left(\frac{1^3}{3} - \frac{(-1)^3}{3} \right) - \left(\frac{1^6}{6} - \frac{(-1)^6}{6} \right) =$
 $= -(0) + 2 \left(\frac{2}{3} \right) - (0) = \frac{4}{3}.$

16 CHAPITRE 2 SOMMATION DE NOMBRES ET DE FONCTIONS.

Exemple 2. $\int_{-\pi}^{\pi} (-2 \cos x + e^{-x} - x^2) dx = -2 \int_{-\pi}^{\pi} (\cos x) dx + \int_{-\pi}^{\pi} e^{-x} dx - \int_{-\pi}^{\pi} x^2 dx =$

$$= -2 [-\sin x]_{-\pi}^{\pi} + [-e^{-x}]_{-\pi}^{\pi} - \left[\frac{1}{3}x^3\right]_{-\pi}^{\pi} =$$

$$= -2 \{(-\sin \pi) - (-\sin(-\pi))\} + \{(-e^{\pi}) - (-e^{-\pi})\} - \left\{\frac{1}{3}\pi^3 - \frac{1}{3}(-\pi)^3\right\} =$$

$$= -e^{\pi} + e^{-\pi} - \frac{2}{3}\pi^3.$$

Relation de Chasles. Si c est un nombre quelconque et $f(x)$ est définie sur les trois intervalles $[a, c]$, $[c, b]$ et $[a, b]$, alors on a la relation suivante dite relation de chales.

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx. \quad (2.8)$$

Deuxième partie

Statistiques descriptives

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 la 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, la deuxième étant un prolongement de la première :

Statistique descriptive, a pour but de recueillir les données et les résumer de façon synthétique et efficace. Elle utilise pour cela des représentations de données sous forme de graphiques, de tableaux et d'indicateurs numériques (par exemple des moyennes). Elle permet de dégager les caractéristiques essentielles du phénomène étudié et de suggérer des hypothèses pour une étude ultérieure plus sophistiquée.

Statistique inférentielle Le rôle de la statistique inférentielle 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.

La statistique descriptive est dite univariée ou " à une dimension ", lorsqu'il s'agit de l'étude d'un seul caractère, sinon elle est dite multidimensionnelle.

Chapitre 3

Statistique unidimensionnelle

3.1 Définition

3.1.1 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.

3.1.2 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. Les techniques utilisées pour assurer cette représentativité s'appellent "théorie de l'échantillonnage".

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, repartition selon le sexe, selon les tranches d'âges, selon le niveau d'instruction...)

3.1.3 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

3.2 Natures des variables statistiques et échelles de mesures.

Les données qualitatives définissent des échelles soit nominales soit ordinales. L'échelle nominale comporte un certain nombre de catégories, dont la seule propriété est qu'elles sont toutes différentes les unes des autres (sexe, nationalité, type de diplôme, etc.). Dans le cas d'une échelle ordinale, en revanche, les catégories qui la composent sont munies d'une structure d'ordre, établie en fonction d'un critère donné (de moins à plus "quelque chose" : origine sociale, opinion plus ou moins favorable, stade de développement d'une maladie,...).

Les données quantitatives se réfèrent à des variables d'intervalle ou de rapport. Dans le premier cas, on dispose d'échelles au sein desquelles la comparaison d'intervalles est possible (il est possible de déterminer si deux intervalles sont ou ne sont pas de même étendue). Dans le deuxième cas, l'échelle permet non seulement la comparaison d'intervalles, mais également la comparaison de rapports (il est possible de déterminer si deux rapports sont ou ne sont pas égaux). La différence fondamentale entre ces deux types d'échelles est liée au statut de la valeur nulle : sur une échelle d'intervalle, le zéro est situé de manière arbitraire, comme pour la mesure des températures

3.2 NATURES DES VARIABLES STATISTIQUES ET ÉCHELLES DE MESURES. 23

par exemple (échelles Celsius et Fahrenheit). Sur une échelle de rapport, en revanche, le zéro a une signification précise, puisqu'il désigne l'absence du caractère considéré (âge, salaire, taille, vitesse, etc.).

3.2.1 Exemples.

Caractère quantitatif discret

On s'intéresse au nombre de modules obtenus par chaque étudiant de la première année universitaire écoulée de la faculté de médecine. La population statistique est l'ensemble des étudiants de la première année de la faculté de médecine, un étudiant de la première année est une unité statistique, le caractère statistique étudié est le nombre de modules obtenus, c'est un caractère quantitatif discret.

