Factors Affecting Life Expectancy:

A Regression Analysis

A Student Project in

VEE Exam Course: Regression Analysis

Mary Jeane Y. dela Cerna

Summer 2016

delmj@sunlife.com

Introduction

One interesting question in mind is, “How long can I live?” Knowing how long one can live, or at least have an idea about it may change people’s way of living. People nowadays have grown more conscious of their health, the environment their living in and their lifestyle. According to Max Roser, life expectancy has increased rapidly since the Enlightenment. In a pre-modern poor world, life expectancy was around 30 years in all regions of the world. In the early 19^{th} century, life expectancy started to increase in the early industrialized countries while it stayed low in the rest of the world. Over the last decades, global inequality of life expectancy has decreased. Since 1900, the global average life expectancy has more than doubled, and is now approaching 70 years. [1]

Life expectancy refers to the number of years a person is expected to live base on the statistical average. It differs by geographical area, and depends on several variables such as gender, lifestyle, access to healthcare, economic growth rate, environment, etc. The concept of life expectancy has been used in different fields such as plant or animal ecology, and actuarial science.

This project aims to determine the factors that could mostly explain the variation of life expectancy per country through regression analysis. The explanatory variables considered are real GDP growth rate, health care index, safety index, pollution index, traffic index, literacy rate and gender. The response variable is the actual life expectancy by gender. Data obtained are from year 2015. Real GDP growth rate is the rate of growth of the value of all final goods and services produced within a state in a given year. [2] Estimates suggest that countries with higher GDP have a higher life expectancy. [1] Health care index refers to the estimation of the overall quality of the health care system, health care professionals, equipment, staff, doctors, cost, etc. Safety index suggests that if a city has high safety index, then it is considered very safe. Pollution index is an estimation of the overall pollution in the city wherein the biggest weight is given to air pollution. Traffic index is a composite index of time consumed in traffic due to job commute, estimation of time consumption dissatisfaction, CO_2 consumption estimation in traffic and overall inefficiencies in the traffic system. Literacy rate is based on people aged 15 or over who can read and write. [3]

Methodology

Data Gathering

The data are obtained through Wikipedia.org and Numbeo.com. Since the following sources allow anyone to alter its content, the data may not be credible. However, the data has been gathered and still used for the purpose of this project. Sixty countries are sampled and the data are summarized in *"Final Data"* sheet found in the excel file.

Data Utilization

Different models are constructed to analyze better which of the variables explain the most of the variation. Only one model is chosen that best represents in determining the factors affecting life expectancy of a country based on the following measures: R^2, $\widetilde{R^2}$, standard error, F-value and p-values for the variables considered. The regression statistics and values of the coefficients are derived using Excel. The first model constructed includes all the variables mentioned above, and then reducing the explanatory variables one by one based on the acceptable variable's p-value. The ideal p-value for all variables is < 0.05. The variable with the highest p-value is removed first and then assess the impact to the residuals, standard error, and F-statistics. Reducing the variables aims to reduce the residuals. However, in some cases, removing one of the explanatory variables leads to the increase in R^2 but decrease in $\widetilde{R^2}$. The best model is chosen if the modification of the model results to a minimal increase in $\widetilde{R^2}$ but decreases the standard error, improves the F-value and improves the p-values.

Regression Model

$$Y = \beta_0 + \beta_1 D + \beta_2 X_1 + \beta_3 X_2 + \beta_4 X_3 + \beta_5 X_4 + \beta_6 X_5 + \beta_7 X_6$$

where,

$Y = Life\ expectancy\ (in\ years)$
$D = Dummy\ variable\ for\ gender\ (1 - Male, 0 - Female)$
$X_1 = Real\ GDP\ growth\ rate\ (in\ \%)$
$X_2 = Health\ care\ index$
$X_3 = Safety\ index$
$X_4 = Pollution\ index$
$X_5 = Traffic\ index$
$X_6 = Literacy\ rate$

Results and Analyses

The following are the summary output per model derived using Excel.

