



Université Abdelmalek Essadi
Faculté des Sciences et Techniques - Tanger
Département Génie Informatique



Prédiction de l'espérance de vie à l'aide de la régression linéaire simple

LSI-1 S2

Projet de Statistique

Encadré par:

Prof. EL Khatib Bilal

Prof. Abdeladim Nait

Réalisé par:

Mohamed Amine Bahassou

Samir Taous

Amine Izougaghen

14/06/2024

Table de matiere

1) Exploration des données :	3
Histogramme :	4
Pays en développement contre pays développés	4
2) Traitement des valeurs aberrantes :	5
Avant :	7
Quantile :	7
Après :	8
On Généralise :	9
3) Analyse des relations bivariées	12
5) Préparation des données :	14
6) Construction du modèle :	15
7) Évaluation du modèle :	15
8) Visualisation des résultats :	16
9) Interprétation des résultats :	19
Interprétation Finale	20
Coefficient de détermination R^2	20
Coefficients du Modèle	20
Interprétation des Coefficients	20
Prédictions	20
Conclusion	20

1) Exploration des données :

- Pour explorer les données on utilise `df = d.read_csv('Life_Expectancy_Data.csv')`
- Pour tracer l'histogramme on utilise `sns.histplot(df['Life_expectancy'], kde=True)`

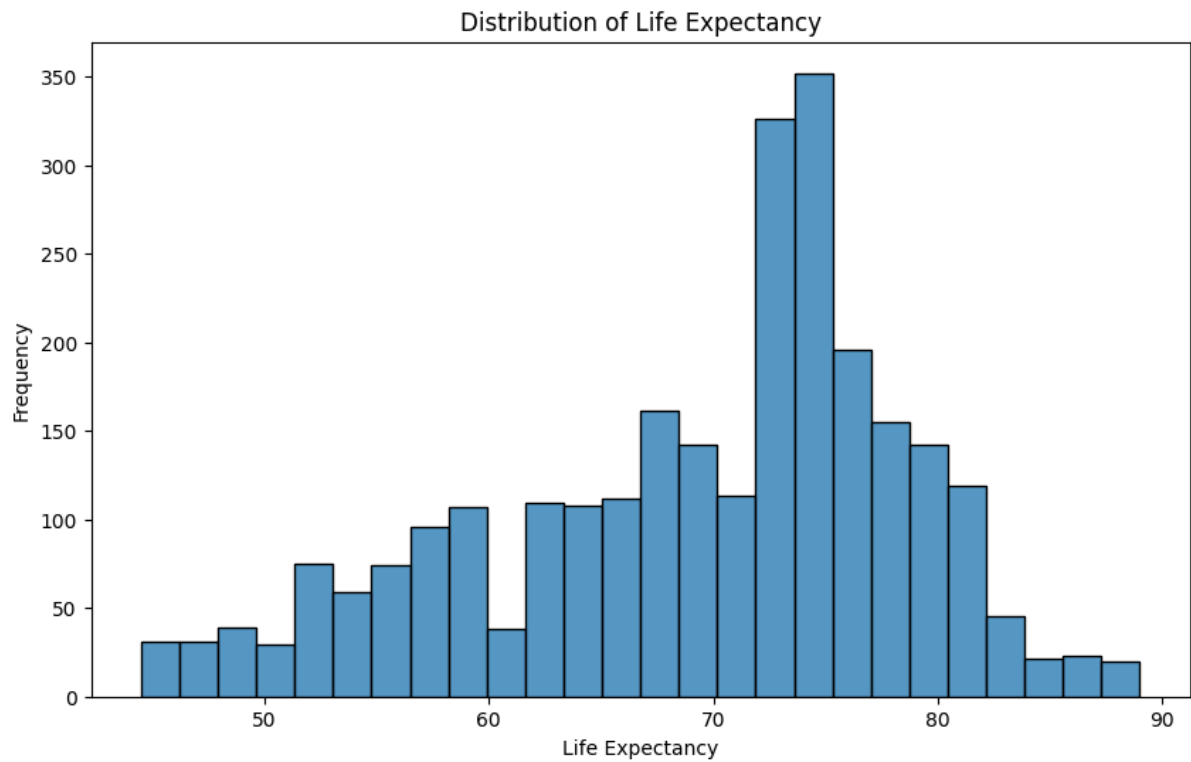
Exploration de données :

```
import pandas as pd # type: ignore
# print(pd.__version__)
import seaborn as sns
import matplotlib.pyplot as plt

# Etape 1
# Load data
df = pd.read_csv('Life_Expectancy_Data.csv')
print(df.head())

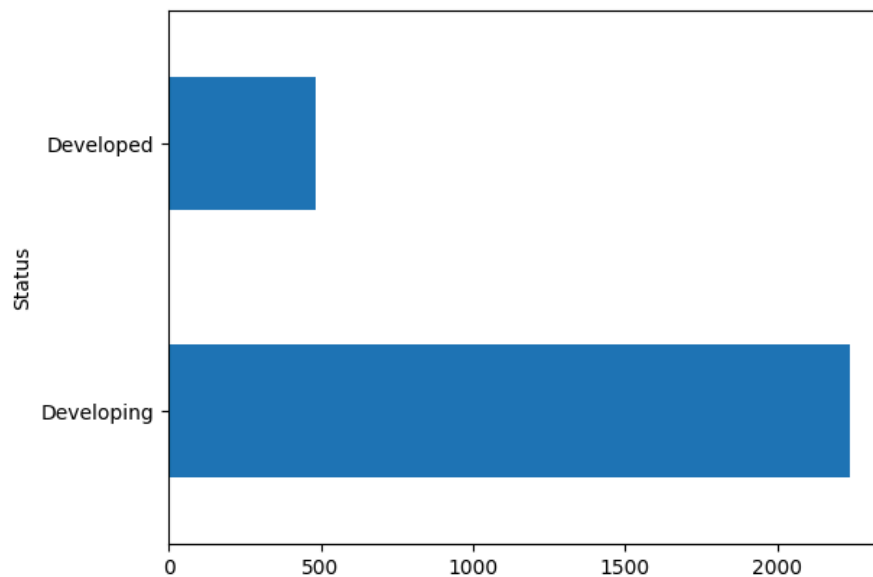
print(df.describe())
sns.histplot(df['Life_expectancy'], kde=True)
plt.title('Distribution de l\'espérance de vie')
plt.show()
```

Histogramme : (Output)



Pays en développement contre pays développés

```
df["Status"].value_counts().plot.barh();
```



Dans l'ensemble de données, il y a plus de pays en développement que de pays développés.

