World Health Organization

Analysis of Effective Elements in Global Life Expectancy

Hamed Soleimani

Dec 2023

Table of Contents

Introduction

Life expectancy (LE) is a statistical measure representing the average number of years a person can expect to live, based on current mortality rates. It serves as a key indicator of a population's overall health and well-being. The significance of LE lies in its reflection of societal health, healthcare quality, and socioeconomic conditions. Analyzing effective elements such as a country's development status, Body Mass Index (BMI), total healthcare expenditure, HIV/AIDS prevalence, income composition of resources, and schooling provides critical insights into factors influencing LE. Understanding these elements not only helps identify health disparities between developed and developing countries but also informs targeted interventions, policies, and resource allocations to improve global health outcomes and promote longevity (Manton., 2007; Naimark, 2017).

1. World Health Organization (WHO): A Historical Overview

The WHO was established on April 7, 1948, as a specialized agency of the United Nations, with the primary goal of promoting international public health. Headquartered in Geneva, Switzerland, WHO operates as a global authority on health matters, coordinating efforts and setting international health standards (WHO, 2023b). Its commitment to improving LE is evident through responsibilities such as conducting epidemiological research and prioritizing maternal and child health (WHO, 2023a; 2023b).

Based on the data provided on “Ourworldindata” (Figure1), the trend of LE between 2000 and 2014 indicates a comprehensive analysis of demographic patterns, reflecting the evolving landscape of population health over this specific timeframe ((Dattani, et al., 2023).

In this research, we conducted a compelling investigation into the impacts of elements shaping global life expectancy. Recognizing the immense business and economic implications of health outcomes, our commitment extends beyond delivering statistical insights; it aims to unearth strategic insights that empower WHO to illuminate the path towards a healthier, more prosperous world in alignment with WHO's vision of health for all.

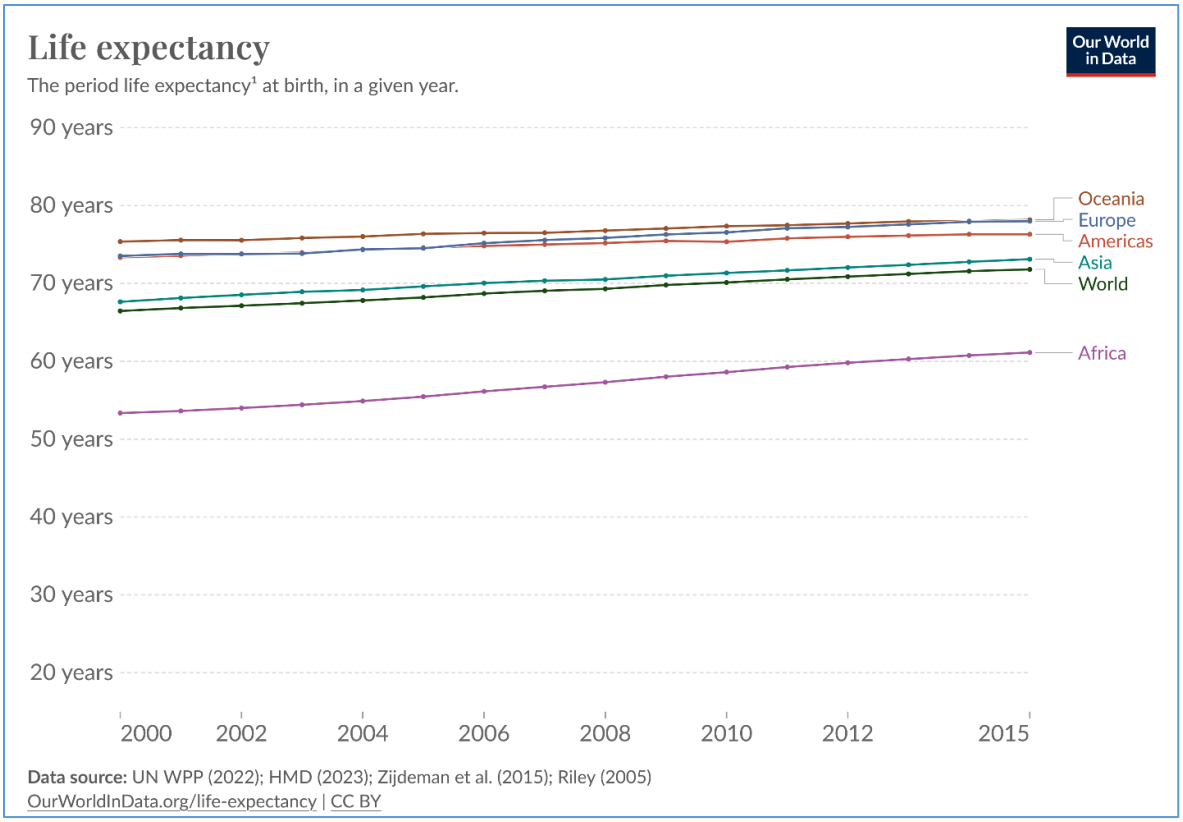


Figure1 Life Expectancy between 2000 and 2014

2. Impact of Key Elements on Life Expectancy: Statistical Analysis

Our objective is to conduct a comprehensive analysis of factors influencing LE across different countries. This study aims to investigate whether a significant difference in LE exists between developed and developing countries. Furthermore, we have identified 5 critical elements—Status of development, BMI, Total expenditure on health, HIV/AIDS prevalence, Income composition of resources, and Schooling—as potential determinants of LE. We examine the impact of these factors on LE using data provided by the WHO for the years 2000 to 2014. Finally, we provide a forecasting analysis to predict the trend in LE for next 3 years. To achieve this, we propose employing different statistical methods including "Descriptive statistics" as well as "Inferential Statistics" to assess the impact of these elements on LE.

A. Descriptive Statistics:

Descriptive statistics are short summaries that help understand a group of numbers. They include average values like mean, median, and mode, telling us where most numbers cluster, and measures of spread like standard deviation, showing how much the numbers vary. They're useful for getting a quick snapshot of data (Hayes, 2023).

- Status: Analyzing demographic characteristics and health indicators in both developing and developed countries.
- BMI: Examining the distribution of BMI across populations.
- Total Expenditure: Investigating healthcare spending patterns globally.
- HIV/AIDS: Assessing the prevalence and incidence rates of HIV/AIDS.
- Income Composition: Exploring economic factors influencing health outcomes.
- Schooling: Investigating educational levels and their correlation with LE.

