# Remarque :

Pour chaque notebook/exercice, expliquer via des commentaires les lignes de codes.

À la fin de chaque exercice :

* téléchargez le fichier jupyter et
* exporter fichier jupyter en fichier pdf ou html
* puis télécharger le sur OneDrive

# Exercice 1 :

Nous utiliserons *la méthode des deux écarts-types* avec Python pour identifier les valeurs aberrantes d’un ensemble de données représentant les valeurs d’un test de certification.

data : [55, 60, 61, 58, 59, 62, 65, 59, 58, 120, 62, 63, 61, 58, 57, 61, 64, 63, 130, 60, 59]

1. Déclarer la liste dans python
2. Calculer la Moyenne et l'Écart-type
3. Définir les Seuils
4. Identifier les valeurs aberrantes
5. Afficher la liste de données
6. Afficher la moyenne et l'Écart-type
7. Afficher les seuils
8. Afficher les valeurs aberrantes
9. Créer un graphique
10. Ploter les données sur ce graphique
11. Ploter la droite représentant la moyenne avec une couleur verte via la méthode plt.axhline(..)
12. Ploter la droite représentant le seuil supérieur avec une couleur verte via la méthode plt.axhline(..)
13. Ploter la droite représentant le seuil inférieur avec une couleur verte via la méthode plt.axhline(..)
14. Ploter les légendes
15. Afficher maintenant le graphique

# Exercice 2 :

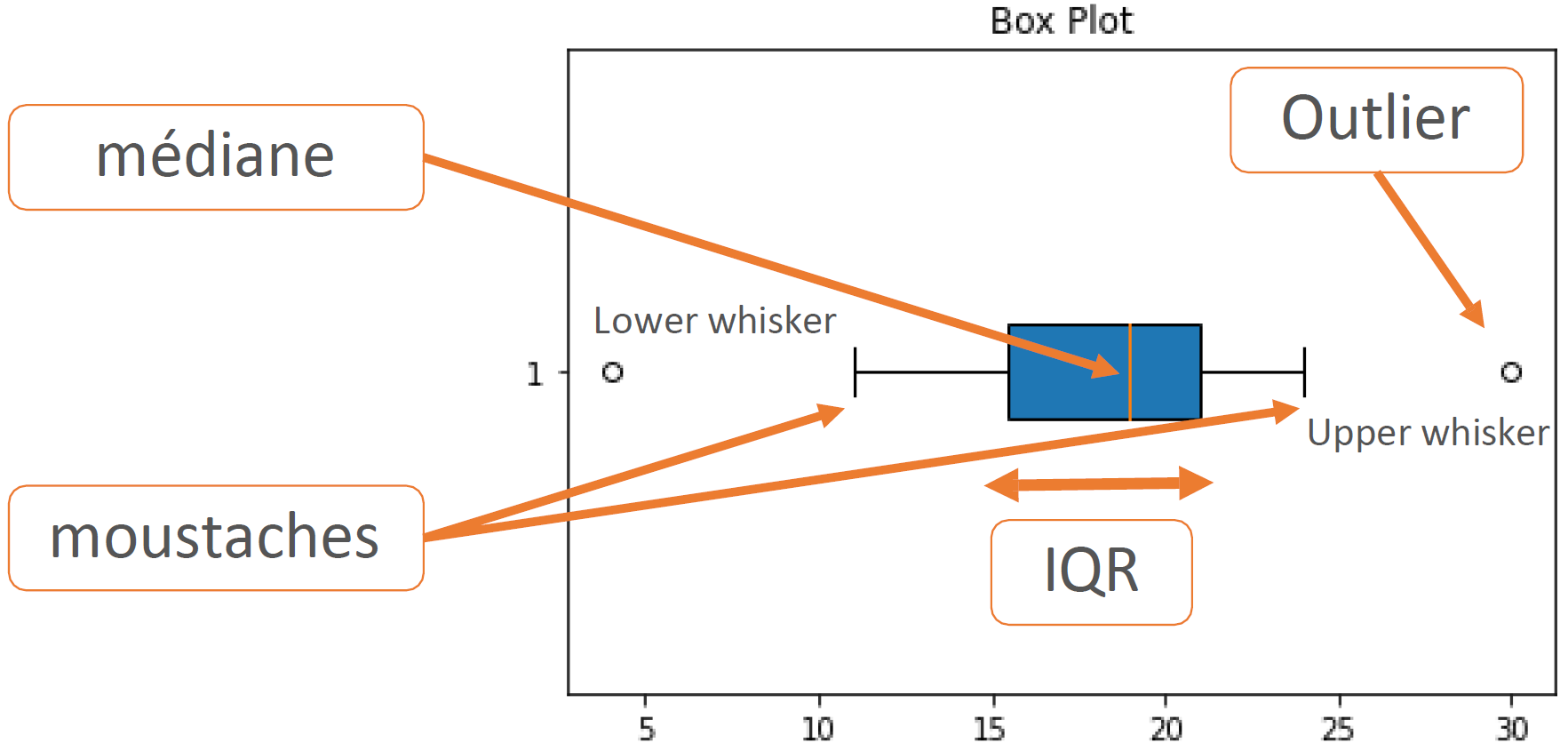
Nous utiliserons *la méthode IQR (InterQuartile Range)* avec Python pour identifier les valeurs aberrantes d’un ensemble de données représentant les valeurs d’un test de certification.