Model 1:

$$Y = \beta_0 + \beta_1 D + \beta_2 X_1 + \beta_3 X_2 + \beta_4 X_3 + \beta_5 X_4 + \beta_6 X_5 + \beta_7 X_6$$

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.760848937
R Square	0.578891105
Adjusted R Square	0.552571799
Standard Error	3.989583049
Observations	120

ANOVA

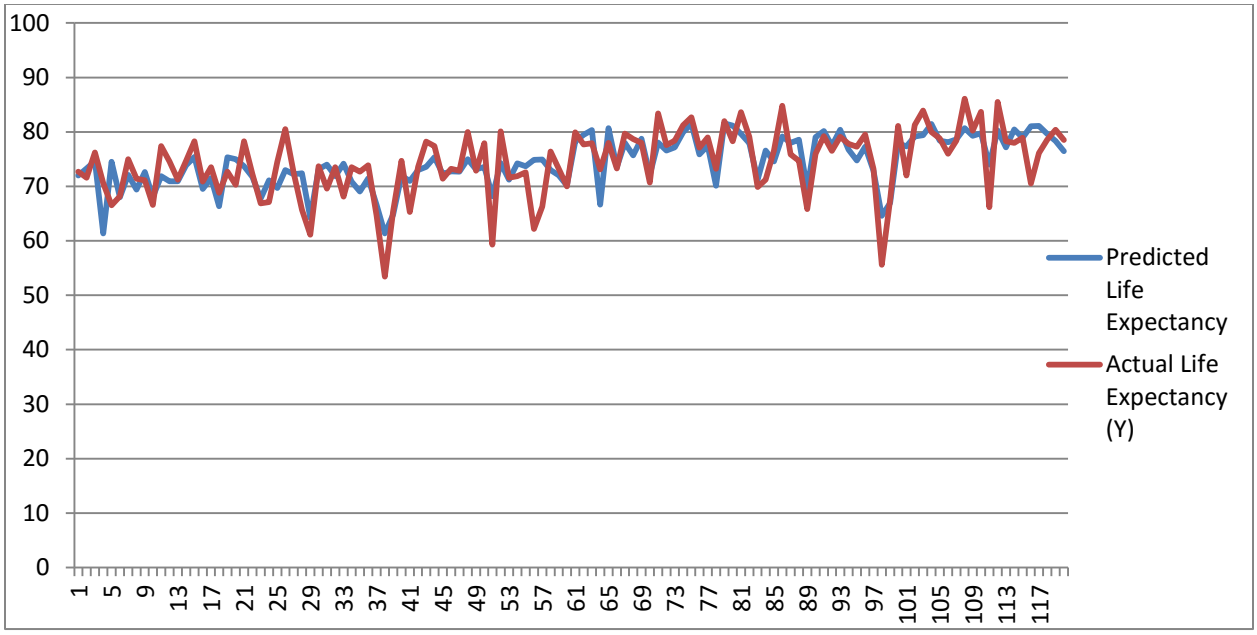
	df	*SS*	*MS*	*F*	*Significance F*
Regression	7	2450.6174	350.088193	21.99492291	1.85655E-18
Residual	112	1782.6786	15.91677291		
Total	119	4233.2959			

	Coefficients	*andard Erı*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	53.65169395	5.2662801	10.18777833	1.20682E-17	43.21723505	64.08615285	43.21723505	64.08615285
GDP Growth Rate	-0.339588955	0.1350657	-2.514250292	0.013349207	-0.607204309	-0.0719736	-0.607204309	-0.0719736
Health Care Index	0.090204283	0.0357099	2.526031012	0.012933549	0.019449719	0.160958847	0.019449719	0.160958847
Safety Index	0.099135972	0.0319733	3.100589255	0.002443478	0.035785039	0.162486905	0.035785039	0.162486905
Pollution Index	0.010198842	0.0245961	0.414652443	0.679189494	-0.038535222	0.058932906	-0.038535222	0.058932906
Traffic Index	-0.007634726	0.0093352	-0.817846797	0.415181191	-0.026131138	0.010861686	-0.026131138	0.010861686
Literacy Rate	15.26561239	4.8153088	3.170225031	0.001964822	5.724695314	24.80652946	5.724695314	24.80652946
Gender	-6.198134261	0.7422866	-8.350055837	2.03074E-13	-7.668879939	-4.727388583	-7.668879939	-4.727388583

Model 1 has a standard error of 3.9896 and the adjusted R^2 is 55.2572%. This implies that only 55.2572% of the variance is explained by Model 1. Looking at the p-value of the coefficients, since *Pollution Index* and *Traffic Index* has p-value greater than 0.05, these suggest that the two variables have no significant predictive capability.

