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

Mesures de Tendance Centrale

1. La Mode (Mode)

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

2. La Médiane ($\tilde{x}$) 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 ($\bar{x}$)

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

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

Mesures de Dispersion

1. Étendue (R)

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

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

2. Variance (σ^2)

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

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

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

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

3. Écart-type (σ) Racine carrée de la variance :

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

4. Intervalle interquartile (IQR) Différence entre le troisième quartile (Q3) et le premier quartile (Q1). Ces quartiles ont été calculés à l'aide de la fonction suivante (lien).

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

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

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)

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

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

Moyenne des valeurs x (dates) :

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

$$\bar{x} = \frac{0 + 1 + ... + 1704 + 1705}{57} = \frac{48561}{57} \approx 851,95\,\$$$

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

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)

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

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)

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

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

$$y = mx + b$$

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