Calculs des Tendances Centrales, Mesures de Dispersion et Corrélation pour le Boeuf à Ragoût

Mesures de Tendance Centrale

1. La Mode (Mode) La mode est la valeur qui apparaît le plus fréquemment dans un ensemble de données. Dans cet ensemble, la mode est :

$$\text{Mode} = 18,04\,\$ \quad \text{(apparaît deux fois)}$$

2. La Médiane ($\tilde{x}$) La médiane est la valeur centrale lorsque les données sont classées par ordre croissant (ou décroissant). S'il y a un nombre pair de points de données, c'est la moyenne des deux valeurs centrales. Le nombre total de données est $n = 57$, ce qui est impair. La médiane est donc la valeur à la position :

$$\text{Position de la médiane} = \frac{n+1}{2} = \frac{57+1}{2} = 29$$

Après tri des données avec Python, il suffit de regarder à la position 29 :

$$\tilde{X} = 18,25\,\$$$

3. La Moyenne ($\bar{x}$) La moyenne est calculée en divisant la somme de toutes les valeurs de données par le nombre total de valeurs. La moyenne est calculée comme suit :

$$\bar{x} = \frac{\sum x}{n}$$

$$\bar{x} = \frac{13,85 + 14,38 + ... + 21,79 + 22,29}{57} = \frac{1023,61}{57} \approx 17,96\,\$$$

Mesures de Dispersion

1. Étendue (R) L'étendue est la différence entre la valeur maximale et la valeur minimale :

$$R = \max(x) - \min(x)$$

$$R = 22,29 - 13,85 = 8,44\,\$$$

2. Variance (σ^2) La variance est la mesure de l'écart entre les nombres d'un ensemble de données. La variance est calculée comme suit :

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

$$\sum (x_i - \bar{x})^2 = (13,85 - 17,96)^2 + (14,65 - 17,96)^2 + \cdots + (21,79 - 17,96)^2 = 254,41989$$

$$n - 1 = 57 - 1 = 56$$

$$\sigma^2 = \frac{254,41989}{56} \approx 4,54\,\$^2$$

3. Écart-type (σ) L'écart-type est la racine carrée de la variance :

$$\sigma = \sqrt{\sigma^2} = \sqrt{4,54} \approx 2,13\,\$$$

4. Écart interquartile (IQR) L'écart interquartile est la différence entre le troisième quartile (Q3) et le premier quartile (Q1). Elle est donné par :

$$\text{IQR} = Q3 - Q1$$

$$\text{IQR} = 19,49 - 16,20 = 3,29\,\$$$

5. Cotes Z (Z-scores) Les cotes Z standardisent une valeur x par rapport à la moyenne et à l'écart-type :

$$Z = \frac{x - \bar{x}}{\sigma}$$

Par exemple, pour $x = 18,04\,\$$:

$$Z = \frac{18,04 - 17,96}{2,12} = \frac{0,08}{2,13} \approx 0,04$$

Corrélation et Régression Linéaire

1. Coefficient de Corrélation (r) Le coefficient de corrélation mesure la force et la direction de la relation linéaire entre deux variables. La formule est donnée par :

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Les valeurs x seront la différence en jours entre la date du prix actuel (y) et le 1er janvier 2020 pour chaque prix.

Moyennes : $\bar{x} = 851,95$ (*automatiquement calculée en utilisant Python, voir ligne 44 de correlation.py*), $\bar{y} = 17,96$.

Numérateur :

$$\begin{aligned}\sum(x_i - \bar{x})(y_i - \bar{y}) &= (0 - 851,95)(13,85 - 17,96) + \dots \\ &\quad + (1705 - 851,95)(21,79 - 17,96) \\ &= 46526,83449999999\end{aligned}$$

Dénominateur :

$$\begin{aligned}\sqrt{\sum(x_i - \bar{x})^2 \cdot \sum(y_i - \bar{y})^2} &= \sqrt{(0 - 851,95)^2 \cdot (13,85 - 17,96)^2 + \dots} \\ &\quad + \sqrt{(1705 - 851,95)^2 \cdot (21,79 - 17,96)^2} \\ &= \sqrt{3634675275,680535} \\ &= 60288,268142985624\end{aligned}$$

Coefficient :

$$r = \frac{46526,83449999999}{60288,268142985624} \approx 0,77$$

2. Pente de la Régression Linéaire (m) La pente de la régression linéaire, m, est calculée comme suit :

$$m = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

Moyennes : $\bar{x} = 851,95$, $\bar{y} = 17,96$

Numérateur (déjà calculé) :

$$\sum(x_i - \bar{x})(y_i - \bar{y}) = 46526,83449999999$$

Dénominateur :

$$\sum(x_i - \bar{x})^2 = (0 - 851,95)^2 + \cdots + (1705 - 851,95)^2 = 14286116,842500001$$

Pente :

$$m = \frac{46526,83449999999}{14286116,842500001} = 0,0032567866420906313 \approx 0,003$$

3. L'ordonnée à l'origine de la Régression Linéaire (b) L'ordonnée à l'origine est calculée comme suit :

$$b = \bar{y} - m \cdot \bar{x}$$

Moyennes : $\bar{x} = 851,95$, $\bar{y} = 17,96$

Calcul :

$$b = 17,96 - (0,003 \cdot 851,95) = 15,40415 \approx 15,40$$

L'ordonnée à l'origine est $b = 15,40$.

4. Équation de la Régression Linéaire L'équation finale de la régression linéaire est donnée par :

$$y = mx + b$$

En substituant les valeurs calculées :

$$y = 0,003x + 15,40$$

Résumé des Résultats

- Mode : $18,04\,\$$
- Médiane : $18,25\,\$$
- Moyenne : $17,96\,\$$

- Étendue : $8,44\,\$$
- Variance : $4,54\,\$^2$
- Écart-type : $2,12\,\$$
- Écart interquartile : $3,29\,\$$
- Coefficient de Corrélation : $0,77$
- Équation de la Régression Linéaire: y = 0,003x + 15,40