B. Inferential Statistics:

Inferential statistics assess the significance of a test result, determining whether differences exist between samples. This significance is often quantified using p-values. The selection of the appropriate statistical test for a dataset depends on the nature of the analyzed data—whether it is binary, nominal, ordinal, interval, or ratio—the distribution of the dataset (normal or not), and whether the analysis investigates potential differences between samples or relationships between variables (Marshall and Jonker, 2011).

- ANOVA (Analysis of Variance): Examining variations in LE across different levels of categorical variables (Jones, 2023).
- Correlation Analysis: Assessing the strength and direction of relationships between LE and each identified element.
- Regression Analysis: Evaluating the combined impact of multiple factors on LE.

Test of Normality

Evaluating the normality of data is crucial before undertaking any formal statistical analysis to avoid making incorrect conclusions (Steinskog, Tjøstheim and Kvamstø, 2007). Several tests are available to check whether a sample is derived from a population with a normal distribution. Kolmogorov-Smirnov test and Shapiro-wilk test are widely used to identify the normality of distribution within a dataset (Drezner, et al., 2010; A Anaeth, 2019). The Shapiro-Wilk test relies on correlations within given observations and their corresponding normal scores(Das and Imon, 2016).

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Life expectancy	.134	2307	<.001	.953	2307	<.001

a. Lilliefors Significance Correction

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Status	Statistic	df	Sig.	Statistic	df	Sig.
Life expectancy	Develope	.076	420	<.001	.979	420	<.001
	Developi	.134	1887	<.001	.945	1887	<.001

a. Lilliefors Significance Correction

Table1 Tests of normality

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Life expectancy	2307	36.3	89.0	69.322	9.7004	-.674	.051	-.178	.102
IS_Developed	2307	.00	1.00	.1821	.38597	1.649	.051	.720	.102
BMI	2307	1.4	77.1	38.095	19.8107	-.237	.051	-1.322	.102
Total expenditure	2307	.37	14.39	5.8730	2.39582	.235	.051	-.257	.102
HIV/AIDS	2307	.1	20.8	12.097	3.3527	-.492	.051	.512	.102
Income composition of resources	2307	.000	.945	.62927	.214783	-1.115	.051	1.203	.102
Schooling	2307	.0	20.7	12.092	3.3462	-.497	.051	.500	.102
Valid N (listwise)	2307								

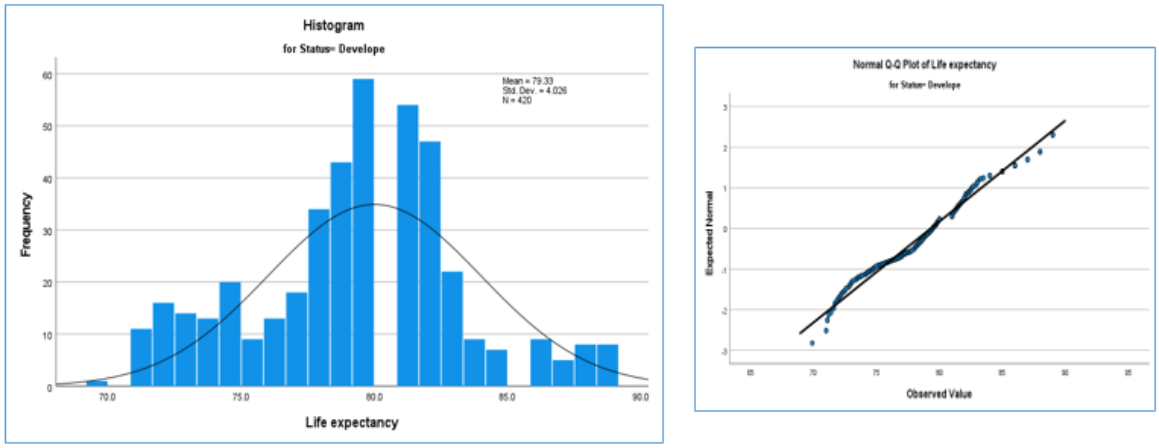
Table2 Descriptive statistics

In evaluating the normality of our sample distribution (Table1) both tests present p-value of <.001 which are less than 0.05. This provides statistical evidence that the data used deviates from normality. However, as noted by Field (2013), these tests, particularly with larger sample sizes like ours (2300), might yield statistically significant results even when deviations from normality are marginal. To gain a more comprehensive understanding, we turn to visual tools such as histogram and Q-Q plots, as recommended by Rafiqul (2018). The histograms for our data (Figure2) suggest the assumption of normality test has been met as the there is a symmetrical peak in the middle.

Understanding the implications of the Central Limit Theorem, we recognize that with an increasing sample size, the sampling distribution tends to approach normality, bringing the mean of the sampling distribution closer to the population mean. In our case, despite the statistical tests indicating non-normality, the visual examination and theoretical considerations suggest a potential convergence towards normality.

When considering skewness, the accepted range for normality, as suggested by some researchers (West et al., 1996), lies within [-2, 2]. Our statistical descriptive results (Table2) reveal a skewness of -.674, comfortably within the acceptable range, reinforcing the notion that, in practice, our data can be considered reasonably normal.

Status = Developed



Status = Developing

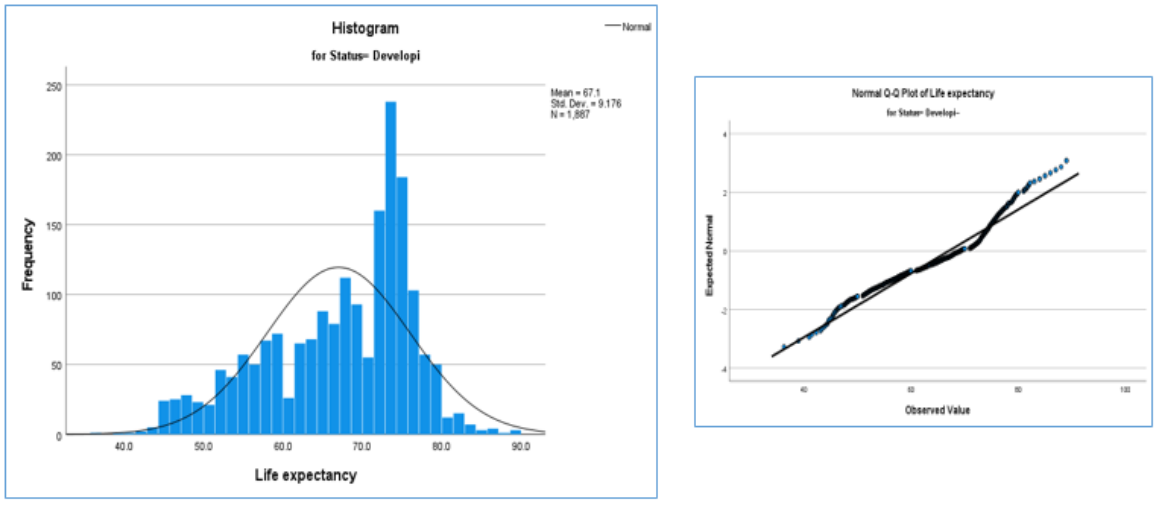


Figure2 Histogram and Q-Q plot

T-Test

We employed a T-Test to rigorously examine whether a significant difference exists in LE between developed and developing countries.

