https://doi.org/ 10.48047/AFJBS.6.5.2024. 9355-9367

African Journal of Biological Sciences

STATISTICAL ANALYSIS OF FACTORS RELATED TO THE SUSTAINABLE DEVELOPMENT GOALS AND THEIR IMPACT ON LIFE EXPECTANCY IN INDIA

Dr Sowbhagya S Prabhu* and Rajuva

Department of Statistics
Rajagiri college of Social Sciences
Kalamassery, Kerala, India
*Corresponding author Email:sowbhagyasprabhu@gmail.com

Article History
Volume 6, Issue 5, 2024
Received: 22 May 2024
Accepted: 03 Jun 2024
doi: *10.48047/AFJBS.6.5.2024. 9355-9367*

Abstract: This research aims to conduct a comprehensive statistical analysis to explore the intricate relationship between the Sustainable Development Goals (SDGs) and life expectancy in India. As the nation strives to achieve the SDGs outlined by the United Nations, understanding the impact of these goals on the overall well-being of its population, particularly life expectancy, is crucial. The study will employ statistical methods to analyse extensive datasets encompassing indicators related to each SDG and life expectancy trends in India. Key variables will include indicators such as GDP, GNI, School enrolment, Unemployment etc. By utilizing regression analysis and other statistical techniques, the research will seek to identify significant correlations and assess the extent to which progress in SDGs contributes to improvements in life expectancy.

Keywords: Life Expectancy, Socioeconomic Factors, Sustainable development Goals

- **Introduction**

The Sustainable Development Goals (SDGs) represent a global commitment to end poverty, protect the environment, and promote universal peace and prosperity by 2030, which was adopted by the UN in 2015.The Sustainable Development Goals (SDGs), which have 17 interrelated targets, recognize the inherent connection between different aspects of development. They emphasize the necessity of a coordinated strategy, acknowledging that developments in one area have a substantial influence on results in other areas—thus highlighting the necessity of striking a balance between social, economic, and environmental sustainability.

Global governments have committed to provide the most marginalized groups the highest priority in order to ensure that progress is inclusive and advantageous to all. The Sustainable Development Goals (SDGs) provide a comprehensive framework for addressing gender inequality, ending discrimination against women and girls, and preventing hunger, poverty, and diseases like AIDS.

A country's life expectancy is a reflection of its social, economic, and health situations as well as its healthcare system. Life Expectancy is the average expected life span (as determined by statistics) of a person or other entity from the year of birth to the present. It is frequently employed as a gauge of the general growth of a nation. The phenomena of Life Expectancy is essential to understanding a nation's overall state, particularly in regard to mortality and the economy. Over the 20th and 21st centuries, life expectancy has significantly increased in high-income, developed nations. The primary focus of Health science has always been Life Expectancy. The frequency of HIV, death rates, healthcare resources, and results were the health-related predictors of Life Expectancy.

Numerous researches have looked into the factors that affect life expectancy in both developed and developing nations. The six primary factors of life expectancy for 95 developing countries were examined by Rogers and Wofford in their study, "Life expectancy in less developed countries: socio-economic development or public health." They discovered that factors influencing emerging countries' life expectancy include urbanization, population growth associated with agriculture, the percentage of illiterates, access to clean drinking water, average daily caloric intake, and the number of doctors per population.

The study conducted by Nandi DC, Hossain MF, Roy P, Ullah MS (2023) examined the relationship between life expectancy and socioeconomic variables for the Sustainable Development Goals (SDG) in Bangladesh through path analysis. The findings indicated that while all factors have an impact on life expectancy, GNI, employment rate, and age dependency ratio are the primary determinants of LE. Age dependency ratios can be lowered and life

expectancies raised with the help of employment opportunities and GDP growth. We can observe that there are variations in the variables deemed significant for explaining life expectancy from the article A Statistical Analysis Concerning The Sustainable Development Goals and Life Expectancy. This can be related to the degree of development in the particular nation. For instance, in developing nations, enhancing basic infrastructure could be crucial, but if it already exists, it might not be as vital.

