

Statistiques

Droite de régression

Il existe plusieurs méthodes pour estimer une droite de régression.

Méthode graphique

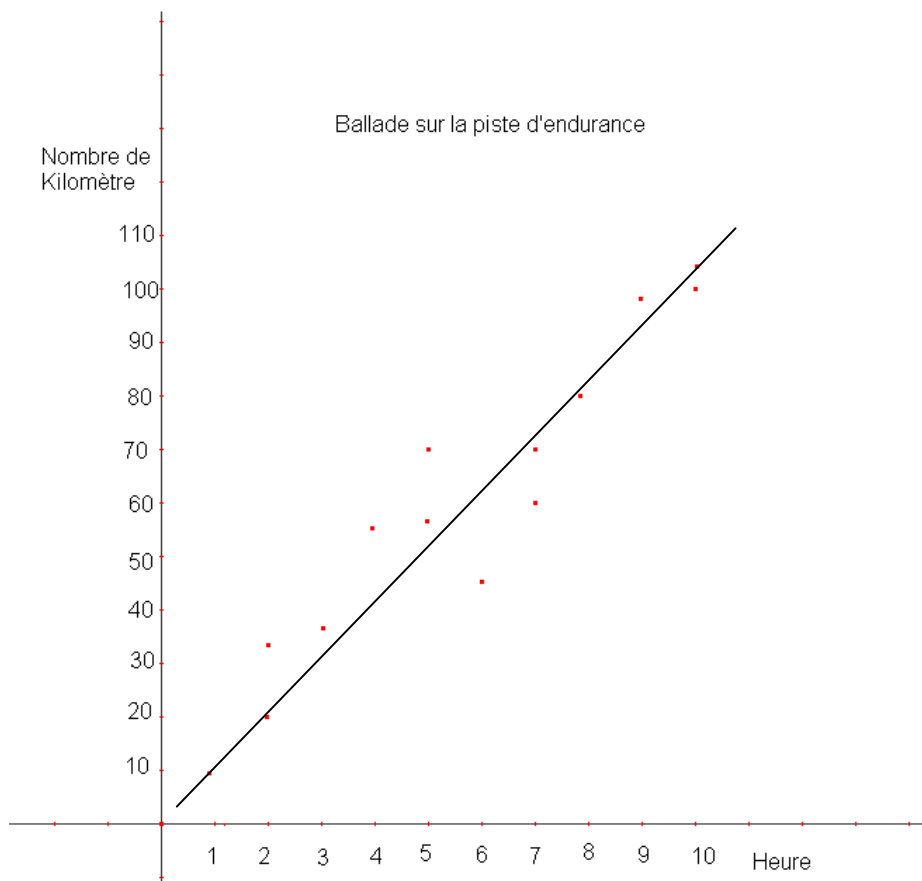
Il suffit de tracer une droite ayant le même nombre de point de chaque côté et nous prenons les coordonnées aux extrémités de cette droite.

Exemple 1 :

Voici le nombre de kilomètre parcourus par des cyclistes en fonction du nombre d'heure passé sur la piste durant une course d'endurance.

La première coordonnée représente le nombre d'heure et la deuxième coordonnée représente le kilométrage parcouru.

(1,9) (6,45) (2,20) (5,71) (5,57) (9,98) (4,56) (10,104) (3,36) (8,75)(7,60) (7,71) (2, 31) (10, 97)



Prenons les coordonnées (2, 20) et (10, 104)

Statistiques

Trouvons l'équation de cette droite.

L'équation de la droite est

$$y = ax + b$$

Trouvons la pente (taux de variation)

$$a = \frac{y_2 - y_1}{x_2 - x_1} = \frac{104 - 20}{10 - 2} = \frac{84}{8} = 10,5$$

$$y = 10,5x + b$$

Pour trouver le paramètre b, utilisons la coordonnée (2, 20)

$$20 = 10,5 \cdot (2) + b \rightarrow 20 = 21 + b \rightarrow 20 - 21 = b \rightarrow b = -1$$

L'équation de la droite de régression est $y \approx 10,5x - 1$

Un cycliste qui roule pendant 13 heures aura parcouru environ 135,5 km.

$$y \approx 10,5x - 1$$

$$y \approx 10,5(13) - 1$$

$$y \approx 135,5$$

Attention!

La marge d'erreur reste assez élevée, car la droite que tu auras dessinée ne sera probablement pas la même que celle d'une autre personne.

