Introduction aux probabilités et statistique descriptiv - les paramètres de dispersion et les paramètres de forme

Rahmani Nacer - **Département de Mathématiques**

07/03/2023

5.Indicateurs de dispersion

Ces paramètres ont pour objectif dans le cas d'un caractère quantitatif de caractériser la variabilité des données dans l'échantillon.
Les indicateurs de dispersion fondamentaux sont

1. la variance observée
2. écart-type observé.
3. écart moyen.
4. Coefficient de variation.
5. L'interquartile, interdécile et intercentile.

- Soit un échantillon de n valeurs observées $x_1, x_2, \ldots., x_i, \ldots., x_n$ d'un caractère quantitatif X et soit $\bar{x}$ sa moyenne observée. On définit la variance observée notée V ou S^2 comme la moyenne arithmétique des carrés des écarts à la moyenne.

$$V = S^2 = \frac{1}{n} \sum_{i=1}^{k} n_i \left(x_i - \bar{X}\right)^2$$

- Soit un échantillon de n valeurs observées $x_1, x_2, \ldots., x_i, \ldots., x_n$ d'un caractère quantitatif X et soit $\bar{x}$ sa moyenne observée. On définit la variance observée notée V ou S^2 comme la moyenne arithmétique des carrés des écarts à la moyenne.

$$V = S^2 = \frac{1}{n} \sum_{i=1}^{k} n_i \left(x_i - \bar{X}\right)^2$$

- La formule de la variance qui résulte du théorème de Koenig est donc:

$$V = S^2 = \frac{1}{n} \sum_{i=1}^{k} n_i x_i^2 - \bar{X}^2$$

- Dans le cas de données regroupées en k classes d'éffectif n_i (variable continue regroupée en classes), la formule de la variance est la suivante:

$$V = S^2 = \frac{1}{n}\sum_{i=1}^{k} n_i \left(c_i - \bar{X}\right)^2 ,$$

ou c_i est le centre de classe

L'écart-type observé correspond à la racine carrée de la variance observée:

$$\sigma = \sqrt{V} = S = \sqrt{S^2}$$

.

- Soit un échantillon de n valeurs observées $x_1, x_2, \ldots, x_n$ d'un caractère quantitatif X et soit $\bar{x}$ sa moyenne observée. On déffinit l''écart moyen notée $E.M$ comme la moyenne arithmétique des écarts à la moyenne.

$$E.M = \frac{1}{n} \sum_{i=1}^{k} n_i \left| x_i - \bar{X} \right|$$

.

- Soit un échantillon de n valeurs observées $x_1, x_2, \dots, x_n$ d'un caractère quantitatif X et soit $\bar{x}$ sa moyenne observée. On déffinit l"écart moyen notée $E.M$ comme la moyenne arithmétique des écarts à la moyenne.

$$E.M = \frac{1}{n}\sum_{i=1}^{k} n_i \left|x_i - \bar{X}\right|$$

- Dans le cas de données regroupées en k classes d'éffectif n_i (variable continue regroupée en classes), la formule est la suivante:

$$E.M = \frac{1}{n}\sum_{i=1}^{k} n_i \left|c_i - \bar{X}\right|$$

.

La variance et l'écart-type observée sont des paramètres de dispersion absolue qui mesurent la variation absolue des données indépendamment de l'ordre de grandeur des données.

Le coefficient de variation noté $C.V.$ est un indice de dispersion relatif prenant en compte ce biais et est égal à:

$$C.V = \frac{\sigma}{\bar{X}}100$$

Exprimé en pour cent, il est indépendant du choix des unités de mesure permettant la comparaison des distributions de fréquence d'unité différente.

L'intervalle interquartile est une mesure de la variation qui n'est pas influencée par les valeurs extrêmes. Sa définition est simple: l'intervalle interquartile mesure l'étendue des 50% de valeurs situées au milieu d'une série de données classées.

$$IQ = Q_3 - Q_1$$

L'intervalle interdécile mesure l'étendue des 80% de valeurs situées au milieu d'une série de données classées.

$$ID = D_9 - D_1$$

L'intervalle interquartile mesure l'étendue des 98% de valeurs situées au milieu d'une série de données classées.

$$IC = C_{99} - C_1$$

6.Paramètres de forme

Nous définissons les paramètres de forme pour une variable statistique quantitative, discrète ou continue, à valeurs réelles.