As stated in the essay by Radhika Bhanja1 and Koel Roychowdhury, Assessing the Progress of India towards Sustainable Development Goals by 2030, Even while India appeared to have done pretty well on some indicators at the national level, a thorough investigation conducted at the sub-national level showed that many of the states are falling significantly short of the benchmarks in those sustainable development goal indicators. At the sub-national level, the indicators of deliberate homicide and access to facilities for safe drinking water are only partially met, compared to the national level. At the subnational level, the transitional indicators for India have done mediocrely. For most states, the maternal mortality rate and undernourishment indices have shown dismal performance.

We may conclude from all of these studies that many factors influence life expectancies in various contexts and nations. India's life expectancy in 2023 is predicted to be 70.42 years, up 0.33% from 2022. In India, the life expectancy in 2022 was 70.19 years, up 0.33% from 2021. India's life expectancy increased by 0.33% from 2020 to 69.96 years in 2021. In order to make progress and achieve a larger increase in life expectancy in the future, it can be helpful to identify which socioeconomic factors impacting life expectancy need to be studied. Due to their strong relationship to the sustainable development goals, improvements in these socioeconomic determinants may eventually result in advancements toward the SDGs.

- **Objectives and methodology**

Finding the variables that affect life expectancy is the aim of this research. To determine whether life expectancy and public health are impacted by some, all, or none of the SGDs that are covered by the probable regressors. Given that the SDGs are meant to be a guide for global peace and prosperity, it is worthwhile to look at this. Thus, it becomes sense to assume that the focus will be on public health, which in turn affects life expectancy. If is determined that any or all of the above factors significantly affect life expectancy, this knowledge might be helpful in prioritizing the steps that should be taken to enhance public health.

We employed a variety of statistical approaches and software, Python and SPSS, to ensure that the analysis was accurate and simple. Numerous fundamental statistical functions are offered by these two statistical programs. The statistical program SPSS are used for analysis

and model fitting. Microsoft Excel was also utilized for this research.

Any significant and valuable research project starts with its data source. The sociodemographic and health factors that had the significant effects on Life Expectancy in the previous studies were considered for this study. The World Bank's World Development Indicators are the source of the data. In order to demonstrate the data analysis, we used 14 variables. Table 1 lists the initially selected variables' names and identifiers.

Variables	**Identifiers**
Life expectancy at birth	y
Prevalence of undernourishment	x1
Food production index	x2
Mortality rate, under-5	x3
Prevalence of HIV	x4
Fertility rate, total	x5
Labor force, female	x6
People using at least basic drinking water services	x7
Age dependency ratio	x8
Population density	x9
GNI per capita, PPP	x10
GDP per capita, PPP	x11
Unemployment	x12
School enrollment	x13

Table 1: Variable names and identifiers

- **Statistical analysis**

The univariate, bivariate and multivariate analysis are carried out in this research. Univariate analysis was conducted to assess the selected variables concerning maximum, minimum, mean, standard deviation (SD), median, and standard error of the mean (SE mean). This approach proves valuable as the study variables are measured across diverse units.

For bivariate analysis, Pearson correlation analysis was employed. This allowed for the derivation of correlation coefficients (r) to explore the direction, strength, and significance of linear associations between the study variables. To further understand the average relationship among significant predictor variables, Backward elimination multiple linear regression analysis was utilized. This method aimed to examine the collective impact of these predictors on the outcome variable.

