Analysis of Health Care Expense over Life Expectancy

Ajit Kumar
School of Computing
Dublin City University
Dublin,Ireland
ajit.kumar7@mail.dcu.ie

Nikhil Mittal
School of Computing
Dublin City University
Dublin,Ireland
nikhil.mittal2@mail.dcu.ie

***Abstract*—This study focuses on the analysis of health care expenditures per capita over life expectancies in the developed countries between 2000 to 2015. There is no direct relationship that the country having higher health care expenditures should higher upward trajectory towards life expectancy. Certain factors like infant deaths, alcohol, BMI, quality of life also play an important role in reducing or increasing the average life of country population.**

Keywords—life expectancies, health care expense, outlier, upward trajectory, BMI, alcohol, infant deaths

I. INTRODUCTION

Health is a multidimensional phenomenon that goes beyond political financial, cultural and national relations. It is also regarded as a measure of social and economic growth. Good health is the key to a quality life. Life expectancy has been considered as an important attribute for health system development and human well-being. Technological advancement in the field of health care and the presence of good quality medicines have led to the longer life of the people as compared to the people living in the past.

All developed countries around the world are investing heavily in their health care, and as medical spending increases, the population lives longer in an upward trend. But by taking a flatter trajectory, the USA stands out. There's far more spending on health care than many developed countries, yet American people don't live longer. This growth has contributed to significant disparity between the United States and other rich countries. In this research, we will take a detailed analysis of health care expenses and life expectancy and try to find out the factors which are co-responsible for increasing or decreasing the life expectancy of country.

II. RELATED WORK

In [1] authors have analyzed the relationship between birth life expectancy and health care spending, considering social development and educational expenses, as well as the quality and quantity of health care and lifestyles provision. Expenditure on health care is not the main determinant of life expectancy at birth, but expenditure on social protection is.

Final Model of Forward Linear Regression Analysis

Indicators	Standardized Beta	Significance Level	Adjusted Explained Variance
Social protection	0.416	.000	84%
Number curative beds per 100,000 inhabitants	−0.193	.042	
Infant standardized mortality rate	−0.233	.013	
% unmet health care needs	−0.232	.008	
Alcohol consumption in liters per inhabitant	−0.280	.004	

The regression analysis shows that citizens have a significant longer life expectancy in countries that spend a high percentage of their GDP on social protection, have fewer healing beds and low infant mortality, whose citizens report less unsatisfied by the health care needs and drink less alcohol.

In [2] author tries to find out whether life expectancy is related to environmental health or not. Analysis is performed on two models a set of independent variable and multivariate variable. For this author have used data of 156 countries which mainly comprises of attributes like Access to safe drinking water, per capita gross domestic product, literacy, calories available as percentage of needs and per capita public health expenditure. A linear regression model is used to estimate the role of different attributes on life expectancy at birth and correlation matrix are used getting the relationship between the variable. The correlation matrix of all the five variables are summarized in table1.The result of multivariate regression analysis are shown on table2

Table 1. Correlation matrix of selected indicators

	GDP	CAL	LIT	SDW	PHE	LE
GDP	1					
CAL	0.224	1				
LIT	0.413	−0.094	1			
SDW	0.367	−0.130	0.667	1		
PHE	0.795	0.156	0.349	0.293	1	
LE	0.559	0.035	0.832	0.695	0.452	1

Table 2. Results of the multivariate linear model

Variable	Regression coefficients	*t*-value	*p*-value	95% CI
GDP	0.00029	3.229	0.0015	0.0001–0.0004
CAL	0.0152	1.826	0.069	−0.0012–0.0316
LIT	0.2439	11.249	<0.001	0.2011–0.2868
SDW	0.0830	4.297	<0.001	0.0448–0.1212
PHE	−0.00034	−0.138	0.889	−0.0051–0.0044

From above figures and results author have concluded that life expectency is a very good indicator of environmental health.

In [3] author tries to examine is there is any relation between health care expenditure and national life expectency. In addition to healthcare expenditure, government expenditure, concentration of doctors in an area, and literacy rate as independent variables are also being considered. Various regression model was used first for the sample of developing countries and then for the

sample of developed country. To find is there is any noticeable difference between the trend of two. In the initial analysis health expenditures show exponential growth with life expectancy as shown in Figure 1.