data : [55, 60, 61, 58, 59, 62, 65, 59, 58, 120, 62, 63, 61, 58, 57, 61, 64, 63, 130, 60, 59]

1. Déclarer la liste dans python
2. Calculer la Moyenne et l'Écart-type
3. Calculer les quartiles Q1, Q2, Q3 avec la fonction np.percentille de numpy
4. Calculer IQR
5. Calculer les Seuils supérieurs et inférieurs de la méthode IQR
6. Calculer les Seuils supérieurs et inférieurs de la méthode de Tukey
7. Identifier les valeurs aberrantes via la méthode IQR
8. Identifier les valeurs aberrantes via la méthode de Tukey
9. Afficher la liste de données
10. Afficher la moyenne et l'Écart-type
11. Afficher les seuils de la méthode IQR
12. Afficher les seuils de la méthode de Tukey
13. Afficher les valeurs aberrantes en se basant sur la méthode IQR
14. Afficher les valeurs aberrantes en se basant sur la méthode IQR
15. Sachant que la syntaxe *data[~np.isin(data, iqr\_outliers)]* renvoient la liste des éléments data qui n’existent pas dans iqr\_outliers, créer un ensemble de données *data\_correct* dans les éléments ne sont pas des iqr\_outliers ni tukey\_outliers
16. Créer un graphique de nuage de points
17. Ploter les éléments de *data\_correct avec une couleur (ex. bleu)*
18. Ploter la droite représentant la moyenne avec une *couleur (ex. verte)* via la méthode plt.axhline(..)
19. Ploter la droite représentant le seuil supérieur IQR avec une couleur verte via la méthode plt.axhline(..)
20. Ploter la droite représentant le seuil inférieur IQR avec une couleur verte via la méthode plt.axhline(..)
21. Ploter la droite représentant le seuil supérieur de Tukey avec une couleur orange via la méthode plt.axhline(..)
22. Ploter la droite représentant le seuil inférieur de Tukey avec une couleur orange via la méthode plt.axhline(..)
23. Ploter les outliers IQR avec une couleur rouge
24. Ploter les outliers de Tukey avec une couleur orange
25. Ploter les légendes
26. Afficher maintenant le graphique

# Exercice 3 : box plot

L’image illustre le concept de diagramme à moustaches.

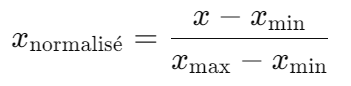


En se basant sur le vecteur data de l’exercice précédent, répondez aux questions suivantes :

1. Créer une figure de (8 , 6)
2. Créer un diagramme à moustaches horizontal
3. ajouter une ligne verticale qui représente la moyenne des données sur le box plot via plt.axvline(…) (bleu)
4. ajouter une ligne verticale qui représente la moyenne des données sur le box plot via plt.axvline(…) (vert)
5. ajouter une ligne verticale qui représente le seuil inférieur Q1 sur le box plot via plt.axvline(…) (orange)
6. ajouter une ligne verticale qui représente le seuil supérieur Q3 sur le box plot via plt.axvline(…) (orange)
7. ajouter les points outliers sur le graphique via une boucle for (couleur rouge)
8. Ajouter le titre, les axes de x et de y
9. Afficher le box plot

# Exercice 4 :

1. Importer pyplot de matplotlib, numpy, norm de scipy.stats
2. En utilisant la fonction np.random.normal(loc,scale,size) créer une *data* en générant une distribution normale de 1000 valeurs, ayant une moyenne 50 et un écart-type de 10
3. Calculer des Z-scores de la *data*
4. *Créer deux histogrammes avec couleurs différentes :*
   1. *Un pour l’ensemble data*
   2. *Un pour les z-scores*
5. Un z-score de 2 indique combien de déviation par rapport à la moyenne ?
6. On suppose que les valeurs aberrantes sont en dessous et au-dessus -3 et +3 respectivement.
   1. Identifier les valeurs normales
   2. Identifier les valeurs aberrantes
7. Ploter un histogramme
   1. Valeurs normales avec une couleur ex. bleu
   2. Valeurs aberrantes avec une couleur ex. rouge
   3. Définir le titre, l’axe des x et y pour cet histogramme
8. Qu’est-ce que la standardisation ? et pourquoi ?
9. Après la standardisation des données, les données auront quelle moyenne et écart-type ?
10. Qu’est-ce que la normalisation ? et pourquoi ?
11. Après la normalisation des données, les données auront quelle moyenne et écart-type ?
12. On veut normaliser les données data dans un intervalle de [0,1], proposer un code en utilisant la formule :



1. Ploter l’histogramme de données normalisées.