Group Statistics

	is_dev_1	N	Mean	Std. Deviation	Std. Error Mean
Life expectancy	Developed	420	79.325	4.0258	.1964
	Developing	1887	67.095	9.1764	.2112

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						95% Confidence Interval of the Difference	
		F	Sig.	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Life expectancy	Equal variances assumed	307.041	<.001	26.744	2305	<.001	<.001	12.2302	.4573	11.3334	13.1270
	Equal variances not assumed			42.397	1502.114	<.001	<.001	12.2302	.2885	11.6644	12.7961

Table3 T-Tests

Our objective is to identify variables that impact LE. Initially, we compare LE between developed and developing countries. We conducted an independent T-Test to compare the means of LE between these two groups (developed and developing countries). The results, as shown in the Table3, indicate a significance level of <0.05, leading to the rejection of the null hypothesis. This implies insufficient evidence to support the assumption that the mean LE in developed and developing countries is the same. In other words, the LE in developed and developing countries is not equal. Moreover, positive t-statistic suggests that the mean of the first group (Developed Countries) is higher.

The hypotheses are represented as follows:

$$\text{Null Hypothesis} \rightarrow \mathbf{H_0}: \boldsymbol{\mu_{Developed} = \mu_{Developing}}$$

$$\text{Alternative Hypothesis} \rightarrow \mathbf{H_A}: \boldsymbol{\mu_{Developed} \neq \mu_{Developing}}$$

Regression

This study aims to identify the effects of BMI, Total Expenditure, HIV/AIDS prevalence, Income Composition of Resources, Schooling, and the Status of countries on LE (LE).

There are multiple variables that affect LE, and multiple regression is a statistical method employed to analyze the variability of a dependent variable, in this case, LE. This analysis considers information provided by independent or predictor variables such as BMI, HIV, Schooling, and others (Kraha et al., 2012). The model for multiple regression is formulated as follows (Uyanık and Güler, 2013):

$$\boldsymbol{y = \beta_0 + \beta_1 X_{1+} \beta_2 X_2 + \cdots + \beta_n X_n}$$

y = dependent variable

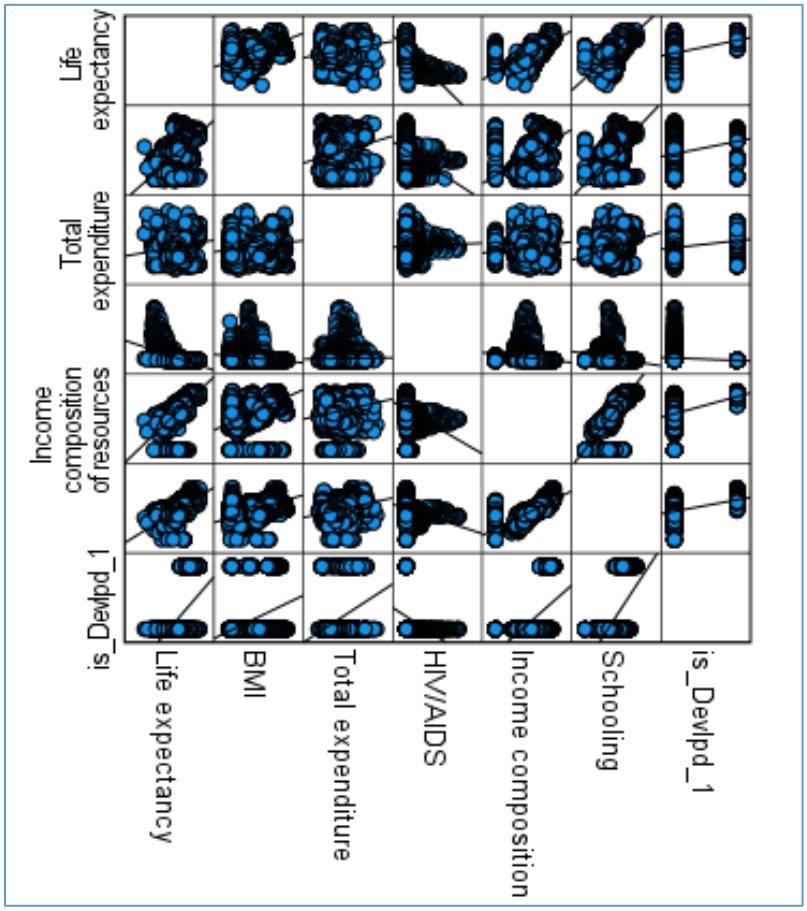
$$X_i = independent\ variable$$
$$\beta_i = parameter$$

The hypothesis for identifying if the independent variables have any significant effect on LE is as below:

$$\boldsymbol{Null\ Hypothesis: H_1: \beta_1 = \beta_2 = \cdots = \beta_n = 0}$$

$$\boldsymbol{Null\ Hypothesis: H_1: \beta_1 = \beta_2 = \cdots = \beta_n \neq 0}$$

A scatterplot indicates the relationship between two quantitative variables measured against the same variable (Moore et al, 2013). Graph1shows the relationship between LE and BMI, total expenditure, HIV/AIDS prevalence, income composition of resources, schooling, and status of countries. One can quickly observe a strong relation between LE and income composition of resources. This can also be seen between schooling and LE. BMI can also be interpreted to have moderately good relationship with LE. As can be seen, the relationship between LE and HIV/AIDS is negative while there is a positive relationship with all other variables.



Graph 1 Scatter Plot Matrix between LE and independent variables

Table4 indicates that schooling has the most significant impact on LE, with Pearson correlation value of 0.75. This followed closely is the income composition of resources with correlation value of 0.725. A moderate positive correlation of LE and BMI with a correlation value of 0.581 is also identified. On the other hand, the correlation between LE and total expenditure is very weak with a value of 0.181. The only negative correlation is found between LE and HIV/AIDS, with a correlation value of t -0.576.

