

TP4 - Régression linéaire

Exercice 1 (à faire sous R).

On dispose de données concernant l'âge (age), le kilométrage en milliers de kilomètres (km) et le prix en milliers d'euros (prix) pour un échantillon de voitures d'occasion du même type.

age	5	4	6	5	5	5	6	6	7	7
km	92	64	124	97	79	76	93	63	111	143
prix	7.8	9.5	6.4	7.5	8.1	9	6.1	9.7	6.4	4.4

On s'intéresse aux couples de variables (age, prix) et (km, prix). Le but étant, si le lien est fort, de modéliser la variable prix en fonction d'une des deux variables explicatives km ou age.

1. Quelle est la première chose à effectuer avant tout calcul ? Effectuez-la et commenter le résultat obtenu.
2. Quelle est la deuxième étape à franchir ?
3. En fonction des résultats précédemment obtenus, choisir la variable explicative à garder puis :
 - (a) Effectuer la régression linéaire sous R. Tracer la droite de régression sur le nuage de points.
 - (b) Donner l'équation de la droite de régression et interpréter ses coefficients.
 - (c) Vérifier numériquement et graphiquement que la droite de régression passe par le point central des données (dit centre de gravité).
 - (d) Peut-on s'appuyer sur le modèle pour faire de la prédiction ?
Si oui, déterminer le prix que l'on peut prédire pour une voiture dont le kilométrage est de 150.000 km.

Exercice 2 (avec R).

On cherche à étudier l'évolution du montant des achats réalisés en ligne par les ménages français. Le tableau ci-dessous donne une estimation du montant de ces achats sur sept années consécutives, de 1998 à 2004, codées par les rangs 0 à 6 :

Rang de l'année	0	1	2	3	4	5	6
Montant d'achats en millions d'euros	75	260	820	1650	2300	4000	5300

1. (a) Pour répondre au but de l'étude, quelle variable devons-nous identifier comme variable explicative ? Comme variable à expliquer ?
 (b) Tracer le nuage de points. Commenter.
 (c) Donner le coefficient de corrélation linéaire entre x et y . Interpréter le résultat obtenu.
 (d) Donner l'équation de la droite de régression de y en x
 (e) Tracer la droite de régression (en bleu) sur le nuage de points.
 (f) Quelle prévision du montant d'achats peut-on faire pour l'année 2005 ? Est-elle fiable ?
2. On considère la nouvelle variable réponse $z = \sqrt{y}$.
 (a) Répondre aux questions b) à f) de la question 1.
 (b) En déduire une expression de y en fonction de x , puis une prévision du montant d'achats pour l'année 2005.