

# TD5: Statistiques bivariées (2)

Statistiques descriptives avec R

Avril 2023

## Exercice 1 :

*Testez vos acquis sur l'ajustement linéaire pour des variables quantitatives. Dites si les assertions suivantes sont vraies ou fausses.*

- 1. Un nuage de points en forme de boule indique l'existence d'une corrélation linéaire.
- 2. L'ordonnée à l'origine se calcule en faisant le rapport entre une covariance et une variance.
- 3. Le coefficient de corrélation linéaire est un nombre sans dimension.
- 4. Le coefficient de détermination est le coefficient de corrélation élevé au carré.
- 5. Le principe des moindres carrés repose sur la minimisation de la somme des carrés des écarts mesurés parallèlement à l'axe des ordonnées.
- 6. La droite des moindres carrés passe dans certains cas par le point moyen.
- 7. Lorsque le coefficient de corrélation linéaire est proche de zéro, il existe une forte corrélation linéaire.
- 8. On peut calculer la droite de régression de Y en X et la droite de régression de X en Y.

## Exercice 2

Démontrer que le coefficient de corrélation linéaire d'une série statistique double  $(x_i, y_i)$  (avec  $i \in \{1, \dots, n\}$ ) n'est pas affecté par un changement de variable du type :

- $x'_i = \frac{x_i - k}{h}$  où  $k$  et  $h$  sont les constantes ( $h \neq 0$ )
- $y'_i = \frac{y_i - g}{l}$  où  $g$  et  $l$  sont les constantes ( $l \neq 0$ )

On rappelle que :

- $\bar{X}' = \frac{\bar{X} - k}{h}$  où  $k$  et  $h$  sont les constantes ( $h \neq 0$ )
- $\bar{Y}' = \frac{\bar{Y} - g}{l}$  où  $g$  et  $l$  sont les constantes ( $l \neq 0$ )

## Exercice3 : Moindres carrés pondérés

On réalise une enquête auprès de 1000 ménages. On s'intéresse à la liaison entre "le nombre d'enfants" et "les dépenses annuelles de fournitures scolaires".

- X : Nombre d'enfants à charge du ménage.
- Y : Dépenses annuelles en fournitures scolaires (en dizaine d'euros).

$x_i; y_i$	de 0 à 4	de 4 à 10	de 10 à 20	de 20 à 40
1	322	12	2	0
2	14	230	116	36
3	0	0	20	248

- 1. Calculer le coefficient de corrélation linéaire.
- 2. Calculer l'équation de la droite de régression de Y en X.

#### Exercice 4 : Moindres carrés sur données individuelles

Pour étudier la relation entre une série X et une série Y, on donne :

$$\sum_{i=1}^{10} x_i = 73 \quad \sum_{i=1}^{10} y_i = 64 \quad \sum_{i=1}^{10} x_i^2 = 565 \quad \sum_{i=1}^{10} y_i^2 = 430 \quad \sum_{i=1}^{10} x_i y_i = 465$$

- 1. Calculer la moyenne et la variance de chacune des deux variables et la covariance des deux variables
- 2. Calculer le coefficient de corrélation linéaire.

#### Exercice 5: Moindres carrés. Relation inverse entre X et Y

Le tableau ci-dessous donne l'observation des prix (Y) et des quantités disponibles (X) d'un produit sur un marché.

Y est donnée en euros et X est un nombre de produits.

$x_i$	$y_i$
2	98
4	90
6	84
12	72
15	64
20	62
22	55
24	54
28	45
30	40

Table 1: Prix et quantités d'un produit

- 1. Représenter graphiquement le nuage de points.
- 2. Calculer les paramètres de la droite des moindres carrés :  $Y = aX + b$  .  
Représenter la droite de régression et le point moyen.
- 3. Calculer le coefficient de corrélation linéaire.

- 4. Calculer les erreurs  $\hat{u}_i$  pour chacune des valeurs  $x_i$  de la variable X à l'aide des estimateurs des paramètres a et b. Calculer la variance de l'erreur.
- 5. Quel prix peut-on attendre si la quantité disponible est de 5, puis si la quantité disponible est de 26 ?