- **Results and discussion**

Univariate Analysis

Variables	Minimum	Maximum	Mean	Median	Standard Deviation	1st Quartile	3rd Quartile
Life expectancy at birth	58.652	70.91	64.99775	65.204	3.860651984	61.691	68.03025
Prevalence of undernourishment	12.9	22	16.52857143	16.52857	2.26902606	15.25	16.54642857
Food production index	47.88	123.84	77.9178125	70.72	23.09907521	59.7775	97.0725
Mortality rate, under-5	30.6	126.5	74.66875	72.75	30.04584065	48.375	99.8
Prevalence of HIV	0.1	0.6	0.3375	0.3	0.138540782	0.2	0.425
Fertility rate, total	2.031	4.045	2.951875	2.911	0.628646306	2.381	3.463
Labor force, female	20.98563563	26.78427181	24.83060092	25.2405	1.535373318	23.86732503	25.9192532
People using at least basic drinking water services	79.87927122	92.72465388	86.39742641	86.39743	3.277322277	84.75925144	88.13361811
Age dependency ratio	38.05008458	65.51317234	52.48935592	52.84938	8.640696836	45.28129646	60.02187555
Population density	292.7670835	473.4187327	388.1663454	391.3326	55.80496126	341.9424051	435.6131544
GNI per capita, PPP	1190	7220	3485	3060	1905.364244	1807.5	4995
GDP per capita, PPP	1204.242059	7367.994665	3522.448222	3079.129	1931.699652	1823.714004	5059.5105
Unemployment	6.51	10.195	7.8145	7.8785	0.766693908	7.2545	8.32775
School enrollment	91.30232239	108.3117065	98.22464052	97.76905	5.700167667	93.14939308	101.6793926

Table2: Descriptive Statistics of variables

From Table 2 we can observe minimum, maximum, mean, standard deviation, 1st and 3rd quartiles of all the variables involved in this study. The lowest life Expectancy of 58.658 was recorded in 1990 and the highest recorded life expectancy was 70.91 in 2019 and the approximate value of average life expectancy in India is calculated as 65. The prevalence of undernourishment was recorded minimum in the year 2017 and maximum in the year 2004.

The highest food production index of 123.84 was recorded in the year 2021.The mortality rate (under 5) was maximum during the year 1990 and we can see mortality rate gradually declining in the years after 1990. Fertility rate was maximum in the year 1990. The maximum labour force was observed in the year 2005 and minimum in the year 2018.People using at least basic drinking water was minimum during 2000 and maximum in 2021. Age dependency ratio and population density was minimum in 2021 and maximum in 1990. GNI and GDP was minimum in 1990 and maximum in 2021. Unemployment was minimum in 2020 and maximum in 2019. School Enrolment was minimum in the year 1991 and maximum in 2008.

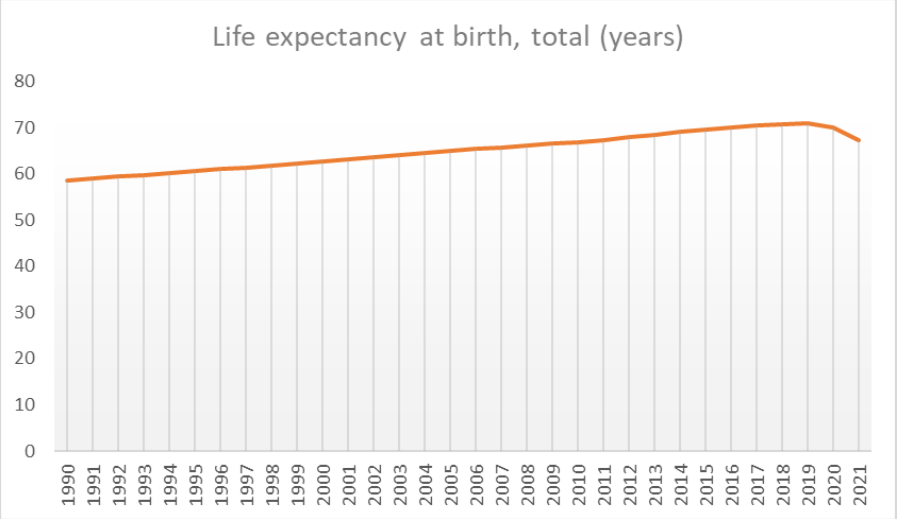

Figure 1: Life Expectancy at birth

The population's overall mortality rate is indicated by the average life expectancy at birth. It captures the overall pattern of death that impacts people of all ages, including adults, kids and teens, and senior citizens. Figure 1 shows the evolution of life expectancy from 1990 to 2021 in India. The lowest life Expectancy of 58.658 was recorded in 1990 and the highest recorded life expectancy was 70.91 in 2019.The life expectancy started to decline after 2019 and was recorded at 67.24 in 2021.