Looking at the graph below, there are underestimation and overestimation of life expectancy for some countries. This suggests that an improvement of the model is needed.

Figure 1 Predicted vs Actual Life Expectancy (Model 1)



Model 2:

$$Y = \beta_0 + \beta_1 D + \beta_2 X_1 + \beta_3 X_2 + \beta_4 X_3 + \beta_6 X_5 + \beta_7 X_6$$

SUMMARY OUTPUT

Regression Statistics		
Multiple R	0.760423987	*(0.000425)*
R Square	0.57824464	*(0.000646)*
Adjusted R Square	0.55585055	*0.003279*
Standard Error	3.97493835	*(0.014645)*
Observations	120	*0.000000*

ANOVA

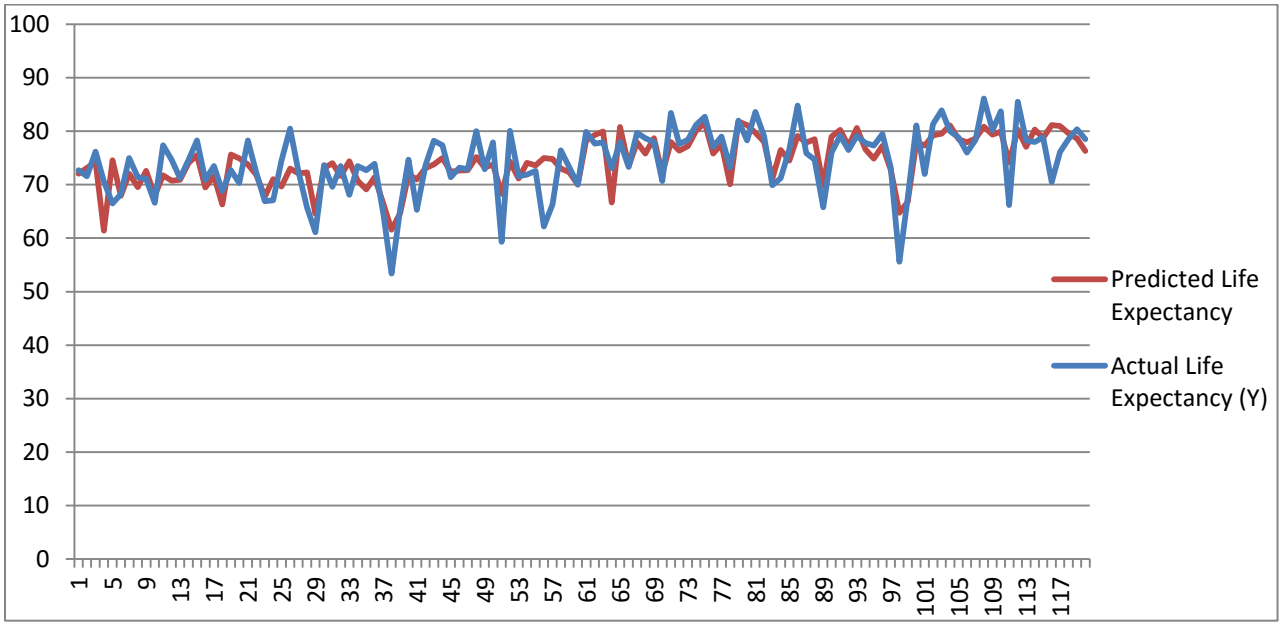
	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	2447.880675	407.9801124	25.82130566	3.77534E-19
Residual	113	1785.415242	15.80013489		
Total	119	4233.295917			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	54.4506685	4.883125607	11.15078187	6.33772E-20	44.77631595	64.12502104	44.77631595	64.12502104
GDP Growth Rate	-0.32796294	0.131638486	-2.491391001	0.014176802	-0.588762523	-0.067163356	-0.588762523	-0.067163356
Health Care Index	0.085956467	0.034083357	2.521948418	0.013063557	0.018431191	0.153481744	0.018431191	0.153481744
Safety Index	0.097045648	0.031457457	3.084980694	0.002559686	0.034722752	0.159368544	0.034722752	0.159368544
Traffic Index	-0.006198033	0.008636524	-0.717653626	0.47445188	-0.023308545	0.010912479	-0.023308545	0.010912479
Literacy Rate	15.23679407	4.797133314	3.176229026	0.001923793	5.732807769	24.74078038	5.732807769	24.74078038
Gender	-6.197278837	0.739558972	-8.379695299	1.65149E-13	-7.662478547	-4.732079128	-7.662478547	-4.732079128