Correlations

		Life expectancy	BMI	Total expenditure	Income composition of resources	HIV/AIDS	Schooling	IS_Developed
Life expectancy	Pearson Correlation	1	.581**	.181**	.725**	-.576**	.750**	.487**
	Sig. (2-tailed)		<.001	<.001	<.001	<.001	<.001	<.001
	N	2307	2307	2307	2307	2307	2307	2307
BMI	Pearson Correlation	.581**	1	.195**	.523**	-.247**	.570**	.310**
	Sig. (2-tailed)	<.001		<.001	<.001	<.001	<.001	<.001
	N	2307	2307	2307	2307	2307	2307	2307
Total expenditure	Pearson Correlation	.181**	.195**	1	.168**	.026	.262**	.271**
	Sig. (2-tailed)	<.001	<.001		<.001	.207	<.001	<.001
	N	2307	2307	2307	2307	2307	2307	2307
Income composition of resources	Pearson Correlation	.725**	.523**	.168**	1	-.254**	.797**	.490**
	Sig. (2-tailed)	<.001	<.001	<.001		<.001	<.001	<.001
	N	2307	2307	2307	2307	2307	2307	2307
HIV/AIDS	Pearson Correlation	-.576**	-.247**	.026	-.254**	1	-.230**	-.157**
	Sig. (2-tailed)	<.001	<.001	.207	<.001		<.001	<.001
	N	2307	2307	2307	2307	2307	2307	2307
Schooling	Pearson Correlation	.750**	.570**	.262**	.797**	-.230**	1	.530**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001		<.001
	N	2307	2307	2307	2307	2307	2307	2307
IS_Developed	Pearson Correlation	.487**	.310**	.271**	.490**	-.157**	.530**	1
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	
	N	2307	2307	2307	2307	2307	2307	2307

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

Table4 Correlations

As shown in Table5, some predictors in our regression model exhibit high correlation, which has the potential to affect the reliability of individual coefficients. The Variance Inflation Factor (VIF) is used to quantify how much the variance of an estimated regression coefficient increases when predictor variables are correlated. According to Katrutsa and Strijov (2017), a VIF less than 5 and condition scores below 30 suggest no significant multicollinearity. Upon examining the VIF scores and condition index in Table7, it is evident that the maximum VIF is 3.278, and all condition indexes are below 30. Therefore, the regression model does not exhibit problematic collinearity issues. In Table6, where higher absolute B values signify greater importance of the corresponding independent variable, the variable with the highest B is deemed the most important.

Correlations

		Life expectancy	BMI	Total expenditure	Income composition of resources	HIV/AIDS	Schooling	IS_Developed
Life expectancy	Pearson Correlation	1	.581**	.181**	.725**	-.576**	.750**	.487**
	Sig. (2-tailed)		<.001	<.001	<.001	<.001	<.001	<.001
	N	2307	2307	2307	2307	2307	2307	2307
BMI	Pearson Correlation	.581**	1	.195**	.523**	-.247**	.570**	.310**
	Sig. (2-tailed)	<.001		<.001	<.001	<.001	<.001	<.001
	N	2307	2307	2307	2307	2307	2307	2307
Total expenditure	Pearson Correlation	.181**	.195**	1	.168**	.026	.262**	.271**
	Sig. (2-tailed)	<.001	<.001		<.001	.207	<.001	<.001
	N	2307	2307	2307	2307	2307	2307	2307
Income composition of resources	Pearson Correlation	.725**	.523**	.168**	1	-.254**	.797**	.490**
	Sig. (2-tailed)	<.001	<.001	<.001		<.001	<.001	<.001
	N	2307	2307	2307	2307	2307	2307	2307
HIV/AIDS	Pearson Correlation	-.576**	-.247**	.026	-.254**	1	-.230**	-.157**
	Sig. (2-tailed)	<.001	<.001	.207	<.001		<.001	<.001
	N	2307	2307	2307	2307	2307	2307	2307
Schooling	Pearson Correlation	.750**	.570**	.262**	.797**	-.230**	1	.530**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001		<.001
	N	2307	2307	2307	2307	2307	2307	2307
IS_Developed	Pearson Correlation	.487**	.310**	.271**	.490**	-.157**	.530**	1
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	
	N	2307	2307	2307	2307	2307	2307	2307

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

Table5 Correlations for collinearity diagnosis

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	48.491	.445		109.024	<.001		
	Total expenditure	.051	.042	.013	1.209	.227	.886	1.129
	HIV/AIDS	-.674	.018	-.391	-37.822	<.001	.907	1.102
	BMI	.066	.006	.135	10.996	<.001	.646	1.547
	Income composition of resources	10.556	.756	.234	13.967	<.001	.346	2.894
	Schooling	1.021	.052	.352	19.777	<.001	.305	3.278
	IS_Developed	1.982	.299	.079	6.629	<.001	.684	1.463

a. Dependent Variable: Life expectancy

Table6 Coefficient for collinearity diagnosis

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions (Constant)	BMI	Total expenditure	HIV/AIDS	Income composition of resources	Schooling	IS_Developed
1	1	5.084	1.000	.00	.00	.00	.00	.00	.00	.01
	2	1.016	2.237	.00	.00	.00	.61	.00	.00	.12
	3	.616	2.874	.00	.01	.00	.26	.00	.00	.64
	4	.139	6.045	.01	.45	.43	.09	.00	.00	.00
	5	.095	7.326	.05	.46	.38	.00	.08	.02	.00
	6	.035	12.019	.76	.04	.17	.03	.34	.01	.20
	7	.015	18.157	.18	.03	.01	.00	.57	.97	.03

a. Dependent Variable: Life expectancy

Table7 collinearity diagnostics

Following the linear regression analysis as depicted in Table8, it was determined that the model's capability to forecast the dependent variable is represented by R^2=0.777, with an adjusted R^2 value of 0.777. Considering the coefficients, it can be said that the independent variables predict nearly 78% of the LE. However, since the significance of ANOVA in Table9 is 0.000, we reject the null hypothesis which means that one or more that one of the parameters is not equal to zero. In summary, one can claim that the model adopted explains well the relationship between independent and dependent variables.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.882[a]	.777	.777	4.5821

a. Predictors: (Constant), IS_Developed, HIV/AIDS, Total expenditure, BMI , Income composition of resources, Schooling

Table8 Model summary

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	168700.920	6	28116.820	1339.200	<.001[b]
	Residual	48289.045	2300	20.995		
	Total	216989.965	2306			

a. Dependent Variable: Life expectancy

b. Predictors: (Constant), IS_Developed, HIV/AIDS, Total expenditure, BMI , Income composition of resources, Schooling