Bivariate Analysis

Pearson Correlation Coefficient is used to examine the strength and direction. It is also used to examine the linear relationship between the variables. The value of the Pearson correlation coefficient, r ranges between -1 and 1:

- r=1 implies a perfect positive linear relationship, meaning that as one variable increases, the other variable also increases proportionally.
- r=−1 signifies a perfect negative linear relationship, indicating that as one variable

increases, the other variable decreases proportionally.

- r=0 suggests no linear relationship between the variables

Variables	y	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13
Life expectancy at birth (y)	1	-.447*	.939**	-.983**	-0.233	-.983**	-.613**	.521**	-.976**	.982**	.947**	.945**	.509**	.605**
Prevalence of undernourishment (x1)		1	-.530**	.395*	.532**	.411*	.785**	-.731**	.438*	-.397*	-.522**	-.522**	0.263	-0.101
Food production index (x2)			1	-.961**	-.380*	-.964**	-.736**	.710**	-.982**	.966**	.996**	.996**	.397*	.479**
Mortality rate, under-5 (x3)				1	0.213	.999**	.580**	-.527**	.995**	-1.000**	-.966**	-.965**	-.542**	-.622**
Prevalence of HIV (x4)					1	0.226	.646**	-.696**	0.292	-0.22	-.405*	-.408*	0.252	-0.071
Fertility rate, total (x5)						1	.585**	-.547**	.995**	-.999**	-.968**	-.967**	-.536**	-.631**
Labor force, female (x6)							1	-.850**	.642**	-.588**	-.733**	-.734**	0.106	0.016
People using at least basic drinking water services (x7)								1	-.596**	.538**	.701**	.704**	-0.114	0.069
Age dependency ratio (x8)									1	-.996**	-.986**	-.986**	-.496**	-.577**
Population density (x9)										1	.970**	.969**	.537**	.610**
GNI per capita, PPP (x10)											1	1.000**	.376*	.501**
GDP per capita, PPP (x11)												1	.374*	.498**
Unemployment(x12)													1	.581**
School enrollment (x13)														1

