

# Université de Toulouse

---

## *Compte Rendu*

---

*TP1 : Description monodimensionnelle et  
bidimensionnelle de données.*

✗ Module : *Analyse statistique de données*

✗ Réalisé par :

- KABOU Abdeldjalil
- GHENAI Islam

## Sujet 1 : Description monodimensionnelle et bidimensionnelle de données.

### 1) Première étude de la réflectance de matériaux :

#### a. **Données considérées :**

*Question :* Après avoir chargé les données, déterminer le nombre d'individus et le nombre de variables mis en jeu dans ces données.

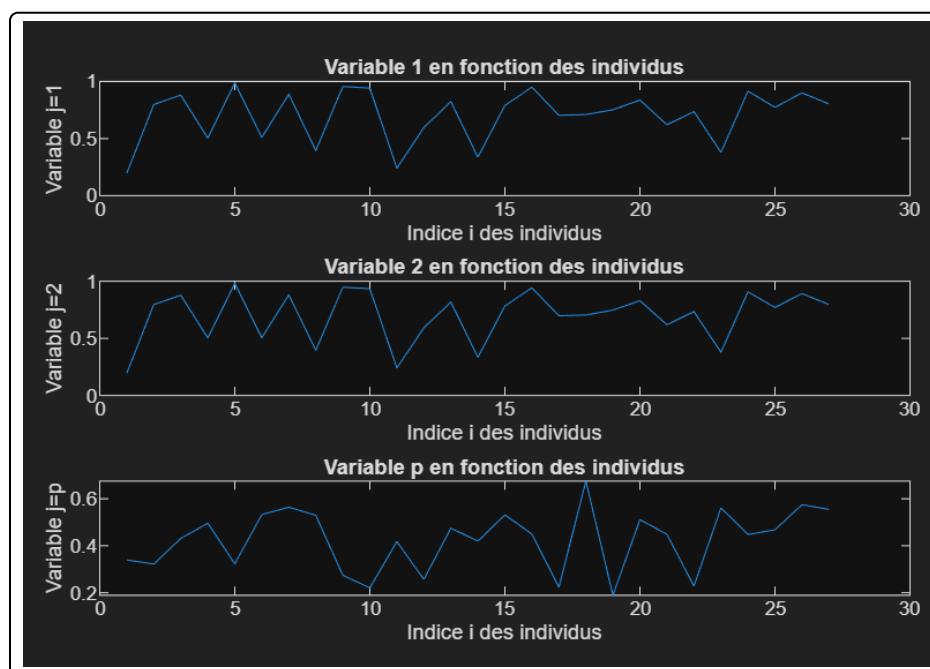
Réponse : il faut regarder la taille de la matrice X :

$$n = \text{nb de lignes} \rightarrow \text{nb d'individus} = 27$$

$$p = \text{nb de colonnes} \rightarrow \text{nb de variables} = 1790$$

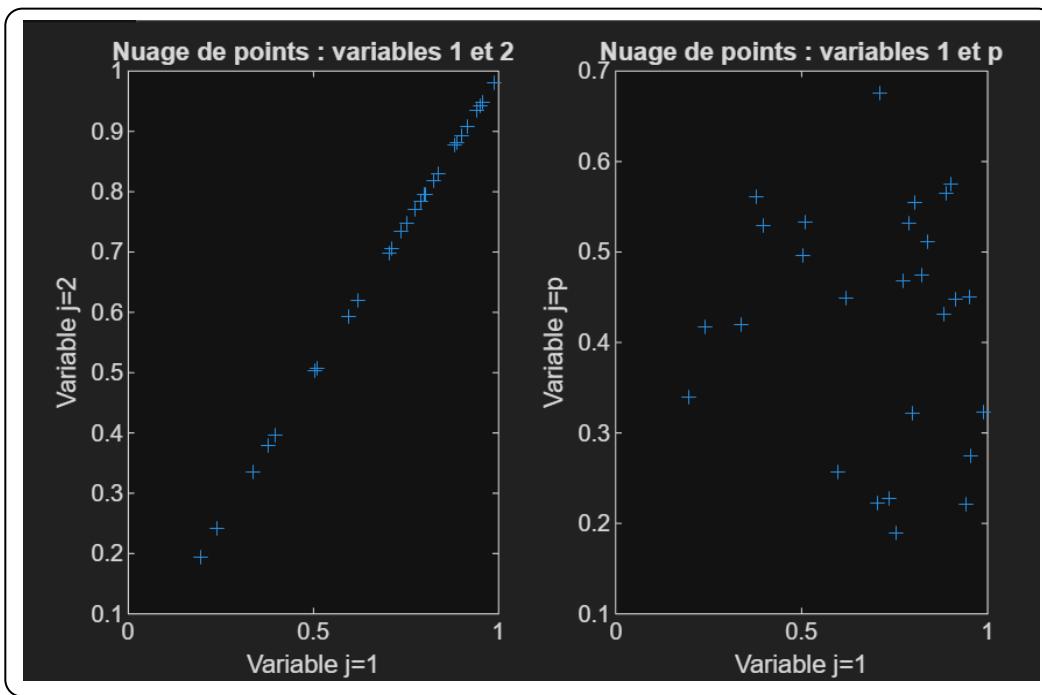
#### b. **Représentation partielle des données :**

