Puja Sridhar
Ps1393

BACKGROUND AND PREDICTIVE ANALYSIS

With an initial study of publicly accessible datasets, my interest is to investigate the factors influencing life expectancy in various countries, utilizing the WHO life expectancy dataset. As the importance of global health continues to grow, it is valuable to predict life expectancy using different socio-economic indicators. These features help in directing the distribution of resources, implementing health interventions, and making policy choices, rendering this research essential for international health organizations, governments, and public health experts.

Life expectancy is a crucial measure of the general health and welfare of a population. This research suggests **estimating countries' life expectancy by considering their GDP per capita, adult mortality rate, and health expenditure per capita, using country as the unit of analysis.**

Life expectancy is greatly influenced by the reliance on technology and the availability of resources like education and healthcare. Analyzing these factors allows us to discover important patterns and connections that influence both policy and practice within the health sector.

MODEL

This study will employ a multi-linear regression model. The model will use three independent variables to predict a single dependent variable. The proposed variables, all of which are numeric, are as follows:

Dependent Variable (A):

- Life Expectancy (years)

Independent Variables (B, C, D):

- GDP per capita (US Dollars)
- Adult Mortality rate (Numerical)
- Health expenditure per capita (US Dollars)

Unit of Analysis (U): Country-wise

DATA SOURCE

Life Expectancy, GDP per capita, Adult Mortality Rate, Health Expenditure per capita:

Source: World Health Organization (WHO) Life Expectancy Dataset

URL: [WHO Life Expectancy Dataset]

The proposed model can be represented by the following regression equation:

Life Expectancy$_i$ = β_0 + β_1 (GDP per capita)$_i$ + β_2 (Adult literacy rate)$_i$ + β_3(Health expenditure per capita)$_i$ + ϵ

Where,

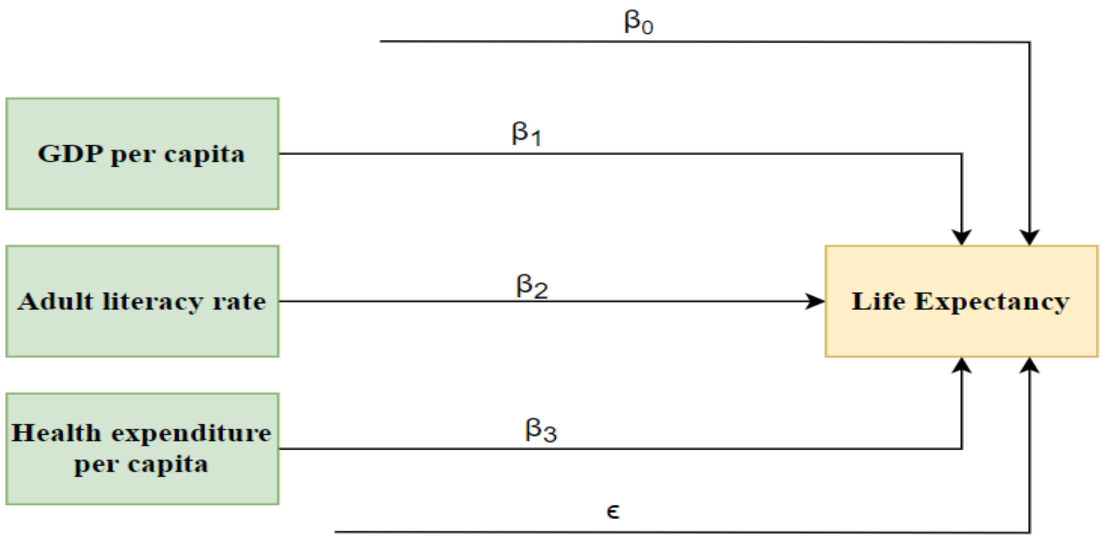
β_0 is the intercept,

β_1, β_2, β_3 are the coefficients for each predictor

ϵ is the error term

DIAGRAM

The diagram below aims to visualize the proposed model for this research effort:



Unit of Analysis - country-year

NEXT STEP IN ANALYSIS

Firstly, Handling missing values by employing imputation methods such as mean/mode substitution, regression imputation, or multiple imputation is crucial. Following data cleaning, Visual tools such as histograms, box plots, and scatter plots are used to explore relationships between variables. Descriptive statistics summarize data to offer initial insights into the dataset, leading to a better grasp of the features and possible relationships.

