Economic Indicator Analysis Report 2020

721words

1 Analysis of Life Expectancy Indicators

1.1 Global Average Life Expectancy

Based on the dataset, the calculated average life expectancy worldwide is 72.67 years.

1.2 Variations in Life Expectancy across Continents

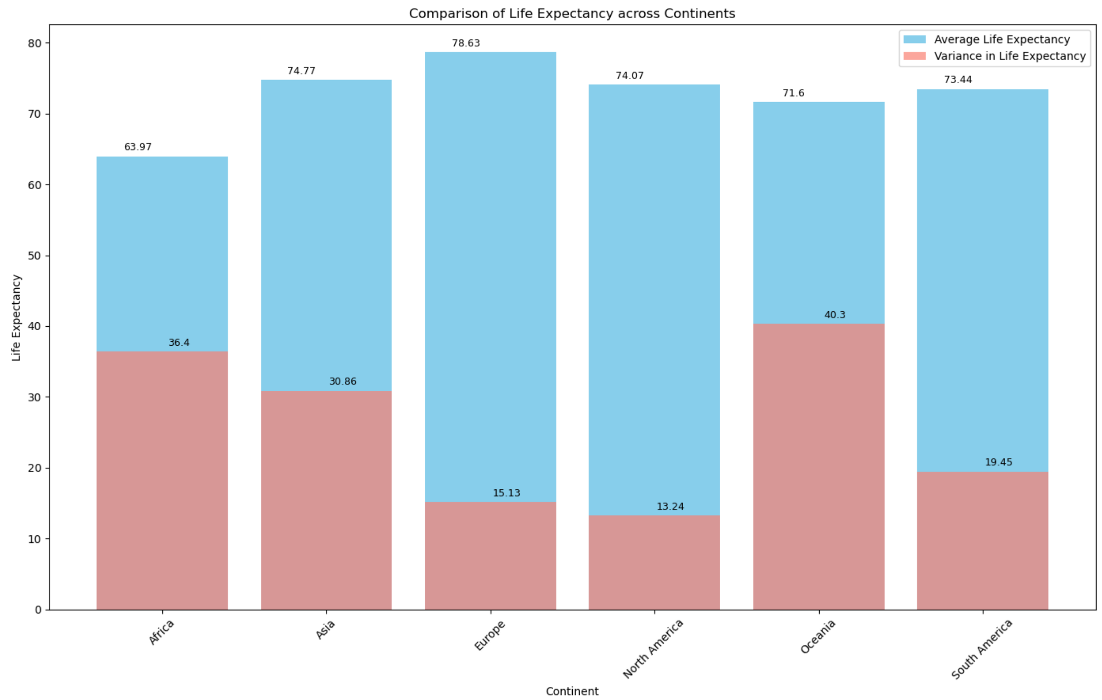

Figure 1: Bar Chart- Comparison of Life Expectancy Metrics across Continents

As seen in Figure 1, Europe stands out with the highest average life expectancy of 78.63 years, while Africa exhibits the lowest, at 63.97 years. However, Oceania displays significant variance, seeing 40.3 years, which indicates considerable disparities among countries within the Oceania region.

1.3 Symmetry of Life Expectancy Distribution

According to Figure 2, it is evident that the distribution of life expectancy is not symmetrical. The negative skew presented by the histogram indicates the unevenness in life expectancy distribution among countries or regions, where some exhibit significantly higher average life expectancies than others. This disparity might be influenced by various factors, including disparities in healthcare, socio-economic conditions, and overall health statuses among different nations.