Table9 ANOVA

From the result of the regression analysis, the following is the presented regression equation:

$$Life\ Expectancy = 48.491 + 0.066(BMI) + 0.051(total\ expenditure) + 10.556(income\ composition\ of\ resources + 1.021(schooling) + 1.982(is\_devlpd\_1) - 0.674(HIV|AIDS)$$

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	48.491	.445		109.024	<.001		
	Total expenditure	.051	.042	.013	1.209	.227	.886	1.129
	HIV/AIDS	-.674	.018	-.391	-37.822	<.001	.907	1.102
	BMI	.066	.006	.135	10.996	<.001	.646	1.547
	Income composition of resources	10.556	.756	.234	13.967	<.001	.346	2.894
	Schooling	1.021	.052	.352	19.777	<.001	.305	3.278
	IS_Developed	1.982	.299	.079	6.629	<.001	.684	1.463

a. Dependent Variable: Life expectancy

Table10 Multiple Regression Analysis Results

However, the significance level of total expenditure is more than 0.05 which means it is not significant. As a result, it should be deleted from the model and the do regression analysis again.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	48.699	.410		118.736	<.001		
	BMI	.066	.006	.136	11.135	<.001	.651	1.536
	HIV/AIDS	-.672	.018	-.389	-37.890	<.001	.917	1.091
	Income composition of resources	10.471	.753	.232	13.913	<.001	.349	2.869
	Schooling	1.030	.051	.355	20.149	<.001	.311	3.213
	IS_Developed	2.046	.294	.081	6.952	<.001	.706	1.417

a. Dependent Variable: Life expectancy

Table11 Regression Analysis Results after changing the model variables

The new model is shown below:

$$Life\ Expectancy = 48.699 + 0.066(BMI) + 10.471(income\ composition\ of\ resources + 1.030(schooling) + 2.046(is\_devlpd\_1) - 0.672(HIV|AIDS)$$

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.882[a]	.777	.777	4.5825

a. Predictors: (Constant), IS_Developed, HIV/AIDS, BMI, Income composition of resources, Schooling

Table12 New Model summary-After removing Total Expenditure variable

As shown in Table4, the correlation between the variable 'Total Expenditure' was very weak, suggesting that this variable did not significantly contribute to explaining the variation in the dependent variable, as reflected by the unchanged adjusted R2 value (Table12). As expected, removing it from the model resulted in no change in the adjusted R2 value.

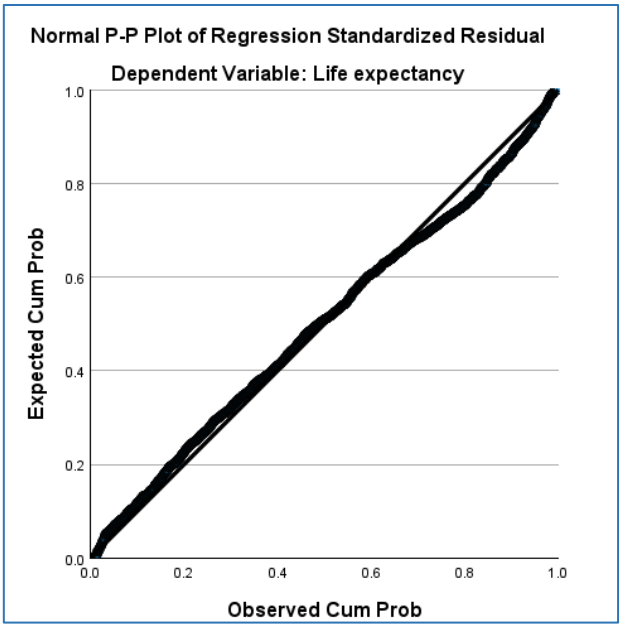


Figure3 P-P Plot

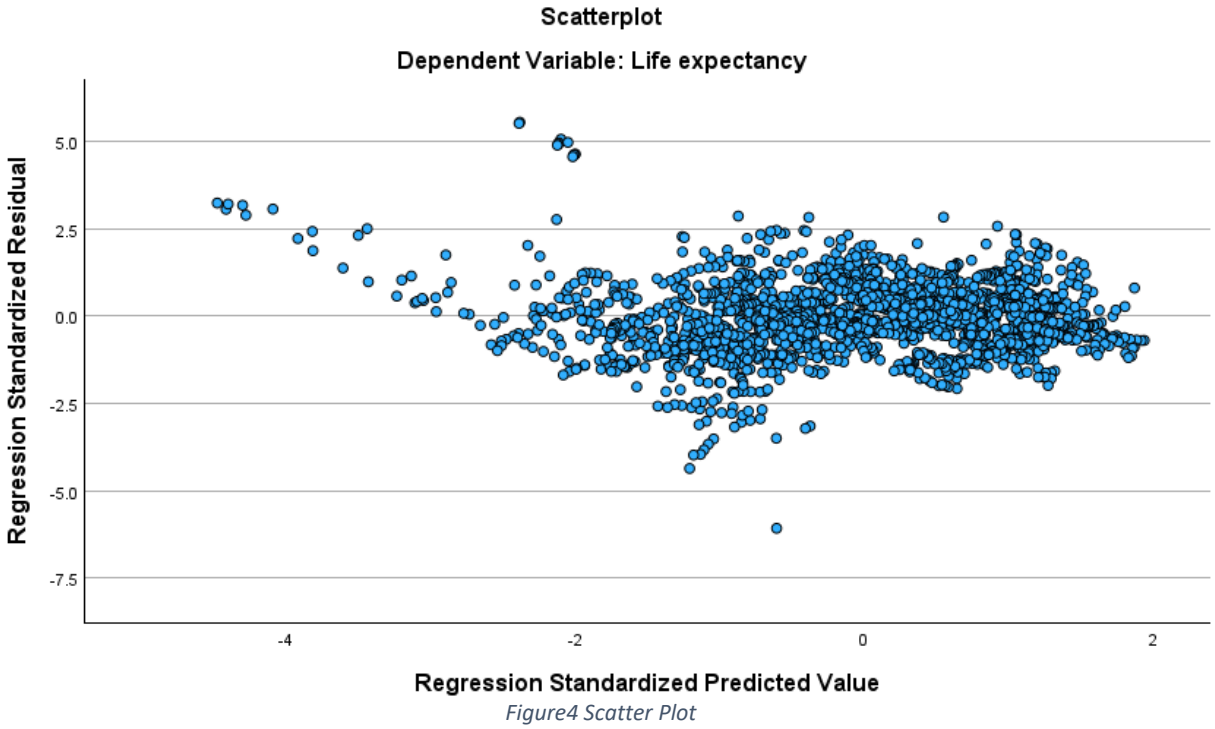


Figure4 Scatter Plot

3. Forecasting LE of countries according to their development status

To forecast LE of developed and developing countries, this report utilizes data of 2000 to 2014 from WHO datasets.