Correlation analysis then assesses the strength and direction of relationships between life expectancy and GDP per capita, adult literacy rate, and health expenditure per capita. Developing a correlation matrix aids in the visualization and interpretation of relationships among all features included. Followed by Regression analysis using multi-linear regression to model the relationship between life expectancy and the predictors. The regression equation is

Life Expectancy$_i$ = β_0 + β_1 (GDP per capita)$_i$ + β_2 (Adult literacy rate)$_i$ + β_3(Health expenditure per capita)$_i$ + ϵ

The coefficients (β_0,β_1,β_2,β_3) are estimated using the dataset. Finally, evaluating the model's performance involves using metrics like Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).

CURATION PROCESS

Even though the datasets on their own are helpful and provide information, they need to be properly curated as a whole to guarantee accuracy in analysis. Maintaining a consistent format for each variable is crucial for the accuracy of this analysis. All input and output variables are sourced from similar worldwide databases that cover multiple countries. Below is a comprehensive overview of the curation method employed for this task:

The first step involved collecting data from reliable sources. The primary dataset was the WHO Life Expectancy Dataset, which includes variables such as life expectancy, GDP per capita, adult mortality rate, and health expenditure per capita. This data was sourced from the World Health Organization, United Nations Population Division, World Bank, Our World in Data, and other reputable databases, covering the period from 2000 to 2015 and including data from 179 countries. Ensuring comprehensive and accurate data collection was crucial to the study.

Afterward, the data was cleaned to address any missing values and maintain uniformity. This required eliminating missing data for countries without any preceding or subsequent values, as using imputation in these instances would be inappropriate. Examples of countries like Iran, Bahamas, etc. For other missing values, imputation was performed by changing them to the mean of three values belonging to the same country. A consistency check was performed to ensure that each variable remained in a uniform format across all datasets.

The next steps involved fine-tuning and structuring the data. Territorial information was excluded to concentrate solely on countries, guaranteeing the absence of any major gaps in data. Outliers or extreme values that could affect the analysis were also removed. The final dataset was organized into a user-friendly format, with clear labels and color coding: purple for units of analysis, green for input variables, and orange/yellow for the output variable. A secondary sheet in the Excel workbook shows the raw data.

[DataSet2]

Statistics

		AdultMortality	infantdeaths	percentageexpenditure	GDP	Population	Lifeexpectancy
N	Valid	2928	2938	2938	2490	2286	2928
	Missing	10	0	0	448	652	10

