ENICARTHAGE
المدرسة الوطنية للمهندسين بقرطاج
Ecole Nationale d'Ingénieurs de Carthage

Introduction aux Statistiques

Enseignants :Aloui.M/Bacha.I/Debbech.A **AU :2019/2020**

Classe :1Ing-Info .

1 Définition

La statistique est une branche mathématique qui a pour but dans un premier temps de rassembler une série de données et de les présenter (Statistique descriptive).

Dans un deuxième temps, ces données sont interpretées afin d'entirer des conclusions et d'effectuer des précisions éventuelles (Statistique inférentielle).

2 Applications

L'interprétation et l'utilisation des données statistique se retrouve dans de trés nombreux domaines dont notamment,

✓ Les sciences humaines

✓ Les sciences économiques

✓ Les média

✓ La gestion des entreprises

✓ La recherche médicale

✓ La météorologie

3 Variables statistiques

Au départ, on se base généralement sur les résultats d'une enquête. La variable (ou le caractètre)étudiée dans l'enquête peut être de différents types

1)Données ou variables chronologiques : ce sont des variables dont on analyse l'évolution de leurs valeurs en fonction du temps.

Exemples :

✓Chiffres d'affaires d'une société au cours des années

✓ Recencement 1966 à 2014

2)Données ou variables non chronologiques : Ce sont dont on n'étudie pas l'évolution en fonction du temps.

Exemples :

Tailles, poids, vote, notes d'un certain d'un ensemble de personnes à un moment donné.

Dans ce cas, la variable (ou le caractère) étudiée lors de l'enquête peut être qualitative, quantitative, discrète ou continue.

a)Variables qualitatives :

Une variable qualitative est une variable qui exprime une qualité et une variable dont les valeurs ne sont ni mesurable, ni repérables (Ce ne sont pas des nombres).

Les valeurs prise par une valeur qualitative sont appelées des "modalités" qui porte des noms.

a) C'est la raison pour laquelle on parle aussi de variables **nominales**

Exemple :

Couleurs, profession, marques de voitures.

b)Les modalités d'une variable qualitative sont parfois ordonnés. Dans ce cas, on parle de variables ordonale.

Exemples : Les grades au examens (Passable, Mention assez bien, Mention bien, Mention trés bien)

Les degrés d'une brulure premier, deuxième, troisième degré.

c)Les variables binaires : il s'agit d'un type particulier de variables qualitatives qu'on peut également appeler variables dichotomiques.Ces variables ne peuvent pas être hiérarchisées et elles ne possèdent que deux modalités.

Exemple. Sexe :homme, femme

2)variables quantitatives Une variable quantitative est une variable dont les valeurs sont mesurables ou repérables par des nombres.

Exemple : Salaire, température, taille, poids, ...

a)les variables quantitatives discrètes :elles ne peuvent prendre que des valeurs numériques entières par la façon que le passage d'une valeur à une autre est discontinu.

Exemple :

Nombre de frères et soeurs (il est impossible d'avoir 3.4 frères)

Tailles de cent personnes .

b) les variables quantitatives continues : Une variable quantitative continue est une variable qui peut prendre toutes les valeurs possibles dans un intervalle (un grand nombre de valeurs possibles).

Exemples :

Taile d'une population, poids.

Résumé : $Variables \begin{cases} \checkmark Quantitatives & Discretes \\ & continues \\ \checkmark Qualitatives & nominales \\ & Ordinales \\ & Binaires \\ \checkmark Temporelles & Date \\ & Heure \end{cases}$

3.1 *La collècte des données*

Deux méthodes consistent à étudier un caractère statistique comme par exemple le recencement d'une population

✓ La première consiste à étudier le caractère sur la population.

✓ La deuxième consiste à étudier le caractère sur un ensemble d'individus qui représentent une population.Cette méthode est appelée échantillonnage.L'ensemble d'individus représentatifs est appelé échantillon.

Remarques

1) La première solution est plus précise si elle est traitée dans un délai tel qu'on ne change pas la population.