Réponse 1 :



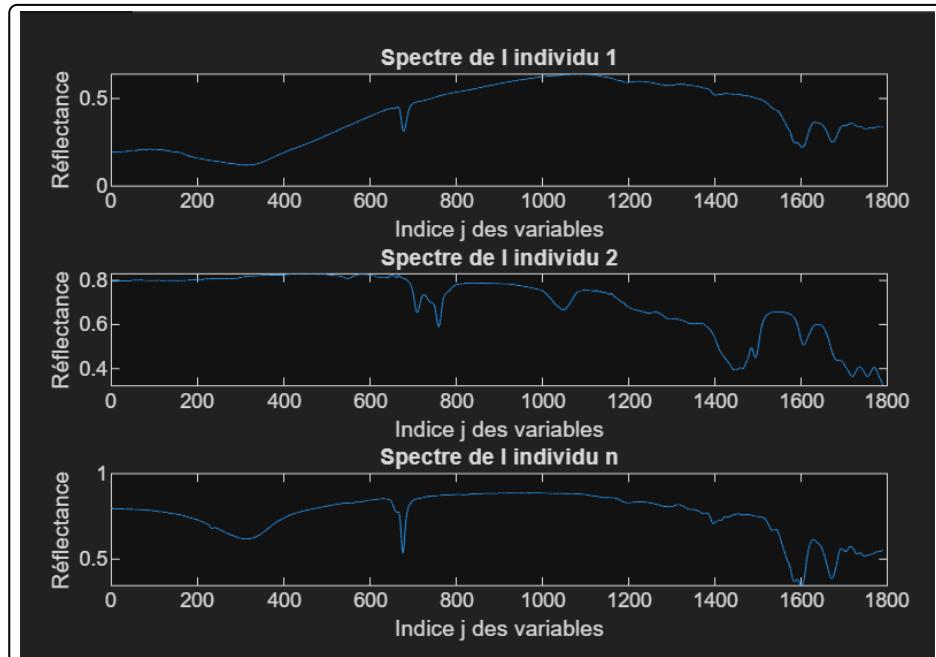
*Commentaires :* Le degré de réflectance des matériaux pour des longueurs d'onde proches est globalement similaire, ce qui explique que les courbes obtenues pour les variables d'indices 1 et 2 soient très proches. Par contre, pour une longueur d'onde beaucoup plus grande (variable d'indice p), la forme de la courbe change nettement, ce qui est cohérent d'un point de vue physique.

Réponse 2 :



Commentaires : Les nuages de points correspondant à deux longueurs d'onde proches (variables 1 et 2) présentent une forme quasi-linéaire avec une pente positive, ce qui traduit une forte corrélation entre ces deux variables. Par contre, lorsque l'on compare deux longueurs d'onde très différentes (variables 1 et p), le nuage de points apparaît plus dispersé, ce qui indique une corrélation plus faible.

