

TP4

Régression linéaire multiple (1) : premiers pas

Objectifs du TP

- Réaliser et interpréter une analyse descriptive de données quantitatives multivariées.
- Interpréter des résultats obtenus par régression linéaire simple.
- Mettre en œuvre une régression linéaire multiple.

Consommation de glaces

On s'intéresse à la consommation hebdomadaire de glaces aux États-Unis, sur 30 semaines réparties entre le 18 Mars 1950 et le 11 Juillet 1953. Le jeu de données 'glaces.txt' contient, pour chaque semaine d'observation :

- **Periode** : le numéro de la semaine (un entier entre 1 et 30),
- **Conso** : la consommation (exprimée en pintes par habitant),
- **Prix** : le prix des glaces (exprimé en dollars),
- **Salaire** : le salaire hebdomadaire (exprimé en dollars),
- **Temp** : la température (exprimée en degrés Fahrenheit).

Analyse descriptive

1. Importer les données 'glaces.txt'.
2. Faire une analyse descriptive des données.
R propose un large choix pour cela : résumés numériques (`summary`), histogrammes (`hist`), boxplots (`boxplot`), fonction de répartition (`ecdf`), graphes pair à pair (`pairs`), matrice de corrélation (`cor`), clustering hiérarchique (avec la fonction `heatmap`), etc.
Parmi toutes ces possibilités, sélectionnez un petit nombre de sorties vous permettant de dire quelque chose d'intéressant.

Quelques modèles univariés

3. Essayer d'expliquer la consommation de glaces à l'aide de la température uniquement. La température a-t-elle un effet sur la consommation de glaces ? En plus de la réponse à cette question, sélectionner deux résultats qui vous semblent intéressants et les interpréter brièvement.

Note : si cela est judicieux, on pourra mettre en œuvre toutes les idées vues lors des trois premières séances de TP.

4. Même question en remplaçant la température par le prix des glaces.
5. Parmi les quatre covariables disponibles, laquelle explique le mieux la consommation de glaces ? Interpréter.

Régression multiple

On considère maintenant un modèle de régression linéaire qui explique la consommation de glaces à l'aide de toutes les covariables : **Période**, **Prix** et **Salaire** et **Temp**. On appelle ce modèle « le modèle complet ».

6. Écrire ce que signifie (mathématiquement) le modèle complet.
Que vaut p , le nombre de covariables, pour le modèle complet ?
7. Ajuster les données au modèle complet. Interpréter brièvement le résultat obtenu.
8. Comparer la qualité du modèle complet aux modèles « simples » considérés aux questions 3 à 5. Interpréter.