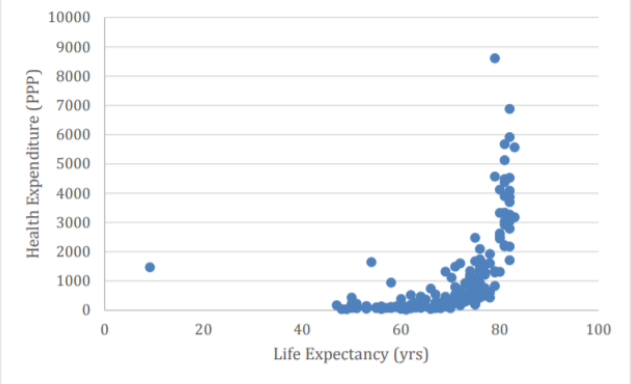

Figure 1. Health Expenditure vs. Life Expectancy

Author concluded that there would be a positive correlation between healthcare expenditure and life expectency which indicates that expenditure will increase then expectancy will also increase.This trend have shown positive impact in developed country whereas healthcare expenditure is an insignificant life expectancy factor in developing country.

III. Dataset and Exploratory Analysis

The Global Health observatory (GHO) data repository which comes under World health organization (WHO) keeps the track of wellbeing status and numerous other related health features. This dataset is publicly available for the for the purpose of health data analysis. The dataset consists of WHO) aggregated data of different countries of the world. The data for each country is an average of many indicator. Essentially, the data are multiple variable in a country-specific time series. Two different type of dataset is being used for this analysis one is health expenditure per annum in USD of 35 different countries of the world from year 2000 to 2015.

	LOCATION	INDICATOR	SUBJECT	MEASURE	FREQUENCY	Year	Cost
0	AUS	HEALTHEXP	TOT	USD_CAP	A	2000	2156.8080
1	AUS	HEALTHEXP	TOT	USD_CAP	A	2001	2271.4553
2	AUS	HEALTHEXP	TOT	USD_CAP	A	2002	2426.4793
3	AUS	HEALTHEXP	TOT	USD_CAP	A	2003	2546.2803
4	AUS	HEALTHEXP	TOT	USD_CAP	A	2004	2745.6316
...	...	...	...	...	...	...	...
555	LVA	HEALTHEXP	TOT	USD_CAP	A	2011	1086.9471
556	LVA	HEALTHEXP	TOT	USD_CAP	A	2012	1148.7918
557	LVA	HEALTHEXP	TOT	USD_CAP	A	2013	1219.4164
558	LVA	HEALTHEXP	TOT	USD_CAP	A	2014	1295.0128
559	LVA	HEALTHEXP	TOT	USD_CAP	A	2015	1369.9330

Fig 1. Health care expense

The second data set used is life expectancy of country from year 2000 to 2015 which contains information about death by diseases, infant mortality, death by alcohol.
Different columns of data are of different type for example country and year are of nominal data type, whereas rest of the columns like life expectancy, infant mortality, alcohol is ratio data type.

	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	...	Polio	Total expenditure	Diphtheria	HIV/AIDS	
0	AUS	2015	Developed	82.8	59	1	NaN	0.000000	93.0	74	...	93	NaN	93	0.1	56554.38
1	AUS	2014	Developed	82.7	6	1	9.71	10769.363050	91.0	340	...	92	9.42	92	0.1	62214.69
2	AUS	2013	Developed	82.5	61	1	9.87	11734.853810	91.0	158	...	91	9.36	91	0.1	67792.33
3	AUS	2012	Developed	82.3	61	1	10.03	11714.998580	91.0	199	...	92	9.36	92	0.1	67677.63
4	AUS	2011	Developed	82.0	63	1	10.30	10986.265270	92.0	190	...	92	9.20	92	0.1	62245.12
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
555	LVA	2004	Developed	71.0	26	0	8.81	80.032837	98.0	0	...	99	6.51	99	0.1	6351.81
556	LVA	2003	Developed	78.0	29	0	8.24	478.064491	98.0	0	...	98	6.15	98	0.1	5134.95
557	LVA	2002	Developed	73.0	219	0	7.44	376.457019	98.0	0	...	98	6.29	97	0.1	4132.34

Fig 2. Life expectancy datasets.