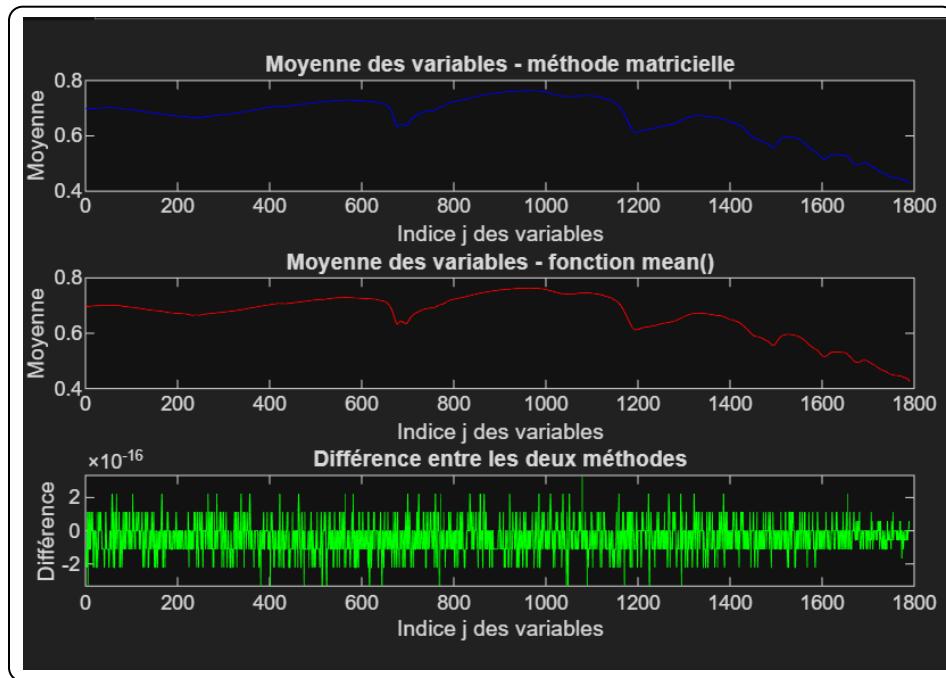
Réponse 3 :



*Commentaires :* Les spectres montrent que chaque individu a une courbe différente. Certaines zones se ressemblent, mais on observe aussi des variations nettes selon l'individu. Cela est normal car chaque matériau réfléchit la lumière d'une façon particulière selon la longueur d'onde.

### c. Analyse monodimensionnelle :

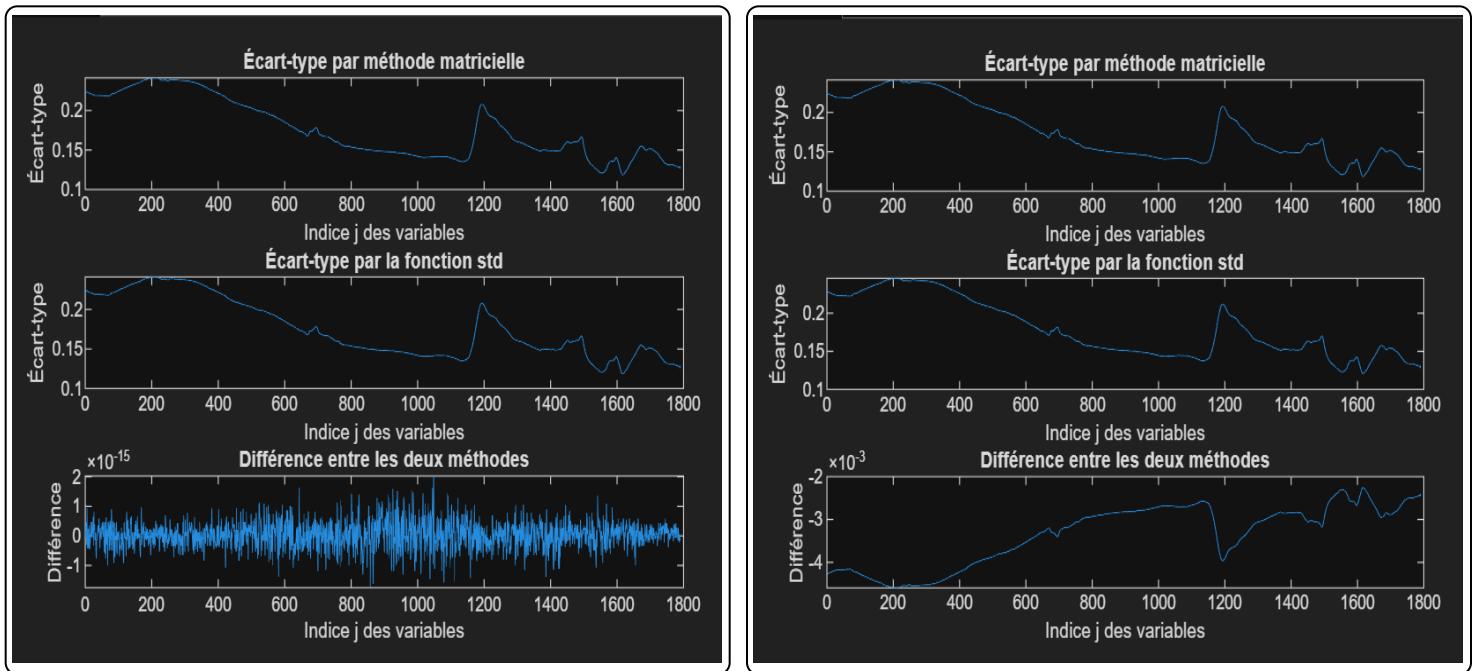
#### a) Caractéristiques de tendances centrales :