Caractère quantitatif continu

Si au lieu du nombre de modules obtenus, on s'intéresse à la taille de chaque étudiant, alors le caractère statistique "la taille" est un caractère quantitatif continu.

Caractère qualitatif nominal

Si au lieu du nombre de modules obtenus, on s'intéresse à la couleur des yeux de chaque étudiant, alors le caractère statistique "couleur des yeux" est un caractère qualitatif nominal.

Caractère qualitatif ordinal

Si on considère le stade de la maladie (niveau de gravité de la maladie) chez un ensemble de cancéreux, alors on a affaire à un caractère statistique ordinal.

3.2.2 Echelles de mesures : Recapitulation par des exemples

Variables qualitatives nominales, **échelle nominale**, exemples : le nom des journaux, le métier ou le nom.

Variables qualitatives ordinales, **échelle ordinale**, exemples : le rang, stade d'une maladie.

Variables quantitatives discrètes, **échelle de rapport**, exemples : nombre d'enfants, nombre d'habitants.

Variables quantitatives discrètes, **échelle d'intervalle**, exemples : date de naissance, heure d'arrivée

Variables quantitatives continues, **échelle de rapport**, exemples : distance, durée.

Variables quantitatives continues, **échelle d'intervalle**, exemple : température.

3.3 La variabilité et l'incertain en biologie

3.3.1 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 de complication d'un état pathologique, et 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 :

$$\text{variabilité totale} = \text{variabilité biologique} + \text{variabilité métrologique.}$$

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.

$$\text{variabilité biologique} = \text{variabilité intra-individuelle} + \text{variabilité inter-individuelle.}$$

La variabilité intra-individuelle peut être observée lors de la mesure de la performance d'un athlète qui n'est pas capable des mêmes performances à chaque essai, mais qui se différencie des autres athlètes (variabilité inter-individuelle). En général, la variabilité intra est moindre que la variabilité inter.

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

variabilité métrologique = variabilité expérimentale + variabilité appareil de mesure

3.3.2 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. Dans sa démarche, le médecin doit en particulier, prendre en compte la variabilité naturelle des grandeurs qu'il manipule, pour distinguer ce qui est normal de ce qui est pathologique (décision à propos d'un patient) et pour évaluer la qualité d'un nouvel examen, ou d'une nouvelle thérapeutique (décision thérapeutique). La compréhension des méthodes statistiques, de leur puissance et de leurs limites, est essentielle pour un médecin de nos jours. Tout résultat de recherche médicale résulte d'une expérimentation (clinique ou biologique) qui s'appuie sur une méthodologie statistique rigoureuse, et dont les résultats sont analysés en termes statistiques.

3.4 Représentation des données

Il y a trois principales façons de représenter les données concernant un caractère statistique : représentation brute (ou en vrac), dans un tableau statistique et la représentation graphique.

On développe dans ce qui suit, par des exemples, les différentes façons de représenter des données, selon la nature du caractère statistique.

3.4.1 Caractère qualitatif

Pour un caractère qualitatif, on présente les données dans un tableau à plusieurs lignes. Dans la première ligne, on reporte les différentes modalités, et dans la deuxième ligne les effectifs correspondants. On peut également ajouter d'autres lignes, par exemple une ligne pour les fréquences et une autre pour les pourcentages correspondants.

Il y a plusieurs façons de représenter graphiquement un caractère qualitatif, mais les plus utilisées sont les diagrammes en secteurs circulaires ou les diagrammes en bandes.

Exemple

On reprend l'exemple de la couleur des yeux et on suppose que le nombre des étudiants est 200 et qu'il y a quatre couleurs différentes.

Les données sont représentées dans un tableau statistique simple

Couleur des yeux : x_i	noire	noisette	bleu	verte	Total
effectif : n_i	120	30	30	20	200

Ce tableau peut être développé, en y ajoutant, par exemple, des lignes pour les fréquences relatives et les pourcentages relatifs. On obtient alors le tableau statistique suivant.

Couleur des yeux : x_i	noire	noisette	bleue	verte	Total
effectif : n_i	120	30	30	20	200
fréquences relatives	$\frac{120}{200}$	$\frac{30}{200}$	$\frac{30}{200}$	$\frac{20}{200}$	1
pourcentages relatifs	60	15	15	10	100

La représentation en bandes (en TD)

La représentation en secteurs circulaires (en TD)

3.4.2 Caractère quantitatif discret

Dans le cas d'un caractère discret, on collecte la valeur du caractère chez chaque individu de la population et on représente ces valeurs de deux façons :

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

La représentation des données groupées par valeurs consiste à représenter les données dans un tableau avec au moins deux lignes, dans la première ligne on inscrit les valeurs et dans la deuxième les effectifs correspondants. Parfois on ajoute d'autres lignes, par exemple pour les effectifs cummulés, les fréquences,...

