

Résumé des notes de cours

Chapitre 4

A Le tableau à double entrée et le nuage de points

Le tableau à double entrée et le nuage de points permettent d'observer s'il existe un lien, appelé **corrélacion**, entre deux variables.

Intensité

Ce lien peut varier de nul à très fort en passant par faible, moyen et fort, selon l'alignement des données sur :

une diagonale (tableau à double entrée)

ou

une droite (nuage de points)

Sens

Si la diagonale est décroissante, la corrélation est positive.
Si la diagonale est croissante, la corrélation est négative.

Sens

Si la droite est croissante, la corrélation est positive.
Si la droite est décroissante, la corrélation est négative.

Tu dois être en mesure de bâtir un tableau à double entrée ainsi qu'un nuage de points, et de les interpréter par la suite, en spécifiant le sens et l'intensité de la corrélation.

Tableau à double entrée

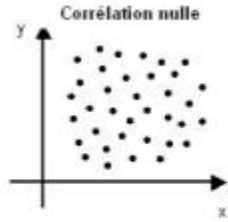
Intensité	Faible	Moyenne	Forte	Très forte
Sens				
Positive				
	Plus x augmente, plus y augmente	Plus x augmente, plus y augmente	Plus x augmente, plus y augmente	Plus x augmente, plus y augmente
Négative				
	Plus x augmente, plus y diminue	Plus x augmente, plus y diminue	Plus x augmente, plus y diminue	Plus x augmente, plus y diminue

Nuage de points

Corrélation positive

$x \nearrow y \nearrow$

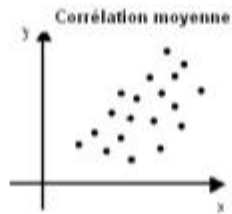
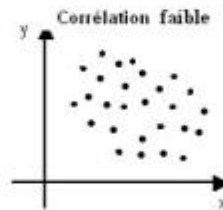
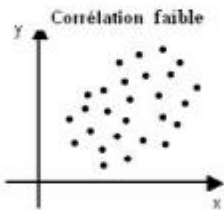
Le nuage de points est croissant, la corrélation est positive



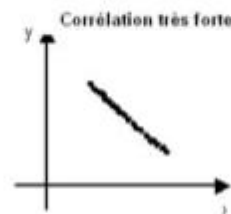
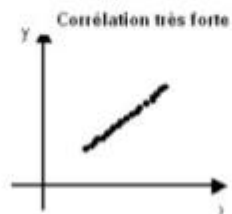
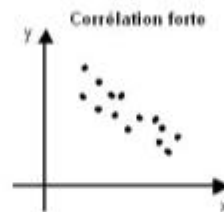
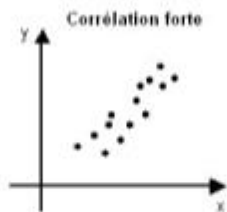
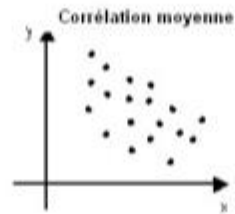
Corrélation négative

$x \nearrow y \searrow$

Le nuage de points est décroissant, la corrélation est négative



Plus le nuage de points se rapproche d'une droite, plus la corrélation est forte.



B Estimation du coefficient de corrélation (r)

$$r \approx \pm \left(1 - \frac{p}{g} \right) \quad \text{où } p \text{ est la mesure du petit côté du rectangle}$$

et g la mesure du grand côté

Qualification et interprétation du coefficient de corrélation

Corrélation positive	Description (Lien entre les deux variables)	Corrélation négative
Près de 0	Corrélation linéaire nulle	Près de 0
Près de 0,5	Corrélation linéaire faible	Près de -0,5
Près de 0,75	Corrélation linéaire moyenne	Près de -0,75
Près de 0,87	Corrélation linéaire forte	Près de -0,87
Près de 1	Corrélation linéaire très forte	Près de -1
Exactement 1	Corrélation linéaire parfaite	Exactement -1

C Droite de régression et estimation d'une valeur

La méthode de la droite de Mayer

Cette méthode permet de trouver l'équation d'une droite passant par deux points représentatifs d'une distribution.

Il faut avoir accès à la liste des données pour utiliser cette méthode.

Voici les étapes à réaliser:

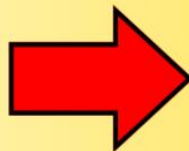
- Ordonner la distribution des données en ordre croissant selon la variable indépendante.
- Séparer la distribution en deux parties égales. Si le nombre de données est impair, la donnée centrale sera comprise dans les deux parties.
- Calculer la moyenne des données de chaque groupe pour obtenir deux points $P_1(x_1, y_1)$ et $P_2(x_2, y_2)$.
- Établir l'équation de la droite $y=ax + b$ passant par ces deux points.
- Pour estimer une valeur de **x** ou de **y**, remplacer la variable de la valeur connue dans l'équation et isoler l'autre variable.

D Écart moyen

L'écart moyen est une mesure de dispersion qui indique la moyenne des écarts de chacune des données par rapport à la moyenne de la distribution.

L'écart moyen se calcule ainsi:

$$\text{Écart moyen} = \frac{\text{somme des écarts à la moyenne}}{\text{nombre total de données}}$$



Attention!
Les écarts à la moyenne sont toujours des valeurs positives.

E Rang centile

Pour attribuer un rang centile à une donnée, il faut :

- 1) Placer les données en ordre croissant.
- 2) Calculer le nombre de données inférieures ou égales à la donnée.
- 3) Calculer le nombre total de données de la distribution.
- 4) Calculer le rang centile avec la formule :

$$R_{100}(x) = \frac{(\text{nombre de données inférieures à cette donnée} + 0,5(\text{nombre de données égales à cette donnée})) \times 100}{\text{Nombre total de données}}$$

- 5) Si le nombre n'est pas un entier, on l'arrondit à l'entier supérieur.

Pour déterminer la donnée (x) dont le rang centile est connu, il faut :

- 1) Placer les données en ordre croissant.
- 2) Calculer le nombre total de données de la distribution.
- 3) Calculer la position de la donnée:

$$X = \left(\frac{\text{Rang centile}}{100} \right) \times \text{Nombre de données}$$

- 4) Si le nombre trouvé n'est pas un entier, on l'arrondit à l'entier inférieur.
- 5) Chercher la donnée dans la distribution ordonnée située à la position trouvée(x), à partir du moins bon résultat.