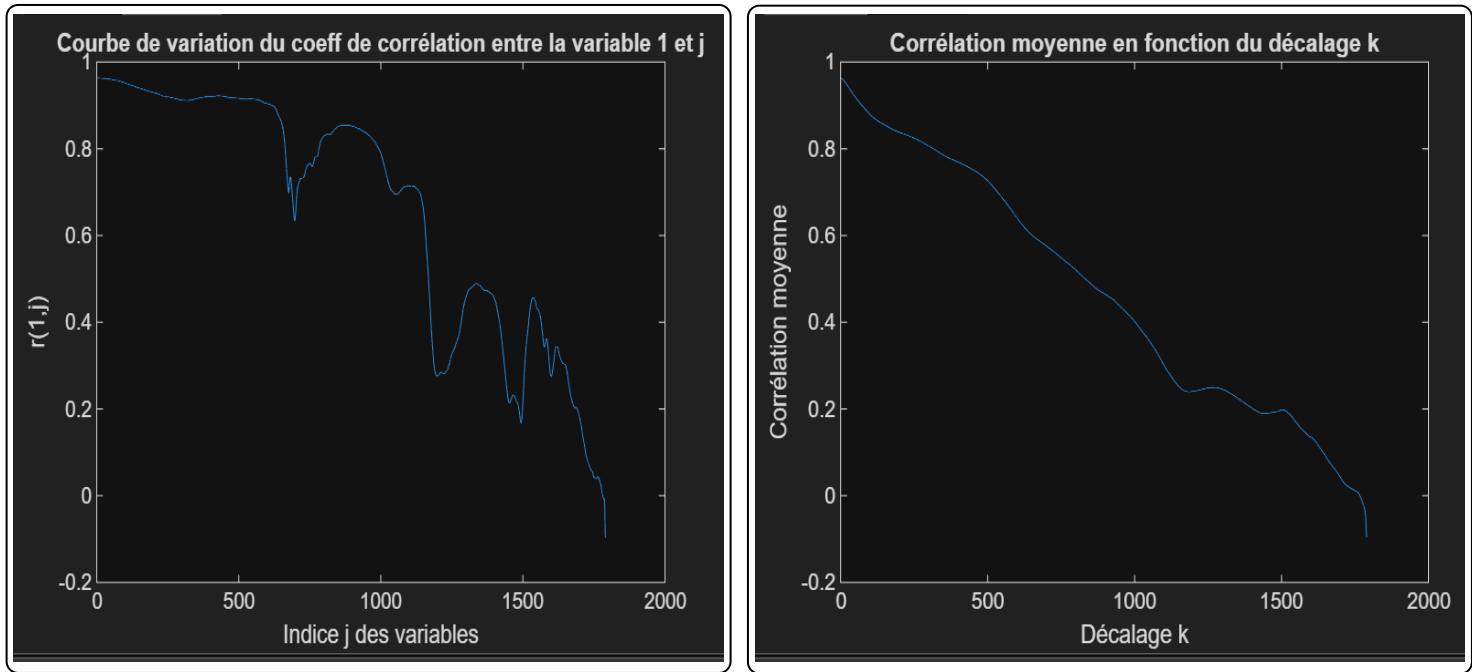
*Commentaires :* Les deux courbes sont visuellement identiques. La différence observée est de l'ordre de  $10^{16}$ , ce qui correspond uniquement aux erreurs d'arrondi liées aux calculs numériques. Cela confirme que les deux méthodes sont cohérentes.

