

Corrélation linéaire

Régression linéaire

Élaboré par Afef Ben Zine El Abidine

Objectif du module:

- Étudier la dépendance entre deux variables quantitatives: diagramme de dispersion puis le coefficient de corrélation linéaire.
- Mesure de la force de cette dépendance : la régression linéaire et le coefficient de détermination.

Corrélation linéaire

On a deux variables quantitatives, notées X et Y.

X: variable indépendante

Y: variable dépendante

Exemple: suite à une enquête auprès de PME de la région de Lanaudière, on a obtenu l'information suivante concernant le salaire annuel (Y) de 16 cadres intermédiaires en fonction du nombre d'années d'expériences (X)

Nombre d'années (X)	13	4	6	7	5	3	12	8	10
Salaire (Y)	54 100	47 100	49 200	50 200	48 000	45 900	53 100	50 000	51 600

Questions :

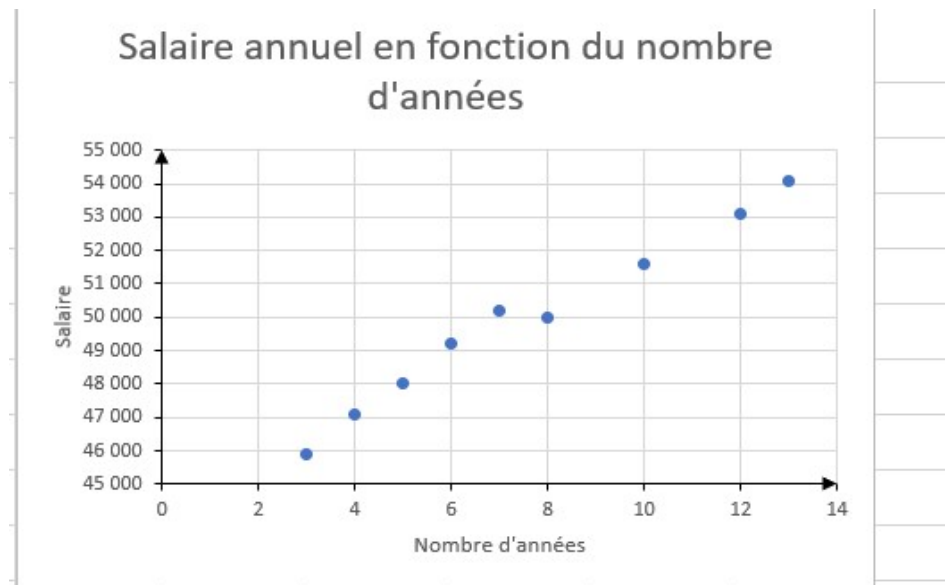
- Existe-il une relation entre X et Y?
- Si Oui, comment se fait la variation de X par rapport à celle de Y? Est-ce qu'elles ont tendance à varier toutes deux dans le même sens ou en sens contraire?
- Quelle est l'intensité de cette relation. Est-elle forte ou faible?
- Est-ce qu'on peut estimer la valeur de Y à partir de celle de X?

Diagramme de dispersion (ou nuage de points)

C'est la représentation graphique dans le plan cartésien de l'ensemble de paires de données (x, y) provenant de l'étude de deux variables quantitatives.

Pour obtenir ce graphique sur Excel:

- Sélectionner la plage de données;
- Dans Graphiques, faites Insertion Nuages de points
- Ajouter les axes et les titres du graphique