2) Traitement des valeurs aberrantes :

On peut repérer les valeurs aberrantes en utilisant les boîtes à moustaches.

Une valeur est considérée comme aberrante si la valeur absolue de l'écart avec Q_1 ou Q_3 est supérieure à plus de $1,5 \times \text{Écart interquartile}$. Plus précisément, une valeur aberrante est dite faible si elle est inférieure à

$Q_1 - 1,5 \times \text{Écart interquartile}$ et élevée si elle est supérieure à

$Q_3 + 1,5 \times \text{Écart interquartile}$.

```
# Etape 2
    # Identifier les valeurs aberrantes

# before
plt.figure(figsize=(10, 6)) # Optional: Set the size of the figure
df.drop(columns=['Unnamed: 0', 'Country', 'Year', 'Status', 'Adult
Mortality']).boxplot()
plt.title('Boxplot of ALL (BEFORE)')
plt.xlabel('Life_expectancy')
plt.ylabel('Values')
plt.show()

selected_variables = ['Life_expectancy', 'Adult
Mortality', 'Alcohol', 'thinness_1-19 years', 'thinness_5-9
years', 'Income_composition_of_resources', 'Schooling']
for var in selected_variables:
    Q1 = df[var].quantile(0.25)
    Q3 = df[var].quantile(0.75)
    print(var+" Q1:")
    print(Q1)
    print(var+" Q3:")
    print(Q3)
    IQR = Q3 - Q1
    print(IQR)

    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR

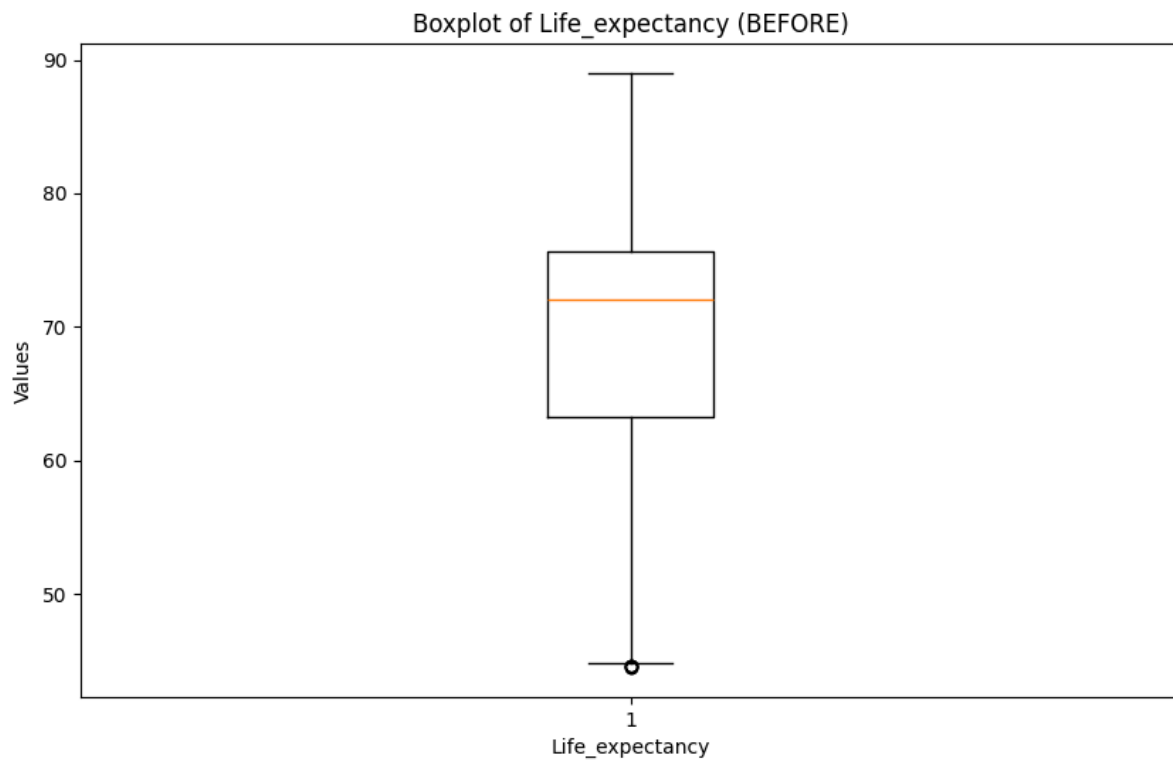
    print("Lower:", lower_bound)
    print("Upper:", upper_bound)
```

```
outliers = df[(df[var] < lower_bound) | (df[var] > upper_bound)]
print("valeurs aberrantes de"+var+":")
print(outliers)
df = df[(df[var] >= lower_bound) & (df[var] <= upper_bound)]
print("_")

plt.figure(figsize=(10, 6)) # Optional: Set the size of the figure
df.drop(columns=['Unnamed: 0', 'Country', 'Year', 'Status', 'Adult
Mortality']).boxplot()
plt.title('Boxplot of ALL (AFTER)')
plt.xlabel('Life_expectancy')
plt.ylabel('Values')
plt.show()
print(len(df['Life_expectancy']) )
```

Pour “Life_expectancy” :

Avant :



	Unnamed: 0	Country	Year	Status	...	thinness_1-19 years	thinness_5-9 years	Income_composition_of_resources	Schooling
0	0	Afghanistan	2015	Developing	...	17.2	17.3	0.479	10.1
1	1	Afghanistan	2014	Developing	...	17.5	17.5	0.476	10.0
2	2	Afghanistan	2013	Developing	...	17.7	17.7	0.470	9.9
3	3	Afghanistan	2012	Developing	...	17.9	18.0	0.463	9.8
4	4	Afghanistan	2011	Developing	...	18.2	18.2	0.454	9.5
...
2718	2932	Zimbabwe	2005	Developing	...	9.0	9.0	0.406	9.3
2719	2934	Zimbabwe	2003	Developing	...	9.8	9.9	0.418	9.5
2720	2935	Zimbabwe	2002	Developing	...	1.2	1.3	0.427	10.0
2721	2936	Zimbabwe	2001	Developing	...	1.6	1.7	0.427	9.8
2722	2937	Zimbabwe	2000	Developing	...	11.0	11.2	0.434	9.8