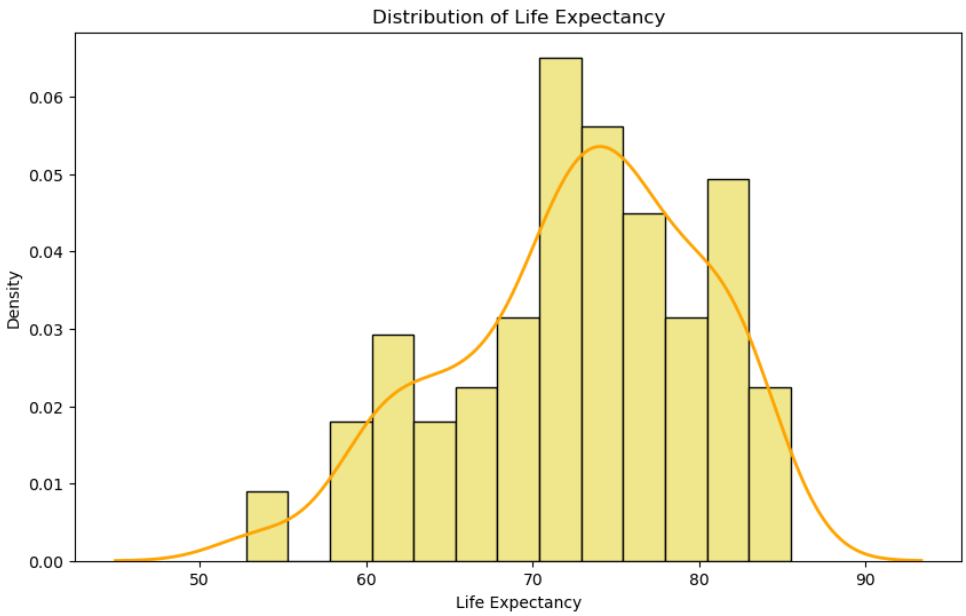

Figure 2: Histogram with Density Plot- Distribution of Life Expectancy

2 Hypothesis Testing: Average Life Expectancy Verification at 99% Confidence Interval

n	177
μ	72.67
s	7.36
H0	μ=70
Ha	μ≠70
t stat	4.825784830727795
P value	3.0106331354163777e-06

Based on the assumption of normal distribution for life expectancy values and using pandas to calculate the sample deviation, we aim to determine if our data supports the suggestion of an expected average life expectancy of 70 years. Hence, applying a t-test is appropriate. Through hypothesis testing, we reject null hypothesis (P value:3.0106331354163777e-06< significant level:0.01). The data refutes the claim that the average life expectancy is 70 years.

3 Simple Regression Analysis

3.1 Relationship between Life Expectancy and GDP per Capita

	LifeExpect	*GDPPC*
LifeExpect	*1*	
GDPPC	*0.619943*	*1*

Figure 3: Correlation matrix

Based on Figure 3, there is a very strong correlation of 0.619943 between GDP per capita and average life expectancy. Looking further at the scatter chart, which we take average life expectancy as the dependent variable and GDP per capita as the independent variable, indicates a positive trend. Higher GDP often corresponds to increased life expectancy, but not consistently. It is interesting to note that there is a high variation of life expectancy within low GDP per capita, suggesting other factors may also have an impact on life expectancy.