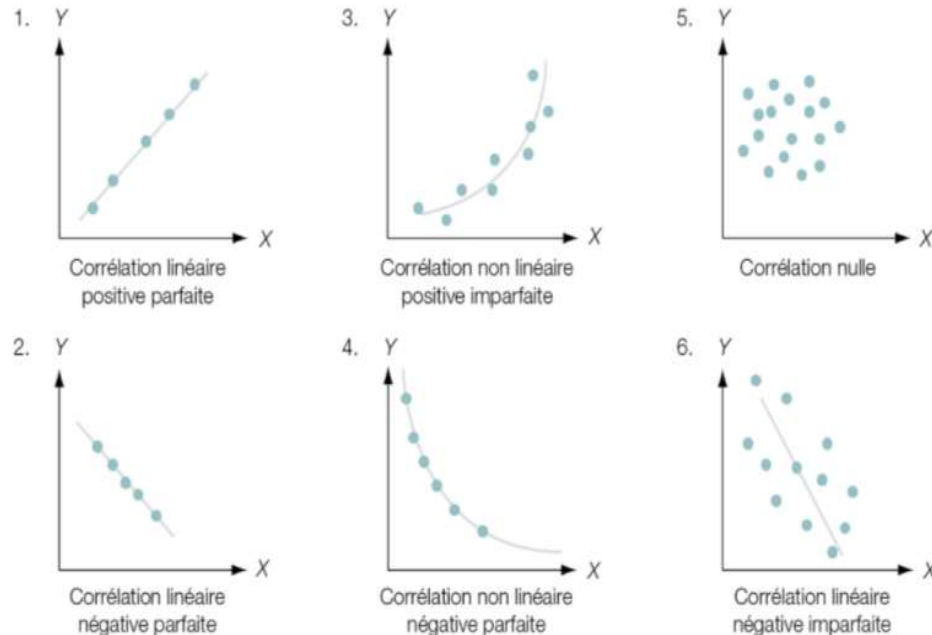


Interprétation: ?

Caractéristiques de la corrélation, p. 302

On dit qu'il y a une corrélation (dépendance) entre deux variables quantitatives X et Y si elles ont généralement tendance à varier toutes les deux dans le même sens ou en sens contraire. La corrélation entre X et Y se caractérise par : la **forme**, le **sens** et l'**intensité**.

Forme	Linéaire (p. ex. 1)
	Non linéaire (p. ex. 3)
Sens	Positif (p. ex. 1)
	Négatif (p. ex. 2)
Intensité	Parfaite (p. ex. 1)
	Imparfaite (p. ex. 6)
	Nulle (5)



Le coefficient de corrélation linéaire (noté r)

- Le diagramme de dispersion permet une analyse qualitative de la tendance à une relation entre les variables X et Y.
- Le coefficient de corrélation linéaire appelé aussi coefficient de Pearson permet de mesurer quantitativement la force de la corrélation **linéaire** entre les deux variables.

Formule du coefficient de corrélation linéaire (noté r)

Coefficient de corrélation linéaire

$$r = \frac{\sum xy - n \bar{x} \bar{y}}{(n-1)s_x s_y}$$

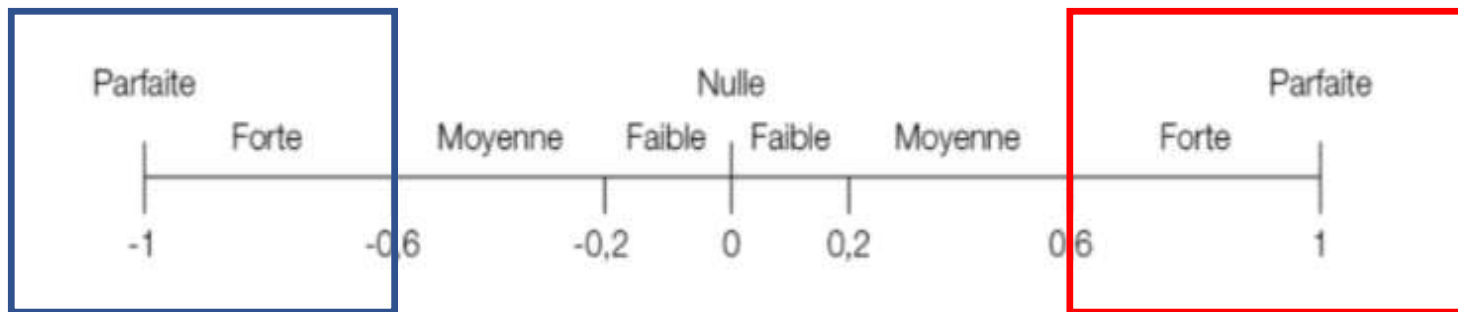
$\sum xy$ représente la somme des produits de chaque valeur de la variable X par la valeur correspondante de la variable Y

n correspond au nombre de couples (x, y)


