Statistique Discreptive

Population => designe humain, objet, des villes
entreprise, pays.

Caractère statistique => designe une grandeur
ou un attribut

Modalité => des valeurs ou des reponces
par exemple : { Renaut, Ford, peujot {Marque de voiture}
{ 15 ; 11,25 ; 3 {note des etudiant}

Type de caractère ==> Caratère qualitatif

Caracterère quantitatif

- discret
  (nombre entier)
- continue
  (des intervale [,[ )

Représentation des données

Série statistique 2 1 0 3 2 2 4 4

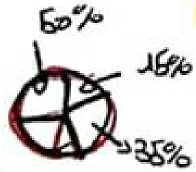
Tableaux statistique

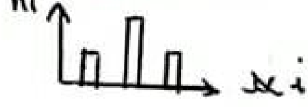
modalité x_i	0	1	2
n_i	1	1	2
$\tilde{n}_i$	1	$1+1$ 2	$2+2$ 4
f_i	1/4	1/4	2/4
$\tilde{f}_i$	1/4	2/4	4/4 = 1

=> (n)

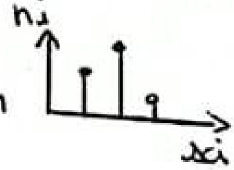
Représentation graphique

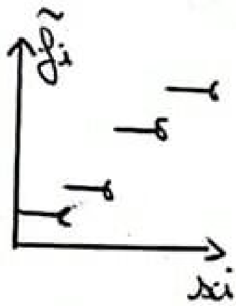
*** Caractère qualitatif :** $\theta_i = f_i \times 360$

- Secteur circulaire 

- Tuyaux d'orgue 

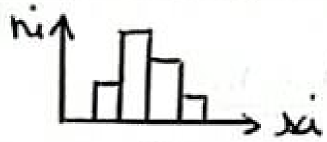
*** Caractère quantitatif discret :**

- Diagramme de baton 

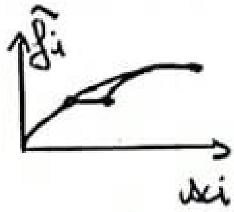
- Courbe cumulative
courbe en escalier 

tère quantitatif continue

- Histogramme



Courbe cumulative



Fonction de repartition (cas discret)

$$f(x) = \begin{cases} 0 & \text{si } x < x(1) \\ f_{i\,cum} & \text{si } x(i) \leq x < x(i+1) \\ & \text{pour } i = 1, \dots n-1 \\ \vdots & \\ 1 & \text{si } x \geq x(n) \end{cases}$$

Fonction de répartion (cas continue)

$$f(x) = \begin{cases} 0 & \text{si } x < a_i \\ \underbrace{\tilde{f}_{i-1}}_{} + \dfrac{x - a_i}{\text{amplitude}} \underbrace{f_i}_{} & \text{si } a_i \leq x \leq a_{i+1} \\ 1 & \text{si } x \geq a_{i+1} \end{cases}$$

frequence cumulé de classe precedente

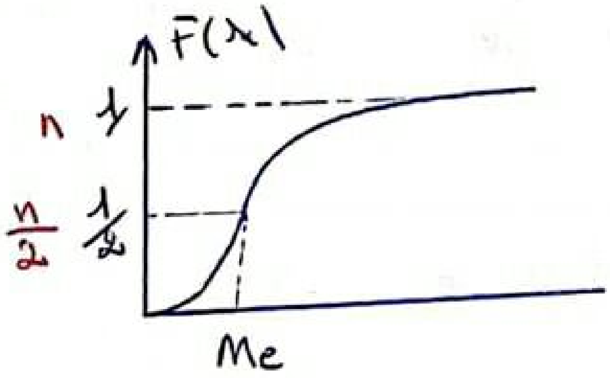
frequence de classe

Cas continue

$$Me = a_i + \frac{\frac{n}{2} - n_{i-1}\,cum}{n_i}\,(a_{i+1} - a_i)$$

La classe médiane $[a_i, a_{i+1}[$

graphiquement



Les quantile — **cas discret**

$$q_\alpha = \begin{cases} \dfrac{x_{n\alpha} + x_{n\alpha+1}}{2} & n\alpha \text{ entier} \\ x_{[n\alpha]+1} & n\alpha \notin \mathbb{N} \end{cases}$$

cas continue

La classe $[a_i, a_{i+1}[$

$$q_\alpha = a_i + \frac{(a_{i+1} - a_i)}{n_i}\,(n\alpha - n_{i-1}\,cum)$$

Remarque

* Pour $\alpha = \frac{1}{2}$ q_α = Mediane (Me)