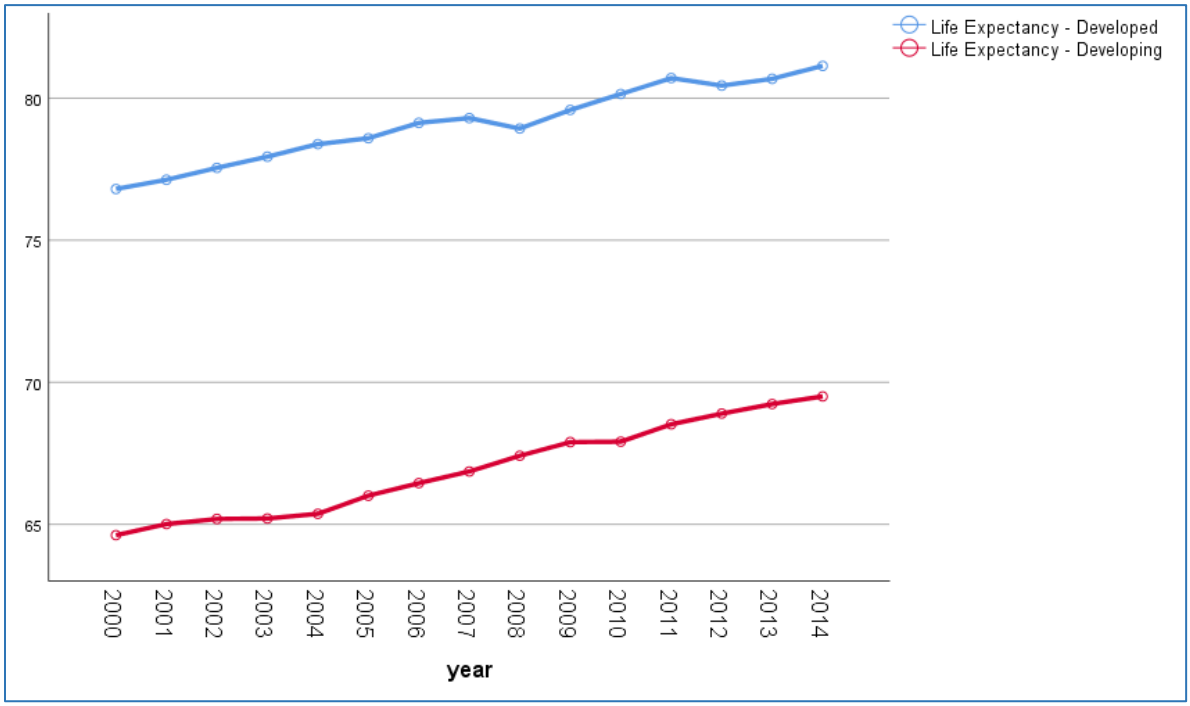


Figure5 Life Expectancy between 2000 and 2014 in Developed and Developing countries

As depicted in Figure 5 above, there is a slight yearly increase in the LE of developing countries from 64.6 in 2000 to 69.5 in 2014. In contrast, the LE of developed countries increases from 76.8 in 2000 to 79.3 in 2008, but then experiences a decrease in 2009 to 78.9, followed by a rise again to 81.13 in 2014. Utilizing 15 years of historical data, we aim to predict our dependent variable, i.e., the trend in LE, up to the year 2018.

To predict LE, we use the 'Expert Modeler' tool in SPSS, and the results are interpreted based on the model's findings and related literature. However, before delving deep into the prediction part, we inspect the status of our data in terms of non-stationarity of the mean. If our data is non-stationary, we initially need to apply differencing once or twice to remove the non-stationarity of our trend and then proceed with our analysis. The non-stationarity of data can be determined using autocorrelations and partial autocorrelations.

Autocorrelations

Series: Life Expectancy - Developed

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic Value	df	Sig.[b]
1	.771	.234	10.838	1	.001
2	.567	.226	17.140	2	.000
3	.394	.217	20.446	3	.000
4	.202	.208	21.389	4	.000
5	.044	.198	21.439	5	.001
6	-.089	.188	21.662	6	.001
7	-.133	.177	22.224	7	.002
8	-.220	.166	23.994	8	.002
9	-.360	.153	29.493	9	.001
10	-.419	.140	38.450	10	.000
11	-.427	.125	50.076	11	.000
12	-.359	.108	61.010	12	.000
13	-.293	.089	71.944	13	.000

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

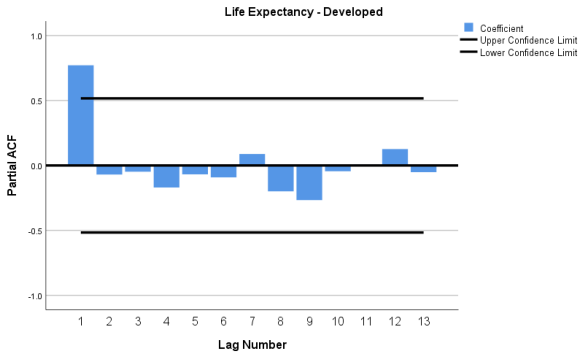
Autocorrelations

Series: Life Expectancy - Developing

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic Value	df	Sig.[b]
1	.818	.234	12.196	1	.000
2	.636	.226	20.141	2	.000
3	.450	.217	24.453	3	.000
4	.251	.208	25.917	4	.000
5	.067	.198	26.032	5	.000
6	-.106	.188	26.348	6	.000
7	-.249	.177	28.326	7	.000
8	-.355	.166	32.917	8	.000
9	-.416	.153	40.283	9	.000
10	-.429	.140	49.669	10	.000
11	-.408	.125	60.281	11	.000
12	-.345	.108	70.399	12	.000
13	-.263	.089	79.241	13	.000

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.



