Stat 1 :

*** fréquence (tout)**

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

*** effectif corrigé (Continue)**

$$eff_c = \frac{n_i}{a_i}$$

*** hauteur (Continue)**

$$h_i = C \times \frac{f_i}{a_i}$$

PGDC $\{a_i\}$

*** effectif cumulé (qualitatif)**

$$eff_{cum} = f_{cum} \times N$$

*** Mode (Continue)**

$$M_o = x_i + a_i \frac{\Delta_1}{\Delta_1 + \Delta_2}$$

*** Médiane (discrète)**

Si N paire

$$M_e = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

Si N impaire

$$M_e = x_{\frac{n+1}{2}}$$

*** Moyenne (qualitative)**

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{k} x_i n_i \text{ (discrète)}$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{k} c_i n_i \text{ (continue)}$$

$$\overline{X + k} = \bar{X} + k$$

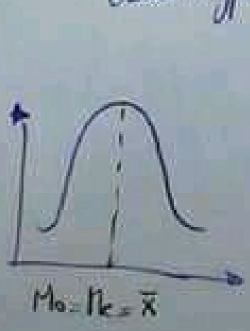
$$\overline{k \times X} = k \times \bar{X}$$

*** Coefficient d'asymétrie de Pearson :**

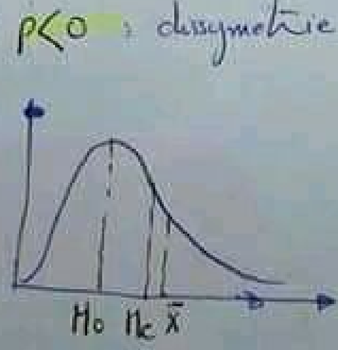
Pour analyser la position du mode et de médiane relativisée par la dispersion de la série :

$$P = \frac{\text{moyenne} - \text{mode}}{\text{écart type}}$$

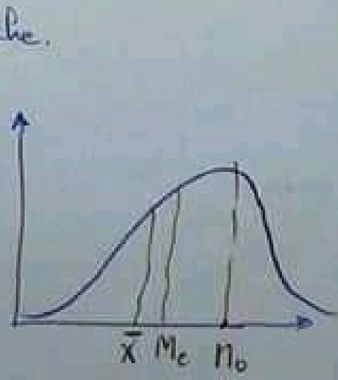
- $P = 0$: asymétrie parfaite.
- $P > 0$: dissymétrie à droite.
- $P < 0$: dissymétrie à gauche.



$P = 0$



$P > 0$



$P < 0$

②

* étendue :

$e = max(x) - Minimum(x)$

* les quantiles (Discrète)

Si N est divisible par 4 :

$$Q_1 = \frac{x_{\frac{n}{4}} + x_{\frac{n}{4}+1}}{2}$$

$$Q_3 = \frac{x_{\frac{3n}{4}} + x_{\frac{3n}{4}+1}}{2}$$

Si N n'est pas divisible par 4 :

$Q_1 = x_i$ où i plus petit entier $\geqslant \frac{n}{4}$.

$Q_3 = x_i$ où i plus petit entier $\geqslant \frac{3n}{4}$.

* écart absolut :

$$e = \frac{1}{n} \sum_i n_i |x_i - \bar{x}|$$

* Variance

*

$$V_x = \frac{1}{n} \sum_i n_i (x_i)^2 - (\bar{X})^2 \text{ (discret)}$$

$$V_x = \frac{1}{n} \sum_i n_i (c_i)^2 - (\bar{X})^2 \text{ (continue)}$$

* ecart type :

$\sigma = \sqrt{V}$ Plus $\sigma \uparrow$, plus les valeurs du caractère sont dispersées autour de la $\bar{x}$.

~~σ↑ plus~~ Plus $\sigma \downarrow$, plus " " " " regroupés " " (homogènes).

* Coefficient de Variation.

$C_v = \frac{\sigma_x}{\bar{X}}$ % → Plus $C_v \uparrow$, plus grande la dispersion.

STAT 2:

* Covariance:

$$Cov(XY) = \overline{XY} - \overline{X}.\overline{Y}$$

$$\overline{XY} = \sum_{i=1}^{p}\sum_{j=1}^{k} n_{ij} x_i y_j$$

$\overline{X}$, $\overline{Y}$ } Moyennes marginales

$Cov(XY) > 0$

⇒ les deux variables X et Y évoluent dans le même sens.