Country	Year	Status	Life expec	Adult Mor	infant dea	Alcohol	percentag	Hepatitis	Measles	BMI	under-five	Polio	Total expe	Diphtheri	HIV/AIDS	GDP	Populatior	thinn
Afghanistan	2015	Developin	65	263	62	0.01	71.27962	65	1154	19.1	83	6	8.16	65	0.1	584.2592	33736494	
Afghanistan	2014	Developin	59.9	271	64	0.01	73.52358	62	492	18.6	86	58	8.18	62	0.1	612.6965	327582	
Afghanistan	2013	Developin	59.9	268	66	0.01	73.21924	64	430	18.1	89	62	8.13	64	0.1	631.745	31731688	
Afghanistan	2012	Developin	59.5	272	69	0.01	78.18422	67	2787	17.6	93	67	8.52	67	0.1	669.959	3696958	
Afghanistan	2011	Developin	59.2	275	71	0.01	7.097109	68	3013	17.2	97	68	7.87	68	0.1	63.53723	2978599	
Afghanistan	2010	Developin	58.8	279	74	0.01	79.67937	66	1989	16.7	102	66	9.2	66	0.1	553.3289	2883167	
Afghanistan	2009	Developin	58.6	281	77	0.01	56.76222	63	2861	16.2	106	63	9.42	63	0.1	445.8933	284331	
Afghanistan	2008	Developin	58.1	287	80	0.03	25.87393	64	1599	15.7	110	64	8.33	64	0.1	373.3611	2729431	
Afghanistan	2007	Developin	57.5	295	82	0.02	10.91016	63	1141	15.2	113	63	6.73	63	0.1	369.8358	26616792	
Afghanistan	2006	Developin	57.3	295	84	0.03	17.17152	64	1990	14.7	116	58	7.43	58	0.1	272.5638	2589345	
Afghanistan	2005	Developin	57.3	291	85	0.02	1.388648	66	1296	14.2	118	58	8.7	58	0.1	25.29413	257798	
Afghanistan	2004	Developin	57	293	87	0.02	15.29607	67	466	13.8	120	5	8.79	5	0.1	219.1414	24118979	
Afghanistan	2003	Developin	56.7	295	87	0.01	11.08905	65	798	13.4	122	41	8.82	41	0.1	198.7285	2364851	
Afghanistan	2002	Developin	56.2	3	88	0.01	16.88735	64	2486	13	122	36	7.76	36	0.1	187.846	21979923	
Afghanistan	2001	Developin	55.3	316	88	0.01	10.57473	63	8762	12.6	122	35	7.8	33	0.1	117.497	2966463	
Afghanistan	2000	Developin	54.8	321	88	0.01	10.42496	62	6532	12.2	122	24	8.2	24	0.1	114.56	293756	
Albania	2015	Developin	77.8	74	0	4.6	364.9752	99	0	58	0	99	6	99	0.1	3954.228	28873	
Albania	2014	Developin	77.5	8	0	4.51	428.7491	98	0	57.2	1	98	5.88	98	0.1	4575.764	288914	
Albania	2013	Developin	77.2	84	0	4.76	430.877	99	0	56.5	1	99	5.66	99	0.1	4414.723	289592	
Albania	2012	Developin	76.9	86	0	5.14	412.4434	99	9	55.8	1	99	5.59	99	0.1	4247.614	2941	
Belize	2008	Developin	69.6	181	0	7.22	51.25232	94	0	45.7	0	94	5.9	94	0.1	447.228	36165	
Belize	2007	Developin	69.6	181	0	7.24	69.63051	96	0	45	0	97	4.76	96	0.6	4324.876	29847	
Belize	2006	Developin	69.4	184	0	6.48	387.3325	98	0	44.4	0	98	4.4	98	0.6	4187.378	29747	
Belize	2005	Developin	69	191	0	6.25	365.7999	96	0	43.8	0	96	4.45	96	0.8	3933.332	283277	
Belize	2004	Developin	68.7	197	0	6.2	325.6807	97	0	43.2	0	97	4.39	97	0.8	3831.538	27689	
Belize	2003	Developin	68.4	21	0	5.67	312.7996	96	0	42.6	0	95	4.53	96	1.5	3679.995	26913	
Belize	2002	Developin	68.5	199	0	4.99	262.8299	97	0	42	0	93	4.38	89	0.1	3556.562	26226	
Belize	2001	Developin	68.2	21	0	4.9	251.6587	96	0	41.4	0	96	4.5	96	0.4	3419.276	254984	
Belize	2000	Developin	68.3	196	0	4.79	219.024	76	0	4.8	0	91	3.98	91	0.3	3364.424	247315	
Benin	2015	Developin	60	249	25		0	82	55	25.7	39	78		82	1	783.9479	1575952	
Benin	2014	Developin	59.7	252	25	0.01	90.12207	78	786	25.2	39	74	4.59	78	1.1	943.6866	1286712	
Benin	2013	Developin	59.5	251	25	0.01	87.40804	77	637	24.6	39	73	4.59	77	1.2	915.2675	14451	
Benin	2012	Developin	59.3	251	25	0.01	9.804075	8	288	24.1	39	8	4.86	8	1.3	837.9551	972916	
Benin	2011	Developin	59.1	251	25	1.4	110.2634	75	426	23.5	39	77	5.37	75	1.4	825.9428	94682	
Benin	2010	Developin	58.7	254	25	1.33	94.25738	76	392	23	39	77	4.95	76	1.4	757.696	9199259	
Benin	2009	Developin	58.4	259	25	1.16	71.01399	79	1461	22.5	39	8	4.46	79	1.6	793.4524	894476	
Benin	2008	Developin	57.6	278	25	1.28	7.61543	75	928	22	39	77	4.2	75	1.8	82.15135	8696916	
Benin	2007	Developin	57.1	283	25	1.12	7.492818	82	341	21.5	39	82	4.55	82	2	76.53542	8454791	
Benin	2006	Developin	56.8	284	25	1.19	75.91429	74	176	21	39	76	4.75	74	2	625.8392	8216896	
Benin	2005	Developin	56.5	285	25	1.13	7.106997	7	210	2.5	39	73	4.73	7	2.1	61.79998	7982225	
Benin	2004	Developin	56.1	285	25	1.15	10.73628	75	262	2.1	39	74	4.56	72	2.1	583.4935	7754	