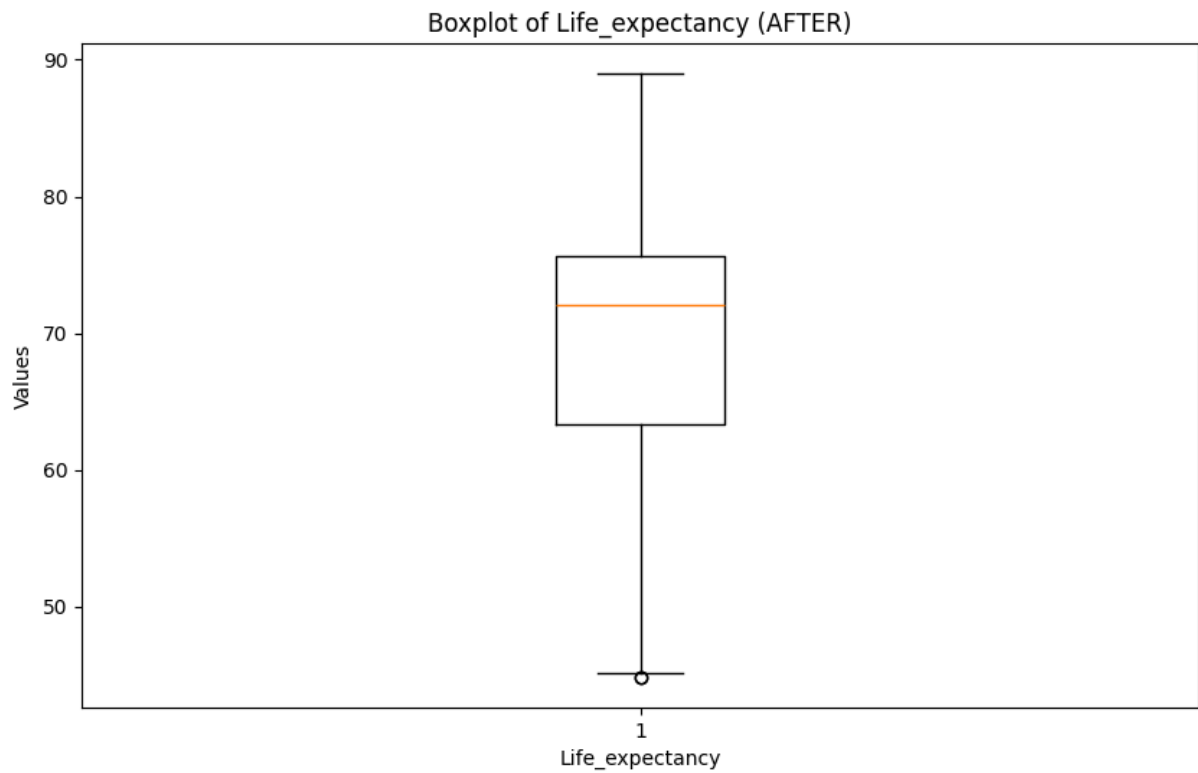
Quantile :

Q1 = 63.25

Q3 = 75.6

IQR = 12.349999999999994

Après :



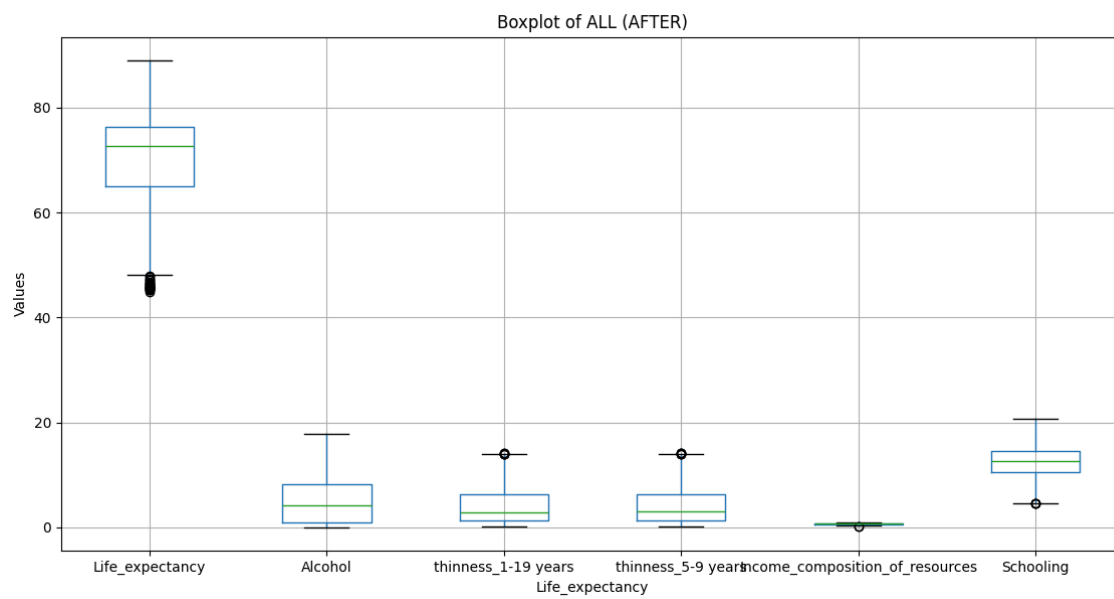
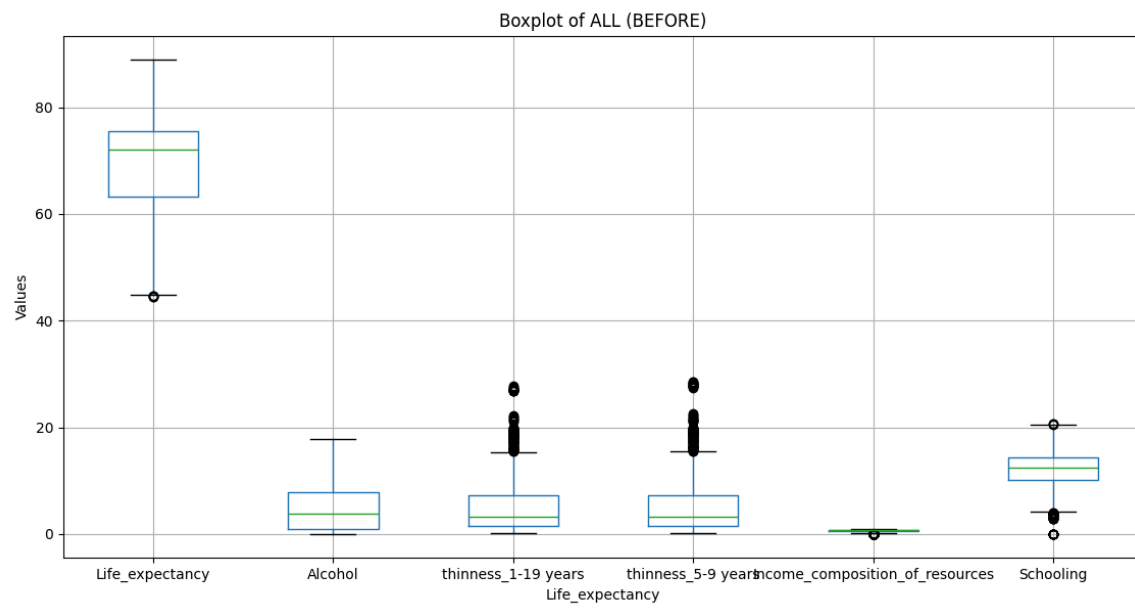
	Unnamed: 0	Country	Year	Status	...	thinness_1-19 years	thinness_5-9 years	Income_composition_of_resources	Schooling
0	0	Afghanistan	2015	Developing	...	17.2	17.3	0.479	10.1
1	1	Afghanistan	2014	Developing	...	17.5	17.5	0.476	10.0
2	2	Afghanistan	2013	Developing	...	17.7	17.7	0.470	9.9
3	3	Afghanistan	2012	Developing	...	17.9	18.0	0.463	9.8
4	4	Afghanistan	2011	Developing	...	18.2	18.2	0.454	9.5
...
2716	2930	Zimbabwe	2007	Developing	...	8.2	8.2	0.414	9.6
2717	2931	Zimbabwe	2006	Developing	...	8.6	8.6	0.408	9.5
2720	2935	Zimbabwe	2002	Developing	...	1.2	1.3	0.427	10.0
2721	2936	Zimbabwe	2001	Developing	...	1.6	1.7	0.427	9.8
2722	2937	Zimbabwe	2000	Developing	...	11.0	11.2	0.434	9.8