To improve the fit, *Pollution Index* is removed in the model. Looking at the summary output, adjusted R^2 slightly increased and the standard error slightly decreased. The F-value has

improved as well. Thus, Model 2 is better than Model 1. Graphically, prediction on the response variable has also improved.

Figure 2 Predicted vs Actual Life Expectancy (Model 2)



Model 3:

$$Y = \beta_0 + \beta_1 D + \beta_2 X_1 + \beta_3 X_2 + \beta_4 X_3 + \beta_7 X_6$$

SUMMARY OUTPUT

Regression Statistics		
Multiple R	0.759158996	*(0.001265)*
R Square	0.576322381	*(0.001922)*
Adjusted R Square	0.557740029	*0.001889*
Standard Error	3.966474344	*(0.008464)*
Observations	120	*0.000000*

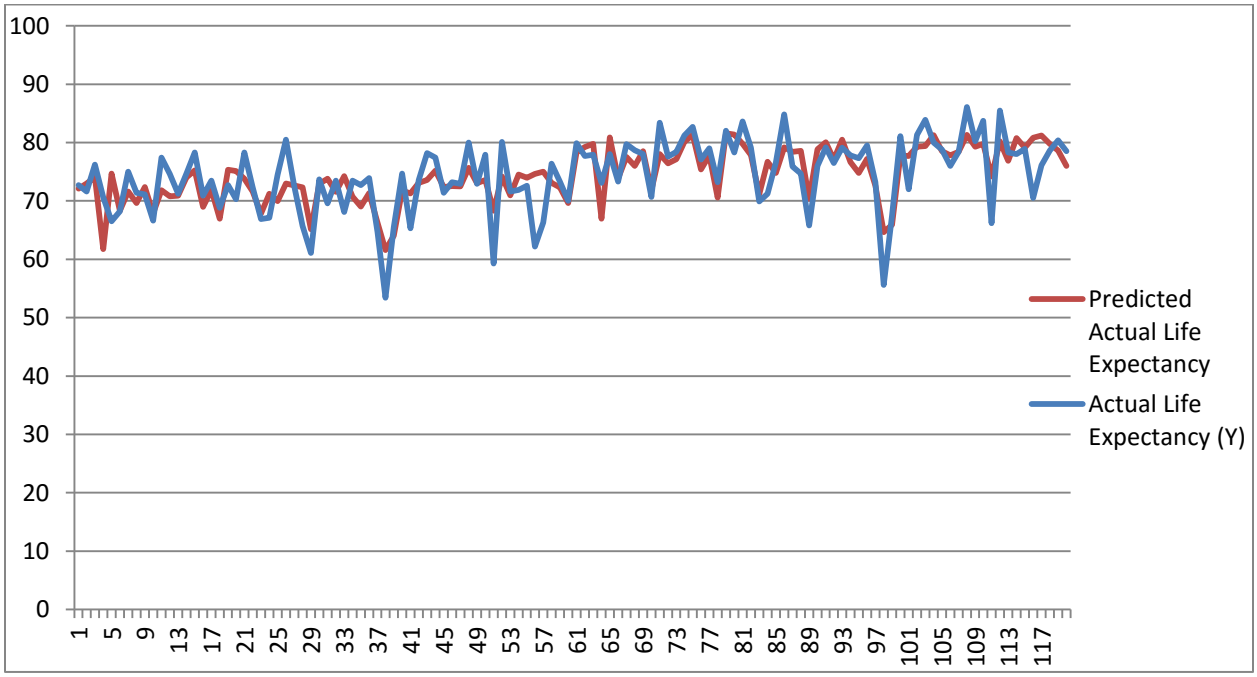
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	2439.743183	487.9486366	31.01450184	8.21392E-20
Residual	114	1793.552734	15.73291872		
Total	119	4233.295917			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	52.16041836	3.688141443	14.14273806	7.68445E-27	44.85423857	59.46659815	44.85423857	59.46659815
GDP Growth Rate	-0.334667987	0.131026912	-2.554192742	0.011962468	-0.59423129	-0.075104684	-0.59423129	-0.075104684
Health Care Index	0.084000818	0.033901901	2.477761287	0.014686925	0.016841413	0.151160222	0.016841413	0.151160222
Safety Index	0.107934608	0.027497659	3.925229032	0.000148676	0.053461958	0.162407257	0.053461958	0.162407257
Literacy Rate	16.2323842	4.582381154	3.542347015	0.000575745	7.15472243	25.31004598	7.15472243	25.31004598
Gender	-6.226831271	0.736839303	-8.450731718	1.0796E-13	-7.686504256	-4.767158287	-7.686504256	-4.767158287