- \bar{x} représente la moyenne des valeurs de la variable X
- \bar{y} représente la moyenne des valeurs de la variable Y
- s_x est l'écart type corrigé de la variable X
- s_y est l'écart type corrigé de la variable Y

Propriétés du coefficient de corrélation linéaire

- Sans unités
- Compris entre -1 et 1.
- Si la valeur de r proche de 1 ($>0,6$) : corrélation parfaite positive forte
- Si la valeur de r proche de -1 ($<-0,6$) : corrélation parfaite négative forte



Calcul du coefficient de corrélation exemple diapositive 3

Démarche de calcul coefficient de corrélation					
$\sum xy - n\bar{x}\bar{y}$	75244,4				
$(n - 1)s_x s_y$	76002,2				
r	0,99		0,990029362		
$r = \frac{\sum xy - n\bar{x}\bar{y}}{(n-1)s_x s_y}$ <p>r = 0,99</p> <p>Interprétation: Il y a une très forte corrélation linéaire positive entre le nombre d'années d'expériences et le salaire annuel des cadres.</p>					

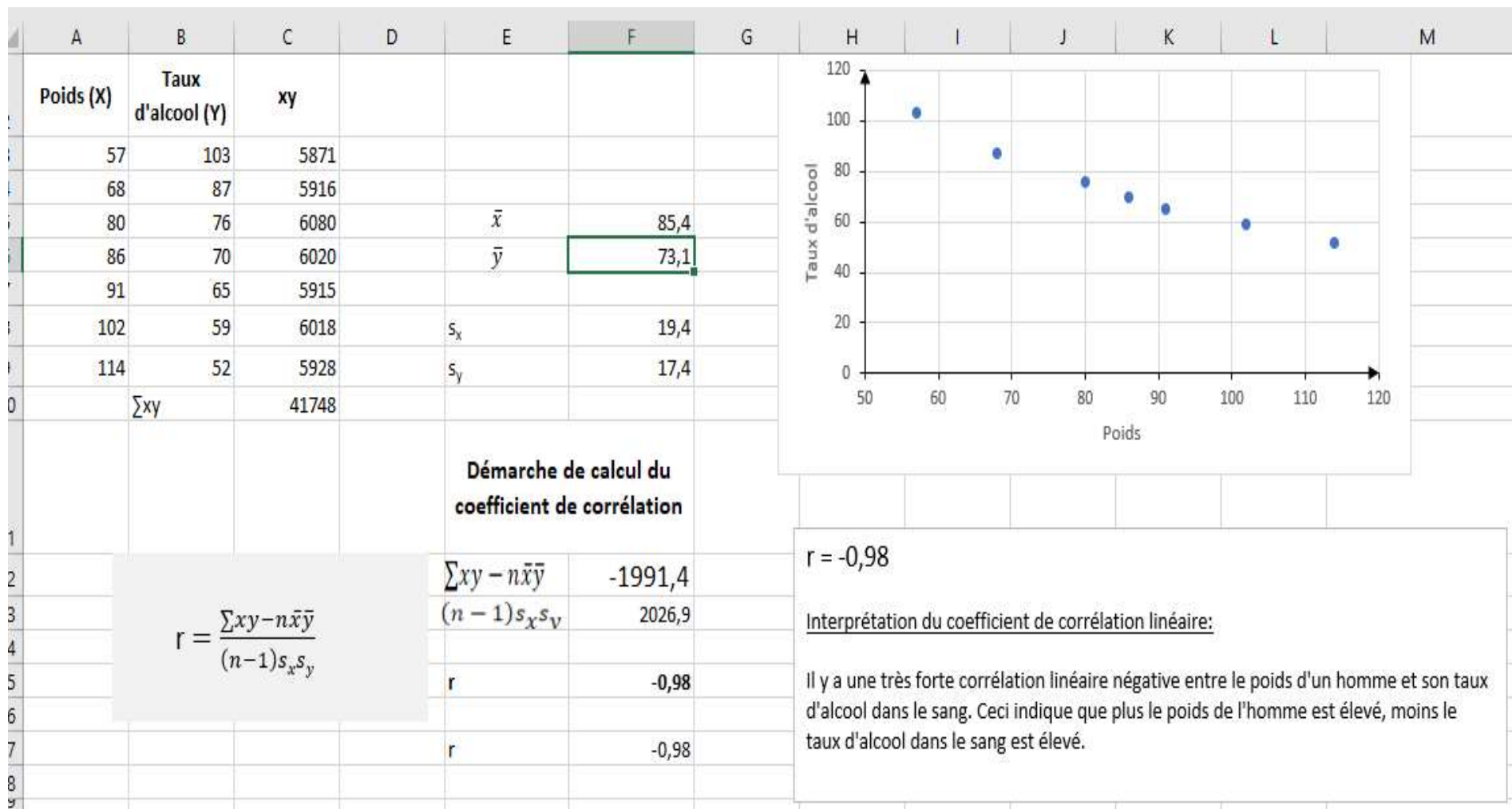
Exercice 1

Dans la mise en situation, on obtient les résultats suivants pour le taux d'alcool sanguin chez les hommes. Calculer le coefficient de corrélation entre les variables X et Y et commenter. Pour obtenir une plus grande précision, conserver au moins deux décimales dans les calculs intermédiaires.

Hommes

Taux d'alcool dans le sang en fonction du poids après la consommation de trois bières

X: Poids	57 kg	68 kg	80 kg	86 kg	91 kg	102 kg	114 kg
Y: Taux d'alcool (en mg/100 ml)	103	87	76	70	65	59	52



Régression linéaire

- Lorsque la corrélation linéaire entre X et Y s'avère significative, on peut envisager d'établir l'équation de la liaison linéaire existant entre les deux variables. On recherche la droite qui s'ajuste le mieux aux observations: **Droite de régression**.
- La droite de régression est celle qui permet de rendre minimum la somme des carrées des écarts des valeurs observées y_i à la droite (Droite déterminée selon la méthode des moindres carrés)