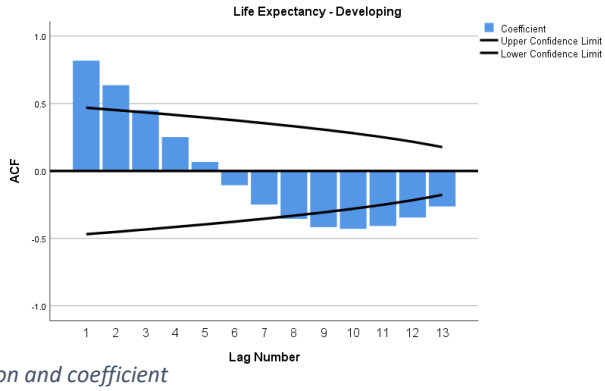


Figure6 AutoCorrelation and coefficient

As indicated in the autocorrelation tables above, the correlation between the LE of developed countries and a lagged version of itself at lag 1 is 0.771. This suggests a strong positive correlation between LE at time t and its t-1 value. Similarly, the autocorrelation at lag 2 is 0.567, which is also strong and positive. However, as we progress further down the lag line, the autocorrelation gradually decreases. This pattern is consistent with the autocorrelation of developing countries, which is 0.818 at lag 1 and 0.636 at lag 2, gradually decreasing for subsequent lags.

The null hypothesis in this context posits that the autocorrelation up to a certain lag is equal to 0. However, according to the Box-Ljung statistic, the p-value for all lags in both developed and developing countries is less than the significance level. This indicates evidence of non-zero autocorrelation, leading us to reject the null hypothesis (Zach, 2020).

Additionally, the autocorrelation function (ACF) diagrams for developing countries exhibit a gradual decay of autocorrelation between lags. However, in developed countries, the pattern is different. These results suggest that our assumption of having autocorrelations of zero for all lags is not met. Moreover, the results

for developing countries suggest that we need to apply at least one differencing. By doing so, we might be able to use ARIMA for our prediction. Before proceeding, we also check the results of partial autocorrelation.

Partial Autocorrelations

Series: Life Expectancy - Developed

Lag	Partial Autocorrelation	Std. Error
1	.771	.258
2	-.070	.258
3	-.048	.258
4	-.170	.258
5	-.069	.258
6	-.092	.258
7	.088	.258
8	-.199	.258
9	-.267	.258
10	-.044	.258
11	.006	.258
12	.126	.258
13	-.052	.258

Partial Autocorrelations

Series: Life Expectancy - Developing

Lag	Partial Autocorrelation	Std. Error
1	.818	.258
2	-.101	.258
3	-.123	.258
4	-.169	.258
5	-.111	.258
6	-.135	.258
7	-.096	.258
8	-.076	.258
9	-.043	.258
10	-.013	.258
11	-.030	.258
12	.021	.258
13	-.012	.258

Table13 Partial AutoCorrelations

Partial autocorrelation of a LE's time series data measures the correlation of it with its lagged version as a control for other lags effect. As the table above indicates, we have a strong positive correlation of 0.771 for developed countries and 0.818 for developing ones for the first lags. But at the values suddenly plummets down to the end. This pattern continues for all other lags and partial autocorrelation fluctuates between positive and negative values for both groups.

Our null hypothesis is that no partial autocorrelation exists between variable and the lagged variable. Considering the first lags with high values, we reject the null hypothesis in favor of alternative ones as there is a significant difference from zero and a strong positive correlation between the variable and its lag. Whereas, for lags 2 to 13 of both tables, the partial autocorrelations are not different from zero by significant amount and as so it can be assumed that no correlation exists between variables and their lags after controlling of all other variables' effect in the model.

Finally using expert "Expert Modeler" in SPSS, a forecast of LE from 2015 to 2018 for both developing and developed countries is made.

Model Description

			Model Type
Model ID	Life Expectancy - Developing	Model_1	ARIMA(0,1,0)

Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	Model Fit statistics: R-squared	Ljung-Box Q(18): Statistics	Ljung-Box Q(18): DF	Ljung-Box Q(18): Sig.	Number of Outliers
Life Expectancy - Developing-Model_1	0	6.661E-16	.984	.	0	.	0

Table14 ARIMA Model statistics

The "Expert Modeler" applies ARIMA (autoregressive integrated moving average) which is a generalized model of ARMA (autoregressive moving average) to predict future LE of developing countries. ARIMA model with a difference of one i.e. ARIMA (0,1,0) which is known as "random walk", is proposed by the modeler.

According to Model Statistics, stationary R-squared is **6.6613E-16**. The R-squared value is equal to **0.984**, which shows that the model explains **98.4%** of the variation of the response variable. The Ljung-Box Q statistic is **0** with **18 degrees of freedom** and the **p-value is 0**. Therefore, it signifies that there is strong evidence of autocorrelation in the model residuals and they exhibit serial correlation (Frost, 2023). Hence, time series can be applied to data for doing the prediction.

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Life Expectancy - Developing-Model_1	Life Expectancy - Developing	No Transformation	Constant	.349	.053	6.524	.000
			Difference	1			

Table15 ARIMA model parameters

Forecast

Model		2015	2016	2017	2018
Life Expectancy - Developing-Model_1	Forecast	69.85070964	70.19943245	70.54815525	70.89687806
	UCL	70.28280628	70.81050938	71.29656860	71.76107136
	LCL	69.41861299	69.58835551	69.79974191	70.03268477

For each model, forecasts start after the last non-missing in the range of the requested estimation period, and end at the last period for which non-missing values of all the predictors are available or at the end date of the requested forecast period, whichever is earlier.