Snippets of the WHO Life Expectancy Dataset (Before)

UNIT OF ANALYSIS		INPUT FEATURES			OUTPUT FEATURE
Country	Year	Adult Mortality	Percentage Expenditure	GDP	Life Expectancy
Afghanistan	2000	321	10.424960000	114.5600000	54.8
Afghanistan	2001	316	10.574728200	117.4969800	55.3
Afghanistan	2002	3	16.887350910	187.8459500	56.2
Afghanistan	2003	295	11.089052730	198.7285436	56.7
Afghanistan	2004	293	15.296066430	219.1413528	57.0
Afghanistan	2005	291	1.388647732	25.2941299	57.3
Afghanistan	2006	295	17.171517510	272.5637700	57.3
Afghanistan	2007	295	10.910155980	369.8357960	57.5
Afghanistan	2008	287	25.873925360	373.3611163	58.1
Afghanistan	2009	281	56.762216820	445.8932979	58.6
Afghanistan	2010	279	79.679367360	553.3289400	58.8
Afghanistan	2011	275	7.097108703	63.5372310	59.2
Albania	2000	11	91.711540520	1175.7889810	72.6
Albania	2001	14	96.205570780	1326.9733900	73.6
Albania	2002	15	104.516915700	1453.6427770	73.3
Albania	2003	18	14.719288820	189.6815570	72.8
Albania	2004	17	221.842800000	2416.5882350	73.0
Croatia	2008	116	2425.403891000	15893.8656000	76.0
Croatia	2009	19	2160.380199000	14157.1441600	76.3
Croatia	2010	16	206.886818200	1355.7458600	76.6
Croatia	2011	14	1913.356642000	14539.1842100	77.0
Croatia	2012	14	1851.713262000	13235.9775700	77.1
Croatia	2013	97	1899.107385000	13574.7490000	77.7
Croatia	2014	97	1884.098811000	13467.4682700	77.8
Croatia	2015	95	0.000000000	11579.6674000	78.0
Cuba	2000	115	49.340077560	2741.1154200	76.9
Cuba	2001	115	322.586314900	2832.1888930	76.7
Cuba	2002	19	334.767210700	2994.3399880	77.7
Cuba	2003	18	60.341451240	3192.6693780	77.4
Cuba	2004	17	4.484254579	339.7162560	77.3
Cuba	2005	19	518.935922100	3779.5770000	77.2
Cuba	2006	14	523.472428400	4669.6916000	78.0
Cuba	2007	14	750.714764100	5184.4942270	78.1
Cuba	2008	12	708.615796600	5376.4476220	77.9
Cuba	2009	11	818.877101600	5484.7763000	78.1
Cuba	2010	98	787.280816300	5676.1414300	78.0
Cuba	2011	92	102.064578400	675.9243600	78.8
Cuba	2012	96	742.196199200	6425.9411190	78.7

Snippets of the WHO Life Expectancy Dataset (After)

SPSS STATISTICAL ANALYSIS

Correlations

		Adult Mortality	Percentage Expenditure	GDP	Life Expectancy
Adult Mortality	Pearson Correlation	1	-.254**	-.297**	-.687**
	Sig. (2-tailed)		<.001	<.001	<.001
	N	2465	2465	2465	2465
Percentage Expenditure	Pearson Correlation	-.254**	1	.900**	.409**
	Sig. (2-tailed)	<.001		<.001	<.001
	N	2465	2465	2465	2465
GDP	Pearson Correlation	-.297**	.900**	1	.464**
	Sig. (2-tailed)	<.001	<.001		<.001
	N	2465	2465	2465	2465
Life Expectancy	Pearson Correlation	-.687**	.409**	.464**	1
	Sig. (2-tailed)	<.001	<.001	<.001	
	N	2465	2465	2465	2465