La représentation graphique se fait par le diagramme en bâtons (**en TD**).

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.

Cette suite est une représentation brute de la série des nombres d'enfants des vingt familles.

La représentation groupée par valeurs est

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 cummulés (croissants) et une ligne pour les fréquences. On obtient alors 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
effectifs cummulés ($n_i cum$)	3	5	8	12	14	17	20	
fréquences (f_i)	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{4}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{3}{20}$	1

On peut également ajouter d'autres lignes pour y faire des calculs concernant le calcul de paramètres comme la moyenne m ou la variance σ^2 .

La représentation des données groupées par valeurs peut aussi se faire par une suite de couples $(x_1; n_1)$, $(x_2; n_2)$, ..., $(x_k; n_k)$, où les x_i sont les valeurs de la série et les n_i les effectifs. Néanmoins la représentation dans un tableau est la plus commode, soit pour les calculs, soit pour éviter les subtilités.

3.4.3 Caractère quantitatif représenté en classes de valeurs.

Exemple

Les notes en mathématiques des élèves d'une classe de vingt élèves sont comme suit :

les notes de 4 élèves sont ≥ 6 et < 9 , les notes de 7 élèves ≥ 9 et < 12 , les notes de 6 élèves sont ≥ 12 et < 15 , et 3 élèves avaient une note ≥ 15 et < 18 .

La représentation de la série des notes de ces élèves en classes de valeurs d'amplitude 3 est

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

La représentation graphique d'un caractère, dont les valeurs sont représentées en classes et non en valeurs discrètes, se fait par un histogramme (**en TD**). 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).

3.5 Les différents types de tableaux statistiques

Il y a trois types de tableaux en statistiques.

3.5.1 Tableaux de données.

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

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.

3.5 LES DIFFÉRENTS TYPES DE TABLEAUX STATISTIQUES 29

N°	sexe		français			moyenne générale			
	M	F	A	B	C	[7; 10[[10; 13[[13; 16[[16; 19[
1	x			x		x			
2		x		x			x		
3	x		x		x		x		
4		x			x			x	
5	x				x		x		
6		x		x				x	
7		x	x				x		
8	x		x				x		
9		x		x			x		
10	x		x					x	
11		x	x						x
12		x		x			x		
13		x		x				x	
14		x	x			x			
15	x			x					x

dans ce tableau, on a utilisé les conventions suivantes.

Pour le sexe, M = sexe masculin, F = sexe féminin.

Pour la langue française, A = bon, B = moyen, C= faible.

3.5.2 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.

moyennes générales (x_i)	[7; 10[[10; 13[[13; 16[[16; 19[total
effectifs (n_i)	2	7	4	2	15

3.5.3 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

3.5.4 Utilisation des effectifs cummulés.

La ligne des effectifs cummulés dans un tableau statistique est utilisée pour trouver la valeur d'une observation dont on fixe le numéro : x_1 est la valeur des observations numéro 1, 2, ..., n_1cum ,

x_2 est la valeur des observations numéro $n_1cum + 1, n_1cum + 2, \dots, n_2cum$,

...

x_r est la valeur des observations numéro $n_{r-1}cum + 1, n_{r-1}cum + 2, \dots, n_rcum$.

3.5 LES DIFFÉRENTS TYPES DE TABLEAUX STATISTIQUES 31

Exemple.

Les données concernant le nombre d'enfants pour 200 familles sont présentées dans le tableau suivant

nombres d'enfants x_i	3	4	5	6	Total
effectif n_i de familles ayant x_i enfants	120	30	30	20	200

On ajoute à ce tableau une ligne pour les effectifs cummulés, on obtient alors le tableau statistique suivant.

nombre d'enfants : x_i	3	4	5	6	Total
effectif : n_i	120	30	30	20	200
effectifs cummulés	120	150	180	200	

Dans cet exemple, on a

Le nombre des valeurs (ou modalités) est $r = 4$

Le nombre total d'observations ou l'effectif total est

$$\begin{aligned} N &= \sum_{i=1}^4 n_i \\ &= 120 + 30 + 30 + 20 \\ &= 200. \end{aligned}$$

Les quatre modalités sont $x_1 = 3$, $x_2 = 4$, $x_3 = 5$, et $x_4 = 6$.

$$n_{1cum} = n_1 = 120,$$

$$n_{2cum} = n_{1cum} + n_2 = 120 + 30 = 150,$$

$$n_{3cum} = n_{2cum} + n_3 = 150 + 30 = 180,$$

$$n_{4cum} = n_{3cum} + n_4 = 180 + 20 = 200.$$

le numéro 61 est compris entre 1 et $n_{1cum} = 120$, donc la valeur de l'observation numéro 61 est égale à 3.