Initailly the data set was not clean ,so the dataset was cleaned by finding the null .These are done by finding the inplicit null value which are of no use and they are transformed into explicit null values after that they are dropped. Some of the column are also dropped which are not used in this analysis.Some of the variables are all given different name so that it can be easily understood.After this outliers were removed from the columns of infant deaths and body mass index(BMI) as some of the values in like 0 and 1 in the cloumn of infant mortility rate and BMI of 75 or above these outlier are detected with help of boxplot and after this these are removed because they are not making any sense.

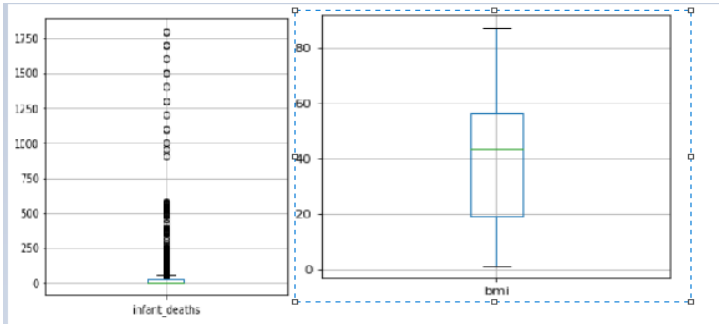

Fig 3. Boxplot of infant death and BMI

Mean of different column have been calcualted and there respective standard deviation have also been calcualted in order to check the spread of the value mean.Continuity of various variable are checed by using univariate technique by plotting the histogram of the different columns.

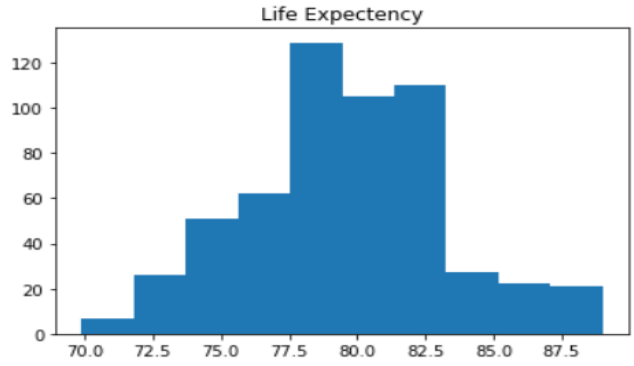

Fig 4. Histogram of life expectancy(target variable)

Bivariate analysis is being perform to show the relationship between the two variable . In this various varibale are being compared with the target varibale(life expectancy) in order to find realtion ship between them by using correlation techniques.

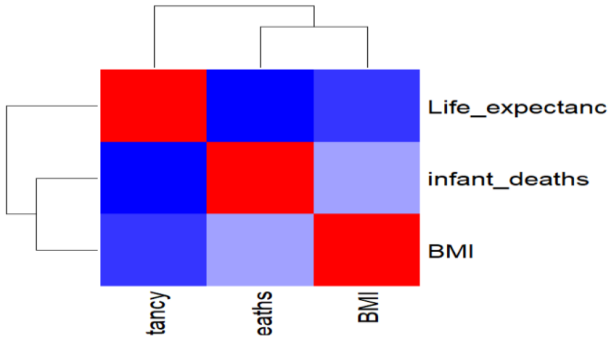

Fig 5 . Heatmap of various varible with target variable

IV. Hypothesis and Research Question

In this analysis we aim to answer following research question.(1) life expectancy increases with higher healthcare (2) Does life expectancy is directly proportional to healthcare expense of USA. (3) How does infant mortality and BMI affect life expectancy.

V. Methods Used and Why

For first hypothesis mean value of health care expense and life expectancy is used and dataset have year wise healthcare and life expectancy of different countries so to compare both mean value suited best. For second hypothesis we have used two sample t test which is used to compare the mean value of two sample. This method is used because the dataset is normally distributed and symmetric about the mean and they are continues and two sample t test comes under normal distribution As life expectancy and healthcare are continuous and two sample test are used for comparison of two numerical variable .for third hypothesis pearson's correlation coefficient is used .This is used because it calculates the strength of relationship between the two continuos variables and in this we are trying to find that is there is any relation between infant mortality, life expectancy and obesity. This will analyse by the value of correlation coefficient. For the fourth hypothesis linear regression model is used to for prediction of life expectancy and health care expense.

