I / Statistiques descriptives : Generalités

I) Déf et Objet

La statistique est la science qui a pour objet de recueillir un ensemble de données relatives à un certain phénomène et d'exploiter cette information par l'analyse et expliquer le phénomène étudié

II / Notions de population, caractère.

Déf 1 : Une population est un ensemble fini, noté I $(card(I)<\infty)$
un élément de la population est dit unité ou individu.
Le cardinal de I noté $n<\infty$ est dit taille de la population

Déf 2
Un caractère (ou variable) est une application définie de de I dans un ensemble E.

On note : $X : I \longrightarrow E$
$\qquad\qquad w_i \longmapsto X(w_i) = x_i$

$I = \{w_1, \dots, x_n\}$

$x_1, \dots, x_n$ est dit echantillon ou suite d'observations.
les differentes observations de la suite $x_1, \dots, x_n$ sont dites modalite

Rq : selon la nature de E, une variable peut etre quantitative ($E \subset \mathbb{R}$ mesurable) ou qualitative (non mesurable)
si $E \subset \mathbb{Z}$, X est dite variable quantitative discrete (denombrable)
si $E = [a, b]$ X est dite " " continue (non denombrable)
La variable qualitative peut être ordinal (on peut ordonner les m…) ou nominal (cas contraire)

Exp
Soit une population I formée de 10 étudiants de 1ère année MI
on considère les 4 variables suivantes :

$I = \{\omega_1, \dots, \omega_{10}\}$ taille de la population est $n = 10$

a/ $X_1 : I \longrightarrow E_1 = \{F, M\}$ sexe de l'etudiant

X_1 qualitative nominale, les modalites F et M

b) $X_2 : I \longrightarrow E_2 = \{TB, B, AB, S.M\}$ mension obtenue au bac

X_2 qualitative ordinale, les modalite TB, B, AB et S.M.

c/ $X_3 : I \longrightarrow E_3 = \{0, 1, 2, 3\}$ nombre de frères et soeurs.

X_3 : quantitative discrète

d/ $X_4 : I \longrightarrow E_4 = [155 \quad 175]$ taille en cm.

X_4 : quantitative continue.

Déf3 : Effectif et fréquence.

Soit l'échantillon de taille n, $x_1', \dots, x_n''$

on note par $x_1, \dots, x_p$ les modalités

- l'effectif d'une modalite x_i est le nombre d'observations de la suite $x_1', \dots, x_n'$ égales à x_i. On note n_i
- la fréquence de x_i est le rapport noté $f_i = \frac{n_i}{n}$

Rqs

- $1 \leq n_i \leq n$ $\qquad \sum_{i=1}^{p} n_i = n$.
- $0 \leq f_i \leq 1$ $\qquad \sum_{i=1}^{p} f_i = 1$

Déf4 : Tableau statistique ou distribution statistique.

$X : I = \{\omega_1, \dots, \omega_n\} \longrightarrow E = \{x_1, \dots, x_p\}$

La distribution statistique de X est donnée par le tableau suivant

Modalité	x_1	x_2	— —	x_p
Effectif n_i	n_1	n_2	- — —	n_p
Freq f_i	f_1	f_2	- — —	f_p

Déf5 : Mode

Le mode d'une distribution est la modalite x_i ayant la plus grande effectif n_i (ou frequence f_i)

II | Variable qualitative

I | Introduction

Soit une population I de taille n.

X une variable qualitative.

$x_1, \ldots, x_p$ le modalités de X.

On note par n_i l'effectif de x_i pour $i = 1, \ldots, p$

et par f_i la fréquence $\left(f_i = \frac{n_i}{n}\right)$

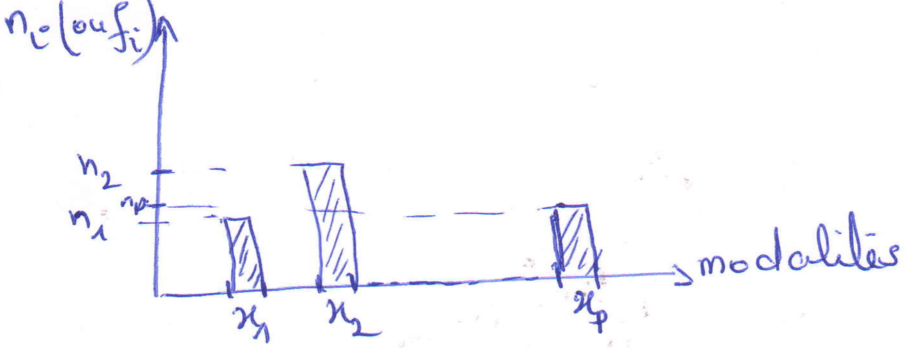
La distribution statistique

x_i	x_1	x_2	— — —	x_p
n_i	n_1	n_2		n_p
f_i	f_1	f_2		f_p

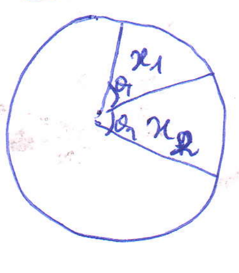
II | Representation graphique.