2) La première solution est plus couteuse.

3) La deuxième solution est plus rapide.

Types d'échantillonages

On a deux types d'échantillonages

✓ Un échantillonage non probabiliste

Ces méthodes sont aussi appelées méthode empiriques ou par choix raisonné.

Le choix des individus qui représente l'échantillon dans ce cas n'obéit pas au hasard.Cette méthode est définie selon des critères de faisabilité, de ressemblance à la population-cible et de critères subjectifs dépendants du choix des enquêteurs.

✓ Un échantillonage probabiliste

Dans ce cas tous les individus de l'échantillon sont choisis au hasard parmi la population donc admettent une probabilité non nulle d'être sélectionnés pour faire partie de l'échantillon.

L'enquêteur n'intervient pas sur le choix de l'échantillon.

Les informations recueillis sur l'échantillon peuvent être inférées à ce moment là sur la population source.

1)Vocabulaires

✓ Population :C'est l'ensemble étudié

✓ Individu :C'est un élément de la population

✓ Effectif total :C'est le nombre total d'individus

✓ Caractère :C'est la propriété étudiée .On distingue les caractères discrets qui ne peuvent prendre qu'un nombre fini de valeurs(Notes à un devoir,...) et les caractères continus dont on regroupe les valeurs par intervalles (Tailles, durée d'écoutes...)

2)Série statistique associée à un caractère discret :

2.1)Classement des données :

Définition : On appelle série statistique la donnée simultanée (dans un Tableau)des valeurs du caractère étudié notée x_i rangées dans l'ordre croissant et des effectifs notés n_i de ces valeurs.

Remarque

A la place des effectifs n_i, on peut utiliser les fréquences $f_i = \frac{n_i}{N}$ où (N représente l'effectif total) ou les fréquences en pourcentages $f_i = \frac{n_i}{N}100$

Exemple 1

Les notes sur 20 obtenues, lors d'un devoir de mathématique dans une classe sont les suivantes.

10 8 11 9 12 10 8 7 9 10 11 12 10 8 9 10 11 10 9 10

.

la population étudiée est la classe et les individus sont les élèves. L'effectif total est égal à 20 et la note obtenue au devoir est le caractère discret que l'on étudie.

La série statistique définie par les effectifs est la suivante :

Valeurs des caracteres x_i	7	8	9	10	11	12
Effectifs n_i (nombre d'élèves ayant la note x_i)	1	3	4	7	3	2

La série statistique définie par les fréquences en pourcentage est la suivante :

Valeurs des caracteres x_i	7	8	9	10	11	12
Fréquence en pourcentage $f_i = \frac{n_i}{N}100$	$\frac{1}{20}100$	$\frac{3}{20}100$	$\frac{4}{20}100$	$\frac{7}{20}100$	$\frac{3}{20}100$	$\frac{2}{20}100$
	5%	15%	20%	35%	15%	10%

n_i L'effectif de x_i N L'effectif Total $= 20$

2.2)Effectifs cumulés :

Définition : L'effectif cumulé croissant d'une valeur x est la somme des efféctifs des valeurs $y/\ y \leq x$.

L'effectif cumulé décroissant d'une valeur x est la somme des efféctifs des valeurs $y/\ y \geq x$.

Exemple La série statistique (x_i, n_i et effectifs cumulés $\nearrow$) est donnée par :

Valeurs des caracteres x_i	7	8	9	10	11	12
Effectifs n_i	1	3	4	7	3	2
Effectifs cumulés $\nearrow$	1	$1+3=4$	4+4= 8	8+7= 15	15+3=18	18+2= 20

On commence par la première valeur de n_i donc 4 élèves leurs notes ≤ 8

La série statistique (x_i, n_i et effectifs cumulés $\searrow$) est donnée par :

Valeurs des caracteres x_i	7	8	9	10	11	12
Effectifs n_i	1	3	4	7	3	2
Effectifs cumulés $\searrow$	19+1= 20	16+3= 19	12+4=16	5+7=12	5=2+3	2