La valeur de l'observation numéro 140 est égale à 4, car 140 est dans l'intervalle des entiers $[n_{1cum} + 1; n_{2cum}[= [121; 150[$.

3.6 Caractéristiques de tendance centrale

3.6.1 La moyenne arithmétique.

Définition

La moyenne arithmétique noté \bar{x} ou m , d'une série statistique simple représentée en données groupées en valeurs " $(x_1, n_1), (x_2, n_2), \dots, (x_p, n_p)$ " est donnée par la formule

$$\bar{x} = \frac{\sum_{i=1}^p n_i x_i}{\sum_{i=1}^p n_i}. \quad (3.1)$$

Ici, p est le nombre des valeurs de la série statistique X , n_i est le nombre de répétition de la valeur x_i .

Remarque.

Si chaque valeur apparait une seul fois, ou si on répète dans le numérateur chaque x_i , n_i fois, la formule (3.1) devient

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N}, \quad (3.2)$$

où N est le nombre total des termes de la série statistique, c'est-à-dire le nombre d'observations.

Exemple.

Soit les deux séries statistiques

$$\begin{aligned} X &: 1, 5; -1; 2; -1; 1, 5; 2; 3; 1, 5; 1, 5; 1; -1 \\ Y &: 3; 2; 4; 1; 1; -2; 8; 7; 0; 10; -3 \end{aligned}$$

Les deux séries statistiques sont exprimées en représentation brute.

Pour la série X , puisque plusieurs valeurs de la série se répète, il vaut mieux utiliser la représentation des données groupées par valeurs.

x_i	-1	1	2	3	1.5
n_i	3	1	2	1	4

Pour la série Y , on ordonne les valeurs dans un ordre non décroissant,

$$Y : -3; -2; 0; 1; 2; 3; 4; 7; 10.$$

La moyenne arithmétique de X est donnée par la formule

$$\begin{aligned} \bar{x} &= \frac{\sum_{i=1}^5 n_i x_i}{\sum_{i=1}^5 n_i} \\ &= \frac{(3 \times -1) + (1 \times 1) + (2 \times 2) + (1 \times 3) + (4 \times 1.5)}{3 + 1 + 2 + 1 + 4} \\ &= \frac{11}{11} = 1. \end{aligned}$$

La formule (3.2) s'écrit ici

$$\begin{aligned} \bar{x} &= \frac{\sum_{i=1}^N x_i}{N} \\ &= \frac{(-1) + (-1) + (-1) + 1 + 2 + 2 + 3 + 1.5 + 1.5 + 1.5 + 1.5}{11} \\ &= \frac{11}{11} = 1. \end{aligned}$$

Pour calculer la moyenne \bar{y} de la série Y , puisque chaque valeur de la série apparait uniquement une seule fois, on utilise la formule (3.2) pour avoir

$$\begin{aligned} \bar{y} &= \frac{\sum_{i=1}^N y_i}{N} \\ &= \frac{(-3) + (-2) + 0 + 1 + 2 + 3 + 4 + 7 + 10}{9} \\ &= \frac{22}{9} = 2.44. \end{aligned}$$

Cas de données groupées en classes de valeurs. Si la série est représentée en classes de valeurs, on accepte une approximation de la moyenne par la même formule (3.1), où

$$x_i = \frac{e_{i-1} + e_i}{2}$$

est le centre de la classe $[e_{i-1}, e_i[$.

3.6.2 Exemple

Soit une série statistique X , dont les données sont représentées par le tableau suivant, où C_i représente la classe (i.e. l'intervalle) $[e_{i-1}, e_i[$

c_i	$[-2; 0[$	$[0; 2[$	$[2; 4[$	$[4; 6[$	$[6; 8[$
n_i	3	2	2	1	3

Pour calculer la moyenne on ajoute une ligne dans laquelle on exprime les centres de classe, qu'on note x_i et on utilise la formule (3.1).