1. Coefficient d'asymétrie.
2. Coefficient d'aplatissement

Le coefficient d'asymétrie de **Pearson** fait intervenir le mode *Mo*: quand il existe, il est définie par:

$$P = \frac{\bar{X} - Mo}{s}$$

Le coefficient d'asymétrie de **Yule** fait intervenir la médiane et les quartiles, il est défini par:

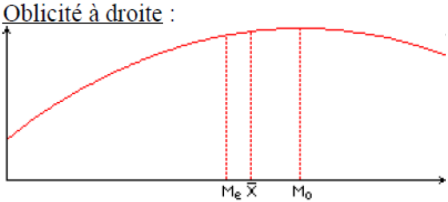
$$Y = \frac{Q_1 + Q_3 - 2Me}{2\left(Q_3 - Q_1\right)}$$

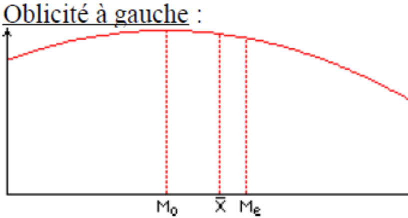
Le coefficient d'asymétrie de **Fisher** fait intervenir les moments centrés, il est défini par:

$$F = \frac{\mu_3}{s^3} = \frac{\mu_3}{\sigma^3}$$

$$\mu_r = \frac{1}{n}\sum_{i=1}^{k} n_i\left(x_i - \bar{X}\right)^r$$

Lorsque le coefficient d'asymétrie est positif, la distribution est plus étalée à droite : on dit qu'il y a oblicité à gauche.
Lorsque le coefficient d'asymétrie est négatif, la distribution est plus étalée à gauche : on dit qu'il y a oblicité à droite.





Là encore plusieurs définitions sont possibles.

- Le coefficient d'aplatissement de **Pearson** est:

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

$$F_2 = \frac{\mu_4}{\mu_2^2} - 3$$

Là encore plusieurs définitions sont possibles.

- Le coefficient d'aplatissement de **Pearson** est:

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

- Le coefficient d'aplatissement de **Yule** est:

$$F_2 = \frac{\mu_4}{\mu_2^2} - 3$$

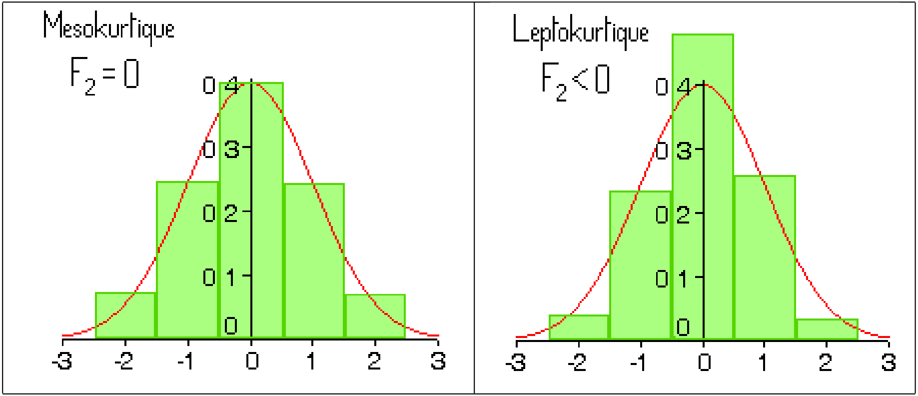
- Si F_2 est égal à 0, le polygone statistique de la variable réduite a le même aplatissement qu'une courbe en cloche, on dit que la variable est **mésokurtique**.

- Si F_2 est égal à 0, le polygone statistique de la variable réduite a le même aplatissement qu'une courbe en cloche, on dit que la variable est **mésokurtique**.
- Si F_2 est > 0, le polygone statistique de la variable réduite est moins aplati qu'une courbe en cloche, on dit que la variable est **leptokurtique**.

- Si F_2 est égal à 0, le polygone statistique de la variable réduite a le même aplatissement qu'une courbe en cloche, on dit que la variable est **mésokurtique**.
- Si F_2 est > 0, le polygone statistique de la variable réduite est moins aplati qu'une courbe en cloche, on dit que la variable est **leptokurtique**.
- Si F_2 est < 0, le polygone statistique de la variable réduite est plus aplati qu'une courbe en cloche, on dit que la variable est **platykurtique**.

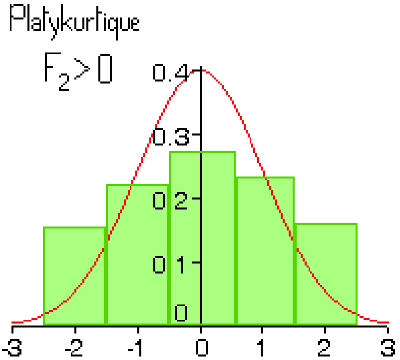
Paramètres de forme

6.2 Coefficient d'aplatissement:



Paramètres de forme

6.2 Coefficient d'aplatissement:



Exemple: Taille des étudiants de première année

Classes	**Effectifs**	**Fréquences:** $f_i = \frac{n_k}{n}$
$[150, 155[$	2	$2/28$
$[155, 160[$	5	$5/28$
$[160, 165[$	3	$3/28$
$[165, 170[$	7	$7/28$
$[170, 175[$	5	$5/28$
$[175, 180[$	2	$2/28$
$[180, 185[$	3	$3/28$
$[185, 190[$	1	$1/28$
$[190, 195[$	2	$2/28$