Chapitre 1

statistique unidimentionnelle

statistique (méthode scientifique) ⇒
- réunir les données
- organiser
- analyser
- résumer
- synthetiser l'information.

Population :

L'ensemble des éléments sur lesquels porte l'étude.

- humains - objets - villes - pays
- entreprise - logements - étudiants ....

Elément de la population :
- → individus
- → unité statistique

x_i	n_i.
x_p	n_{p1}

Caractère :

(ce qu'on veut étudier)

marque de voiture — nombre d'absence — notes — Taux

Modalité :

valeurs ou réponse possible prise par le caractère

- Renault
- Ford
- ⋮

0
1
2
⋮

Nature du caractère

Quantitatif
mesurable

disret
0, 1, 2, 3

continue
]-∞, +∞[
0,55 ; 1,5

Qualitatif
non mesurable

$F(x) = a_n \cdot x^n$

Représentation des données

- présentation brute
- tableau numérique
- graphiques
- résumé numérique

Fréquence absolues, relative et cumulées:

Effectif (Fréquence absolue) "n_i"

$$(N) \quad n = \sum_{i=1}^{p} n_i$$

Fréquence (Fréquence relative) f_i

$$f_i = \frac{n_i}{n_T}$$

Pourcentage % $= f_i \times 100$

Effectif cumulé croissant et f cumulé

$$n_i^c = \sum_{i=1}^{p} n_i \qquad f_i^c = \frac{n_i^c}{n_T}$$

x_p

n_j

Représentation graphique :

C. qualitatif

Diagramme à secteur circulaire

Comenbert

$$\theta_i = f_i \times 360^\circ$$

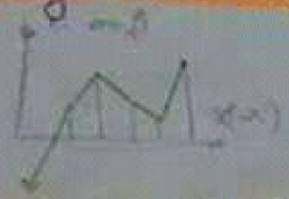
Diagramme en barre

- en bande
- Tuyau d'orgue

C. quantitatif

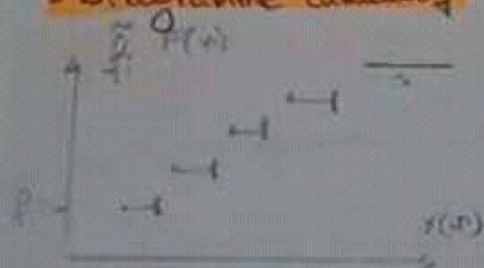
Discret

- Diagramme en bâtons



polygone des effectifs

- Diagramme cumulatif



$$F(x) = \begin{cases} 0 & \text{si } x < x_1 \\ \tilde{f}_i & x_i \leq x < x_{i+1} \\ 1 & \text{si } x \geq x_n \end{cases}$$

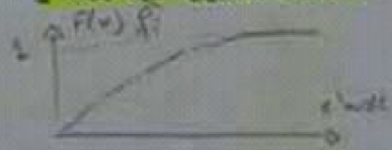
(Fonction de répartition)

Continue

- Histogramme

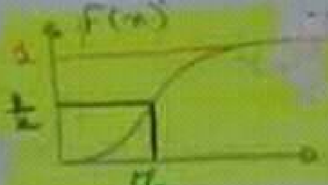
polygone des eff

- Courbe cumulative



$$F(x) = \begin{cases} 0 & \text{si } x < a_1 \\ \frac{x-a_1}{a_2-a_1} f_1 & \text{si } a_1 \leq x < a_2 \\ f_1 + \frac{x-a_2}{a_3-a_2} f_2 & \text{si } a_2 \leq x < a_3 \\ \sum_{j=1}^{i-1} f_j + \frac{x-a_i}{a_{i+1}-a_i} & a_i \leq x < a_{i+1} \\ 1 & \text{si } x \geq a_{p+1} \end{cases}$$

Médiane : (M_e)



La valeur statistique qui divise la série ordonnée en deux parts égales.

↓

Cas discret

$$M_e = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{si } n \text{ est impair} \\ \left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+1}{2}\right)} \right) & \text{si } n \text{ est pair} \end{cases}$$

$\frac{n}{2}$ = valeur ↓ n

Graphiquement => M_e => courbe (n_i dont l'ordonnée est $n/2$)

↓

Cas Continue

=> méthode d'interpolation

$$M_e = a_i + \frac{(a_{i+1} - a_i)}{n_i} \left(\frac{n}{2} - \tilde{n}_{i-1} \right)$$

$\tilde{n}_{i-1} < \frac{n}{2}$ $\qquad$ $\check{n}_i \geq \frac{n}{2}$ $\qquad$ $[a_i ; a_{i+1}[$ ⇓ classe médiane

Mode : (M_o)