VI. Results and Findings

We have taken sample of 10 developed country of the world and for the first hypothesis and came to conclusion that a country with greater health care expense have greater life expectancy. As the expense increase life expectancy also increase. But USA act as an outlier because among the top 10 developed country USA expense is health care expense is far higher than the any developed country present in the sample and the life expectancy is lower than the other country. Which led to second hypothesis that does life expectancy is directly related to health care expense. In order to get the answer of second hypothesis we found that distribution of dataset by using dnorm function of R and found that the data is normally distributed over the mean.

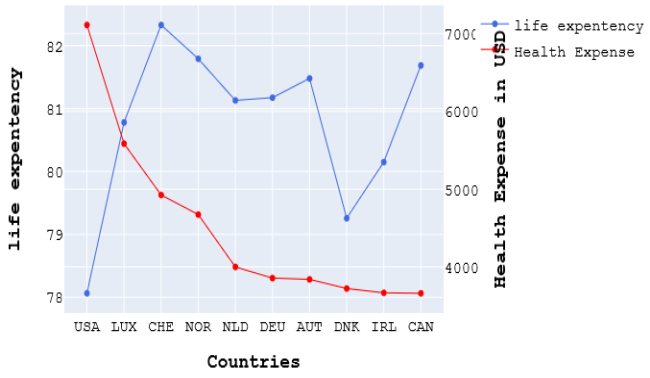

Fig 6 . Health care Vs life expectancy

For distribution of the dataset is found by using dnorm function of the R and after visualizing the result a bell curved is form which show that the data is normally distributed.

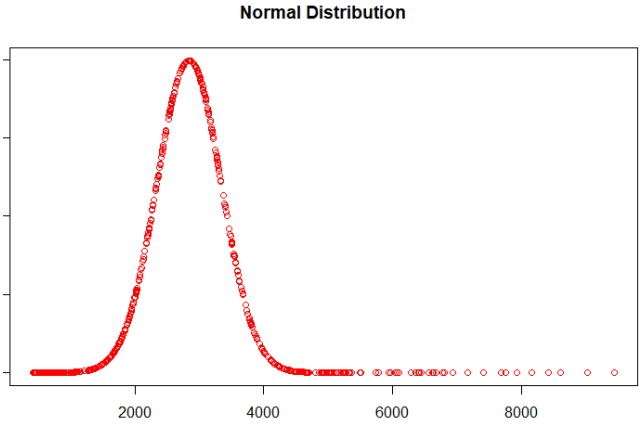

Fig 7 . Normal distribution of health care expense

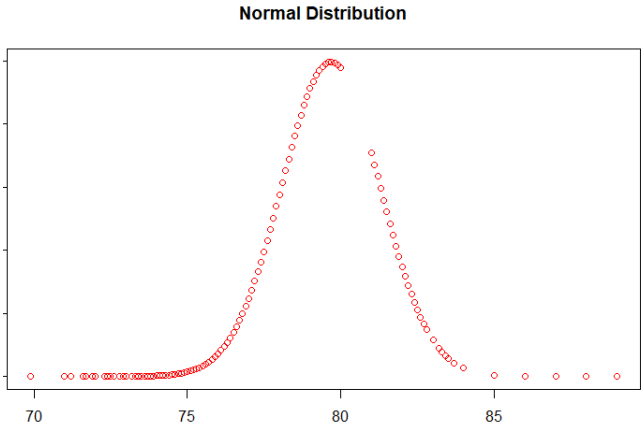

Fig 8 . Normal Distribution of life expenctancy

For getting the results of we have use two sample test and result of two sample alternative hypothesis show true difference in means is not equal to 0 on the basis output we can say that health care cost and life expectancy are not related as there mean are not same so it can such that one can increase and other can decrease and vice versa thus from two sample test we conclude our second hypothesis.

```
        Welch Two Sample t-test

data:  usa_health$Cost and usa_life$Life.expectancy
t = 18.871, df = 15, p-value = 7.312e-12
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 6237.952 7826.541
sample estimates:
mean of x mean of y
7110.3091   78.0625
```

Fig 9. Output of two sample test

For third hypothesis we have used correaltion matrix to find the relation ship between the variable and plotted the correlation matrix using heat map.The coefficient correlation for life expectancy and BMI is -0.182 where as correlation coefficient of life expectancy and infant death is -0.38 as our both the correlation coeffcient is negative which shows that both infant death and BMI are inversaly related to the target variable life expectancy. Which means that if infant death increases life expectency decreases and if life expectancy increases infant death decreses and same thing is true in case of BMI it is also inversaly related to life expactancy which conclude our third hypothesis.

```
        Pearson's product-moment correlation

data:  bmi$Life_expectancy and bmi$BMI
t = -1.0672, df = 33, p-value = 0.2936
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.4862928  0.1603622
sample estimates:
       cor
-0.1826449
```

Fig 10 . Correlation coefficient of life expectancy and BMI.

