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

Rahmani Nacer - **Département de Mathématiques**

14/04/2024

5.Dispersion parameter (spread)

These parameters aim, in the case of a quantitative nature, to characterize the variability of the data in the sample..

The fundamental dispersion indicators are

1. The variance
2. The standard deviation.
3. The .mean deviation.
4. The Coefficient of variation.
5. The interquartile, interdecile and interpercentile.

- 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

- The observed standard deviation corresponds to the square root of the observed variance:

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

- Let us be a sample of n observed values $x_1, x_2, \ldots, x_n$ of a quantitative character X and let $\bar{x}$ be its observed mean. We define the mean deviation denoted $E.M$ as the arithmetic mean of the deviations from the mean

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

- Let us be a sample of n observed values $x_1, x_2, \ldots, x_n$ of a quantitative character X and let $\bar{x}$ be its observed mean. We define the mean deviation denoted $E.M$ as the arithmetic mean of the deviations from the mean

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

- In the case of data grouped into k classes of frequency n_i (continuous variable grouped into classes), the formula is as follows:

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

- Observed variance and standard deviation are absolute dispersion parameters that measure the absolute variation of the data independent of the order of magnitude of the data.

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

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

- Observed variance and standard deviation are absolute dispersion parameters that measure the absolute variation of the data independent of the order of magnitude of the data.
- The coefficient of variation denoted $C.V.$ is a relative dispersion index taking into account this bias and is equal to:

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

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

- The interquartile range is a measure of variation that is not influenced by extreme values. Its definition is simple: the interquartile range measures the extent of the 50% of values located in the middle of a series of classified data.

$$IQ = Q_3 - Q_1$$

The interpercentile range measures the range of the middle 98% of values in a classified data series.

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

- The interquartile range is a measure of variation that is not influenced by extreme values. Its definition is simple: the interquartile range measures the extent of the 50% of values located in the middle of a series of classified data.

$$IQ = Q_3 - Q_1$$

- The interdecile interval measures the extent of the 80% of values located in the middle of a series of classified data.

$$ID = D_9 - D_1$$

The interpercentile range measures the range of the middle 98% of values in a classified data series.

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

6.Shape parameters

We define the shape parameters for a quantitative statistical variable, discrete or continuous, with real values.

1. **Skewness coefficient.(Asymmetry)**
2. **Kurtusis coefficient (heaviness of tail)**

1. The Pearson Skewness coefficient involves the *Mo* mode: when it exists, it is defined by:

$$S_k = \frac{\bar{X} - Mo}{\sigma}$$

- Si S_k est égal à 0, the reduced variable has the same flattening as a bell curve, we say that the variable is normal.

6.Shape parameters

6.1 Skewness coefficient.

1. The Pearson Skewness coefficient involves the *Mo* mode: when it exists, it is defined by:

$$S_k = \frac{\bar{X} - Mo}{\sigma}$$

- Si S_k est égal à 0, the reduced variable has the same flattening as a bell curve, we say that the variable is normal.
- If $S_k > 0$, When Skewness coefficient is positive, the distribution is more spread out to the right: we say that there is an skew to the left.

1. The Pearson Skewness coefficient involves the *Mo* mode: when it exists, it is defined by:

$$S_k = \frac{\bar{X} - Mo}{\sigma}$$

- Si S_k est égal à 0, the reduced variable has the same flattening as a bell curve, we say that the variable is normal.
- If $S_k > 0$, When Skewness coefficient is positive, the distribution is more spread out to the right: we say that there is an skew to the left.
- If $S_k < 0$, When Skewness coefficient is negative, the distribution is more spread out to the left: we say that there is skew to the right

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

$$Y = \frac{Q_1 + Q_3 - 2Me}{Q_3 - Q_1}$$

- Si Y est égal à 0, the reduced variable has the same flattening as a bell curve, we say that the variable is normal.

