Statistiques à 2 variables – Fiche de cours

1. Série statistique à 2 variables

a. Définition (nuage de points)

On définit une série statistique double en observant deux critères sur une même population de dimension n :

$$x=(x_1,x_2\ldots,x_n) \qquad y=(y_1,y_2\ldots,y_n)$$

Valeur x_i	x_1	x_2	…	x_n
Valeur y_i	y_1	y_2	…	y_n

L'ensemble des points de coordonnées (x_i,y_i), rapportés dans un repère, forme le nuage de points de la série (x,y).

b. Point moyen

Le point moyen G de la série statistique est défini par $G(\overline{x};\overline{y})$:

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{et} \quad \overline{y}=\frac{1}{n}\sum_{i=1}^{n} y_i$$

c. Variance, écart type

Les variances de x et y sont définies par :

$$Var(X)=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2 \qquad Var(Y)=\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2$$

Les écarts type de x et y sont définies par :

$$\sigma(X)=\sqrt{Var(X)} \qquad \sigma(Y)=\sqrt{Var(Y)}$$

d. Covariance

La covariance de x et y est définie par :

$$cov(X,Y)=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})\cdot(y_i-\overline{y})$$

2. Ajustement par la méthode des moindres carrés

a. Coefficient de corrélation

On définit le coefficient de corrélation de X et Y par :

$$r=\frac{cov(X,Y)}{\sigma(X)\cdot\sigma(Y)} \qquad -1\leq r\leq 1$$

Lorsque $|r|\approx 1$ le nuage de points peut être approcher par une droite d'ajustement linéaire

Lorsque $|r|\approx 0$ le nuage de points ne peut pas être approcher par une droite d'ajustement linéaire

b. Ajustement linéaire

La droite de régression linéaire par la méthode des moindres carrés a pour expression :

$$y=ax+b \quad \text{avec} \quad a=\frac{cov(X,Y)}{var(X)} \quad \text{et} \quad b=\overline{y}-a\overline{x}$$

Le point moyen $G(\overline{x};\overline{y})$ est placé sur la droite de régression linéaire

c. Autres ajustements

En posant un changement de variables adapté, il est possible de définir d'autres types d'ajustements :

- exponentiel
- polynomial