On commence par la dernière valeur de n_i donc 4 élèves leurs notes ≤ 8

Déterminer le nombre d'élèves ayant une note ≥ 8 (effectifs cumulés décroissant il faut faire la somme des y tq $y \geq 8$ donc 19 est le nombre d'élèves ayant une note ≥ 8

Déterminer le nombre d'élèves ayant une note ≤ 8 (effectifs cumulés croissant il faut faire la somme des y tq $y \leq 8$ donc 4 est le nombre d'élèves ayant une note ≤ 8

Remarque : Le raisonnement est le même pour déterminer la fréquence cumulée croissante et décroissante

2.3)Représentation graphique :

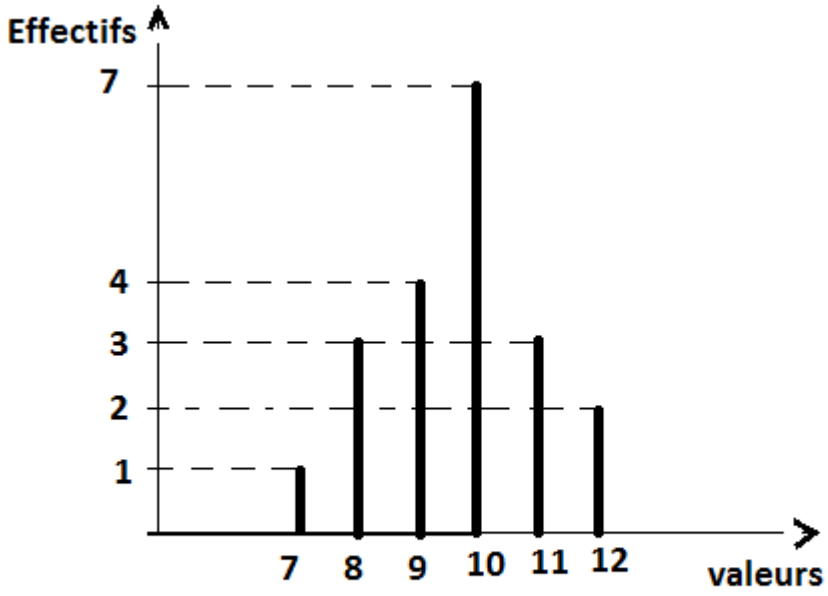
Pour les caractères quantitatifs discrets on utilise le diagramme en bâton.

Dans un repère orthogonal pour chaque valeur de la série discrète on trace un trait verticale dont la hauteur est proportionnelle à l'effectif (dans l'unité choisi).

Exemple

Revenons à la série statistique précédente.

Valeurs des caracteres x_i	7	8	9	10	11	12
Effectifs n_i (nombre d'élèves ayant la note x_i)	1	3	4	7	3	2



2.4 Paramètres de positions :

a)La moyenne :

Définition :

On appelle moyenne d'une série statistique d'effectif total $N = n_1 + n_2 + n_3 + ... + n_k$ le réel $\overline{x} = \dfrac{n_1x_1 + n_2x_2 + n_3x_3 + ... + n_kx_k}{N}$

Exemple

Valeurs des caracteres x_i	7	8	9	10	11	12
Effectifs n_i (nombre d'élèves ayant la note x_i)	1	3	4	7	3	2

$$\overline{x} = \frac{1\times 7 + 3\times 8 + 4\times 9 + 7\times 10 + 3\times 11 + 2\times 12}{20} = 9.7$$

Remarque :

Si on utilise les fréquences $(f_i = \frac{n_i}{N})$. $\overline{x} = f_1x_1 + f_2x_2 + ... + f_kx_k$

Propriétés :

✓ Si on ajoute à toutes les valeurs d'une série statistique le même nombre b, on augmente la moyenne de cette série de b.

✓ Si les valeurs d'une série sont multipliées ou divisées par un même nombre $a(\neq 0)$.La moyenne de cette série est aussi multipliée ou divisée par a.