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

$$Y = \frac{Q_1 + Q_3 - 2Me}{Q_3 - Q_1}$$

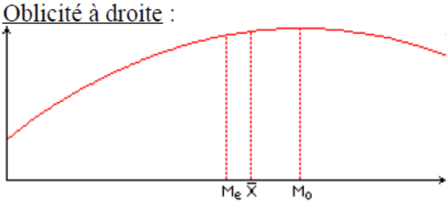
- Si Y est égal à 0, the reduced variable has the same flattening as a bell curve, we say that the variable is normal.
- If $Y > 0$, When Skewness coefficient is positive, the distribution is more spread out to the right: we say that there is an skew to the left.

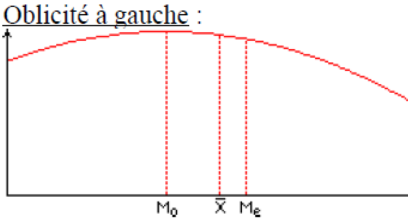
1. e coefficient d'asymétrie de **Yule** fait intervenir la médiane et les quartiles, il est défini par:

$$Y = \frac{Q_1 + Q_3 - 2Me}{Q_3 - Q_1}$$

- Si Y est égal à 0, the reduced variable has the same flattening as a bell curve, we say that the variable is normal.
- If $Y > 0$, When Skewness coefficient is positive, the distribution is more spread out to the right: we say that there is an skew to the left.
- If $Y < 0$, When Skewness coefficient is negative, the distribution is more spread out to the left: we say that there is skew to the right

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.





Here again several definitions are possible.

- The **Pearson** kurtosis coefficient is:

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

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

Here again several definitions are possible.

- The **Pearson** kurtosis coefficient is:

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

- The **Yule** kurtosis coefficient is:

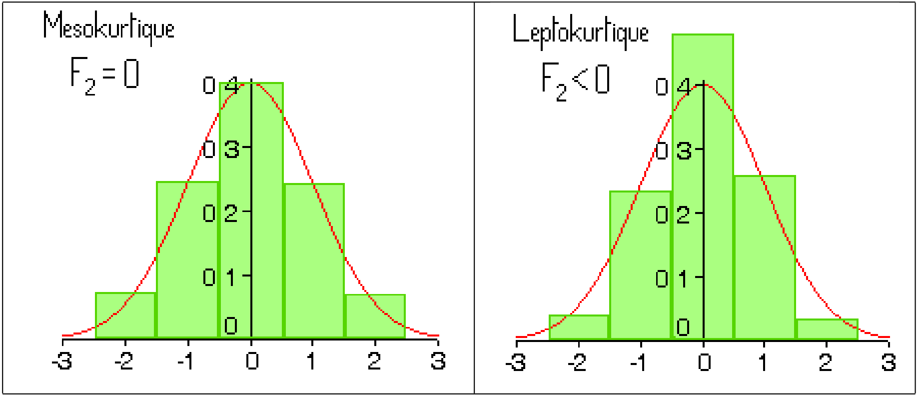
$$F_2 = \frac{\mu_4}{\mu_2^2} - 3$$

- If $F_2 = 0$,the statistical polygon of the reduced variable has the same flattening as a bell curve, we say that the variable is **mesokurtic**.(normal)

- If $F_2 = 0$,the statistical polygon of the reduced variable has the same flattening as a bell curve, we say that the variable is **mesokurtic**.(normal)
- If $F_2 > 0$, the statistical polygon of the reduced variable is less flattened than a bell curve, we say that the variable is **leptokurtic**. (less flattened)

- If $F_2 = 0$,the statistical polygon of the reduced variable has the same flattening as a bell curve, we say that the variable is **mesokurtic**.(normal)
- If $F_2 > 0$, the statistical polygon of the reduced variable is less flattened than a bell curve, we say that the variable is **leptokurtic**. (less flattened)
- If $F_2 < 0$, the statistical polygon of the reduced variable is flatter than a bell curve, we say that the variable is **platikurtic**.(more flattened)

Shape parameters
6.2 Kurtusis coefficient:



Shape parameters

6.2 Kurtusis coefficient: