```python
"""
Script : choix_seuil_outlier.py
Auteur : CISSE Ibrahim
Objectif : Tester plusieurs méthodes de détection d'outliers sur une variable numérique et choisir le seuil le plus pertinent.

Méthodologie :
1. Chargement d'un jeu de données simulé
2. Application de trois méthodes de détection :
   - Écart-type (±2σ)
   - IQR (boîte à moustaches)
   - Quantiles extrêmes (1% et 99%)
3. Comparaison du nombre d'outliers détectés
4. Visualisation des seuils et des points extrêmes
5. Choix du seuil optimal selon le contexte métier
"""

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# 1. Génération de données simulées
np.random.seed(42)
data = np.random.normal(loc=100, scale=15, size=1000)
data = np.append(data, [30, 200, 250])  # Ajout d'outliers
df = pd.DataFrame({"valeurs": data})

# 2. Méthode écart-type
mean = df["valeurs"].mean()
std = df["valeurs"].std()
outliers_std = df[(df["valeurs"] < mean - 2*std) | (df["valeurs"] > mean + 2*std)]

# 3. Méthode IQR
q1 = df["valeurs"].quantile(0.25)
q3 = df["valeurs"].quantile(0.75)
iqr = q3 - q1
outliers_iqr = df[(df["valeurs"] < q1 - 1.5*iqr) | (df["valeurs"] > q3 + 1.5*iqr)]

# 4. Méthode quantiles extrêmes
low = df["valeurs"].quantile(0.01)
high = df["valeurs"].quantile(0.99)
outliers_quant = df[(df["valeurs"] < low) | (df["valeurs"] > high)]

# 5. Comparaison
print("Méthode écart-type : ", len(outliers_std), " outliers")
print("Méthode IQR        : ", len(outliers_iqr), " outliers")
print("Méthode quantiles  : ", len(outliers_quant), " outliers")

# 6. Visualisation
plt.figure(figsize=(10, 6))
plt.boxplot(df["valeurs"], vert=False)
plt.title("Détection des outliers par boîte à moustaches")
plt.xlabel("Valeurs")
plt.grid(True)
plt.show()
```