#### b) Caractéristiques de dispersion :

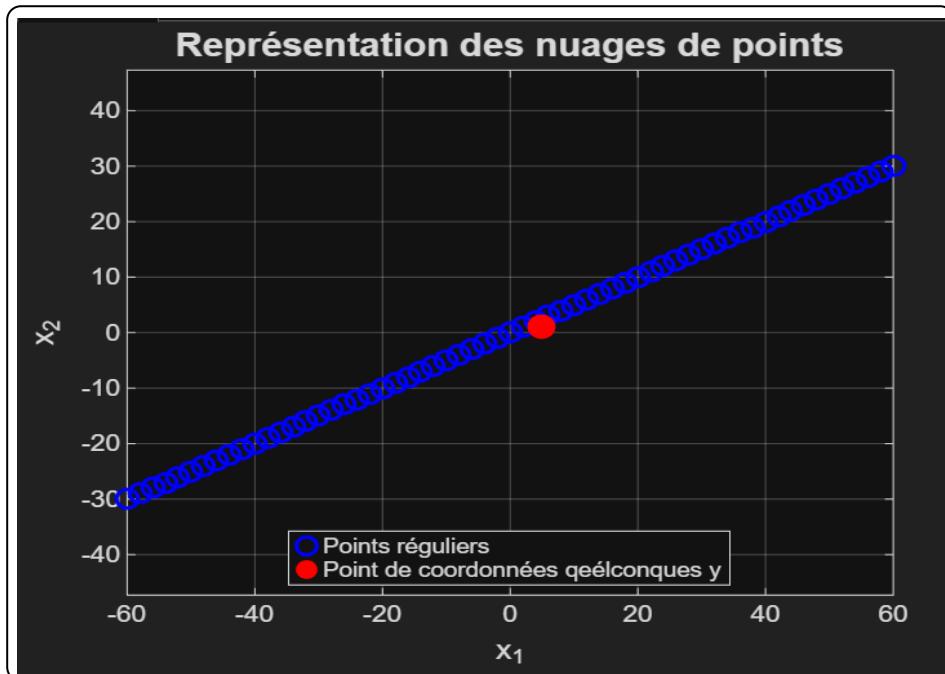
*Commentaires :* Les deux méthodes (formule matricielle et fonction std) donnent des résultats identiques, à une petite erreur numérique près ( $\approx 10^{15}$ ) due aux arrondis. Dans MATLAB, la fonction std(X) calcule par défaut l'écart-type non biaisé (division par  $n-1$ ), alors que si on utilise std(X,1) on obtient l'écart-type biaisé (division par  $n$ ), ce qui correspond à la formule matricielle.