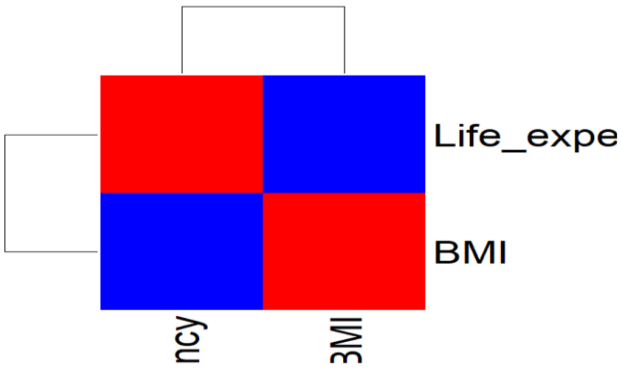

Fig 11 . Heat map of correlation matrix of life expectancy and BMI

```
        Pearson's product-moment correlation

data:  y$Life_expectancy and y$infant_deaths
t = -2.3554, df = 33, p-value = 0.0246
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.63263779 -0.05280092
sample estimates:
       cor
-0.3793721
```

Fig 12 . Correlation coeficient of life expectency and infant deaths

	Life_expectancy	infant_deaths
Life_expectancy	1.00	-0.38
infant_deaths	-0.38	1.00

Fig 13.Correlation matrix of life expectancy and deaths

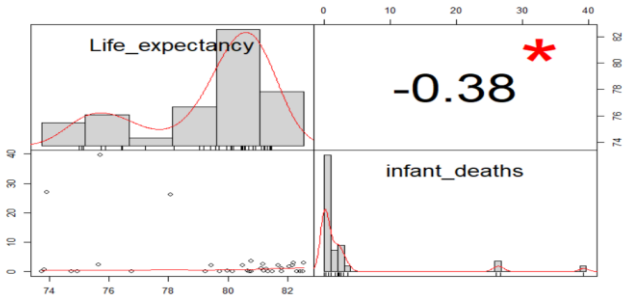

Fig 14.Scatter plot of correlation matrix of life expectency and infant deaths

VII. Conclusion

As a conclusion of this study, there is no direct relationship between the health care expense and life expectancy. Higher health care expense is not only responsible for good quality of life in any country or lower health care expense is not the only cause for the poor quality of life in any country and the same has been proved in hypothesis (2). There as certain factors like infant deaths, obesity is also responsible for low life expectancy. If a country is very rich and has a very high infant death ratio and a very large population of the country is obese then the country will have a low life expectancy as proved in hypothesis (3).

References

[1] W. J. A. van den Heuvel and M. Olaroiu, “How Important Are Health Care Expenditures for Life Expectancy? A Comparative, European Analysis,” J. Am. Med. Dir. Assoc., vol. 18, no. 3, pp. 276.e9-276.e12, 2017.

[2] S. Viscomi, G. Pastore, E. Dama, F. Merletti, and C. Magnani, “Life expectancy as an indicator of outcome,” Epidemiol. Prev., vol. 28, no. 3, pp. 174–177, 2004.

[3] “The Effect of National Healthcare Expenditure on Life Expectancy Natasha Deshpande, Anoosha Kumar, Rohini Ramaswami.”

[4] E. H. DuGoff, V. Canudas-Romo, C. Buttorff, B. Leff, and G. F. Anderson, “Multiple Chronic Conditions and Life Expectancy,” Med. Care, vol. 52, no. 8, pp. 688–694, 2014.

[5] J. Lubitz, L. Cai, E. Kramarow, and H. Lentzner, “Health, life expectancy, and health care spending among the elderly,” N. Engl. J. Med., vol. 349, no. 11, pp. 1048–1055, 2003.

[6] J. S. Sekhon et al., “Costs of Health Care Administration in the United States and Canada [2] (multiple letters),” N. Engl. J. Med., vol. 349, no. 25, pp. 2461–2464, 2003.

[7] https://ourworldindata.org/the-link-between-life-expectancy-and-health-spending-us-focus