**. Correlation is significant at the 0.01 level (2-tailed).

Correlation Scores

Adult Mortality & Life Expectancy ➡ $r = -0.687$

Percentage Expenditure & Life Expectancy ➡ $r = 0.409$

GDP & Life Expectancy ➡ $r = 0.464$

Significance Values

Adult Mortality & Life Expectancy ➡ $p < 0.01$

Statistically significant ($p < 0.05$)

Percentage Expenditure & Life Expectancy ➡ $p < 0.01$

Statistically significant ($p < 0.05$)

GDP & Life Expectancy ➡ $p < 0.01$

Statistically significant ($p < 0.05$)

Correlation Directions

Adult Mortality & Life Expectancy ➡ Negative direction

Percentage Expenditure & Life Expectancy ➡ Positive direction

GDP & Life Expectancy ➡ Positive direction

Correlations: Interpretations and Analysis

Comparing the r-values for each of the independent variable correlations with the dependent Life Expectancy variable, it appears that Adult Mortality has a stronger linear relationship compared to Percentage Expenditure and GDP because their absolute values are closer to zero whereas the absolute value of Adult Mortality is closer to one.

The correlation value (r) of -0.687 indicates a strong negative relationship between **adult mortality and life expectancy**. This suggests that as adult mortality increases, life expectancy decreases. The higher absolute r-value indicates a more linear, and less scattered relationship between adult mortality and life expectancy.

The correlation value of 0.409 indicates a moderate positive relationship between the **percentage expenditure on health and life expectancy**. This means that as the percentage of health expenditure increases, life expectancy tends to increase. The lower absolute r-value indicates a less linear, and more scattered relationship between percentage expenditure on health and life expectancy.

The correlation value of 0.464 indicates a moderate positive relationship between **GDP and life expectancy**. This suggests that as GDP increases, life expectancy also tends to increase. The lower absolute r-value indicates a less linear, and more scattered relationship between GDP and life expectancy.

Comparing the p-values for each of the independent variable and dependent variable relationships, it appears that all the relationships produced a p-value lower than the 0.05 critical value.

The p-value derived from the relationship between **Adult Mortality and Life Expectancy** is less than 0.01, which is below the 0.05 critical value. This implies that there is a high likelihood of this correlation being genuine rather than coincidental. Policymakers should prioritize reducing adult mortality rates through improved healthcare, disease prevention, and better living conditions to enhance life expectancy.

The p-value calculated from the relationship between **Percentage Expenditure and Life Expectancy** is less than 0.01, which is below the 0.05 critical value. This implies statistical significance and indicates a genuine correlation with a low chance of coincidence. Increasing health expenditure is crucial for improving public health outcomes and life expectancy. Policymakers should consider allocating more resources to the health sector.

The p-value of the relationship between **GDP and Life Expectancy** is less than 0.01, which is below the 0.05 critical value. This implies that the correlation is statistically significant and is likely to be genuine rather than coincidental. Increasing health expenditure is crucial for improving public health outcomes and life expectancy. Policymakers should consider allocating more resources to the health sector.

The analysis reveals statistically significant relationships between life expectancy and the three independent variables: adult mortality, percentage expenditure on health, and GDP. Each variable demonstrates a significant correlation with life expectancy, indicating their importance in influencing public health outcomes.

Through bivariate correlation analysis, it is evident that the relationships between life expectancy and the independent variables are significant and meaningful. The p-values for all relationships are below the 0.05 critical value, indicating the statistical significance and genuine nature of these correlations. These findings provide valuable insights for policymakers in formulating strategies to improve public health and life expectancy.

Multiple Linear Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.739[a]	.546	.546	6.5041

a. Predictors: (Constant), GDP, Adult Mortality, Percentage Expenditure

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	75.349	.248		303.321	<.001
	Adult Mortality	-.046	.001	-.602	-42.330	<.001
	Percentage Expenditure	-5.960E-6	.000	-.001	-.043	.966
	GDP	.000	.000	.286	9.082	<.001

a. Dependent Variable: Life Expectancy

Percentage of Variance Explained by the Created Model

- R – Square = 0.546 = 54.6%
  Which means 54.6% of the variance was explained by the created model.

