Enseignant : Sofiane MAAZI

TP2 IS2A3 : Manipulations de base (Python)

24 mars 2021

Exercice 1 : Températures

1. Écrivez une fonction :
   (a) *Fahrenheit(TC)* qui convertit des degrés Celsius en degrés Fahrenheit.
   (b) *Celcius(TF)* qui convertit des degrés Fahrenheit en degrés Celsius.
2. Vérifiez que *Celsius(Fahrenheit(TC))* renvoie bien TC.

Rappel : TF = 32 + 1,8 × TC.

Exercice 2 : Analyse d'un jeu de données

Généralités

1. Installez les modules pandas, matplotlib, scipy et numpy.
2. Importez la table de données.
3. Donnez le nombre d'individus statistiques et de variables.
4. Donnez une commande équivalente à la fonction *str()* de R.
5. Donnez la liste contenant le nom de chacune des variables de la table.

Variables quantitatives

1. Décrivez simultanément toutes les variables quantitatives du jeu de données.
2. Décrivez la variable *Age*.
3. Calculez la moyenne, la médiane, la variance , l'écart-type, le minimum, le maximum , le mode, les quartiles et des quantiles de la variable *Age*.
4. Représentez graphiquement la variable *Age* par un histogramme.
5. Représentez graphiquement la variable *Age* par un diagramme à moustaches.

Variables qualitatives

1. Décrivez la variable *Club*.
2. Donnez les modalités de la variable *Club*.
3. Donnez l'effectif de chacune des modalités de *Club*.
4. Donnez la table des occurences de *Club*.
5. Représentez graphiquement la variable *Preferred Foot* par un diagramme en barres.
6. Représentez graphiquement la variable *Preferred Foot* par un diagramme circulaire.

Croisement de variables

1. Donnez le tableau de croisement des variables *Preferred Foot* et *Real Face*.
2. Représentez sur un même graphique les variables *Preferred Foot* et *Real Face*.
3. Existe-t-il un lien entre *Preferred Foot* et *Real Face* ?

Exercice 3 : Alignements de points

Soit une suite finie $X = \{x_1, x_2, ..., x_n\}$ suposée triée et une suite finie $Y = \{y_1, y_2, ..., y_n\}$ des valeurs d'un phénomène modélisé par une relation $\forall i \in [|1; n|]$, $y_i = ax_i + b$. Le but est de trouver les valeurs de a et b avec la meilleure approximation possible.

Méthode de Wald

1. calculer la médiane $m = \left\lfloor \frac{n}{2} \right\rfloor$
2. séparer X en 2 listes $X_G = \{x_1, x_2, ..., x_m\}$ et $X_D = \{x_{m+1}, ..., x_{n-1}, x_n\}$.
3. séparer Y en 2 listes $Y_G = \{y_1, y_2, ..., y_m\}$ et $Y_D = \{y_{m+1}, ..., y_{n-1}, y_n\}$.
4. calculer la moyenne de chacune des 4 listes notées M_{XG}, M_{XD}, M_{YG} et M_{YD} .
5. déterminer $y = ax + b$ telle que $M_{YG} = a.M_{XG} + b$ et $M_{YD} = a.M_{XD} + b$ (droite qui passe par les 2 barycentres).

Questions :

- Écrire une fonction *separer2(U)* séparant une liste U en 2 listes U_G et U_D comme susmentionné.
- En déduire une fonction *coeffW(X, Y)* qui calcule a et b à partir des listes X et Y passées en paramètres.

La méthode de Wald repose sur le calcul de moyennes. Or la moyenne n'est pas une statistique résistante. La méthode de Wald est donc perturbée par la présence de points (x_i, y_i) aberrants. Un moyen classique de gérer ces points est d'utiliser un calcul de médiane (statistique résistante) plutôt que la moyenne. C'est le principe de la méthode de Tuckey.

Méthode de Tuckey

1. séparer les listes X et Y en 3 parties quasi-égales selon les indices. X est séparée en X_1, X_2 et X_3. Y est séparée en Y_1, Y_2 et Y_3 (avec X_1et Y_1 de même longueur ainsi que X_2 et Y_2 et que X_3 et Y_3).
2. calculer les médianes des 6 parties notées x_1, x_2, x_3, y_1, y_2, y_3.
3. considérer la droite passant par (x_1, y_1) et (x_3, y_3) d'équation $y = ax + b_1$.
4. considèrer la droite de même pente a passant par (x_2, y_2) d'équation $y = ax + b_2$.
5. la droite de Tukey admet $y = ax + b$ avec pour équation avec b= $\frac{b_1+b_2}{2}$.

Questions :

- Écrire une fonction *separer3(U)* séparant une liste U en 3 listes U_G ,U_M et U_D comme susmentionné.
- En déduire une fonction *coeffT(X, Y)* qui calcule a et b à partir des listes X et Y passées en paramètres.

Méthode des moindres carrés

On considère X et Y de même taille n.

La méthode des moindres carrés consiste à déterminer a et b tels que $\sum_{i=1}^{n}(y_i-(ax_i+b))^2$ est minimale.

On admet que la solution est obtenue par $a = \frac{cov(X,Y)}{var(X)}$ et $b = moy(Y) - a.moy(X)$ avec :

1. moy(X)=$\frac{1}{n}\sum_{i=1}^{n} x_i$
2. var(X)=$\frac{1}{n}\sum_{i=1}^{n}(x_i - moy(X))^2 = moy(X^2) - (moy(X))^2$
3. cov(X,Y)=$\frac{1}{n}\sum_{i=1}^{n}(x_i - moy(X))(y_i - moy(Y)) = moy(XY) - moy(X)moy(Y)$
4. $X^2 = (x_1^2, x_2^2, ..., x_n^2)$ et $XY = (x_1y_1, x_2y_2, ..., x_ny_n)$
5. $X = (x_1, x_2, ..., x_n)$ est représentée par une liste de longueur n avec $X[i] = x_{i+1}$.

Questions :

Soient U et V des listes de même longueur.

- Écrire une fonction *moy(U)* qui calcule la moyenne des termes de U.
- Écrire une fonction *produit(U,V)* qui calcule la liste des produits $U[i]V[i]$.
- En déduire des fonctions *var(U)* et *cov(U,V)* qui calculent respectivement la variance de U et la covariance de U et V.
- En déduire une fonction *coeffC(X,Y)* qui à partir de deux listes de même longueur X et Y calcule le couple (a,b) des coefficients a et b déterminés par la méthode des moindres carrés.

Exercice 4 : Moyenne mobile

Soit S une série de nombres. La moyenne mobile sur n valeurs de S est la série M dont chaque valeur est la moyenne de n valeurs consécutives de S. Par définition, la taille de M est $len(S) - n + 1$.

Écrivez une fonction $Mobile(S, n)$ qui renvoie la liste des moyennes mobiles sur n périodes de S . On suppose que S a k éléments avec k $\leq$ n. On peut utiliser la méthode *append()* des listes pour ajouter un élément en fin de liste.

Exemple : la moyenne mobile de [2,3,5,7,11,13, 17] sur 3 périodes, donne une liste de 5 valeurs [3.333, 5, 7.666, 10.333, 13.666]. 3.333 étant la moyenne de [2, 3, 5], 5 étant la moyenne de [3, 5, 7], ect..