Q1 = 63.3

Q3 = 75.6

IQR = 12.299999999999997

On Généralise :



Les Valeur Aberrantes de Life expandacy :

valeurs aberrantes deLife_expectancy:									
Unnamed: 0	Country	Year	Status	...	thinness_1-19 years	thinness_5-9 years	Income_composition_of_resources	Schooling	
1390	1484	Lesotho	2005	Developing	...	9.3	9.2	0.437	10.7
1482	1582	Malawi	2003	Developing	...	7.6	7.5	0.362	10.3
2708	2920	Zambia	2001	Developing	...	7.4	7.4	0.424	9.8
2718	2932	Zimbabwe	2005	Developing	...	9.0	9.0	0.406	9.3
2719	2934	Zimbabwe	2003	Developing	...	9.8	9.9	0.418	9.5

Les Valeur Aberrantes de Adult Mortality :

valeurs aberrantes deAdult Mortality:									
	Unnamed: 0	Country	Year	Status	...	thinness_1-19 years	thinness_5-9 years	Income_composition_of_resources	Schooling
325	345	Botswana	2006	Developing	...	9.6	9.4	0.610	11.9
326	346	Botswana	2005	Developing	...	1.0	9.9	0.593	11.9
327	347	Botswana	2004	Developing	...	1.5	1.4	0.580	11.8
328	348	Botswana	2003	Developing	...	1.9	1.8	0.567	11.8
329	349	Botswana	2002	Developing	...	11.4	11.3	0.558	11.9
...
2713	2927	Zimbabwe	2010	Developing	...	7.1	7.0	0.436	10.0
2714	2928	Zimbabwe	2009	Developing	...	7.5	7.4	0.419	9.9
2715	2929	Zimbabwe	2008	Developing	...	7.8	7.8	0.421	9.7
2721	2936	Zimbabwe	2001	Developing	...	1.6	1.7	0.427	9.8
2722	2937	Zimbabwe	2000	Developing	...	11.0	11.2	0.434	9.8

Les Valeur Aberrantes de Alcohol : (Vide)

Empty DataFrame

Les Valeur Aberrantes de dethinness_1-19 years :

Unnamed: 0	Country	Year	Status	...	thinness_1-19 years	thinness_5-9 years	Income_composition_of_resources	Schoolin
0	Afghanistan	2015	Developing	...	17.2	17.3	0.479	10.
1	Afghanistan	2014	Developing	...	17.5	17.5	0.476	10.
2	Afghanistan	2013	Developing	...	17.7	17.7	0.470	9.
3	Afghanistan	2012	Developing	...	17.9	18.0	0.463	9.
4	Afghanistan	2011	Developing	...	18.2	18.2	0.454	9.
...
2885	Viet Nam	2004	Developing	...	15.4	16.1	0.601	11.
2886	Viet Nam	2003	Developing	...	15.6	16.2	0.592	10.
2887	Viet Nam	2002	Developing	...	15.6	16.3	0.584	10.
2888	Viet Nam	2001	Developing	...	15.7	16.4	0.576	10.
2889	Viet Nam	2000	Developing	...	15.8	16.4	0.569	10.