Valeur du caractère la plus fréquente

Cas discret

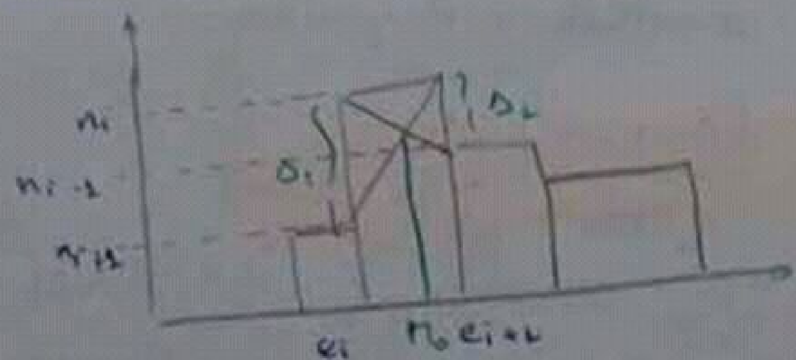
$x_i \Rightarrow n_i$ maximal

f_i maximal

$M_o = 6$

Cas continue :

$$M_o = e_i + \frac{\Delta_1}{\Delta_1 + \Delta_2}(e_{i+1} - e_i)$$



Serie entère

Quantiles :

Discret

$$q\alpha = \begin{cases} \dfrac{x_{[n\alpha]} + x_{[n\alpha+1]}}{2} & \text{si } n\alpha \in \mathbb{N} \\ x_{[n\alpha]+1} & \text{sinon} \end{cases}$$

Continue

$$q\alpha = a_i + \frac{(a_{i+1} - a_i)}{n_i}(n\alpha - \tilde{n}_{i-1})$$

$\alpha = \frac{1}{2}$ — $q\alpha$ = Me

$\alpha = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$ — $q\alpha$: 1er, 2ème, 3ème quartile

$\alpha = \frac{1}{10}, \frac{2}{10}, \dots, \frac{9}{10}$ — $q\alpha$: 1er, 2ème, ..., 9ème décile

→ centiles

$\alpha = \frac{1}{100}, \frac{2}{100}, \dots, \frac{99}{100}$ — $q\alpha$: 1er, 2ème, ..., 99ème

Moyenne :

Serie stat discrète

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{p} n_i x_i$$

n_T

$$\bar{X} = \sum_{i=1}^{p} f_i x_i \qquad f_i = \frac{n_i}{n_T}$$

Données brute

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{p} x_i$$

Serie stat continu

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{p} n_i c_i$$

$c_i \in [a_i ; a_{i+1}[$ $\qquad c_i = \frac{a_{i+1} + a_i}{2}$

Paramètre de dispersion :

Etendue

$E = x_{max} - x_{min}$

Ecart absolu moyen

$$E_{am} = \sum_{i=1}^{p} \frac{n_i |x_i - \bar{x}|}{n}$$

Variance

$$V(x) = \sigma^2 = \frac{1}{n} \sum_{i=1}^{p} n_i x_i^2 - \bar{x}^2$$

Ecart type

$$\sigma = \sqrt{V(x)}$$

Ecart interquantile

Définit par : q_α et $q_{1-\alpha}$

$0 < \alpha \leq \frac{1}{2}$

$I_\alpha = q_{1-\alpha} - q_\alpha$ $\quad q_{\frac{3}{4}}$ $q_{\frac{1}{4}}$

$(q_\alpha ; q_{1-\alpha}) \Rightarrow 100(1-2\alpha)\%$

- $q_{\frac{3}{4}} - q_{\frac{1}{4}}$ contient 50% des valeurs central
- $q_{\frac{9}{10}} - q_{\frac{1}{10}}$ " 80% " " "

Coefficient de variation

$$CV = \frac{\sigma}{\bar{X}}$$

($< 15\%$ en pratique

Chapitre 1
statistique unidimentonnelle

Chapitre 2
Statistique double

Etudier deux (2) caractère X et Y simultanément

Quantitatif	Qualitatif	Quantitatif
Quantitatif	Qualitatif	Qualitatif

Tableau de Contingence

$X \backslash Y$	y_1	y_2	y_j	...	y_K	$n_{i\cdot}$
x_1	n_{11}	n_{12}	n_{1j}	...	n_{1K}	$n_{1\cdot}$
x_2	n_{21}	n_{22}	n_{2j}	...	n_{2K}	$n_{2\cdot}$
x_i	n_{i1}	n_{i2}	n_{ij}	...	n_{iK}	$n_{i\cdot}$
x_p	n_{p1}	n_{p2}	n_{pj}	.	n_{pK}	$n_{p\cdot}$
$n_{\cdot j}$	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot j}$		$n_{\cdot K}$	n

n_{ij} : nombre d'individus qui possède en même temps les caractères x_i et y_j

Notation :

$$n_{i\cdot} = \sum_{j=1}^{k} n_{ij}$$

$$n_{\cdot j} = \sum_{i=1}^{p} n_{ij}$$

$$\sum_{j=1}^{k} n_{\cdot j} = n \qquad \sum_{i=1}^{p} n_{i\cdot} = n$$

Statistique marginale :