✓ Si une population d'effectif N composée d'une partie d'effectif N_1 et de moyenne $\overline{x_1}$ et d'une autre partie d'effectif N_2 et de moyenne $\overline{x_2}$, alors la moyenne de la population Totale est tel que $\overline{x} = \dfrac{N_1\overline{x_1} + N_2\overline{x_2}}{N}$

Exemple :

Si dans une classe les 15 garçons mesurent en moyenne 182 cm et les 20 filles mesurent en moyenne 168 cm alors la taille moyenne d'un élève de cette classe est : $\dfrac{15\times 182 + 20\times 168}{15+20} =$ 174 cm **b)La médiane :**

Définition :

L'idée générale est que la médiane est une valeur du caractère qui partage la population en

deux parties de même effectif.De façon plus précise on appelle médiane d'une série statistique discrète toute valeur M du caractère tel qu'au moins 50% des individus aient une valeurdu caractère $\leq M$ et au moins 50% des individus aient une valeurdu caractère $\geq M$.

Recherche pratique de la médiane :

On range les valeurs du caractère.une par une dans l'ordre croissant.Chaque valeur du caractère doit apparaître un nombre de fois égal à l'effectif correspondant.

✓ Si l'effectif total N est impair, la médiane M est la valeur du caractère situé au milieu, autrement dit si N est impair : M est la valeur du caractère (x_i) de rang $\dfrac{N+1}{2}$

✓ Si l'effectif total N est pair, la médiane M est la demi somme des valeurs de caractère situées au milieu.autrement dit si l'effectif Total N est pair la médiane est la moyenne (demi somme) des valeurs de x_i de rang $\dfrac{N}{2}$ et $\dfrac{N}{2}+1$

Exemple :

x_i	6	8	9	12	13	17
Effectifs n_i	3	1	2	1	3	3

L'effectif Total $N = 13$ impair donc la médiane est la valeur de x_i correspondant à $n_i = \dfrac{13+1}{2} = 7$

On commence à partir de la première valeur de n_i

Dans la série statistique,en passant de la première colonne à la deuxième colonne.

On fait la somme des n_i on trouve 4 de la deuxième à la troixième colonne .On somme les n_i $4+2=6$ (c'est toujours < 7).On passe à la colonne suivante et on fait la somme des n_i $6+1=7$ par suite la médiane $M = 12$

2.4) Les paramètres de dispersion :

Ces paramètres permettent de mesurer la façon dont les valeurs du caractère sont réparties autour de la moyenne et de la médiane.

Les paramètres de dispersion associée à la médiane :

Définition : L'idée générale est de partager la population en 4 parties de même effectif.

étant donnée une série statistique de médiane M, dont la liste est rangée dans l'ordre croissant.En coupant la liste en deux sous séries de même effectif (attention quand l'effectif Total est impair, la médiane ne doit être incluse dans les sous séries).

On appelle premier quartile ,le réel noté Q_1 qui est égal à la médiane de la sous série inférieure.

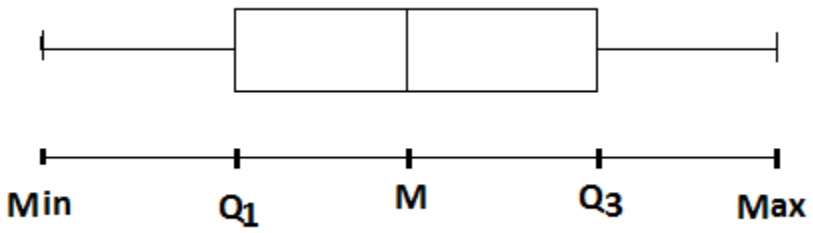
On appelle troisième quartile le réel noté Q_3 égal à la médiane de la sous série supérieure.

L'écart interquartile $Q_3 - Q_1$.

$]Q_1, Q_3[$ est appelé intervalle interquartile .

Définition : Le diagramme en boîte d'une série statistique se construit alors de la façon suivante :

(Les valeurs du caractère sont en abscisse min et max représentant les valeurs minimales et maximales du caractère)



Interprétation :

✓ 25% de la population admet une valeur du caractère entre Min et Q_1.