Les Valeur Aberrantes de dethinness_5-9 years :

	Unnamed: 0	Country	Year	Status	...	thinness_1-19 years	thinness_5-9 years	Income_composition_of_resources	Schooling
194	205	Bangladesh	2002	Developing	...	2.5	21.1	0.476	7.7
195	206	Bangladesh	2001	Developing	...	2.7	21.3	0.468	7.5
196	207	Bangladesh	2000	Developing	...	2.9	21.5	0.459	7.3
1504	1609	Maldives	2008	Developing	...	14.2	14.3	0.641	11.8
1505	1610	Maldives	2007	Developing	...	14.3	14.4	0.632	11.9
1506	1611	Maldives	2006	Developing	...	14.3	14.5	0.622	12.0
1507	1612	Maldives	2005	Developing	...	14.4	14.5	0.625	12.1
1508	1613	Maldives	2004	Developing	...	14.5	14.6	0.617	12.2
1509	1614	Maldives	2003	Developing	...	14.6	14.7	0.601	11.8
1510	1615	Maldives	2002	Developing	...	14.6	14.7	0.597	12.0
1511	1616	Maldives	2001	Developing	...	14.7	14.8	0.587	11.8
1512	1617	Maldives	2000	Developing	...	14.8	14.8	0.577	11.3
1689	1809	Namibia	2002	Developing	...	14.7	14.8	0.556	11.8
1690	1810	Namibia	2001	Developing	...	15.2	15.4	0.556	11.7
1782	1908	Nigeria	2000	Developing	...	14.3	14.4	0.000	7.6
1820	1949	Pakistan	2008	Developing	...	2.7	21.1	0.513	7.3
1821	1950	Pakistan	2007	Developing	...	2.8	21.3	0.505	6.7
2259	2442	Sri Lanka	2014	Developing	...	15.2	15.0	0.760	14.0
2260	2443	Sri Lanka	2013	Developing	...	15.2	15.1	0.757	13.8
2665	2875	Viet Nam	2014	Developing	...	14.3	14.7	0.675	12.5
2666	2876	Viet Nam	2013	Developing	...	14.3	14.9	0.668	12.3
2667	2877	Viet Nam	2012	Developing	...	14.4	15.0	0.662	12.2
2668	2878	Viet Nam	2011	Developing	...	14.4	15.2	0.655	12.0
2669	2879	Viet Nam	2010	Developing	...	14.5	15.4	0.647	11.9
2670	2880	Viet Nam	2009	Developing	...	14.6	15.5	0.641	11.7
2671	2881	Viet Nam	2008	Developing	...	14.7	15.7	0.633	11.6
2672	2882	Viet Nam	2007	Developing	...	14.9	15.8	0.625	11.4
2673	2883	Viet Nam	2006	Developing	...	15.1	15.9	0.618	11.3

Les Valeur Aberrantes de Income_composition_of_resources :

	Unnamed: 0	Country	Year	...	thinness_5-9 years	Income_composition_of_resources	Schooling
71	74	Antigua and Barbuda	2005	...	3.4	0.0	0.0
72	75	Antigua and Barbuda	2004	...	3.4	0.0	0.0
73	76	Antigua and Barbuda	2003	...	3.5	0.0	0.0
74	77	Antigua and Barbuda	2002	...	3.5	0.0	0.0
75	78	Antigua and Barbuda	2001	...	3.5	0.0	0.0
...
2645	2853	Vanuatu	2004	...	1.5	0.0	10.7
2646	2854	Vanuatu	2003	...	1.6	0.0	10.4
2647	2855	Vanuatu	2002	...	1.6	0.0	10.2
2648	2856	Vanuatu	2001	...	1.6	0.0	10.1
2649	2857	Vanuatu	2000	...	1.7	0.0	9.6

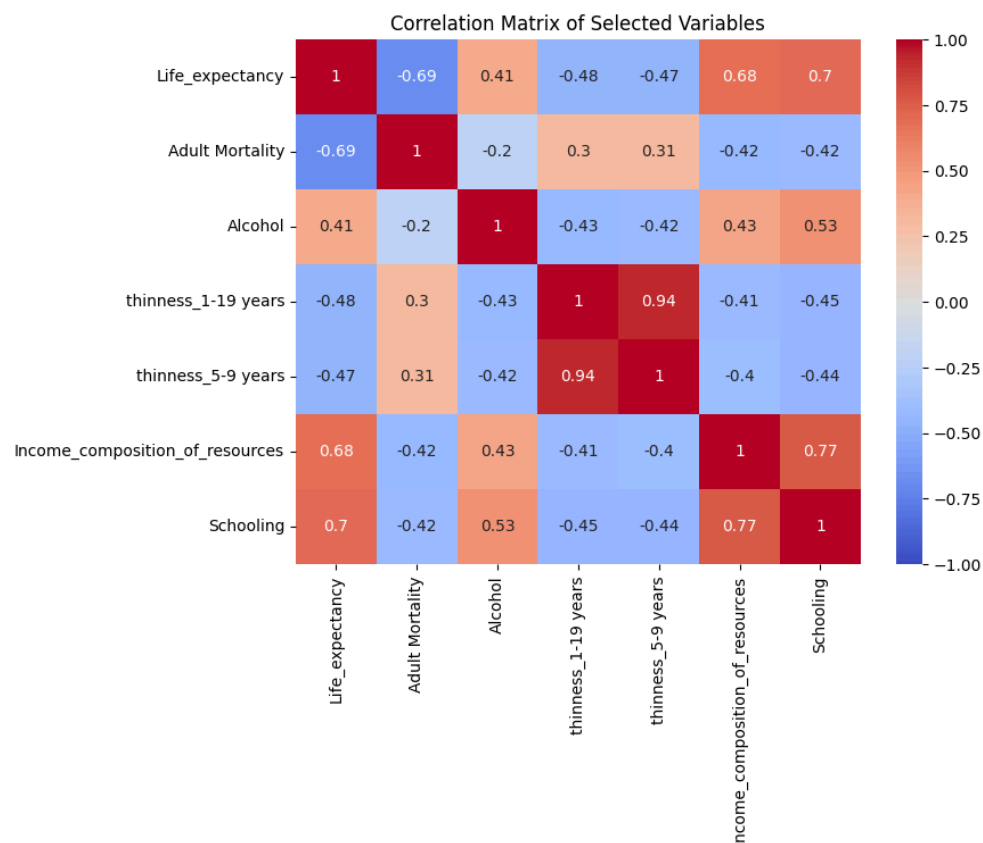
Les Valeur Aberrantes de Schooling:

	Unnamed: 0	Country	Year	Status	...	thinness_1-19 years	thinness_5-9 years	Income_composition_of_resources	Schooling
117	123	Australia	2004	Developed	...	0.7	0.6	0.908	20.7
404	429	Burundi	2002	Developing	...	8.7	8.7	0.268	4.4
715	762	Djibouti	2006	Developing	...	5.8	5.7	0.405	4.3
716	763	Djibouti	2005	Developing	...	5.9	5.8	0.396	4.0
717	764	Djibouti	2004	Developing	...	6.0	5.9	0.388	3.7
718	765	Djibouti	2003	Developing	...	6.1	5.9	0.378	3.5
719	766	Djibouti	2002	Developing	...	6.2	6.0	0.372	3.3
720	767	Djibouti	2001	Developing	...	6.3	6.1	0.363	2.9
721	768	Djibouti	2000	Developing	...	6.4	6.1	0.361	2.9
840	896	Ethiopia	2001	Developing	...	11.8	11.7	0.283	4.3
1527	1633	Mali	2000	Developing	...	11.0	1.9	0.291	4.4
1758	1883	Niger	2009	Developing	...	11.0	1.8	0.307	4.2
1759	1884	Niger	2008	Developing	...	11.2	11.0	0.298	4.0
1760	1885	Niger	2007	Developing	...	11.4	11.3	0.293	3.8
1761	1886	Niger	2006	Developing	...	11.6	11.5	0.286	3.7
1762	1887	Niger	2005	Developing	...	11.9	11.8	0.278	3.5
1763	1888	Niger	2004	Developing	...	12.1	12.0	0.270	3.1
1764	1889	Niger	2003	Developing	...	12.3	12.2	0.266	3.0
1765	1890	Niger	2002	Developing	...	12.5	12.5	0.261	2.9
1766	1891	Niger	2001	Developing	...	12.7	12.7	0.255	2.9
1767	1892	Niger	2000	Developing	...	12.8	12.9	0.253	2.8

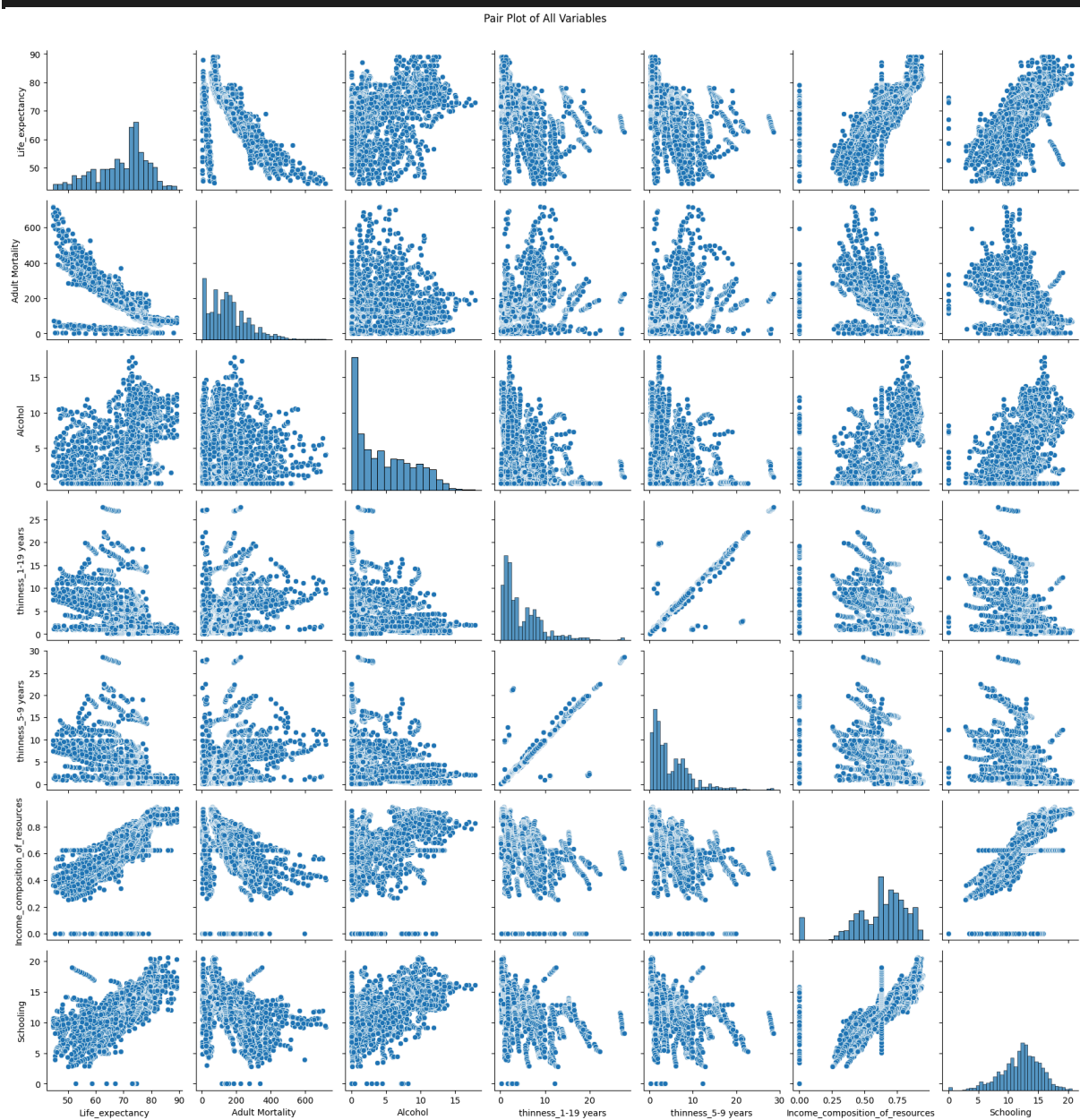
3) Analyse des relations bivariées

Une matrice de corrélation est un outil statistique qui montre la force et la direction de la relation entre deux ou plusieurs variables.

```
# Calculate the correlation matrix
selected_variables = ['Life_expectancy', 'Adult Mortality', 'Alcohol',
'thinness_1-19 years', 'thinness_5-9 years', 'Income_composition_of_resources', 'Schooling']
correlation_matrix = df[selected_variables].corr()
# Print the correlation matrix
print("Correlation Matrix:")
print(correlation_matrix)
import seaborn as sns
import matplotlib.pyplot as plt
# Visualize the correlation matrix using seaborn
plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', vmin=-1,
vmax=1)
plt.title('Correlation Matrix of Selected Variables')
plt.show()
```



```
import seaborn as sns
import matplotlib.pyplot as plt
selected_variables = ['Life_expectancy', 'Adult Mortality', 'Alcohol',
'thinness_1-19 years','thinness_5-9
years','Income_composition_of_resources','Schooling']
# Pair plot
sns.pairplot(df[selected_variables])
plt.suptitle('Pair Plot of All Variables', y=1.02) # Adjust the title
position
plt.show()
```