Removing the *Traffic Index* variable from Model 2, Model 3 shows better fit than model 2. Adjusted R^2 has slightly increased and a slight decreased in standard error. F-value has improved as well. The remaining variables have statistically significant predictive capability. With this, this is the best model by far.

Figure 3 Predicted vs Actual Life Expectancy (Model 3)



Model 4:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.736742668
R Square	0.542789759
Adjusted R Square	0.530965357
Standard Error	4.084776452
Observations	120

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	2297.789672	765.9298907	45.90420081	1.24185E-19
Residual	116	1935.506245	16.68539866		
Total	119	4233.295917			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	50.98326293	3.400435453	14.99315709	6.47476E-29	44.24827205	57.71825381	44.24827205	57.71825381
Safety Index	0.096799479	0.026726808	3.621812166	0.000435252	0.043863668	0.14973529	0.043863668	0.14973529
Literacy Rate	22.81583403	4.069216107	5.606935938	1.41562E-07	14.7562388	30.87542926	14.7562388	30.87542926
Gender	-6.422250007	0.755492999	-8.500740595	7.46839E-14	-7.918599085	-4.925900928	-7.918599085	-4.925900928

Model 4 considers only two explanatory variables, *Safety Index* and *Literacy Index*. Based on the summary output, the model can be considered a good fit. Adjusted R^2 is high and comparable to Model 3 but with higher standard error. F-value also is higher than the Model 3.

Model 5:

SUMMARY OUTPUT

Regression Statistics		
Multiple R	0.743018505	*0.006276*
R Square	0.552076498	*0.009287*
Adjusted R Square	0.53649655	*0.005531*
Standard Error	4.060619712	*(0.024157)*
Observations	120	*0.000000*

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	2337.103186	584.2757964	35.43506706	2.87977E-19
Residual	115	1896.192731	16.48863244		
Total	119	4233.295917			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	49.16453774	3.579654097	13.73443813	5.32954E-26	42.07393191	56.25514357	42.07393191	56.25514357
Health Care Index	0.049024565	0.031749372	1.544111334	0.125308788	-0.01386483	0.111913959	-0.01386483	0.111913959
Safety Index	0.088731333	0.02707767	3.276919076	0.001387885	0.03509568	0.142366986	0.03509568	0.142366986
Literacy Rate	22.01739832	4.078066443	5.398979793	3.65264E-07	13.93953356	30.09526308	13.93953356	30.09526308
Gender	-6.398549773	0.751181956	-8.517975871	7.18263E-14	-7.886496656	-4.910602891	-7.886496656	-4.910602891

Adding *Health Care Index* from Model 4 increases the Adjusted R^2 and decreases the standard error, however, as its p-value suggests, this explanatory variable has no statistically significant predictive power.