* Les quartile : $q_{1/4}$ $q_{1/2}$ $q_{3/4}$

  Q_1 Q_2 Q_3

* Les décile : $D_1 = q_{1/10}$ $D_2 = q_{2/10}$

  jusqu'a $D_9 = q_{9/10}$

* Les centile : $C_1 = 1/100$ $C_2 = q_{2/100}$

  jusqu'a $C_{99} = q_{99/100}$

Les moyennes

1- Moyenne arthmétique

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

$$\overline{X} = \frac{1}{n}\sum x_i$$ => donnée brute

2- Moyenne géometrique (statistique discrete)

$$G = \sqrt[n]{\prod_{i=1}^{p} x_i^{n_i}} = \prod_{i=1}^{p} x_i^{f_i}$$

Moyenne harmonique (série discrete)

$$H = \frac{n}{\sum_{i=1}^{p} n_i \left(\frac{1}{x_i}\right)} = \frac{1}{\sum_{i=1}^{p} f_i \left(\frac{1}{x_i}\right)}$$

4- Moyenne quadratique (" ")

$$Q = \sqrt{\frac{1}{n} \sum n_i x_i^2} = \sqrt{\sum_{i=1}^{p} f_i x_i^2}$$

Remarque

=> Dans les série statistique continue on remplace x_i par le centre de la classe

$$c_i = \frac{a_i + a_{i+1}}{2}$$

=> toujours on a : $H < G < \overline{X} < Q$

=> Si $H = G = \overline{X} = Q$ le nombre de modalité $p = 1$ ou $(x_i)_{1 < i \leq n}$ sont egale

Paramètre de dispersion

a- Etendue

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

b- ecart absolue moyen

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

c- La varience

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

d- Ecart type

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

e- Ecart interquartile

est donnée par : $I\alpha = q_{1-\alpha} - q_\alpha$

L'intervalle interquantile : $[q_\alpha , q_{1-\alpha}]$

contient $100(1-2\alpha)\%$ des valeur centrale

=> L'ecart interquantile $q_{3/4} - q_{1/4}$ contient 50% des valeurs centrale

=> L'ecart interdécile contient 80% des valeurs centrale. $q_{9/10} - q_{1/10}$

fficient de variation

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

Parametre de Forme

a- coefficient d'asymetrie

=> Distribution symétrique $M_o = M_e = \bar{X}$

=> " asymetrique a droite $\bar{X} < M_e < M_o$

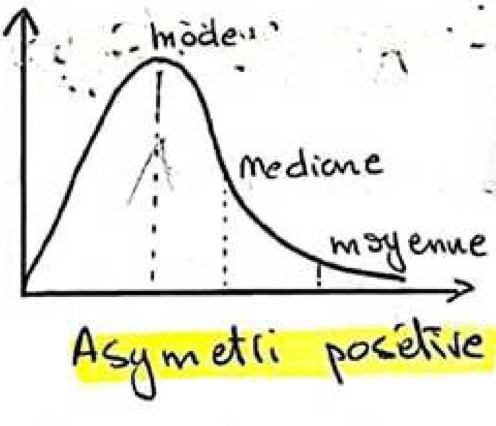
=> " " a gauche $M_o < M_e < \bar{X}$

=> Coefficient d'asymetrie de Fisher :

$$\gamma = \frac{\mu_3(X)}{\sigma^3(X)}$$

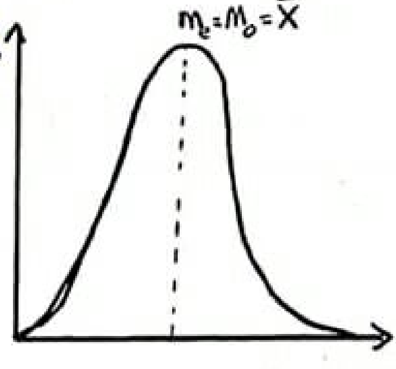
On : $\mu_3(X) = \frac{1}{n} \sum n_i (x_i - \bar{X})^3$ ← moment centre d'ordre 3