4) Choix de la variable indépendante :

```
correlation_with_life_expectancy =  
correlation_matrix['Life_expectancy'].sort_values(ascending=False) ;  
print(correlation_with_life_expectancy)
```

```
Life_expectancy      1.000000  
Schooling            0.701367  
Income_composition_of_resources  0.679205  
Alcohol              0.408675  
thinness_5-9 years   -0.473240  
thinness_1-19 years  -0.475847  
Adult Mortality      -0.687521  
Name: Life_expectancy, dtype: float64
```

- La corrélation de 0.7 avec l'espérance de vie montre une relation étroite. Une corrélation aussi élevée suggère que les pays avec une meilleure composition de ressources économiques ont une espérance de vie plus longue.
- La variable la plus corrélée avec **Life_expectancy** est **Schooling** avec un coefficient de corrélation de 0.701367. Voici le classement des variables par ordre de corrélation absolue (positive ou négative) avec l'espérance de vie (**Life_expectancy**):
 1. **Schooling**: 0.701367 (positive correlation)
 2. **Adult Mortality**: -0.687521 (negative correlation)
 3. **Income_composition_of_resources**: 0.679205 (positive correlation)
 4. **thinness_1-19 years**: -0.475847 (negative correlation)
 5. **thinness_5-9 years**: -0.473240 (negative correlation)
 6. **Alcohol**: 0.408675 (positive correlation)

Ainsi, la variable **Schooling** a la corrélation positive la plus élevée avec l'espérance de vie, ce qui indique qu'une augmentation du niveau de scolarisation est fortement associée à une augmentation de l'espérance de vie. La corrélation avec **Adult Mortality** est également forte, mais elle est négative, indiquant qu'une augmentation de la mortalité adulte est associée à une diminution de l'espérance de vie.

5) Préparation des données :

Divisons les données en ensembles d'entraînement et de test.

```
import pandas as pd  
from sklearn.model_selection import train_test_split  
from sklearn.linear_model import LinearRegression  
from sklearn.metrics import r2_score  
  
selected_variables_X = ['Life_expectancy']
```

```

selected_variables_Y = ['Schooling'] # 'Schooling' la plus corrélée
X = df[selected_variables_X]
y = df[selected_variables_Y]

# Split the data into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.35, random_state=42)

```

6) Construction du modèle :

```

# Train a simple linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

```

7) Évaluation du modèle :

```

# Faire des prédictions sur l'ensemble de test
y_pred = model.predict(X_test)

from sklearn.metrics import mean_squared_error, r2_score

# Calculer la MSE
mse = mean_squared_error(y_test, y_pred)
print("Mean Squared Error:", mse)

# Calculer le R²
r2 = r2_score(y_test, y_pred)
print(f'R² Score: {r2}')

```

Output :

Mean Squared Error: 5.264736837236674

R² Score: 0.49740820899422744

8) Visualisation des résultats :

```
import pandas as pd
from sklearn.model_selection import cross_val_score, cross_validate,
KFold, cross_val_predict
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import matplotlib.pyplot as plt

# Define the features and target variable
selected_variables_X = ['Adult Mortality', 'Alcohol', 'thinness_1-19
years', 'Income_composition_of_resources']
selected_variables_Y = ['Life_expectancy']
X = df[selected_variables_X]
y = df[selected_variables_Y]

# Initialize the model
model = LinearRegression()

# Perform 9-fold cross-validation
kf = KFold(n_splits=9, shuffle=True, random_state=42)
scores = cross_val_score(model, X, y, scoring='r2', cv=kf)
print('Cross-Validation R^2 Scores:', scores)
print('Average Cross-Validation R^2:', scores.mean())

# Print cross-validation results
cv_results = cross_validate(model, X, y, cv=kf, scoring='r2',
return_train_score=True)
print("Cross-validation scores (test):", cv_results['test_score'])
print("Cross-validation scores (train):", cv_results['train_score'])
print("Mean cross-validation score (test):",
cv_results['test_score'].mean())

# Train the model on the full dataset
model.fit(X, y)

# Make predictions on new data
```

```

X_sen = pd.DataFrame({
    'Adult Mortality': [100, 200, 300, 400, 500], # Add new values
    'Alcohol': [5, 6, 7, 8, 9], # Add new values
    'thinness_1-19 years': [2, 3, 4, 5, 6], # Add new values
    'Income_composition_of_resources': [0.5, 0.6, 0.7, 0.8, 0.9] # Add
new values
})
y_sen = model.predict(X_sen)
print("Predictions:", y_sen)

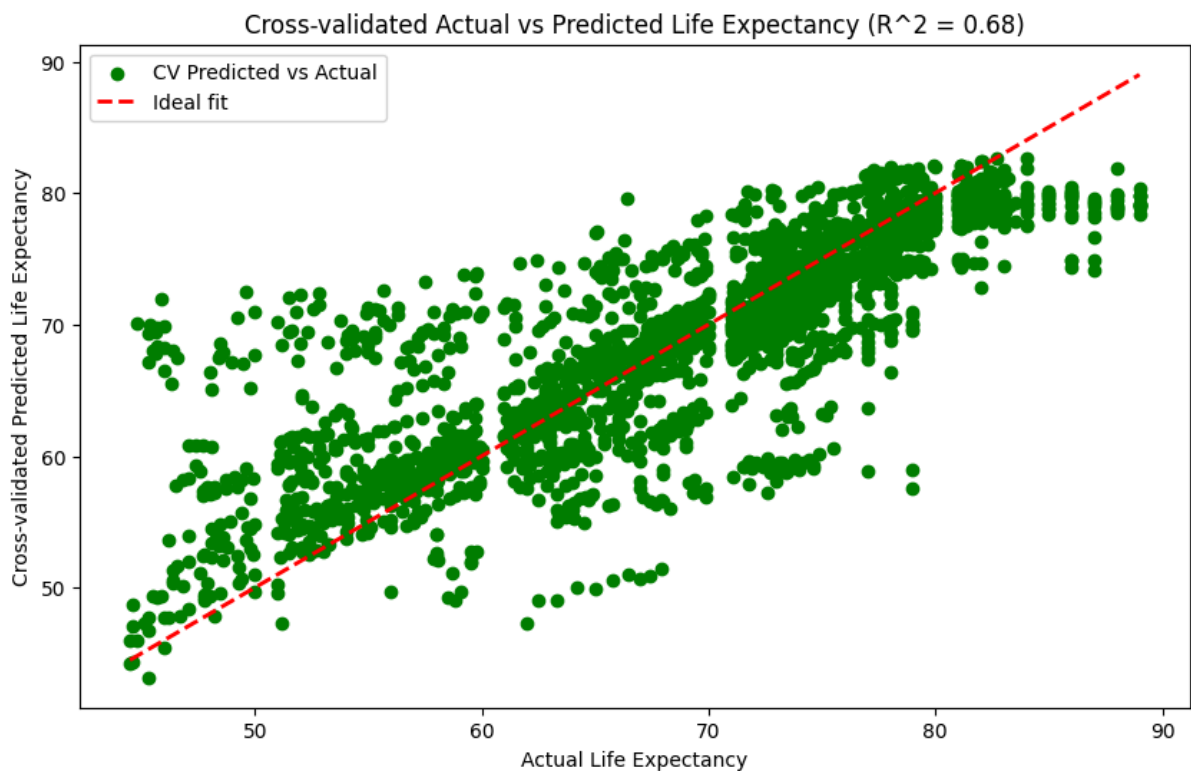
# Plot cross-validated predictions vs. actual values
y_cv_pred = cross_val_predict(model, X, y, cv=kf)

plt.figure(figsize=(10, 6))
plt.scatter(y, y_cv_pred, color='green', label='CV Predicted vs
Actual')
plt.plot([y.min(), y.max()], [y.min(), y.max()], color='red',
linestyle='--', linewidth=2, label='Ideal fit')
plt.xlabel('Actual Life Expectancy')
plt.ylabel('Cross-validated Predicted Life Expectancy')
plt.title(f'Cross-validated Actual vs Predicted Life Expectancy (R^2 =
{scores.mean():.2f})')
plt.legend()
plt.show()