C_i	$[-2; 0[$	$[0; 2[$	$[2; 4[$	$[4; 6[$	$[6; 8[$
n_i	3	2	2	1	3
x_i	-1	1	3	5	7

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^5 n_i x_i}{\sum_{i=1}^5 n_i} \\ &= \frac{(3 \times -1) + (2 \times 1) + (2 \times 3) + (1 \times 5) + (3 \times 7)}{3 + 2 + 2 + 1 + 3} \\ &= \frac{31}{11}.\end{aligned}$$

3.6.3 La médiane

Définition. La médiane d'une série est une valeur qui partage cette série, préalablement ordonnée, en deux sous séries ayant le même effectif. La médiane peut être une valeur de la série ou non.

Calcul de la médiane.

1)- Cas de série représentée par valeurs discrètes Si l'effectif N est impair la médiane M_e est égale à l'observation numéro $\frac{N-1}{2} + 1$:

$$M_e = \left(\frac{N-1}{2} + 1 \right)^{\text{ème}} \text{ obs.} \quad (3.3)$$

Si l'effectif est pair la médiane M_e est la moyenne des deux observations numéro $\frac{N}{2}$ et $\frac{N}{2} + 1$:

$$M_e = \frac{1}{2} \left[\left(\frac{N}{2} \right)^{\text{ème}} \text{ obs.} + \left(\frac{N}{2} + 1 \right)^{\text{ème}} \text{ obs.} \right] \quad (3.4)$$

3.6 CARACTÉRISTIQUES DE TENDANCE CENTRALE 35

2)- **Cas de série représentée par classes de valeurs** On définit d'abord la classe médiane.

Définition de la classe médiane. La classe médiane est la classe contenant l'observation numéro $\frac{N}{2}$ si l'effectif total N est pair, ou l'observation numéro $\frac{N+1}{2}$ si N est impair.

La médiane est un nombre qu'on calcule par La méthode d'interpolation linéaire. Cette méthode suppose que les observations de la classe médiane sont réparties d'une façon uniforme.

Définition de la valeur médiane La médiane M_e est définie par

$$M_e = e_{i-1} + k \frac{\frac{N}{2} - n_{i-1}cum}{n_i cum - n_{i-1}cum}, \quad (3.5)$$

où

e_{i-1} est la borne inférieure de la classe mediane,

k est l'amplitude de la classe médiane.

$n_{i-1}cum$ est l'effectif cummulé de la classe qui précède la classe médiane.

$n_i cum$ est l'effectif cummulé de la classe médiane.

On détermine la classe médiane de la façon suivante.

On ajoute une ligne pour les effectifs cummulés, pour l'utiliser dans la détermination de l'observation numéro $\frac{N}{2}$ ou $\frac{N+1}{2}$.

3.6.4 Exemple

On considère la série suivante, représentée en classes de valeurs.

c_i	$[-2; 0[$	$[0; 2[$	$[2; 4[$	$[4; 6[$	$[6; 8[$
n_i	3	2	2	1	3

Pour déterminer la médiane on commence par ajouter au tableau une ligne des effectifs cummulés :

c_i	$[-2; 0[$	$[0; 2[$	$[2; 4[$	$[4; 6[$	$[6; 8[$
n_i	3	2	2	1	3
$n_i cum$	3	5	7	8	11

L'effectif total est $N = 11$, c'est un entier impair.

$$\frac{N+1}{2} = \frac{11+1}{2} = 6$$

La classe médiane est celle contenant l'observation numéro 6, la classe médiane est donc $[2; 4[$.

La médiane M_e par interpolation linéaire est

$$\begin{aligned} M_e &= e_{i-1} + k \frac{\frac{N}{2} - n_{i-1}cum}{n_i cum - n_{i-1}cum} \\ &= 2 + 2 \frac{5.5 - 5}{7 - 5} = 2.5. \end{aligned}$$

3.6.5 Le mode

Définition Le mode d'une série statistique simple est la valeur qui a le plus grand effectif. Une série peut avoir plusieurs modes.

Cas des données groupées en classes de valeurs Contrairement au cas des données discrètes, dans le cas des données groupées en classes de valeurs, le mode n'est pas toujours l'une des valeurs de la série, on se satisfait dans cette situation à définir la classe modale qui est la classe ayant le plus grand effectif.

3.6.6 Exemple.

Considérons l'exemple

c_i	$[-2; 0[$	$[0; 2[$	$[2; 4[$	$[4; 6[$	$[6; 8[$
n_i	3	2	5	1	3

La classe modale est $[2; 4[$.