La variable qualitative peut être représentée graphiquement par deux graphe.

1| ~~Batons~~. Tuyau d'orgue.



2) Circulaire



θ_i est proportionnelle à f_i

$\sum_{i=1}^{p} \theta_i = 360° \longrightarrow \sum_{i=1}^{n} f_i = 1$

$\theta_i \longrightarrow f_i$

$\theta_i = 360° \times f_i$

Rqs

- Le diagramme en ~~bâtons~~ tuyaux d'orgue est souvent utilisé dans le cas de variable qualitative ordinale.
- Le mode correspond à x_i qui a le plus long ~~baton~~ tuyau

II / Variable ~~qualitative~~ quantitative discrète

I) Introduction

Soit une population de taille n.

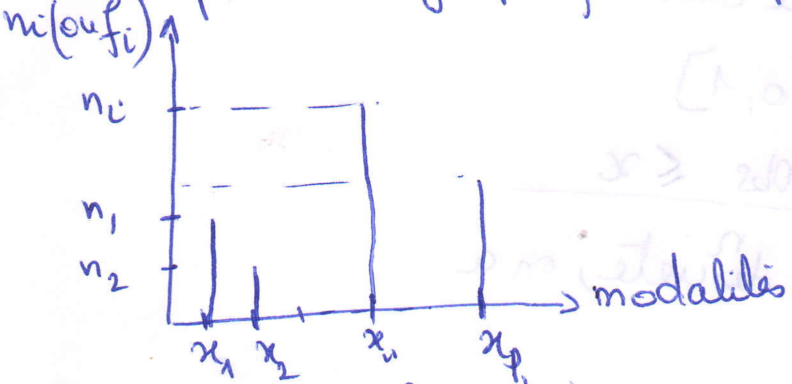
X une variable quantitative discrète : $x_1, \dots, x_p$ les modalités ordonnées dans l'ordre croissant $(x_1 < x_2 < \dots < x_p)$

Tableau statistique

x_i	x_1	— — — —	x_p
n_i	n_1		n_p
f_i	f_1		f_p

II) Représentation graphique.

X est représentée graphiquement par un diagramme en Batons



Rq : Respecter l'échelle sur les axes des x et y.
Le plus long baton correspond au mode.

III / Fréquences Cumulées

Défs

a) On appelle ~~fréquence~~ effectif cumulée croissante d'une modalité x_i et on note $N_i\uparrow$ le nombre d'observations $\leq x_i$: $N_i\uparrow = n_1 + n_2 + \dots + n_i$

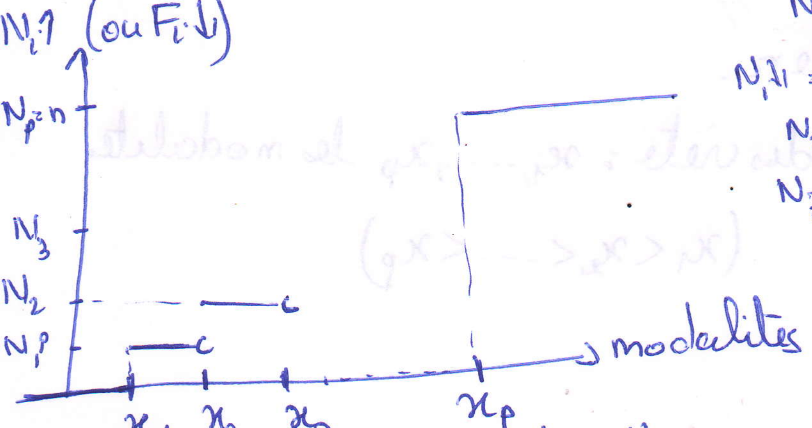
b) La fréquence cumulé croissante d'une modalité x_i, est le rapport $F_i\uparrow = \frac{N_i\uparrow}{n} = f_1 + \dots + f_i$

c) On appelle effectif cumulée decroissante d'une modalité x_i et on note $N_i\downarrow$, le nombre d'observations $\geq x_i$: $N_i\downarrow = n_i + n_{i+1} + \dots + n_p$