Table 3: Pearson Correlation Coefficient between variables

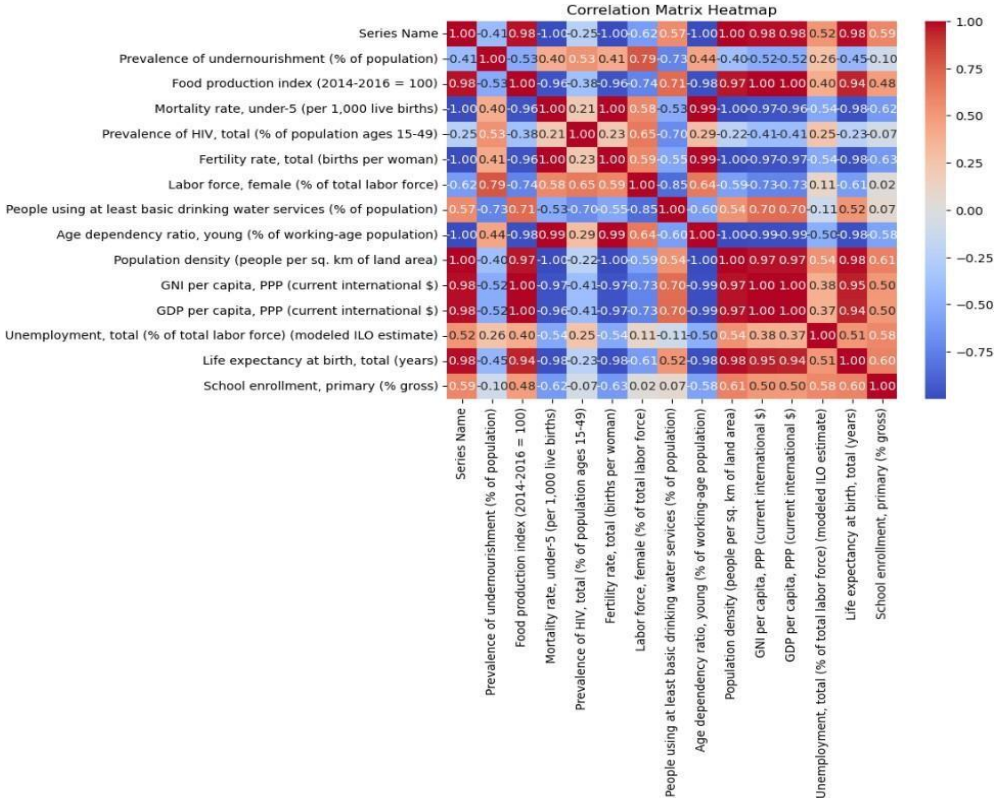

Figure 2: A Simple Correlation Matrix

Table 3 exhibits the correlation between all the variables considered for this study. We can infer from the above table and the correlation matrix that the dependent variable life expectancy is significantly positively related to food production index, people using at least basic drinking water services, population density, unemployment, GDP per capita, GNI per capita, school enrolment. It is negatively related to fertility rate, prevalence of undernourishment, mortality rate (under 5), prevalence of HIV, labour force (female) and age dependency ratio. Life expectancy shows significantly strong correlation with GDP, GNI, population density, fertility, prevalence of undernourishment and food production index. Prevalence of undernourishment was negatively related to food production index people with at least basic drinking water services, population density, GDP, GNI, school enrolment. Both GDP and GNI are positively correlated to food production Index, population density and people with basic drinking water. As a primary analysis we can say by increasing GDP, GNI, food production index, people with basic drinking water services and reducing fertility rate, age dependency ratio we can increase life expectancy

Multiple Linear Regression Analysis

The multiple linear regression model under consideration is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon$$

Where y is the dependent variable (Life Expectancy), $x_1, x_2, \ldots, x_n$ are the predictor variables, β_0 is the intercept term, βi's (i =1, 2, 3,. . ., n) are the unknown regression coefficients and ε is the error term with $N(0,\sigma 2)$ distribution.

Checking for multicollinearity

A small p-value indicates a positive outcome when a model with all potential regressors is summarized. Nevertheless, not every regressor's individual p-value is statistically small, which may indicate multicollinearity. One issue with this study is that the multiple regression analysis's assumptions conflict with the SDGs' theoretical framework (undp.org). The United Nations Development Programme (UNDP) claims that since the SDGs are interconnected, progress made in one area will have an impact on other areas' results. The original dataset's multicollinearity assumption is unquestionably strengthened by this idea.

VIF

Regression analysis uses the Variance Inflation Factor as a metric to identify multicollinearity. When two or more independent variables in a regression model have a high degree of

correlation, this is known as multicollinearity, and it can make it difficult to interpret the model's coefficients and have an impact on how reliable the predictions are. VIF measures how much the model's multicollinearity inflates the variance of an estimated regression coefficient**.** (i) if 0<VIF<5, there is no evidence of a multicollinearity problem; (ii) if 5<VIF<10, there is a moderate multicollinearity problem; and (iii) if VIF>10, there is a serious multicollinearity problem of those variables.

In our dataset the Variance Inflation Factor (VIF) varies a lot for different variables. It amounted to 6260.40 for Age dependency ratio, to 2358.880 for fertility and to 1800.73 for GDP which is indeed very high. Some other predictor variables also showed high VIF values. Variables with high VIF values are removed from the dataset to fit a highly representative model.

Autocorrelation

The Durbin-Watson test was run and the entire model was fitted. The outcome of this test is expressed as a number between 0 and 4, with a value closer to 2 denoting an auto-correlation-free dataset. Since this number was 1.307 for our dataset, we can say that our data set is free of autocorrelation that is there is no systematic relationship between the values of a time series at different lags. In other words, the current value of the series does not depend on its past or future values.

Final Dataset

Variables	**VIF**
Prevalence of undernourishment (% of population)	2.154
Prevalence of HIV, total (% of population ages 15-49)	1.617
Unemployment, total (% of total labour force) (modelled ILO estimate)	2.468
School enrolment, primary (% gross)	1.176
GDP per capita, PPP (current international $)	2.618

Table 4: Selected Variables for the model and its VIF values

Backward Elimination Multiple Regression Analysis

Backward elimination in multiple regression analysis is a method used for model selection and variable reduction. It involves systematically removing independent variables (predictors) from a regression model based on certain criteria, typically p-values or significance levels, to arrive at a more parsimonious and interpretable model. Backward Elimination is employed to select

most significant predictor variables and to fit the model.

