

Statistique descriptive



Moyenne: La moyenne nous permet de résumer l'essentiel de l'information.

$$\frac{1}{n}\sum_{i=1}^{n}x_{i}$$

mean = np.mean(x)

Mesure de la Tendance Centrale Médiane: Valeur qui a 50% des valeurs qui la précèdent et la suivent.

$$\frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

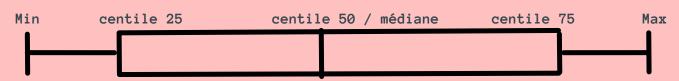
n even

 $x_{int(\frac{n}{2})+1}$

n odd

median = np.median(x)

Centile: Le centile n représente la valeur de l'échantillon qui a n% de l'ensemble des données avant.



p est la proportion des valeurs situées avant le percentile (ex : 0.25, 0.5, 0.99).

percentile = np.quantile(x, p)



Var:

Explique la dispersion des valeurs autour de la moyenne.

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$var = np.var(x)$$

Mesure de Dispersion

Std:

Il nous permet de calculer la volatilité

d'un actif.

 \sqrt{var}

volatility = np.std(x)

Skewness: Donnez des informations sur l'asymétrie de la distribution. skw = scipy.stats.skew(x)

Kurtosis: Donnez des informations sur la masse des valeurs.

kurto=scipy.stats.kurtosis(x)



Statistique descriptive



Mesure de Relation

COV: Quantifier la relation entre deux échantillons (ou variables aléatoires).

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

cov=np.cov(array, rowvar=False)[i][j]

Var Cov Matrix: Matrice organisant la relation entre plusieurs échantillons.

$$\begin{array}{ccccc}
x1 & & & & & & & \\
x1 & & Var(x_1) & \cdots & Cov(x_1, x_n) \\
& \vdots & \ddots & & \vdots \\
& Cov(x_n, x_1) & \cdots & Var(x_n)
\end{array}$$

cov=np.cov(array, rowvar=False)

Correlation Matrix: La corrélation normalise la métrique de la relation entre -1 et 1.

corr = df.corr()