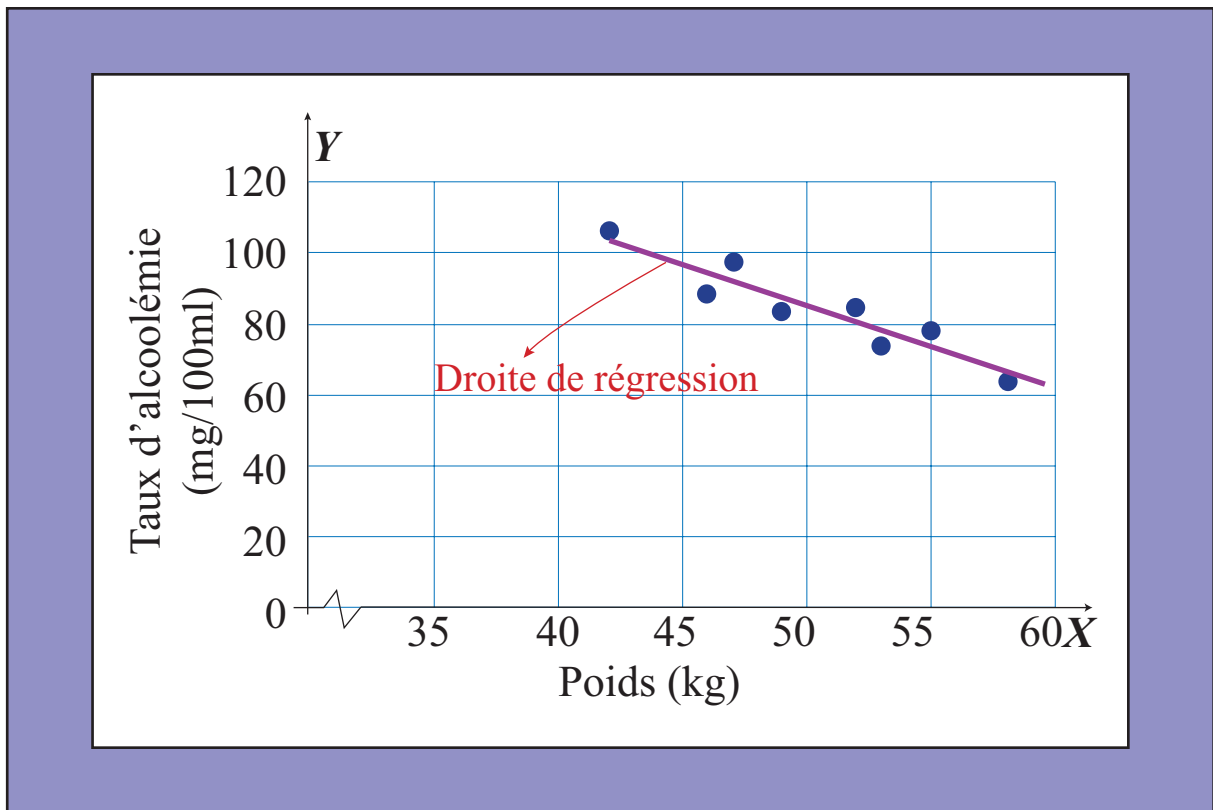


CORRÉLATION-RÉGRESSION, DROITE DE RÉGRESSION



OBJECTIFS

Construire un tableau pour représenter les données de deux variables quantitatives.

Utiliser ce tableau pour calculer le coefficient de corrélation entre les variables.

Mise en situation

Des amies décident de prendre chacune trois bières de 340 ml à 6,1 % en une heure et, une demi-heure plus tard, subir un test pour connaître leur taux d'alcool dans le sang. Le tableau suivant présente les valeurs obtenues.

Prénom	Poids (kg)	Taux (mg/100 ml)
Anaïs	42	106
Béatrice	46	88
Claudine	47	97
Doris	49	84
Elvire	52	85
France	53	74
Guylaine	55	78
Hermance	58	64

Construire un tableau pour calculer les paramètres de la droite de régression, calculer le coefficient de détermination et représenter graphiquement les données et la droite.