=> Si $\gamma > 0$ distribution asymetrique a gauche



Asymetri positive

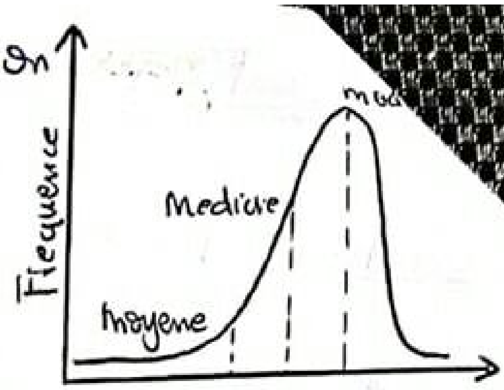
=> Si $\gamma = 0$ distribution symetrique :



Normale (no skew)

=> Si $\gamma < 0$ distribation asymétrique a droite

Asymetrie negative =>



b- coefficient d'applatissement

=> mesure la hauteur d'une distibution

=> Coefficient d'applatissement de Pearson

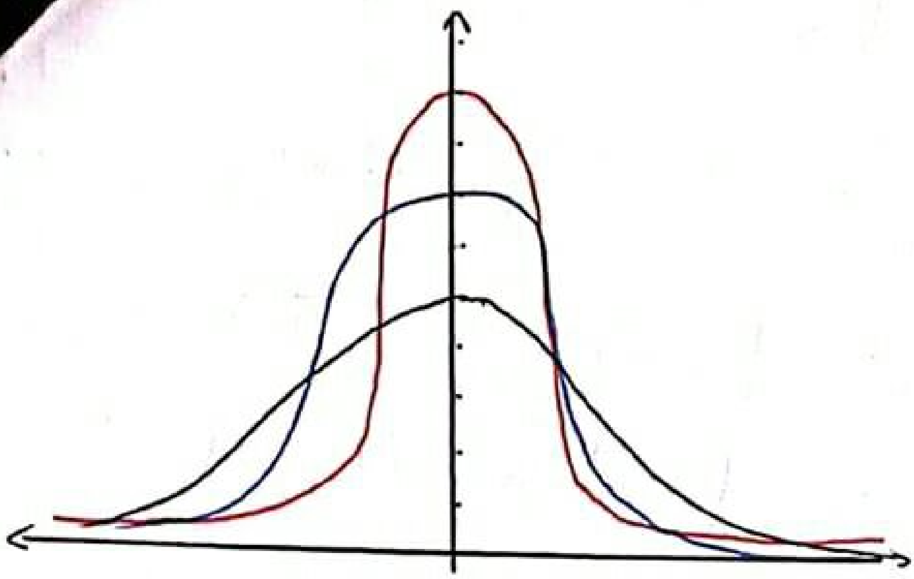
$$K = \frac{\mu_4(x)}{\sigma^4(x)}$$

$$\mu_4(x) = \frac{1}{n} \sum n_i (x_i - \bar{x})^4$$ ← moment centré d'ordre 4

=> Si $K > 3$ distribution (forme pointue) au niveau de la moyenne, extrémité longue et etendue.

=> Si $K < 3$ disstribution plus applatie, extremité courte et resserées

Remarque Dans le cas statistique contine on remplase x_i par c_i



- distribution leptocurtique
- distribution normale (mésokurtique)
- distribution platicurtique

Changement de Variable

Soit X et Y deux Variable statistique et a et b deux nombre réels

$X = ay + b$ $\quad \bar{X} = a\bar{y} + b$ $\quad \sigma_x^2 = a^2 \sigma_y^2$

a = ampetude de la classe

b = milieu de la classe centrale

(Si le nombre de classe est paire, prendre la classe ayant le plus grand effectif)

On obtient :

$$y_i = \frac{x_i - b}{a} \qquad \bar{y} = \frac{1}{n} \sum n_i y_i$$

$$\sigma_y^2 = \frac{1}{n} \sum n_i y_i^2 - \bar{y}^2$$

$$\sigma_y^2 = \frac{1}{n} \sum^{n} n_i \left(y_i - \bar{y}\right)^2$$

$$L'\text{ampétude} = \frac{x_{max} - x_{min}}{\text{nombre de classe}}$$

Statistique Double

Tableau de contengence

X \ Y	[0,5[	[5-10[	[10,15[	[15,20[	n_i
0	2	0	3	0	5
1	0	1	2	3	6
2	0	0	1	1	2
n_J	2	1	6	4	13

n_i => total d'une ligne i (effectif marginal)

n_J => " " " J (" ")

Distribution marginale

X	0	1	2		Y	[0,5[	[5-10[	[15,20[
$n_{i\cdot}$	5	6	2		$n_{\cdot J}$	2	1	6
$f_{i\cdot}$	5/13	6/13	2/13		$f_{\cdot J}$	2/25	1/25	6/25

$$f_{i\cdot} = \frac{n_{i\cdot}}{n} \qquad f_{\cdot J} = \frac{n_{\cdot J}}{n}$$

Les moyenne de $\bar{X}$ et $\bar{Y}$

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

$$\bar{Y} = \frac{1}{n} \sum_{J=1}^{n} n_J \, y_J$$

$$= \sum_{J=1}^{n} f_{\cdot J} \, y_J$$

Calcul des frequence

$* \; f_{xy} = f_{23} = \frac{n_{23}}{n} \Rightarrow n_{ij}$

ligne colonne

le rang de x
le rang de Y

$* \; f_{X=A/Y=B} = \frac{n_{AB}}{n_{\cdot B}}$

$\Downarrow$

la somme de $Y_{\cdot} = B$

Distribution Conditionnelle

Il nous donne par exempele une distribution de $X/Y = 3$:

=> on suit l'orde de X et on prend les $n_{\cdot j}$ de X

=> pour les f_i on derise par la somme de la colonne X.

Calculer une moyenne dans la distribution Conditionnelle

$$(\overline{X}/Y)_{=A} = \sum_{i=1} (X/y)_{=A} \times f(X/Y)_{=A}$$

L'independance

$X \perp\!\!\!\perp Y \Rightarrow f_{ij} = f_i \times f_{\cdot j} \Rightarrow$ il sont independant

$f_{ij} \neq f_i \times f_{\cdot j} \Rightarrow$ il ne sont pas independant

=> Dans une distribution conditionelle :

$\overline{X} = (\overline{X}/Y) = cst \Rightarrow$ indepandant

La covariance entre X et Y (cov(x,y))

Série brute

$$cov(X,Y) = \sigma_{xy}$$

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

$$cov(X,Y) = \left(\frac{1}{n} \sum x_i y_i\right) - (\bar{x}\bar{y})$$

Série groupé

$$cov(X,Y) = \sigma_{xy}$$

$$= \frac{1}{n} \sum^{p} \sum n_{ij} (x_i - \bar{x})(y_j - \bar{y})$$

$$cov(X,Y) = \left(\frac{1}{n} \sum \sum n_{ij} x_i y_j\right) - (\bar{x}\bar{y})$$

$\sum \sum n_{ij} x_i y_j$

x_i	X \ Y	30	50
195	[184,206[	2	1
217	[206-228[	0	1

$\Rightarrow x_i \times 2 \times 30 = \sum \sum n_{ij} y_j$

de correlation

$$r_{xy} = \rho_{xy} = \frac{cov(X,Y)}{\sigma_x \sigma_y}$$

$$Var(X) = \frac{1}{n} \sum x_i^2 n_i - (\bar{x})^2$$

$$Var(Y) = \frac{1}{n} \sum n_i y_i^2 - (\bar{Y})^2$$

$\sigma_x = \sqrt{Var(X)}$ $\quad$ $\sigma_y = \sqrt{Var(Y)}$

Si $\rho_{xy} \begin{cases} \simeq 0 \Rightarrow \text{faible correlation} \\ = 0 \Rightarrow \text{pas de correlation} \\ = \pm 1 \Rightarrow \text{une forte correlation} \end{cases}$