Model 6:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.743979738
R Square	0.553505851
Adjusted R Square	0.53797562
Standard Error	4.054135685
Observations	120

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	2343.154059	585.7885148	35.64054144	2.40233E-19
Residual	115	1890.141857	16.43601615		
Total	119	4233.295917			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	53.59438895	3.722951553	14.39567187	1.71645E-27	46.21993844	60.96883946	46.21993844	60.96883946
GDP Growth Rate	-0.203533919	0.122511696	-1.661342766	0.099368501	-0.44620601	0.039138171	-0.44620601	0.039138171
Safety Index	0.111978979	0.028055812	3.991293435	0.000116077	0.056405815	0.167552143	0.056405815	0.167552143
Literacy Rate	19.64401647	4.467218007	4.397371348	2.45823E-05	10.79531735	28.49271558	10.79531735	28.49271558
Gender	-6.328099889	0.75196441	-8.415424721	1.23502E-13	-7.817596664	-4.838603114	-7.817596664	-4.838603114

Model 6 has included *Real GDP Growth Rate* instead of *Health Care Index* from Model 5 which has p-value of less than 0.05. This model is better than Model 5.

Model 7:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.737972202
R Square	0.544602971
Adjusted R Square	0.528763074
Standard Error	4.094354951
Observations	120

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	2305.465533	576.3663834	34.381725	7.35929E-19
Residual	115	1927.830383	16.76374246		
Total	119	4233.295917			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	53.01540986	4.542704442	11.67045106	3.08704E-21	44.01718664	62.01363307	44.01718664	62.01363307
Pollution Index	-0.015259908	0.022551418	-0.676671792	0.499973009	-0.059929927	0.02941011	-0.059929927	0.02941011
Safety Index	0.090431691	0.02839424	3.184860354	0.001864059	0.034188165	0.146675216	0.034188165	0.146675216
Literacy Rate	22.02762949	4.241825805	5.192959471	9.0384E-07	13.62538895	30.42987003	13.62538895	30.42987003
Gender	-6.398853469	0.758053513	-8.441163267	1.07808E-13	-7.900411584	-4.897295353	-7.900411584	-4.897295353

Replacing *Real GDP Growth Rate* variable with *Pollution Index* does not make the model any better.

Model 8:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.737998413
R Square	0.544641658
Adjusted R Square	0.528803107
Standard Error	4.094181036
Observations	120

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	2305.629305	576.4073264	34.38708859	7.32393E-19
Residual	115	1927.666611	16.76231836		
Total	119	4233.295917			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	53.36404521	4.871916545	10.95339888	1.47671E-19	43.71371618	63.01437424	43.71371618	63.01437424
Traffic Index	-0.006025828	0.008811216	-0.68388147	0.495425744	-0.023479151	0.011427496	-0.023479151	0.011427496
Safety Index	0.087012129	0.03037159	2.864918466	0.004960205	0.026851855	0.147172404	0.026851855	0.147172404
Literacy Rate	21.77728437	4.352129674	5.003822497	2.0406E-06	13.1565531	30.39801565	13.1565531	30.39801565
Gender	-6.391422391	0.758572933	-8.425587197	1.17051E-13	-7.894009377	-4.888835405	-7.894009377	-4.888835405

Same also with the *Traffic Index* variable, it does not improve the model.

Model 9:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.700776331
R Square	0.491087466
Adjusted R Square	0.482388107
Standard Error	4.291093291
Observations	120

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	2078.918565	1039.459283	56.45099081	6.89643E-18
Residual	117	2154.377351	18.41348164		
Total	119	4233.295917			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	50.18414763	3.564659443	14.07824462	6.12853E-27	43.1245264	57.24376886	43.1245264	57.24376886
Literacy Rate	29.40385395	3.823882829	7.689527964	5.0673E-12	21.83085438	36.97685352	21.83085438	36.97685352
Gender	-6.617804398	0.791622536	-8.359797877	1.50038E-13	-8.185571349	-5.050037447	-8.185571349	-5.050037447

Considering only the explanatory variable *Literacy Rate* gives high adjusted R^2 and F-value which explains 48.24% of the variance.

Model 10:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.517624048
R Square	0.267934656
Adjusted R Square	0.255420718
Standard Error	5.146610563
Observations	120

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1134.246683	567.1233416	21.41089927	1.1917E-08
Residual	117	3099.049234	26.48760029		
Total	119	4233.295917			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	78.08613869	0.752603394	103.7546992	5.1351E-117	76.59564711	79.57663027	76.59564711	79.57663027
GDP Growth Rate	-0.325021822	0.139350319	-2.332408175	0.021388537	-0.600997826	-0.049045818	-0.600997826	-0.049045818
Gender	-5.745	0.939638233	-6.114055172	1.3187E-08	-7.605904283	-3.884095717	-7.605904283	-3.884095717

Considering *GDP Growth Rate* contributes only one-fourth of the variation.