Statistiques

Méthode médiane-médiane

Pour estimer la droite de régression à l'aide de la méthode médiane-médiane, il suffit de suivre les étapes suivantes :

- Placer les coordonnées en ordre croissant des abscisses et créer trois groupes égaux. Il doit toujours y avoir le même nombre de coordonnées dans le premier et le dernier groupe.
- Pour chacun des groupes, trouver l'abscisse médiane et l'ordonnée médiane que l'on notera $M_1(x_1, y_1)$, $M_2(x_2, y_2)$, $M_3(x_3, y_3)$.
- Trouver un point P qui correspond à la moyenne des abscisses et à la moyenne des ordonnées en utilisant les trois points M_1 , M_2 et M_3 .
- Il vous reste à trouver la droite de régression passant par le point P et en utilisant les points M_1 et M_3 pour trouver la pente.

Voici les 14 coordonnées en ordre croissants des abscisses.

(1,9)
(2,20)
(2, 31)
(3,36)
(4,56)
(5,57)
(5,71)
(6,45)
(7,60)
(7,71)
(8,75)
(9,98)
(10, 97)
(10,104)

$M_1(2, 31)$

$M_2(5,5; 58,5)$

$P(5,5; 62,17)$

$M_3(9, 97)$

Pour l'ordonnée (y),
classé en ordre
croissant avant de
trouver la médiane.
Observez bien M_2 et
 M_3 .

L'équation de la droite est

$$y = ax + b$$

Trouvons la pente (taux de variation)

Utilisons les coordonnées $M_1(2,31)$ et $M_3(9, 97)$

$$a = \frac{y_2 - y_1}{x_2 - x_1} = \frac{97 - 31}{9 - 2} = \frac{66}{7} = 9,43$$

$$y = 9,43x + b$$

Statistiques

Pour trouver le paramètre b, utilisons la coordonnée P(5,5; 62,17)

$$62,17 = 9,43 \cdot (5,5) + b \rightarrow 62,17 = 51,87 + b \rightarrow 62,17 - 51,87 = b \rightarrow b = 10,3$$

L'équation de la droite de régression est $y \approx 9,43x + 10,3$

Un cycliste qui roule pendant 13 heures aura parcouru environ 132,89 km.

$$y \approx 9,43x + 10,3$$

$$y \approx 9,43(13) + 10,3$$

$$y \approx 132,89$$

Statistiques

Méthode de la droite de Mayer

Pour estimer la droite de régression à l'aide de la méthode de Mayer, il suffit de suivre les étapes suivantes :

- Placer les coordonnées en ordre croissant des abscisses et créer deux groupes égaux (si possible).
- Pour chacun des deux groupes, trouver la moyenne des abscisses et la moyenne des ordonnées que l'on notera $P_1(x_1, y_1)$ et $P_2(x_2, y_2)$.
- Il vous reste à trouver la droite de régression passant par ces deux points.

Voici les 14 coordonnées en ordre croissants des abscisses.

(1,9)	
(2,20)	
(2, 31)	
(3,36)	
(4,56)	
(5,57)	
(5,71)	
(6,45)	
(7,60)	
(7,71)	
(8,75)	
(9,98)	
(10, 97)	
(10,104)	

$P_1(3,14; 40)$

$P_2(8,14; 78,57)$

L'équation de la droite est

$$y = ax + b$$

Trouvons la pente (taux de variation)

Utilisons les coordonnées $P_1(3,14; 40)$ et $P_2(8,14; 78,57)$

$$a = \frac{y_2 - y_1}{x_2 - x_1} = \frac{78,57 - 40}{8,14 - 3,14} = \frac{38,57}{5} = 7,71$$

$$y = 7,71x + b$$

Pour trouver le paramètre b, utilisons la coordonnée $P_1(3,14; 40)$

$$40 = 7,71 \cdot (3,14) + b \rightarrow 40 = 24,21 + b \rightarrow 40 - 23,96 = b \rightarrow b = 15,79$$

L'équation de la droite de régression est $y \approx 7,71x + 15,79$

Statistiques

Un cycliste qui roule pendant 13 heures aura parcouru environ 116,02 km.

$$y \approx 7,71x + 15,79$$

$$y \approx 7,71(13) + 15,79$$

$$y \approx 116,02$$

Résumons :

Un cycliste qui roule pendant 13 heures aura parcouru selon les trois méthodes suivantes :

Méthode graphique $y \approx 135,5$

Méthode médiane-médiane $y \approx 132,89$

Méthode de Mayer $y \approx 116,02$

Le coefficient de corrélation de ce problème est environ **0,81**. La corrélation est donc **positive et moyenne forte**. Cela explique l'écart entre la méthode médiane-médiane et la méthode de Mayer. Cette dernière utilise la moyenne et plus les écarts sont grands, plus la précision diminue.