#### d. Analyse bidimensionnelle :

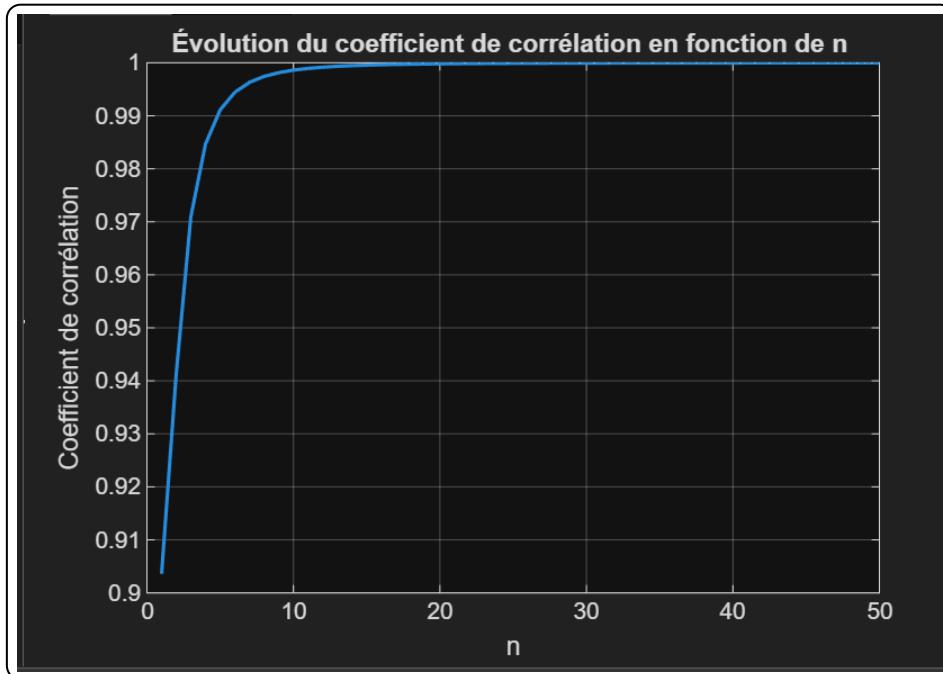


Commentaire : Ce résultat est logique, les variables dont l'indice est proche de 1 présentent un coefficient de corrélation proche de 1, ce qui traduit une forte corrélation. Plus on s'éloigne de la variable 1, plus la corrélation diminue, jusqu'à devenir faible pour les variables très éloignées.

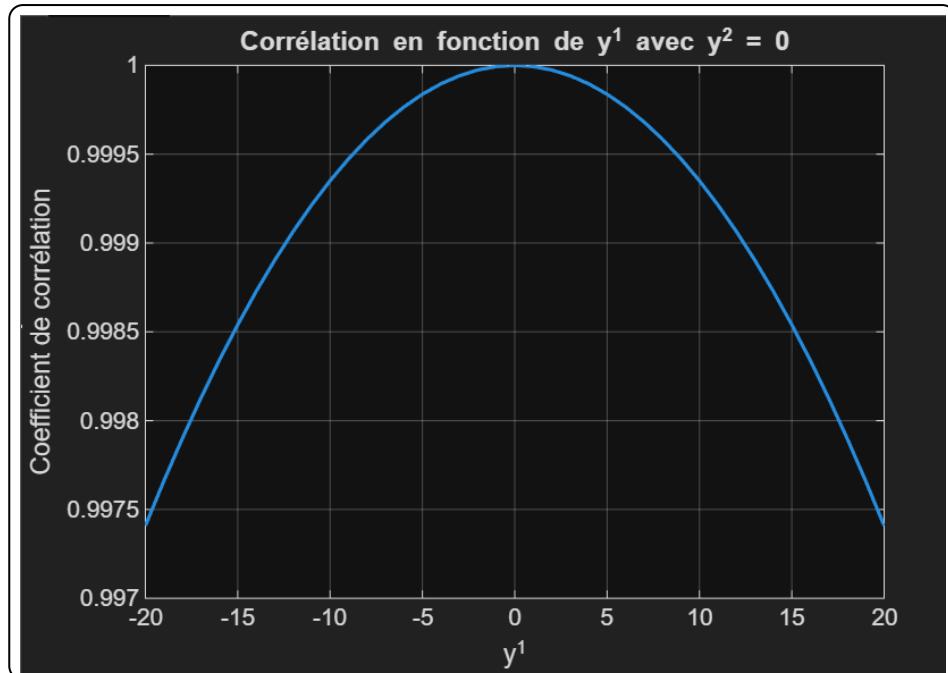
2) Deuxième étude de la réflectance de matériaux3) Etude de données synthétiques

On a choisi :

- $n=30$ ,  $\alpha_1=2$ ,  $\alpha_2=1$  (droite de pente 0.5),
- Point supplémentaire  $y_1=5$ ,  $y_2=1$  .



Commentaire : On observe que le coefficient de corrélation augmente rapidement avec n. Pour de petites valeurs de n, le point supplémentaire perturbe fortement la corrélation. Mais quand n devient grand, l'effet de ce point devient négligeable par rapport aux nombreux points alignés, et la corrélation tend vers 1.



Commentaire : La corrélation est maximale quand le point supplémentaire est sur la droite des points réguliers (ici  $y_1=0$ ), et elle diminue de façon symétrique lorsque  $|y_1|$  augmente, car le point s'éloigne de l'alignement.