

## Statistiques descriptives : Liste des formules utilisées

Cours de L1 de Monsieur Lombardot

Université Paris 1 Panthéon-Sorbonne

### Liste des principales formules utilisées pour l'analyse univariée :

#### Tendance centrale :

$$\text{Mode : } Mo = L + a \times \left[ \frac{e_o - e_i}{(e_o - e_i) + (e_o - e_s)} \right]$$

$$\text{Moyenne : } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\text{Moyenne pondérée : } \bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i$$

Q1, Q2 (médiane), Q3 : interpolation linéaire

#### Dispersion :

Etendue : différence entre les deux valeurs extrêmes

Intervalle inter décile : IIQ = Q3-Q1

Intervalle inter quartile : IID = D9-D1

$$\text{Variance : } V(X) = \frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^2$$

$$\text{Variance pondérée : } V(X) = \frac{1}{n} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2$$

$$\text{Ecart-type : } \sigma(X) = \sqrt{V(X)}$$

$$\text{Coefficient de variation : } C_v = \frac{\sigma}{\bar{x}}$$

$$\text{Le moment simple d'ordre « r » : } m_r = \frac{1}{n} \sum_{i=1}^k n_i x_i^r$$

$$\text{Le moment centré d'ordre « r » : } \mu_r = \frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^r$$

- Pour r=1 les écarts à la moyenne s'annulent, donc  $\mu_1=0$

- Pour  $r=2$  on retrouve la formule de la variance, donc  $\mu_2 = V(X) = m_2 - m_1^2$
- Pour  $r=3$  on peut calculer :  $\mu_3 = m_3 - 3m_1m_2 + 2m_1^3$
- Pour  $r=4$  on peut calculer :  $\mu_4 = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4$

### **Asymétrie :**

Le coefficient d'asymétrie de Yule :

$$s = \frac{(Q3 - Mé) - (Mé - Q1)}{(Q3 - Mé) + (Mé - Q1)}$$

Le premier coefficient d'asymétrie de Pearson :  $s = \frac{\bar{x} - Mo}{\sigma}$

Le second coefficient d'asymétrie de Pearson :  $\beta_1 = \frac{\mu_3}{\mu_2^2}$

Le coefficient d'asymétrie de Fisher :  $\gamma_1 = \frac{\mu_3}{\sigma^3}$

### **Aplatissement :**

Le coefficient d'aplatissement de Pearson :  $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}$

Le coefficient d'aplatissement de Fisher :  $\gamma_2 = \beta_2 - 3 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{\mu_4}{\sigma^4} - 3$

### **Concentration :**

Indice de Gini :  $IG = 2 \times \text{aire de concentration}$

## Liste des principales formules utilisées pour l'analyse bivariée :

### Covariance :

$$COV(x,y) = \frac{1}{n_{..}} \sum_{i=1}^p \sum_{j=1}^q n_{ij} (x_i - \bar{x}) (y_j - \bar{y})$$

$$COV(x,y) = \frac{1}{n_{..}} \sum_{i=1}^p \sum_{j=1}^q n_{ij} (x_i - \bar{x}) (y_j - \bar{y})$$

### Coefficient de corrélation :

$$r = \frac{COV(x,y)}{\sigma(x)\sigma(y)}$$

### Coefficient de détermination :

$r^2$  : pourcentage de variance restituée

### Equation de la droite de régression :

$$a = \frac{cov(xy)}{V(x)}$$

$$b = \bar{y} - a\bar{x}$$