Droite de Regression de Y en X :

On a $Y = aX + b$

$$\begin{cases} a = \dfrac{cov(X,Y)}{Var(X)} \\ b = \bar{Y} - a\bar{X} \end{cases} \Rightarrow Y = aX + b$$

de correlation

$$r_{xy} = \rho_{xy} = \frac{cov(X,Y)}{\sigma_x \sigma_y}$$

$$Var(X) = \frac{1}{n} \sum x_i^2 n_i - (\bar{x})^2$$

$$Var(Y) = \frac{1}{n} \sum n_i y_i^2 - (\bar{Y})^2$$

$\sigma_x = \sqrt{Var(X)}$ $\sigma_y = \sqrt{Var(Y)}$

Si $\rho_{xy} \begin{cases} \simeq 0 & \Rightarrow \text{faible correlation} \\ = 0 & \Rightarrow \text{pas de correlation} \\ = \pm 1 & \Rightarrow \text{une forte correlation} \end{cases}$

Droite de Regression de Y en X :

On a $Y = aX + b$

$$\begin{cases} a = \dfrac{cov(X,Y)}{Var(X)} \\ b = \bar{Y} - a\bar{X} \end{cases} \Rightarrow Y = aX + b$$

Probabilité

=> L'univers, on le note par " Ω "
ou l'espace fondamentale " Ω "

exp * Dans le cas du lancer d'un dé cubique $\Omega = \{1, 2, 3, 4, 5, 6\}$
* Dans le cas du lancer d'une piece de monai $\Omega = \{pile, Face\}$

=> Le cardinale, on le note par $|\Omega|$

Notion d'evenement

Notion	Vacabulaire proba
Ω	Evenement certain
$\emptyset$	" impossible
w	" elementaire
$w \in A$	w réalise A
$A \subset B$	A implique B
$A \cup B$	A ou B
$\overline{A}$	Evenement contraire de A
$A \cap B = \emptyset$	A et B sont incompatible

Opération logique sur les ensemble

* En terme d'evenement $A \cup B$ est réalisé si au moins un des deux evenement A ou B sont réalisé
* En terme d'evenement $A \cap B$ est réalisé lorsque simultanement les deux evenement A et B sont réalisé

Calcul des probabilité

* L'evenement $\bar{A}$:

$$P(\bar{A}) = 1 - P(A)$$

* L'evenement $A \cup B$:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

=> Si A et B sont incompatible

$$P(A \cup B) = P(A) + P(B)$$

* La propabilité $P(A)$:

$$P(A) = \frac{\text{cas favorable}}{\text{cas possible}} = \frac{|A|}{|\Omega|}$$

* $P(A) + P(\bar{A}) = 1$

Tirage de (p) element parmis (n)

et $\longrightarrow \times$

ou $\longrightarrow +$

Simultané	Successive
↓	↓
L'ordre n'est pas important	L'ordre est important
↓	↓
Combinaison C_n^p	Arrangement A_n^p
↓	↓
$C_n^p = \dfrac{n!}{(n-p)!\,p!}$	$A_n^p = \dfrac{n!}{(n-p)!}$

Probabilité conditionnelle

* $P(A \setminus B) = \dfrac{P(A \cap B)}{P(B)}$ } probabilité de A sachant B

* $P(A \cap B) = P(A)P(B)$

* $P(B/A) = \dfrac{P(A \cap B)}{P(A)}$

$$P(\bar{A} \backslash \bar{B}) = \frac{P(\bar{A} \cap \bar{B})}{P(\bar{B})} = \frac{P(\overline{A \cup B})}{P(\bar{B})}$$

$$= \frac{1 - P(A \cup B)}{1 - P(B)}$$

parmi = sachant

Formule de Bayes

$$P(A \backslash B) = \frac{P(B \backslash A) P(A)}{P(B \backslash A) P(A) + P(B \backslash \bar{A}) P(\bar{A})}$$

Formule des probabilité totale

$$P(A \cap \bar{B}) = P(A) - P(A \cap B)$$