✓ 25% de la population admet une valeur du caractère entre Q_1 et M.

✓ 25% de la population admet une valeur du caractère entre M et Q_3.

✓ 25% de la population admet une valeur du caractère entre Q_3 et Max.

Exemple :

Soit la série statistique suivante :

Valeurs des caracteres x_i	7	8	9	10	11	14	16
Effectifs n_i	2	1	1	1	2	1	2

	Sous série inférieure	Sous série supérieure

$N = 10$, N est pair .M est la demi Somme des x_i ayant pour rang $\frac{N}{2}$ et $\frac{N}{2}+1$

x_i ayant pour rang 5 $(2+1+1+1)$ $(x_i = 10)$

x_i ayant pour rang $(2+1+1+1+2 \geq 6)$ $(x_i = 11)$

Donc $M = \frac{10+11}{2} = 10.5$

Donc $M = 10.5$ divise la série en deux sous séries.

L'effectif total de la sous série inférieure $= 5$ qui est impair.

Q_1 est la valeur de x_i ayant pour rang $\frac{N+1}{2}$

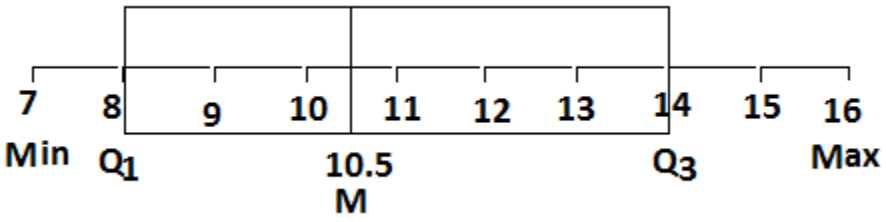
Q_1 a pour rang 3 $Q_1 = 8$

L'effectif total de la sous série supérieure $= 5$ qui est impair.

Q_3 est la valeur de x_i ayant pour rang 3

$Q_1 = 14$

Le diagramme en boite est donné par :



Exemple :

Valeurs des caracteres x_i	6	8	9	12	13	17
Effectifs n_i	3	1	2	1	3	3

	Sous série inférieure	$M = 12$	Sous série supérieure

$N = 3+1+2+1++3+3 = 13$

M est de rang $\dfrac{N+1}{2} = \dfrac{13+1}{2} = 7$

$(3+1+2++1 = 7)$ donc $M = 12$

L'effectif de chaque sous suite est pair $= 6$

Donc Q_1 est la demisomme des deux valeurs ayant pour rang $\dfrac{N}{2}$ et $\dfrac{N}{2}+1$ de la sous série inférieure.

x_i de rang $\dfrac{N}{2} = 3$ donc $x_i = 6$

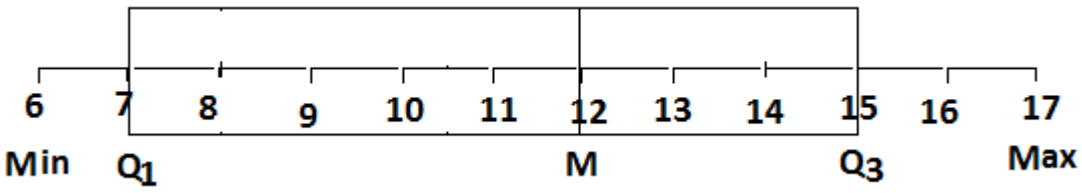
x_i de rang $\dfrac{N}{2}+1 = 4$ donc $x_i = 8$

$Q_1 = \dfrac{6+8}{2} = 7$

Q_3 est la demi somme des deux valeurs de la sous série supérieure ayant pour rang $(\dfrac{N}{2} = 3$ et $(\dfrac{N}{2}+1 = 4$

$Q_3 = \dfrac{13+17}{2} = 15$

Le diagramme en boite de la série est le suivant :


6
7
8
9
10
11
12
13
14
15
16
17
Min
Q_1
M
Q_3
Max