Statistique descriptive avec R
TP 3

Téléchargez au format CSV l'historique d'un portefeuille investi sur le SP500 :
`https://curvo.eu/backtest/fr/indice/sp-500?currency=eur#graphique`.

1. Importez les données et mettez la première colonne au format `Date` au 1er janvier, vous specifierez le type des données à l'importation.

2. Ajoutez la colonne `Perf` définie comme la performance mensuelle du SP500 :

$$\texttt{Perf}_t := \frac{\mathrm{SP500}_t - \mathrm{SP500}_{t-1}}{\mathrm{SP500}_{t-1}}, \quad t \geq 2.$$

Mettez `NA` pour le premier mois où ce n'est pas calculable.

3. Affichez l'histogramme des performances.

4. Définissez une variable `Perf` : le vecteur des performances calculées sans la donnée manquante, puis définissez une variable `batons` : une séquence qui démarre de la valeur minimale de vos données arrondie au centième et s'achève à la valeur maximale arrondie au centième, avec un pas d'un centième.

5. Affichez l'histogramme en utilisant un pas d'un centième pour les bâtons.

6. Calculez la moyenne et l'écart-type des performances.

7. Sur votre histogramme, affichez la densité gaussienne avec la moyenne et la variance empiriques calculées. Pour comparer l'histogramme à la densité, il faut ajouter l'argument `freq= FALSE` à celui-ci.

8. La fonction `density` permet d'obtenir la densité empirique par méthode de noyaux définie par :

$$\widehat{f}(x) := \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - x_i}{h}\right), \quad x \in \mathbb{R},$$

où $(x_i)_{1 \leq i \leq n}$ est l'échantillon observé, h est la *fenêtre* (**R** la calculera de manière optimale) et K est une densité centrée réduite (par défaut, **R** utilise la loi normale).

En utilisant les paramètres par défaut de `density`, affichez la densité empirique des performances et comparez à celle de la loi normale.

9. Nous allons tester si les données sont normales. Nous allons utiliser le test de Jarque-Bera. Calculez le skewness et le kurtosis et commentez par rapport à la loi normale.

10. Le test de Jarque-Bera repose sur le résultat de convergence suivant : si les données $(X_i)_{1\leq i\leq n}$ sont normales et i.i.d.,

$$JB_n := n\left(\frac{S_n^2}{6} + \frac{(K_n-3)^2}{24}\right) \underset{n\to+\infty}{\longrightarrow} \chi_2^2,$$

où S_n est le skewness des données et K_n en est le kurtosis. Calculez JB_n.

11. La p-valeur est la probabilité, sous l'hypothèse nulle $\{H_0$: Les performances sont gaussiennes$\}$, d'obtenir une statistique de test au moins aussi extrême que celle observée, dans la queue de la distribution. Ici, pour le test de Jarque-Bera, elle correspond à $\mathbb{P}(\chi_2^2 > JB_n)$ où JB_n est la valeur du test observée empiriquement. Calculez la p-valeur et concluez.

12. Mettez le calcul du test dans une fonction de **R**.