$Cov(XY) < 0$

⇒ les 2 variables X et Y évoluent dans le sens contraire.

⚠ Si X et Y indépendants ⇒ $Cov(XY) = 0$

mais il peut exister une dépendance non linéaire qui se traduit par $Cov = 0$.

- Propriétés de Cov:

* $V(X) = Cov(XX)$.
* $Cov(XY) = Cov(YX)$.
* $Cov(aX+b) = ab\,Cov(X)$.
* $Cov(aY+b) = ab\,Cov(Y)$.
* $V(X+Y) = V(X) + V(Y) + 2\,Cov(XY)$.

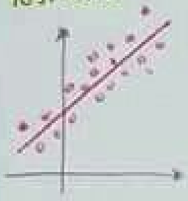
* Coefficient de corrélation:

$$r(X,Y) = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y} \qquad (-1 \leq r \leq 1).$$

- Pour établir l'existence d'un lien linéaire entre X et Y.
- Pour mesurer la force ou l'intensité de ce lien.

* $r > 0$

Correlation Positive

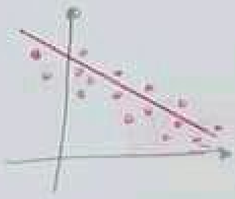


* $r = 1$

Correlation positive Parfaite

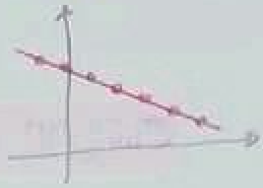
* $r < 0$

Correlation negative.



* $r = -1$

Corrélation negative parfaite.



- Une relation linéaire est considérée forte si

$$\boxed{|r_{X,Y}| \geq 0,9}$$

Elle est dite faible sinon.

* La droite de régression : (droite d'ajustement linéaire)

- Le Centre de gravité de nuage de point $G(\bar{\bar{X}}, \bar{\bar{Y}})$ est le point d'intersection des deux droites d'ajustement.

$$Y = aX + b$$

$$a = \frac{Cov(X,Y)}{V(X)}$$

$$b = \bar{Y} - a\bar{X}$$

$$X = a'Y + b'$$

$$a' = \frac{Cov(X,Y)}{V(Y)}$$

$$b' = \bar{X} - a'\bar{Y}$$

* Indépendance des variables X et Y ;

On Calcul $n_{ij}^{*} = \dfrac{n_{i.} \times n_{.j}}{N}$ (pour le tableau Théorique)

Si Tableau $n_{ij} = Tn_{ij}^{*}$

$\Rightarrow$ X et Y indépendantes.

Si $Tn_{ij} \neq n_{ij}^{*}$

- Il y a dépendance, on mesure l'intensité de la corrélation

Coefficient de cramer

quand $|r_{x,y}| < 9$

$$C_V = \sqrt{\frac{X^2}{n \times \sqrt{(L-1)(C-1)}}}$$

$0 \leq C_V \leq 1$

- Plus est élevé, plus le lien entre les variables est fort.

Avec

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - n_{ij}^{*})^2}{n_{ij}^{*}}$$

$$\chi^2 = \sum_i \sum_j \frac{e_{ij}^{*2}}{n_{ij}}$$

Loi de Bernoulli : ⚠ 2 résultats possibles < Echec / Succès

⑥

p → S = Succès

$q = 1-p$ → $\bar{S}$ = Echecs

$D_X = \{0, 1\}$

fct de distribution :

$$\boxed{P_X(x) = p^x (1-p)^{1-x}}$$

$X = 0$:

$P_X(0) = p^0 (1-p)^1$

$P_X(0) = 1-p$

$X = 1$:

$P_X(1) = p^1 (1-p)^{1-1}$

$P_X(1) = p$

$X \rightsquigarrow B(p)$

x	0	1
$P_i = P(X = x_i)$	$q = 1-p$	p

- Espérence : $\boxed{E(X) = p}$?
- Variance : $V(X) = pq$?

$\boxed{V(X) = p(1-p)}$?