Significant Variables & Coefficients

- Significance
  - Adult Mortality & Life Expectancy ➡ $p < 0.01$
    Statistically significant ($p < 0.05$)
  - Percentage Expenditure & Life Expectancy ➡ $p = 0.996$
    Not Statistically significant ($p > 0.05$)
  - GDP & Life Expectancy ➡ $p < 0.01$
    Statistically significant ($p < 0.05$)

- Unstandardized Coefficients
  An increase of the independent variables by an index of one will reflect a change in the index for the dependent variable in the amount of the specified unstandardized coefficients.
  - Adult Mortality: -0.046 (Negative)
    For every 1 unit of change in Adult Mortality, the dependent variable (Life Expectancy), decreases by -0.046.
  - Percentage Expenditure: -5.960E-6 (Negative)
    For every 1 unit of change in Percentage Expenditure, the dependent variable (Life Expectancy), decreases by -5.960E-6.

- GDP: 0.000
  For every unit of change in the GDP, the dependent variable (Life Expectancy), has no change.

The unstandardized coefficients enable us to calculate the estimated equation to predict Life Expectancy, which is as follows:
ŷ = 75.349 - 0.046 * (Adult Mortality) – 5.960E-6 * (Percentage Expenditure) – 0.000 * (GDP)

- **Standardized Coefficients**
  An increase of the independent variables by a standard deviation of one will reflect a change in standard deviation for the dependent variable in the amount of the specified standardized coefficients.
  - Adult Mortality: -0.602 (Negative)
    For every 1 standard deviation of movement in Adult Mortality, the dependent variable (Life Expectancy), decreases by 0.602 standard deviations.
  - Percentage Expenditure: -0.001 (Negative)
    For every 1 standard deviation of movement in Percentage Expenditure, the dependent variable (Life Expectancy), decreases by 0.001 standard deviations.
  - GDP: 0.286 (Positive)
    For every 1 standard deviation of movement of the GDP, the dependent variable (Life Expectancy), increases by 0.286 standard deviations.

Multiple Linear Regression: Interpretations and Analysis
The model summary provided an R-square value of 0.546, meaning that 54.6% of the variance in the dependent variable (Life Expectancy) is explained by the independent variables (Adult Mortality, Percentage Expenditure, and GDP). R-square values in such models fall between 0% and 100%. Given that more than half of the variance is explained by the model, it suggests a moderately strong fit, indicating that these predictors are meaningful in understanding variations in life expectancy.
Unlike the bivariate correlation results from the previous sections, only two of the three independent variables were statistically significant.

- The p-value derived from the relationship between **Adult Mortality & Life Expectancy** is lesser than the 0.05 critical value. This implies statistical significance and a stronger chance of a genuine result.
- The p-value calculated from the relationship between **Percentage Expenditure & Life Expectancy** was greater than the 0.05 critical value. This implies a greater chance of coincidence, given that it is not statistically significant. Thus, this variable remained the only one of the three independent variables to possess a less genuine and more coincidental relationship with the dependent Life Expectancy variable.
- Finally, the p-value of the relationship between the **GDP & Life Expectancy** was lesser than the 0.05 critical value. This implies statistical significance and a stronger chance of a genuine result.

Examining the coefficients, it appears that for both standardized and unstandardized coefficients, Adult Mortality has a negative relationship. This implies that Life Expectancy decreases in index (for the unstandardized coefficient) and standard deviation (for the standardized coefficient) with an increase in Adult Mortality. In contrast, GDP demonstrated a positive relationship, indicating that Life Expectancy increases with GDP in both index and standard deviation.

Significant Variables & Coefficients

Adult Mortality & Life Expectancy

- **Unstandardized Coefficient:** -0.046 (Negative)
  - For every one-unit increase in Adult Mortality, the dependent variable (Life Expectancy) decreases by 0.046 units.
- **Standardized Coefficient:** -0.602 (Negative)
  - For every one standard deviation increase in Adult Mortality, Life Expectancy decreases by 0.602 standard deviations.

Percentage Expenditure & Life Expectancy

- **Unstandardized Coefficient:** -5.960E-6 (Negative)
  - For every one-unit increase in Percentage Expenditure, the dependent variable (Life Expectancy) decreases by 5.960E-6 units.
- **Standardized Coefficient:** -0.001 (Negative)
  - For every one standard deviation increase in Percentage Expenditure, Life Expectancy decreases by 0.001 standard deviations.