Table16 Forecast

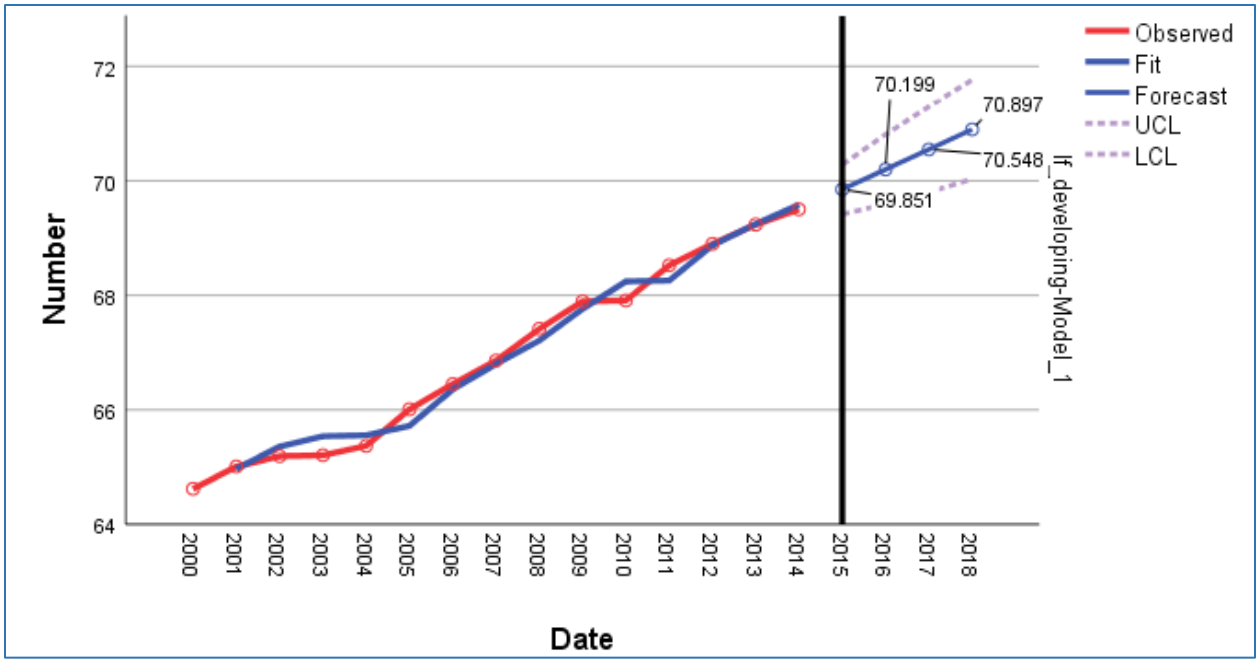


Figure7 Forecasting result

The prediction result indicates that there will be a steady increase in LE of developing countries from 69.8 in 2015 to 70.9 in 2018.

As for developed countries, "Expert Modeler" of SPSS fitted Holt model to do the Forecasting. This model which is also known as "double exponential smoothing method" is a simple technique for predicting time series values with a trend in them (Oyetunji, 2023). For developed countries relates a steady increase in LE over the next four years from 81.48 in 2015 to 82.39 in 2018. The result for both categories seems reasonable and unless unforeseen phenomena occur, LE should see a slight rise in the near future, regardless of development status.

Model Description

			Model Type
Model ID	Life Expectancy - Developed	Model_1	Holt

Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	Model Fit statistics: R-squared	Ljung-Box Q(18): Statistics	Ljung-Box Q(18): DF	Ljung-Box Q(18): Sig.	Number of Outliers
Life Expectancy - Developed-Model_1	0	.680	.967	.	0	.	0

Table17 Holt Model Statistics

References

Das, K. R. and Imon, A. (2016) 'A brief review of tests for normality', *American Journal of Theoretical and Applied Statistics*, 5(1), pp. 5-12.

Dattani, S., Rodés-Guirao, L., Ritchie, H., Ortiz-Ospina, E. and Roser, M, (2023), 'Life Expectancy', *Our World in Data* [Online] [Accessed on 17th December 2023] https://ourworldindata.org/life-expectancy

Drezner, Z., Turel, O. and Zerom, D. (2010) 'A Modified Kolmogorov–Smirnov Test for Normality', *Communications in Statistics - Simulation and Computation*, 39(4), pp. 693-704.

Field, A. (2013) *Discovering Statistics Using IBM SPSS Statistics.* SAGE Publications, p. 188.

Frost, J. (2023) *Autocorrelation and Partial Autocorrelation in Time Series Data*. [Online] [Accessed on 19th December 2023] https://statisticsbyjim.com/time-series/autocorrelation-partial-autocorrelation/

Hayes, A. (2023). 'Descriptive Statistics: Definition, Overview', Types, Example, *Investopedia* [Online] [Accessed on 17th December 2023] https://www.investopedia.com/terms/d/descriptive_statistics.asp

Jones, G.P., Stambaugh,C. Stambaugh, N.and Kathryn E. Huber, K.E., 'Analysis of variance', (2023), *Translational Radiation Oncology*, PP. 171-177.

Katrutsa, A. and Strijov, V. (2017) 'Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria', *Expert Systems with Applications*, 76, pp. 1-11.

Kim, J.H., (2005), Multicollinearity and misleading statistical results ', *Korean Journal of Anesthesiology*, [Online] [Accessed on 21th December 2023] https://ekja.org/upload/pdf/kja-19087.pdf

Kraha, A., Turner, H., Nimon, K., Zientek, L. and Henson, R. (2012) 'Tools to Support Interpreting Multiple Regression in the Face of Multicollinearity', *Frontiers in Psychology*, 3.

Manton, K.G. (2007) 'Life Expectancy', *Encyclopedia of Gerontology*,*Elsevier*, 2, pp. 61-67.

Marshall, G. and Jonker, L. (2011), 'An introduction to inferential statistics: A review and practical guide', Radiography, 17(1), PP. e1-e6.

Mindrila, D. (2018). 'Scatterplots and Correlation ', *From Chapter 4 of Basic practices of statistics. 6th edition. Westgate.edu*. [Online] [Accessed on 19th December 2023] https://www.westga.edu/academics/research/vrc/assets/docs/scatterplots_and_correlation_notes.pdf

Mishra, P., Pandey, C.M., Singh, U., Gupta, A., Sahu, C. and Keshri, A. (2019) 'Descriptive statistics and normality tests for statistical data ', *Annals of cardiac anaesthesia*, 22(1), PP. 67-72.

Naimark, D.M.J. (2017) 'Life Expectancy Measurements', *International Encyclopedia of Public Health*, *Academic Press*, 2, pp. 419-431.

Oyetunji, I. (2023) *Exponential Smoothing: A Beginner's Guide to Getting Started*. [Online] [Accessed on 19th December 2023] https://www.influxdata.com/blog/exponential-smoothing-beginners-guide/

Rafiqul, M. (2018) 'Sample Size and Its Role in Central Limit Theorem (CLT)', *Computational and Applied Mathematics Journal*, 4.

Steinskog, D. J., Tjøstheim, D. B. and Kvamstø, N. G. (2007) 'A Cautionary Note on the Use of the Kolmogorov–Smirnov Test for Normality', *Monthly Weather Review*, 135(3), pp. 1151-1157.

Uyanık, G. K. and Güler, N. (2013) 'A Study on Multiple Linear Regression Analysis', *Procedia - Social and Behavioral Sciences*, 106, pp. 234-240.

World Health Organization, (2023a) *urgent-health-challenges-for-the-next-decade.* [Online] [Accessed on 17th December 2023] https://www.who.int/news-room/photo-story/photo-story-detail/urgent-health-challenges-for-the-next-decade

World Health Organization, (2023b) *who-we-are*. [Online] [Accessed on 17th December 2023] https://www.who.int/about/who-we-are

West, S. G., Curran, P. J. & Finch, J. F. (1996) The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*(1), PP. 16–29.

Zach (2020) *Ljung-Box Test: Definition + Example.* [Online] [Accessed on 19th December 2023] https://www.statology.org/ljung-box-test/