Model 11:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.545112578
R Square	0.297147722
Adjusted R Square	0.285133153
Standard Error	5.042877416
Observations	120

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1257.914239	628.9571194	24.73228343	1.10039E-09
Residual	117	2975.381678	25.43061263		
Total	119	4233.295917			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	69.80325674	2.38894425	29.21929079	1.37074E-55	65.07207784	74.53443564	65.07207784	74.53443564
Health Care Index	0.122164875	0.037648622	3.244869742	0.001532309	0.047603751	0.196725998	0.047603751	0.196725998
Gender	-5.745	0.920699238	-6.239822691	7.23553E-09	-7.568396596	-3.921603404	-7.568396596	-3.921603404

Model 12:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.647208593
R Square	0.418878963
Adjusted R Square	0.40894527
Standard Error	4.585425921
Observations	120

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1773.238604	886.6193019	42.16749657	1.6203E-14
Residual	117	2460.057313	21.02613088		
Total	119	4233.295917			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	68.34864891	1.57590376	43.37108054	6.07E-74	65.22765409	71.46964373	65.22765409	71.46964373
Safety Index	0.163786539	0.026838154	6.102749758	1.39141E-08	0.110634981	0.216938097	0.110634981	0.216938097
Gender	-5.745	0.837180404	-6.862320201	3.43316E-10	-7.402991921	-4.087008079	-7.402991921	-4.087008079

Model 13:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.579476014
R Square	0.335792451
Adjusted R Square	0.324438476
Standard Error	4.90228196
Observations	120

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1421.508813	710.7544063	29.57487977	4.02506E-11
Residual	117	2811.787104	24.03236841		
Total	119	4233.295917			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	83.41270903	1.583814703	52.66569939	2.48969E-83	80.27604701	86.54937106	80.27604701	86.54937106
Pollution Index	-0.097798072	0.023083923	-4.236631329	4.54682E-05	-0.143514573	-0.052081572	-0.143514573	-0.052081572
Gender	-5.745	0.895030138	-6.418778273	3.05093E-09	-7.517560287	-3.972439713	-7.517560287	-3.972439713

Model 14:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.618803421
R Square	0.382917674
Adjusted R Square	0.372369258
Standard Error	4.725175429
Observations	120

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1621.003825	810.5019126	36.30096499	5.4334E-13
Residual	117	2612.292092	22.32728283		
Total	119	4233.295917			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	83.30513846	1.29025985	64.5646212	2.4495E-93	80.74984644	85.86043048	80.74984644	85.86043048
Traffic Index	-0.040835924	0.007682388	-5.315525125	5.16034E-07	-0.05605049	-0.025621358	-0.05605049	-0.025621358
Gender	-5.745	0.862695057	-6.659363531	9.39842E-10	-7.453522354	-4.036477646	-7.453522354	-4.036477646

Summary and Conclusion

The following tables summarized the effect of different models considered to the R^2, adjusted R^2, standard error, F-value and p-values.

SUMMARY PER MODEL

Model	R Square	Adjusted R Square	Standard Error	F	P-value_Intercept	P-value_GDP	P-value_Health Care	P-value_Safety	P-value_Pollution	P-value_Traffic	P-value_Literacy	P-value_Gender
1	0.578891	0.552572	3.989583	21.994923	0.000000	0.013349	0.012934	0.002443	0.679189	0.415181	0.001965	0.000000
2	0.578245	0.555851	3.974938	25.821306	0.000000	0.014177	0.013064	0.002560		0.474452	0.001924	0.000000
3	0.576322	0.557740	3.966474	31.014502	0.000000	0.011962	0.014687	0.000149			0.000576	0.000000
4	0.542790	0.530965	4.084776	45.904201	0.000000			0.000435			0.000000	0.000000
5	0.552076	0.536497	4.060620	35.435067	0.000000		0.125309	0.001388			0.000000	0.000000
6	0.553506	0.537976	4.054136	35.640541	0.000000	0.099369		0.000116			0.000025	0.000000
7	0.544603	0.528763	4.094355	34.381725	0.000000			0.001864	0.499973		0.000001	0.000000
8	0.544642	0.528803	4.094181	34.387089	0.000000			0.004960		0.495426	0.000002	0.000000
10	0.267935	0.255421	5.146611	21.410899	0.000000	0.021389						0.000000
11	0.297148	0.285133	5.042877	24.7322834	0.000000		0.001532309					0.000000
12	0.418879	0.408945	4.585426	42.167497	0.000000			0.000000				0.000000
13	0.335792	0.324438	4.902282	29.574880	0.000000				0.000045			0.000000
14	0.382918	0.372369	4.725175	36.300965	0.000000					0.000001		0.000000
9	0.491087	0.482388	4.291093	56.450991	0.000000						0.000000	0.000000

