

R216 TP2

Covariance : mesure si deux variables augmentent ou diminuent ensemble et si elles semblent dépendantes ou indépendantes.

Corrélation : chiffre entre -1 et 1 qui indique la force du lien linéaire entre deux variables, proche de 1 ou -1 le lien est fort et proche de 0 le lien est faible.

Régression linéaire : droite qui modélise au mieux la relation entre deux variables pour pouvoir prédire une valeur inconnue.

Coefficient de détermination R^2 : nombre entre 0 et 1 qui mesure à quel point la droite de régression représente bien les données, plus c'est proche de 1, mieux c'est.

EXERCICES

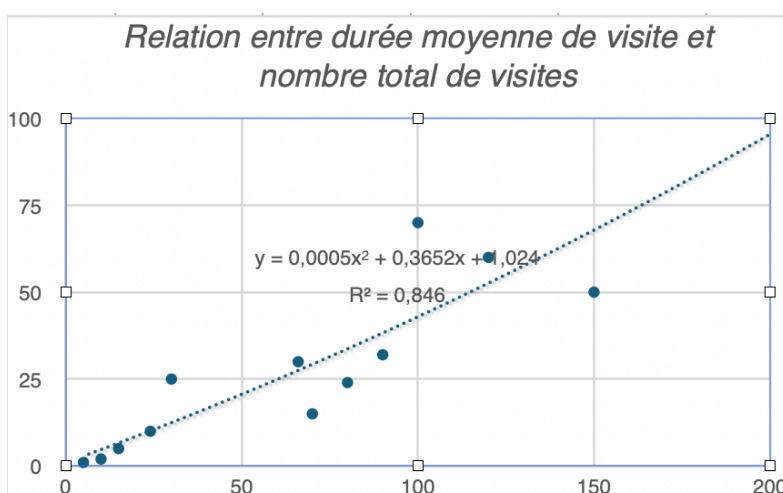
Question 1

La covariance est de 1485,325444.

Le coefficient de corrélation est de 0,917435115, ce qui signifie que quand le nombre total de visites augmente la durée moyenne augmente aussi. Le lien linéaire entre les deux variables est très fort.

Le coefficient de détermination est de 0,84168719, ça veut dire que environ 84 pour cent de la variation de la durée moyenne est expliquée par le nombre de visites. Le modèle linéaire est donc très fiable pour représenter nos données.

Question 2



Graphique : Relation entre durée moyenne de visite et nombre total de visites

Pour cette partie, pense à modifier la courbe de tendance sur ton graphique Excel en choisissant l'option Linéaire pour afficher la vraie droite qui correspond à tes calculs.

Interprétation et prédiction :

Plus le nombre de visites est élevé, plus la durée moyenne prédite est élevée.

L'équation de la droite de régression linéaire est sous la forme $y = ax + b$

$$y = 0,4643x - 1,6699$$

Pour $x = 500$ on a :

$$y = 0,4643 * 500 - 1,6699 = 232,15 - 1,6699 = 230,4801$$

On peut arrondir et dire que pour 500 visites la durée moyenne de visite prédite par le modèle linéaire est d'environ 230,5 à 231 secondes.

Question 3

Pour les modèles non linéaires, le modèle logarithmique est le plus adapté car c'est celui qui a le R^2 le plus élevé, 0,8592, ce qui montre que la relation entre la durée moyenne et le nombre de visites suit plutôt une courbe logarithmique et pas exactement une droite.

Question 4

On imagine que les 13 valeurs de durée ont été relevées jour après jour dans l'ordre donné.

La moyenne glissante sur 5 jours consiste à calculer pour chaque position la moyenne des 5 jours précédents en se décalant d'une case à chaque fois. Dans Excel, on se place sur la cellule du cinquième jour et on utilise une formule de type MOYENNE sur les 5 premières cellules :

$$=MOYENNE(B1:F1)$$

Puis on décale la plage d'une colonne à chaque fois vers la droite pour obtenir toutes les valeurs. La formule pour la cellule d'après devient donc :

$$=MOYENNE(C1:G1)$$

Avec les données du tableau on obtient les moyennes glissantes suivantes pour la durée moyenne en secondes sur 5 jours :

59,8 48,8 84 85 91 81 91,2 69,2 87,2

Question 5

Sur le graphique en fonction du temps, on ajoute une courbe de tendance de type moyenne mobile avec une période de 5. On voit que les valeurs de cette courbe coïncident parfaitement avec les valeurs de la moyenne glissante

calculées dans le tableau. Cela permet de lisser les variations quotidiennes et de mieux observer la tendance générale de l'évolution du temps de visite.