Espérence

$$E[X] = \sum_{x \in D_X} x\, P_X(x) = 0\, P_X(0) + 1\, P_X(1) = 0(1-p) + p = p$$

Variance :

$$V[X] = E[X^2] - E[X]^2 = \sum x^2 P_X(x) - E[X]^2 = \overbrace{0^2 P_X(0^2)}^{0} + \overbrace{1 P_X(1^2)}^{p} - E[X]^2$$

$$= p - p^2 = p(1-p)$$

* Propriétés de $E(X)$ et $V(X)$:

$X \longrightarrow X + b$

$E(X)$: $E(X+b) = E(X) + b$

$V(X+b) = V(X)$

$X \longrightarrow aX$

$E(aX) = aE(X)$

$V(aX) = a^2 V(X)$

$X \longrightarrow aX + b$

$E(aX+b) = aE(X) + b$

$V(aX+b) = a^2 V(X)$

* Loi de Couple :

- Une expérience qui possède 2 Variables aléatoires X et Y. ⑦

on note.

x_i $(1 \leq i \leq n)$ les n valeurs prises par X.

y_j $(1 \leq j \leq p)$ les p valeurs prises par Y.

+ X et Y indépendantes si ?

- $[X = x_i]$ et $[Y = y_j]$ sont indépendants
- $P([X=x_i] \cap [Y=y_j]) = P[X=x_i] \times P[Y=y_j]$.

3 événements formant une partition de l'événement $[X=x_1]$

X \ Y	Y_1	Y_2	Y_3	loi de X
X_1	$X_1 \cap Y_1$	$X_1 \cap Y_2$	$X_1 \cap Y_3$	$P[X=x_1]$
X_2				
Loi de Y	$P[Y=y_1]$			

$P[X=x_1]$ → la somme de la ligne X_1.

$P[Y=y_1]$ → la somme de la colonne y_1

+ X et Y sont indépendantes si $P[X=x_1] \times P[Y=y_1] = X_1 \cap Y_1$

→ $$\boxed{Cov(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j \, P(X = x_i \text{ et } Y = x_j) - E[X]E[Y]}$$

X et Y sont indépendants ⟹ $Cov(X,Y) = 0$

⟹ $E(X,Y) = E(X) \times E(Y)$

mais ⚠ $Cov(X,Y) = 0 \not\Rightarrow$ X et Y sont indépendante.

* Loi Binomiale $\mathcal{B}(n,p)$

- succession d'épreuves de bernouli < identiques / dépendantes. (n fois)
- p probabilité de succès.
- avec remise.

X le nb de succès obtenus en n épreuves de Bernouli

$$X \sim \mathcal{B}(n,p) \longrightarrow \boxed{P(X=k) = C_n^k p^k (1-p)^{n-k}}$$

Espérance

$$\boxed{E(X) = np}$$

Variance

$$\boxed{V(X) = npq}$$

$$X(\Omega) = \{0, 1, \dots, n\}$$

* Loi Hypergéométrique $\mathcal{H}(N,n,p)$

- Population de N individus partagée en deux < individus positifs p / " " négatifs $1-p=q$
  - On prélève un échantillon de "n" personnes
  - sans remise.

$X \sim \mathcal{H}(N,n,p)$ avec
N : la taille de la population.
n : " " de l'échantillon.
p : proportion initiale.

X le nb de personnes positives dans cet échantillon.

$$\boxed{P(X=k) = \frac{C_p^k \, C_{Nq}^{n-k}}{C_N^n}}$$

Espérance

$$\boxed{E(X) = np}$$

Variance

$$\boxed{V(X) = \frac{N-n}{N-1} npq}$$

* Loi géométrique $G(p)$:

- Une épreuve de Bernoulli $\begin{cases} p \text{ (succès)} \\ q \text{ (échec)} \quad (q = 1-p) \end{cases}$
- On renouvle cette épreuve de manière indépendante jusqu'au premier succès.

X le rang du premier succès

$X \sim G(p)$

$$\boxed{P(X=k) = (1-p)^{k-1} \cdot p}$$

Espérance

$$\boxed{E(X) = \frac{1}{p}}$$

$$X(\Omega) = \mathbb{N}^*$$

Variance

$$\boxed{V(X) = \frac{1-p}{p^2} = \frac{q}{p}}$$

* ~~Loi de Poisson~~ ~~P~~(λ) :

Loi de Poisson :

- Un évènement qui se produisent λ fois en moyenne durant un intervalle de temps donné.

$X \rightsquigarrow \mathcal{P}(\lambda)$

"La Probabilité que cet évènement se produit k fois dans une période connue"

$$P(X=k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Espérance

$E(X) = \lambda$

Variance

$V(X) = \lambda$.