It is concluded that Model 3 is the best model among 14 models tested.

BEST MODEL (By Rank)

Model	R Square	Adjusted R Square	Standard Error	F	P-value_Intercept	P-value_GDP	P-value_Health Care	P-value_Safety	P-value_Pollution	P-value_Traffic	P-value_Literacy	P-value_Gender
3	0.576322	0.557740	3.966474	31.014502	0.000000	0.011962	0.014687	0.000149			0.000576	0.000000
2	0.578245	0.555851	3.974938	25.821306	0.000000	0.014177	0.013064	0.002560		0.474452	0.001924	0.000000
1	0.578891	0.552572	3.989583	21.994923	0.000000	0.013349	0.012934	0.002443	0.679189	0.415181	0.001965	0.000000
6	0.553506	0.537976	4.054136	35.640541	0.000000	0.099369		0.000116			0.000025	0.000000
5	0.552076	0.536497	4.060620	35.435067	0.000000		0.125309	0.001388			0.000000	0.000000
4	0.542790	0.530965	4.084776	45.904201	0.000000			0.000435			0.000000	0.000000
8	0.544642	0.528803	4.094181	34.387089	0.000000			0.004960		0.495426	0.000002	0.000000
7	0.544603	0.528763	4.094355	34.381725	0.000000			0.001864	0.499973		0.000001	0.000000
9	0.491087	0.482388	4.291093	56.450991	0.000000						0.000000	0.000000
12	0.418879	0.408945	4.585426	42.167497	0.000000			0.000000				0.000000
14	0.382918	0.372369	4.725175	36.300965	0.000000					0.000001		0.000000
13	0.335792	0.324438	4.902282	29.574880	0.000000				0.000045			0.000000
11	0.297148	0.285133	5.042877	24.7322834	0.000000		0.001532309					0.000000
10	0.267935	0.255421	5.146611	21.410899	0.000000	0.021389						0.000000

Thus, the model that can be used to determine the life expectancy of a country is as follows:

$$Y = 52.1604 - 6.2268D - 0.3347\,X_1 + 0.0840X_2 + 0.1079X_3 + 16.2324X_6$$

where,

$$Y = Life\ expectancy\ (in\ years)$$
$$D = Dummy\ variable\ for\ gender\ (1 - Male, 0 - Female)$$
$$X_1 = Real\ GDP\ growth\ rate\ (in\ \%)$$
$$X_2 = Health\ care\ index$$
$$X_3 = Safety\ index$$
$$X_6 = Literacy\ rate$$

On the average, females live longer than males by 6.23 years. Since this model explains only less than 60% of the variance of the response variable, it is recommended to consider other factors.

References

[1] Max Roser (2016) – ‘Life Expectancy’. *Published online at OurWorldInData.org.* Retrieved from: https://ourworldindata.org/life-expectancy/ [Online Resource]

[2] Wikipedia: https://en.wikipedia.org/wiki/List_of_countries_by_real_GDP_growth_rate

[3] Numbeo: https://www.numbeo.com/health-care/rankings_by_country.jsp

[4] Numbeo: https://www.numbeo.com/traffic/rankings_by_country.jsp

[5] Numbeo: https://www.numbeo.com/crime/rankings_by_country.jsp

[6] Numbeo: https://www.numbeo.com/pollution/rankings_by_country.jsp

[7] Wikipedia: https://en.wikipedia.org/wiki/List_of_countries_by_life_expectancy

[8] Wikipedia: https://en.wikipedia.org/wiki/List_of_countries_by_literacy_rate