Serie $(x_i, n_{i\cdot})_{i=1,\dots,p}$ 1ère statistique marginale

" $(y_j, n_{\cdot j})_{j=1,\dots,p}$ 2ème statistique marginale

Fréquence

$$f_{i\cdot} = \frac{n_{i\cdot}}{n}$$ (caractère x_i)

$$f_{\cdot j} = \frac{n_{\cdot j}}{n}$$ (caractère y_j)

$$f_{ij} = \frac{n_{ij}}{n}$$ (caractère x_i et y_j)

Distribution conditionnelles :

* $(x_i, n_{ij}) \Rightarrow$ Serie stat Conditionnelle de X sachant que Y prend la valeur y_j

* $(y_j, n_{ij}) \Rightarrow$ Serie Stat Conditionnelle de Y sachant que X prend la valeur x_i

Exmpl :

x_i	4	8	10	14	18	tot
n_{i3}	0	3	25	12	1	$n_{\cdot 3} = 44$
f_{i3}	0	0,07	0,64	0,27	0,02	1

$$f_{i3} = \frac{n_{i3}}{n_{\cdot 3}}$$

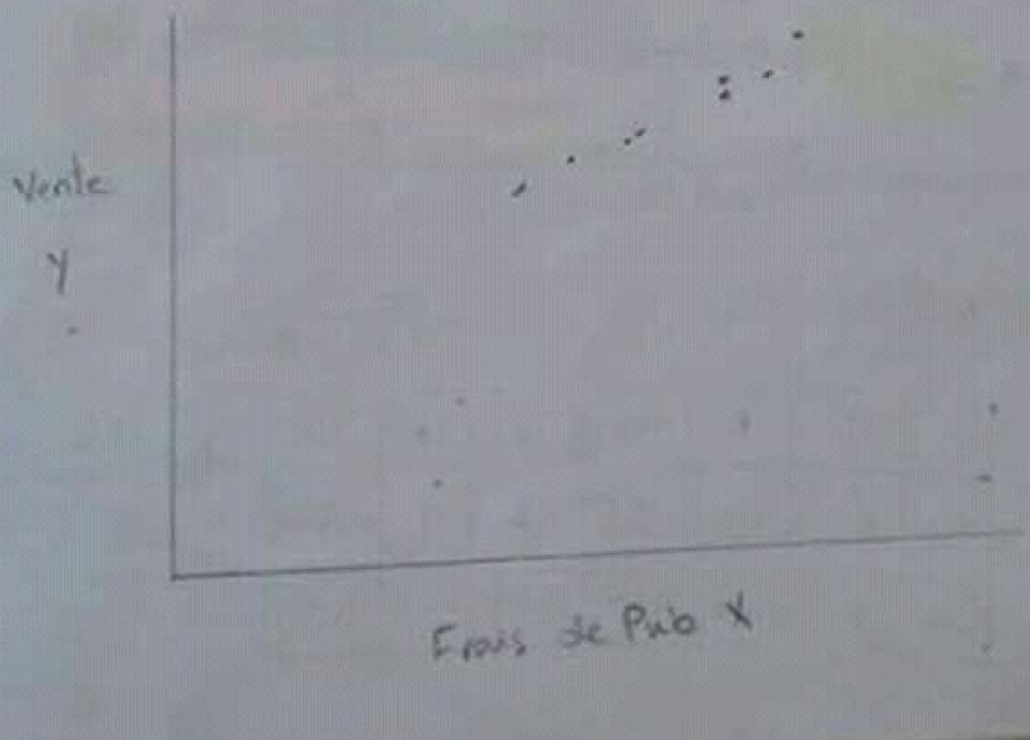
Indépendance entre 2 caractères :

$\boxed{f_{i\cdot} \times f_{\cdot j} = f_{ij}}$ ⇒ indépendantes

$\boxed{f_{i_0\cdot} \times f_{\cdot j_0} \neq f_{i_0 j_0}}$ ⇒ ne sont pas indépendantes

Nuage de point :

nuage de points Frais de Pub/Vente



Covariance entre deux Caractère : $Cov(x, y)$ ou $\sigma_{x,y}$

marginale

$$\boxed{\sigma_{xy} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}$$

$$\boxed{\sigma_{xy} = \left(\sum_{i=1}^{n} \frac{1}{n} x_i y_i \right) - \bar{x}\,\bar{y}}$$

Coefficient de corrélation linéaire :

Cov σ

$$r_{xy} = \frac{Cov(x, y)}{\sigma_x \sigma_y}$$

Propriété : $Cov(x, y) = Cov(y, x)$

$-1 \leq r_{xy} \leq 1$

x et y sont indépendants $\Rightarrow Cov(x, y) = 0$

$U = \alpha_1 X + \beta_1$ $V = \alpha_2 Y + \beta_2$

$Cov(U, V) = \alpha_1 \alpha_2 \, Cov(x, y)$ $r_{UV} = \frac{\alpha_1 \alpha_2}{|\alpha_1 \alpha_2|} r_{xy}$