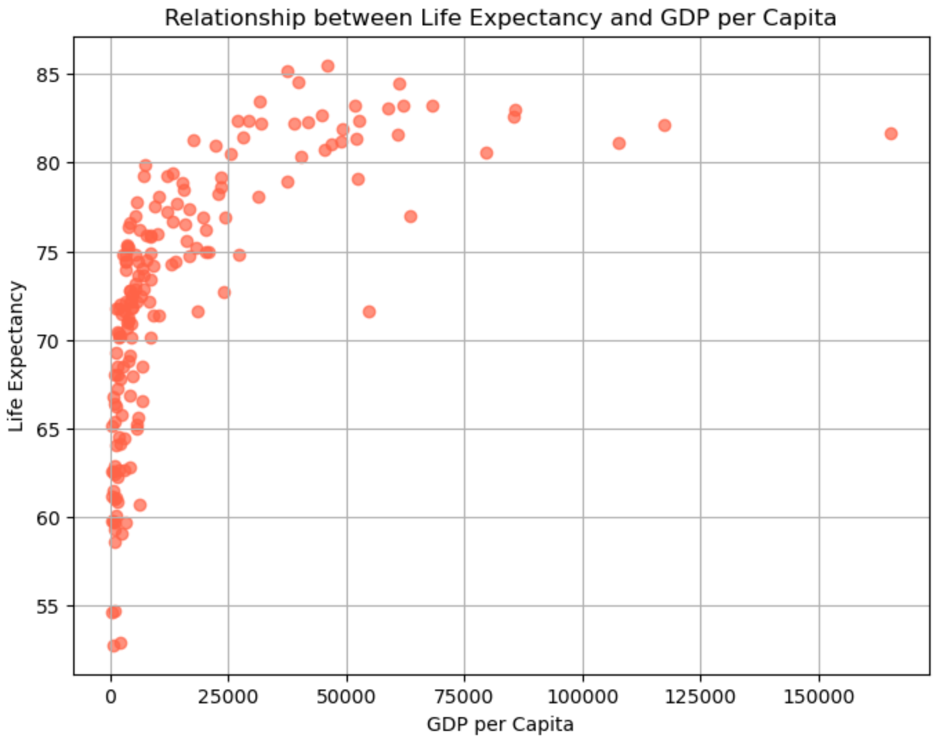

Figure 4: Scatter Chart- Life Expectancy vs GDP per capita

3.2 Simple Linear Regression Model

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             LifeExpect   R-squared:                       0.384
Model:                            OLS   Adj. R-squared:                  0.381
Method:                 Least Squares   F-statistic:                     109.2
Date:                Thu, 23 Nov 2023   Prob (F-statistic):           3.56e-20
Time:                        21:36:42   Log-Likelihood:                -561.07
No. Observations:                 177   AIC:                             1126.
Df Residuals:                     175   BIC:                             1133.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         69.5382      0.529    131.555      0.000      68.495      70.581
GDPPC          0.0002   1.81e-05     10.452      0.000       0.000       0.000
==============================================================================
Omnibus:                       21.907   Durbin-Watson:                   1.977
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               26.132
Skew:                          -0.921   Prob(JB):                     2.12e-06
Kurtosis:                       3.389   Cond. No.                     3.55e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.55e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Figure 5: Linear Regression Life Expectancy vs GDP per capita

Through the simple linear regression, the results indicate a significant relationship between GDP per capita and life expectancy, as the p-value is less than the significant level of 0.01. The coefficient of 0.0002 shows a positive relationship between these two variables. However, with an R-squared value of 0.384, it implies that about 38.4% of the variability in life expectancy can be associated with changes observed in GDP per capita. Therefore, an increase in GDP per capita appears to have a slightly positive impact on life expectancy. This suggests that besides GDP per capita, other factors also influence changes in life expectancy.

4 Multiple Regression Analysis

	GDPPC	*HealthPC$*
GDPPC	*1*	*0.922466*
HealthPC$	*0.922466*	*1*

Figure 6: Correlation matrix

Before conducting regression analysis, it's noted that the CO2kt and HealthPC$ columns contain missing values, accounting for 5.64% and 7.34% of the total samples, respectively. Given the significant correlation between GDP per capita and Health Expenditure per capita, the better way to address these missing values is through imputation. The method of imputation will be decided based on a comparative analysis of correlations after separately employing mean and median imputation.

Imputing the missing data with mean			***Imputing the missing data with median***		
	LifeExpect	*HealthPC$*		*LifeExpect*	*HealthPC$*
LifeExpect	*1*	*0.597936*	*LifeExpect*	*1*	*0.567068*
HealthPC$	*0.597936*	*1*	*HealthPC$*	*0.567068*	*1*

Figure 7: Correlation Comparison

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             LifeExpect   R-squared:                       0.449
Model:                            OLS   Adj. R-squared:                  0.430
Method:                 Least Squares   F-statistic:                     23.13
Date:                Mon, 27 Nov 2023   Prob (F-statistic):           7.23e-20
Time:                        13:38:42   Log-Likelihood:                -551.18
No. Observations:                 177   AIC:                             1116.
Df Residuals:                     170   BIC:                             1139.
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         69.1624      0.540    128.130      0.000      68.097      70.228
GDPPC          0.0001   2.57e-05      4.215      0.000    5.76e-05       0.000
MfgMn$     -7.703e-06    5.4e-06     -1.428      0.155   -1.84e-05    2.95e-06
AgriMn$     3.245e-05   2.79e-05      1.164      0.246   -2.26e-05    8.75e-05
CO2kt       1.893e-06   3.05e-06      0.620      0.536   -4.13e-06    7.92e-06
HealthPC$      0.0015      0.000      4.118      0.000       0.001       0.002
Pop_mn        -0.0144      0.009     -1.557      0.121      -0.033       0.004
==============================================================================
Omnibus:                       15.367   Durbin-Watson:                   2.076
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               17.174
Skew:                          -0.760   Prob(JB):                     0.000187

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.28e+06. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Figure 8: Multiple Linear Regression Life Expectancy vs Given Variables

As depicted in Figure 8, it's evident that the standard error for each variable is minimal, indicating a relatively reliable estimation of coefficients. Adding more variables to the regression, the adjusted R-squared value of 0.430 demonstrates that our model accounts for about 43% of the variability in life expectancy based on the given factors. Moreover, coefficients for Manufacturing Value Added, Agriculture, Forestry, and Fishing Value Added, CO_2 emissions, and Population appear close to zero, implying limited contributions to explaining life expectancy variability, possibly lacking significance. Notably, besides GDP per capita, the coefficient for Current Health Expenditure per capita stands out at 0.0015 with a p-value below the 0.01 threshold, indicating both GDP per capita and Current Health Expenditure per capita play significant roles in explaining their correlation with life expectancy.