TABLEAU ET CHRONOGRAMME

1. On personnalise une feuille Excel que l'on enregistre sous le nom « CorrélationLab ».
2. Dans la plage A10:C10, on écrit l'en-tête du tableau, puis on entre les données dans la plage A11:C18.
3. On sélectionne la plage A10:C18 et on indique au logiciel qu'il s'agit d'un tableau interactif.
4. Dans la plage B20:C20, on fait calculer la somme des valeurs des variables. Dans la plage B21:C21, on fait calculer la moyenne des variables et on leur donne un nom pour l'utiliser dans les calculs.
5. En D10, on écrit « Produit », Excel reconnaît que l'on souhaite ajouter une colonne au tableau. En D11, on définit le produit :

$$\ll =B11*C11 \gg.$$

Excel incrémente automatiquement le produit dans les lignes du tableau.

6. On ajoute une colonne en E pour faire calculer les carrés des écarts à la moyenne,

$$\ll =(B11-moyP)^2 \gg.$$

On en fait calculer la somme en E20 et la variance corrigée en E21.

Remarque

À l'étape 4, les noms suggérés sont « moyP » pour la moyenne des poids et « moyT » pour la moyenne des taux d'alcoolémie.

7. En A23, on écrit « Pente ». En A24 on fait calculer sa valeur et on lui donne le nom « a ».

8. En A25, on écrit « Ordonnée ». En A26 on fait calculer sa valeur et on lui donne le nom « b ».

9. Pour calculer le coefficient de corrélation, on ajoute une colonne en F pour faire calculer les carrés des écarts des taux à la moyenne,

$$\ll = (C11 - \text{moyT})^2 \gg.$$

On en fait calculer la somme en F20 et la variance corrigée en F21.

10. En E22 et F22, on fait calculer les écart-types corrigés des variables.

11. En D24, on fait calculer le coefficient de corrélation et en F24, le coefficient de détermination.

12. On fait calculer les valeurs que prévoit le modèle en définissant en G11

$$\ll = a * B11 + b \gg.$$

13. On sélectionne la plage A11:B18 et on fait représenter graphiquement par un nuage de points. On fait ajouter au graphique les données de la plage G11:G18.

Remarque

La valeur des paramètres de la droite de régression sont calculées à l'aide des valeurs obtenues aux étapes précédentes en les substituant dans les expressions :

$$a = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{(n-1) s_x^2} \quad \text{et} \quad b = \bar{y} - a \bar{x}.$$

Pour la pente a , on inscrit en A24

$$\ll = (D20 - 8 * \text{moyP} * \text{moyT}) / (7 * E21) \gg.$$

Pour l'ordonnée à l'origine b , on inscrit en A26

$$\ll = \text{moyT} - (A24 * \text{MoyP}) \gg.$$

Le calcul du coefficient de corrélation se fait à l'aide des valeurs obtenues aux étapes précédentes en les substituant dans l'expression :

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

soit $\ll = (D20 - 8 * \text{moyP} * \text{moyT}) / (7 * E22 * F22) \gg.$

Remarque

On constate une différence entre les valeurs observées et celles obtenues par le modèle. Puisque le coefficient de détermination est 0,878, cela signifie que la variation du taux d'alcoolémie dépend à 87,8% du poids des participantes.

EXERCICES

Dans une enquête sur les pratiques culturelles au Québec en 2014, le Bulletin de la recherche et de la statistique donne les tableaux suivants. À l'aide du logiciel Excel, construire un tableau pour calculer les paramètres de la droite de régression pour chacune des versions.

a) **Taux de lecture régulière de journaux quotidiens**

Âge	Papier (%)	Numérique (%)
[15; 24[52	46
[15; 24[52	56
[15; 24[48	55
[15; 24[56	45
[15; 24[64	40
[15; 24[69	32
[15; 24[77	19

Source : Bulletin de la recherche et de la statistique. Enquête sur les pratiques culturelles au Québec, 2014.

b) **Heures de visionnement de contenu télévisuel**

Âge	Télévision	Internet
[15; 24[8,4	15,2
[15; 24[11,5	15,2
[15; 24[11,1	14,0
[15; 24[14,2	16,1
[15; 24[17,7	19,4
[15; 24[21,9	23,1
[15; 24[24,6	25,7

Source : Bulletin de la recherche et de la statistique. Enquête sur les pratiques culturelles au Québec, 2014.

c) **Écoute de films, cinéma et internet**

Âge	Cinéma (%)	Internet (%)
[15; 24[92	51
[15; 24[81	60
[15; 24[79	51
[15; 24[69	55
[15; 24[62	59
[15; 24[55	50
[15; 24[38	40

Source : Bulletin de la recherche et de la statistique. Enquête sur les pratiques culturelles au Québec, 2014.