The results of backward elimination analysis are given in Table 5. The prevalence of undernourishment, prevalence of HIV, school enrolment, GDP per capita and unemployment were considered as the predictors. In this analysis, three models (Model 1, Model 2 and Model 3) were employed considering LE as the dependent variable. The VIF for all predictors were less than 5, suggesting that there is no evidence of multicollinearity problem.

Variables	Model 1	VIF	Model 2	VIF	Model 3	VIF
Prevalence of undernourishment (% of population)	-0.162	2.154				
Prevalence of HIV, total (% of population ages 15-49)	4.095	1.617	3.681	1.565	4.406	1.234
Unemployment, total (% of total labour force) (modelled ILO estimate)	0.554	2.468	0.317	1.944		
School enrolment, primary (% gross)	0.078	1.779	0.083	1.767	0.102	1.368
GDP per capita, PPP (current international $)	0.002	2.618	0.002	1.827	0.002	1.633
Adjusted R square	0.932		0.93		0.93	

Table 5: Backward elimination multiple regression models explain the life expectancy

In Model 1 all the predictors are considered significant for the dependent variable life expectancy with adjusted r square value of 0.932. In this model prevalence of undernourishment has negative impact and the remaining predictors express positive impact. In model 2 prevalence of undernourishment is removed as least significant variable compared to other variables. And in model 3, Unemployment and prevalence of undernourishment are removed as least significant variables and the model retains prevalence of HIV, school enrolment and GDP per capita as significant predictors where all the predictors have positive impact on Life Expectancy.

The Final model

The table above provides the final model:

$$y= 46.906+ (4.406)*x4+ (0.002)*x11+(0.102)*x13$$

where y is the dependent variable life expectancy
x4 is the prevalence of HIV
x11 is the GDP per capita
x13 is the school enrolment

The Adjusted R square was given as 0.932 which indicates that the predictors explain approximately 93% of the variation in the dataset.