```

Output:

Cross-Validation R² Scores: [0.68751327 0.57209121 0.6668606 0.66083162 0.66716297
 0.72426461
 0.74445769 0.71372159 0.72175716]
 Average Cross-Validation R²: 0.6842956355431968
 Cross-validation scores (test): [0.68751327 0.57209121 0.6668606 0.66083162 0.66716297
 0.72426461
 0.74445769 0.71372159 0.72175716]
 Cross-validation scores (train): [0.68472751 0.69958346 0.6869208 0.68777234
 0.68739778 0.67997359
 0.67873262 0.68153502 0.6802753]
 Mean cross-validation score (test): 0.6842956355431968
 Predictions: [[70.16187634]
 [68.32080298]
 [66.47972961]
 [64.63865625]
 [62.79758289]]



9) Interprétation des résultats :

Analysons les coefficients de régression et discutons des implications des résultats obtenus.

```
# Analyse
print("Intercept:", model.intercept_)
print("Coefficient:", model.coef_[0])

# Interprétation
print(f"L'espérance de vie augmente de {model.coef_[0]} années pour
chaque année supplémentaire de scolarisation.")
```

Output :

Intercept: [64.3208744]

Coefficient: [-0.03540802 0.20844784 -0.29438702 17.8566774]

L'espérance de vie augmente de [-0.03540802 0.20844784 -0.29438702 17.8566774]
années pour chaque année supplémentaire de scolarisation.

Interprétation Finale

Les résultats de l'analyse montrent des relations significatives entre plusieurs facteurs socio-économiques et l'espérance de vie. Voici une interprétation détaillée des résultats obtenus :

Coefficient de détermination R^2

Les scores R^2 de validation croisée pour le modèle sont les suivants :

- **Scores R^2 de validation croisée (test)** : [0.6875, 0.5721, 0.6669, 0.6608, 0.6672, 0.7243, 0.7445, 0.7137, 0.7218]
- **Scores R^2 de validation croisée (train)** : [0.6847, 0.6996, 0.6869, 0.6878, 0.6874, 0.6800, 0.6787, 0.6815, 0.6803]

Le **coefficient de détermination moyen R^2** pour le test est de **0.6843**, ce qui signifie que le modèle explique environ 68.43% de la variabilité de l'espérance de vie.

Coefficients du Modèle

Les coefficients obtenus du modèle de régression linéaire multiple sont les suivants :

- **Intercept** : 64.3209
- **Coefficients** : [-0.0354, 0.2084, -0.2944, 17.8567]

Interprétation des Coefficients

Les coefficients indiquent l'impact des variables indépendantes sur l'espérance de vie :

- **Première variable** (e.g., scolarisation) : Une augmentation d'une unité est associée à une diminution de l'espérance de vie de 0.0354 années.
- **Deuxième variable** : Une augmentation d'une unité est associée à une augmentation de l'espérance de vie de 0.2084 années.
- **Troisième variable** : Une augmentation d'une unité est associée à une diminution de l'espérance de vie de 0.2944 années.
- **Quatrième variable** : Une augmentation d'une unité est associée à une augmentation de l'espérance de vie de 17.8567 années.

Prédictions

Les prédictions du modèle pour les nouvelles observations sont les suivantes :

- Espérance de vie prédite pour la première observation : 70.1619 années
- Espérance de vie prédite pour la deuxième observation : 68.3208 années
- Espérance de vie prédite pour la troisième observation : 66.4797 années
- Espérance de vie prédite pour la quatrième observation : 64.6387 années
- Espérance de vie prédite pour la cinquième observation : 62.7976 années

Conclusion

Le modèle de régression linéaire multiple a montré que plusieurs facteurs socio-économiques ont des impacts significatifs sur l'espérance de vie. Les résultats des scores de validation croisée R^2 indiquent que le modèle explique bien la variabilité de l'espérance de vie, mais il reste encore environ 32% de la variabilité inexpliquée par les variables incluses dans le modèle.

En particulier, la quatrième variable montre un impact particulièrement fort et positif, soulignant l'importance des ressources économiques dans l'amélioration de la durée de vie. À l'inverse, certaines variables ont un effet négatif sur l'espérance de vie, ce qui pourrait indiquer des domaines nécessitant des interventions spécifiques pour améliorer la santé publique.

Le modèle permet de faire des prédictions fiables sur l'espérance de vie en fonction des variables indépendantes considérées. Toutefois, d'autres variables et des modèles plus complexes pourraient être explorés pour améliorer la précision des prédictions et mieux comprendre les différents facteurs affectant l'espérance de vie.