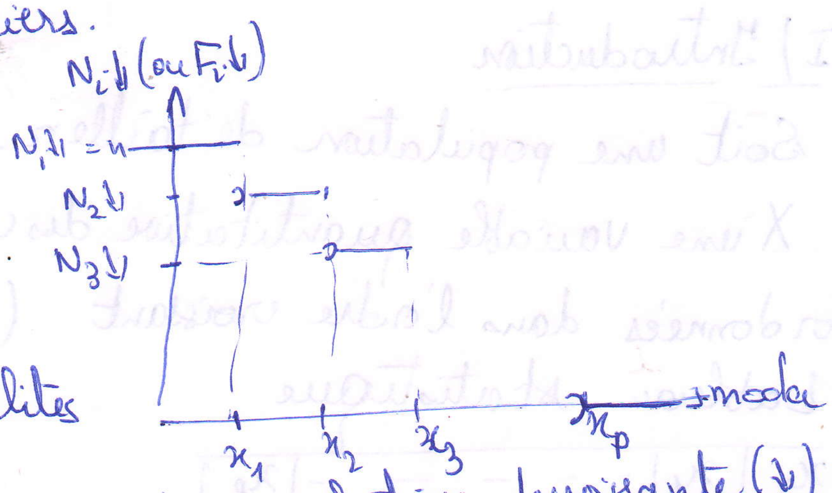
d) La fréquence cumulée decroissante de x_i, est le rapport $F_i\downarrow = \frac{N_i\downarrow}{n}$

IV / Courbes cumulatives

C'est la représentation graphique de effectifs (ou fréquences) cumulées. Courbes en escaliers.



Courbe cumulative croissante (↑)



Courbe cumulative décroissante (↓)

V | Fonction de répartition

Déf

La fonction de répartition d'une distribution est une application

$$F : \mathbb{R} \longrightarrow [0,1]$$

$$x \longmapsto \frac{N^{bre} \text{ d'Obs} \le x}{n}$$

Pour la variable quantitative discrète, on a

$x < x_1$ $\quad F(x) = 0$

$x_i \le x < x_{i+1}$ $\quad F(x) = F_i\uparrow \quad i = 1, \dots, p-1$

$x \ge x_p$ $\quad F(x) = 1$

VI | Paramètres de position et de dispersion

1/ Paramètres de position :

- Moyenne : $\bar{X} = \frac{1}{n} \sum_{i=1}^{p} n_i x_i = \sum_{i=1}^{p} f_i x_i$
- Médiane : si n est impaire $n = 2k+1$ $\quad Med = x_{k+1}$

  si n est paire $n = 2k$ $\quad Med = \frac{x_k + x_{k+1}}{2}$

La médiane Med est la valeur tq 50% des Observations sont $\le Med$

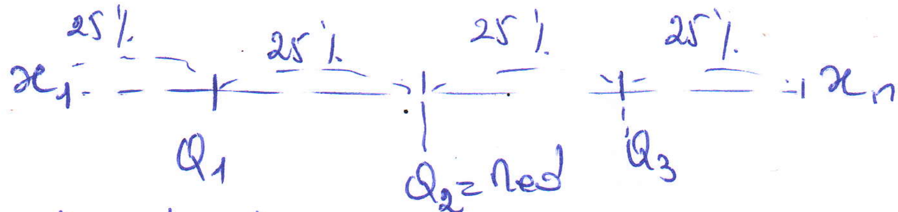
- Mode : noté Mo la modalité ayant la plus grande effectif

- 1^{er} quartile : noté Q_1.

Q_1 est la valeur tq 25% des observations sont $\leq Q_1$.

- $3^{ème}$ quartile : noté Q_3.

Q_3 est la valeur tq 75% des observations sont $\leq Q_3$.



2) Paramètre de dispersion.

- Étendue : $e = \max\limits_{1\leq i\leq n} x_i - \min\limits_{1\leq i\leq n} x_i = x_p - x_1$.

- Variance $\sigma^2 = Var(X) = \frac{1}{n}\sum\limits_{i=1}^{p} n_i(x_i - \bar{X})^2 = \sum\limits_{i=1}^{p} f_i(x_i - \bar{X})^2$

- Écart type $\sigma = \sqrt{Var(X)}$

- Écart interquartile $Q_3 - Q_1$

Proposition

On a les résultats suivants

- Si $y_i = a x_i + b$ $\quad \bar{Y} = a\bar{X} + b \quad (a, b \in \mathbb{R})$

$Var(Y) = a^2\, Var(X)$

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

- Les paramètres de dispersion sont positifs.