Comparison Graphs of significant variables between India and developed Countries

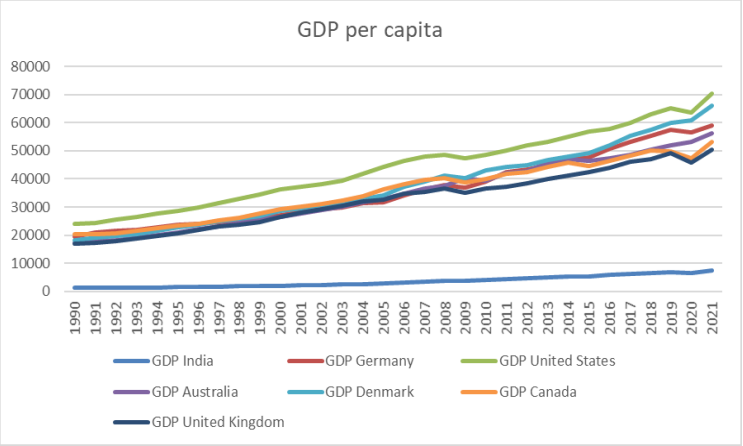

Figure 3: GDP per capita of India and various developed countries

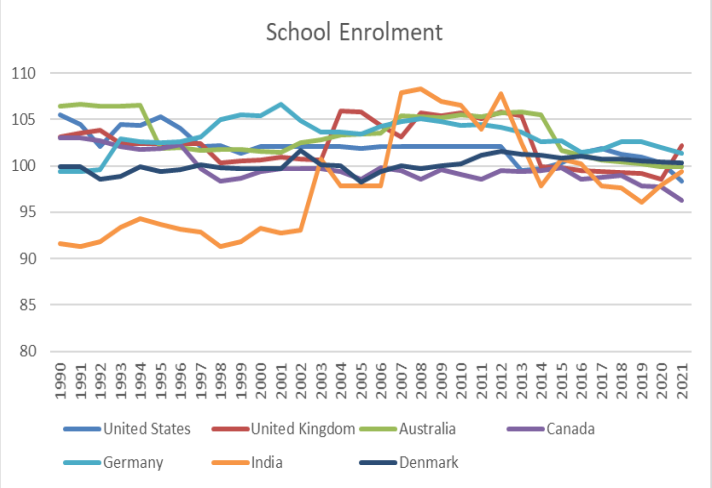

Figure 4: School Enrolment of India and various developed countries

The three important predictor variables for India and other wealthy nations were compared. The HIV prevalence is not plotted since the data are not available. Figure 3 illustrates the stark differences in GDP between India and other wealthy nations. It follows that raising GDP is a necessary step toward extending life expectancy. Figure 4 illustrates how enrolment in schools is nearly identical in India and other nations.

- ## Conclusion

We have examined the variables that impact life expectancy and identified potential predictors of longer life expectancy. It is evident from bivariate analysis that the food production index, school enrolment, GDP, and GNI all have a beneficial effect on life expectancy. Life expectancy rises as social conditions improve as a result of economic development. The GDP per capita and GNI per capita are two indicators of a nation's standard of life. This study clearly shows that India's life expectancy is positively impacted by both GNI and GDP per capita. A higher GDP can extend life expectancy by supplying funds for improved infrastructure, healthcare, education, and sanitation. Increased GDP is frequently associated with higher living standards, easier access to healthcare, and general societal well-being, all of which increase life expectancy and quality of life. The graphs above, which display the GDP of India and other wealthy nations, provide evidence for this.

Life expectancy is significantly impacted by enrolment rates. Education helps people become more knowledgeable about health issues, which promotes healthier lifestyles, better use of healthcare, and illness prevention. Furthermore, those with more education typically earn more money and have better access to resources, both of which increase life expectancy.

By guaranteeing a sufficient supply of food, lowering malnutrition, and enhancing general health, food production indices have an impact on life expectancy. Having access to a diet rich in nutrients promotes healthy physical growth, strengthens immunity, lowers the risk of illness, and lengthens life expectancy.

Age dependency ratio, fertility rate, mortality, female labour force exhibits a negative impact on life expectancy. Life expectancy is influenced by the age dependence ratio, fertility rate, and mortality rate taken together. Life expectancy is positively impacted by reduced dependence ratios, or fewer dependents per working-age population, which are frequently correlated with higher economic production and more resources available for healthcare.

Reduced fertility typically translates into fewer births, which may result in improved resources for each individual, including healthcare, education, and other necessities, ultimately extending life expectancy

Life expectancy is directly impacted by mortality rates, especially reduced ones brought about by improvements in cleanliness, healthcare, and illness prevention. Populations with lower mortality rates typically have longer average lifespans.

Therefore, efforts should be made at national level to increase GDP, GNI, food production index, school enrolment and to decrease mortality rates, fertility rates, labour force and age dependency ration to increase life expectancy.

- **References**

1. Bhanja, Radhika & Roychowdhury, Koel. (2020). Assessing the progress of India towards sustainable development goals by 2030, Journal of Global Resources, 06. 81-91. 10.46587/JGR.2020.v06i02.012.
2. Loft, M. (2021)., A Statistical Analysis Regarding The Sustainable Development Goals and Life Expectancy (Dissertation). Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-311146
3. Mondal, Md. Nazrul. (2015). Sociodemographic and Health Determinants of Inequalities in Life Expectancy in Least Developed Countries. 10.13140/RG.2.1.2864.4003.
4. Nandi DC, Hossain MF, Roy P, Ullah MS. An investigation of the relation between life expectancy & socioeconomic variables using path analysis for Sustainable Development Goals (SDG) in Bangladesh. PLoS One. 2023 Feb 13;18(2):e0275431. doi: 10.1371/journal.pone.0275431. PMID: 36780510; PMCID: PMC9925071.
5. Rogers RG, Wofford S. Life expectancy in less developed countries: socioeconomic development or public health? J Biosoc Sci. 1989 Apr;21(2):245-52. doi: 10.1017/s0021932000017934. PMID: 2722920.
6. United Nations. The 17 Goals. [Online] Available at: https://sdgs.un.org/goals
7. World bank. World development indicators. available form: https://databank.worldbank.org/source/world-development-indicators