GDP & Life Expectancy

- **Unstandardized Coefficient:** 0.000 (No change)
  - For every one-unit increase in GDP, there is no change in Life Expectancy.
- **Standardized Coefficient:** 0.286 (Positive)
  - For every one standard deviation increase in GDP, Life Expectancy increases by 0.286 standard deviations.

These results provide several practical insights. To begin, the model implies that Adult Mortality and GDP are significant predictors of Life Expectancy. The negative relationship with Adult Mortality suggests that reducing adult mortality rates could improve life expectancy. The positive relationship with GDP indicates that economic growth contributes to higher life expectancy, reflecting better access to healthcare, nutrition, and living conditions.

Approaching policy maker perspectives is two-fold. Health policymakers should focus on strategies to reduce adult mortality through improved healthcare access, disease prevention, and health education. Economic policymakers should aim to foster economic growth, as higher GDP is associated with increased life expectancy.

Organizational policy makers, such as those in healthcare systems and economic development agencies, can use these insights to advocate for and implement policies that target these key areas. By addressing adult mortality and promoting economic stability, a substantial impact on life expectancy can be achieved.

Ethical Implications, Study Limitations, and Future Research

Given the accessibility of publicly maintained databases, the dataset used for this study was sourced from WHO. The data, consisting primarily of numerical values, was collected to prevent any practical means of de-anonymizing individuals. This ensures the ethical integrity of the study, avoiding any risk of compromising personal identities. Researchers working with more detailed data may need help maintaining such ethical standards.

Although the study faced no significant ethical limitations, several practical limitations could have affected the consistency and expected outcomes of the results. One key restriction involves the practicality of the outcome variable, life expectancy.

- **Definition Ambiguities:** Life expectancy can be influenced by various factors, including regional healthcare quality and socio-economic conditions. Clear, standardized definitions and measurements are essential for accurate population-level insights.
- **Data Collection Accuracy:** The data was collected through publicly accessible sources, which might not capture the complete picture. More precise data from healthcare providers or government records could yield more accurate and reliable results.

Given additional time and resources, future research could dive deeper into specific aspects to gain more detailed insights. Enhancing the granularity of the variables would be a key focus. Here are some directions for future research:

- **GDP per Capita:** Study the influence of various sectors on GDP and how sector-specific growth is associated with life expectancy. For instance, contrast how growth in the healthcare industry versus the industrial sector impacts life expectancy.
- **Adult Literacy Rate:** Analyze how various education levels (primary, secondary, tertiary) impact health outcomes and life expectancy through adult literacy rates. This could offer valuable information on which levels of education are most influential on health outcomes.
- **Health Expenditure per Capita:** Examine the allocation of health expenditure in more detail. For instance, differentiate between preventive healthcare spending and curative healthcare spending to understand which has a greater impact on life expectancy.

- **Regional Disparities:** Explore the regional differences in life expectancy and the factors that impact them. Understanding the reasons behind the superior performance of some regions can assist in customizing more efficient health policies.

Apart from these specific variable adjustments, obtaining more recent data is crucial. The current study may have used data spanning several years, and more recent data would reflect current trends and changes more accurately. This is particularly relevant in a rapidly changing world where technological advancements and socio-economic shifts occur frequently.

REFERENCES

- Data contains life expectancy, health, immunization, and economic and demographic information about 179 countries from 2000-2015 years. The adjusted dataset has 21 variables and 2.864 rows.
  https://www.kaggle.com/datasets/lashagoch/life-expectancy-who-updated?select=Life-Expectancy-Data-Updated.csv
- United Nations Population Division. World Population Prospects: 2022 Revision; or derived from male and female life expectancy at birth from sources such as: Statistical databases and publications from national statistical offices; Eurostat: Demographic Statistics.
  https://data.worldbank.org/indicator/SP.DYN.LE00.IN?name_desc=false
- India: Life expectancy at birth from 2011 to 2021
  https://www.statista.com/statistics/271334/life-expectancy-in-india/
- Global data and research on life expectancy and related measures of longevity: the probability of death at a given age, the sex gap in life expectancy, lifespan inequality within countries, and more.
  https://ourworldindata.org/life-expectancy
- Lincolnshire County Council Life Expectancy
  https://www.data.gov.uk/dataset/27cb8743-a623-424c-b097-12c398697828/life-expectancy