TP 2

Statistique
Statistiques à deux variables

feuille 2/2

Mots clés : statistisques à deux variables, nuage de points, variance, covariance, écart-type, corrélation, ajustement affine, régression linéaire, méthode des moindres carrés, variance expliquée et variance résiduelle.

On s'intéresse dans cette application à la construction d'un modèle permettant de relier des informations structurales de molécules organiques (les alcanes) à leur température d'ébullition. Un des concepts fondamentaux de la chimie est la dépendance entre les caractéristiques structurales d'une molécule et ses propriétés physico-chimiques. Ceci a été mis en évidence dès le milieu du siècle et a permis de construire des modèles de structure-propriétés.
Come descripteurs, on s'intéresse ici, aux nombre d'atomes de carbone (NA) à l'indice Hosoya (H) et à l'indice de Kier d'ordre 1 (X1). Une sélection d'alcanes avec leurs points d'ébullition ainsi que les valeurs des descripteurs, est dans un fichier joint (alcanes.csv).

1. Importer les données en utilisant les commandes suivantes :
   — import pandas as pd
   — data = $pd.read\_csv('alcanes.csv', sep=';', engine='python')$
   — PE = data.iloc[:,2], NA = data.iloc[:,3], H = data.iloc[:,4], X1 = data.iloc[:,5]
2. (a) Tracer le nuage de point présentant les points d'ébullition en fonction de l'indice Hosoya H.
   (b) Étudier la corrélation entre ces deux variables. Conclure.
   (c) Tracer la droite de régression linéaire de PE en fonction de H
3. Reprendre les questions précédentes en faisant une étude sur « les points d'ébullition en fonction de l'indicede Kier d'ordre 1 ($X1$) ».
4. Faire travailler votre imagination pour d'autres études. Commenter.
5. (a) Construire un modèle linéaire pouvant approximer les points d'ébullitions des alcanes en fonction des autres paramètres. Ceci revient à chercher un modèle sous la forme

$$PE = \alpha\ NA + \beta\ H + \gamma\ X_1 + \gamma,$$

   où α, β, γ et δ sont des réels à déterminer.
   (b) Le modèle parait-il convenable ?

---

1. École d'Ingénieurs de l'Air et de l'Espace et de la Mobilité Durable, 94200 Ivry-sur-Seine