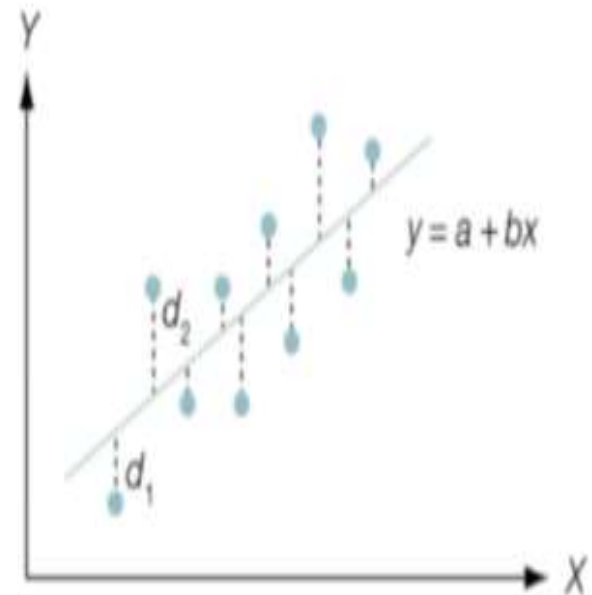
Équation de la droite de régression

Équation de la droite de régression

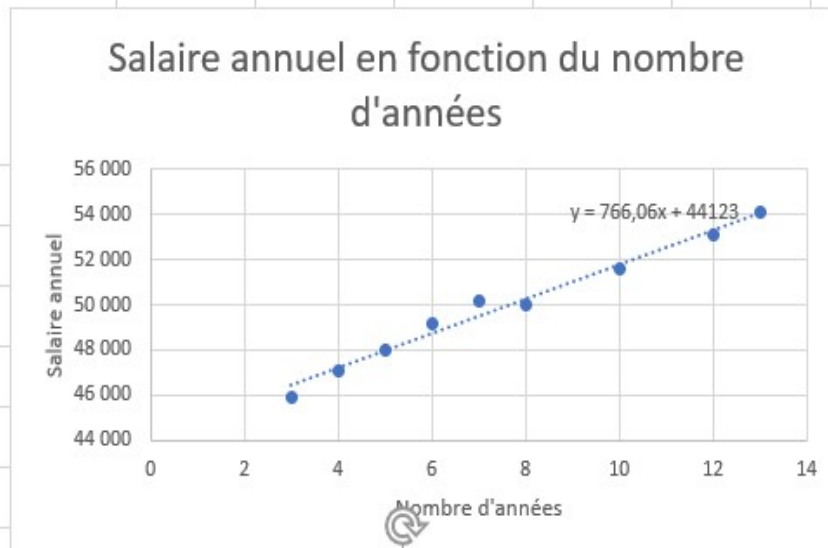
$$y = a + bx$$

On calcule les valeurs de a et de b ainsi :

$$a = \bar{y} - b\bar{x} \qquad b = \frac{\sum xy - n\bar{x}\bar{y}}{(n-1)s_x^2}$$



Droite de régression



Sélectionner courbe de tendance en appuyant sur **+** pour afficher la droite de régression

Ajouter l'option afficher l'équation sur le graphique.

Vous pouvez aussi trouver la valeur de b en sélectionnant la fonction PENTE sur Excel.

Vous pouvez aussi trouver la valeur de a en sélectionnant la fonction ORDONNEE. ORIGINE sur Excel.

La droite de régression :

$$y = a + bx$$

$$a = 44123$$

$$b = 766,06$$

b

766,063348

Le coefficient de détermination (r^2)

- C'est le coefficient de corrélation au carré.
- Il est exprimé en pourcentage.
- Il indique la part de la variation totale de la variable Y expliquée par la droite de régression.

Exercice, p. 300.

Femmes

Taux d'alcool dans le sang en fonction du poids après la consommation de trois bières

X: Poids	45 kg (100 lb)	52 kg (115 lb)	57 kg (125 lb)	68 kg (150 lb)	73 kg (161 lb)	80 kg (176 lb)	91 kg (200 lb)
Y: Taux d'alcool (en mg/100 ml)	152	133	120	101	99	87	76

Source: Éduc'alcool. Boire, conduire, choisir. L'alcool au volant. L'alcool et la loi (0.08), 2014.

1. Tracez le diagramme de dispersion
2. Donner une interprétation de la relation entre le poids des femmes (X) et le taux d'alcool dans le sang (Y) à partir du diagramme de dispersion.
3. Calculer et interpréter la valeur du coefficient de corrélation (r).
4. Représenter graphiquement la droite de régression.
5. Déterminer les coefficients de la droite de régression.
6. Calculer et interpréter la valeur du coefficient de détermination r^2

Travail à faire

- Exercices 5, 6 et 7 p. 310-311.

- Remarque pour les exercices:

Garder deux décimales pour l’affichage des coefficients de corrélation et de détermination.

- Les critères suivants sont essentiels dans les travaux:

- La présentation de la démarche pour le calcul du coefficient de corrélation.
- Exactitude des interprétations des coefficients.