# DISCRIMINANT ANALYSIS OF SOCIOECONOMIC FACTORS OF LIFE EXPECTANCY AT BIRTH IN ASIA

**Mehboob Ali,**
Deprtmenmt of Statistics, Government Post Graduate College, Jauharabad District Khushab - Pakistan.
Email: Statistician_955@yahoo.com
**Maqsood Ali,**
Statistical Officer,Punjab Bureau of Statistics, Lahore - Pakistan.
Email: maqsoodfsd@outlook.com
Humera Razzak,
Department of Statistics, Government College University Faisalabad - Pakistan.

**ABSTRACT:** *Life expectancy (LE) is considered as one of key health outcome and a major indicator of human development as well. Wide ranges of socioeconomic and demographic factors have major impact on life expectancy rate at birth in various countries. Association of several socioeconomic factors with life expectancy at birth and the influencing factors in forty three countries of Asia has been explored in this paper. Less surprisingly the results and discussions obtained in this paper are in agreement with previous research. A close relationship between several socioeconomic variables and life expectancy at birth is found. Discriminant analysis is performed on quantitative secondary data collected from World Population Data Sheet, 2014 which shows that life expectancy at birth is statistically significant and have positive association with four factors. However infant mortality rate, percent of dependent population, births per 1000 population and deaths per 1000 population have negative relationship with life expectancy at birth which shows life expectancy at birth decreases as the values of these factors increases.*

***Keywords:*** *Life Expectancy, Socioeconomic Factors, Discriminant Analysis.*

## INTRODUCTION

Enjoying healthy and long life in presence of better health facilities, good education and governess is a dream of every citizen in every nation [1] and these are the promises likely to be forgotten by every government which they makes while their establishment. Since life expectancy at birth is a major indicator of human development which is somehow associated with level of income yet nations with low income but with skillful demonstration of resources and expansion of human capabilities has achieved better developments [2-4].

Life expectancy rate has been greatly affected by socioeconomic risk factors universally especially in vulnerable communities living in under developed nations. Although statistical data on socioeconomic variables including population growth, health expenditures, gross national income, population density, infant mortality rate, crude death rate, crude birth rate, education, gender equality, income distribution, good governance, and healthy life is inconsistently available yet identifying and summarizing the relationship between life expectancy gap and socioeconomic risk factors mentioned above is of paramount importance. Two major hurdles that go hand in analyzing economic aspects of human development are selection of set of explanatory variables that explains variation in other economic phenomenon and specification of functional form. However third major problem which is addressed in this study is statistical analysis of human indicators in unavailability of data especially in developing countries where human indicators are not maintained on regular basis.

Our study is aimed to analyze the influence of some human and health indicators in the variation in life expectancy in Asia. Necessary data of 43 countries from Asia is obtained from World Population Data Sheet, 2014. Statistical discriminant analysis technique is used. Discriminant analysis technique will determine the set of human and health indicators relevant in determination of other indicators, which determines an economic model for explaining and estimating relationship between variation in life expectancy at birth and socioeconomic risk factors.

## LITERATURE REVIEW

Studies regarding determinants of life expectancy mostly include identifying and correlating factors influencing LE. Limitations regarding to the unavailability of LE data and lack of strong evidence against biological connection between socioeconomic determinants of life expectancy must be kept in mind before reviewing the studies regarding determinants of life expectancy. Factors related to both medical research and policy making has a great influence on life expectancy as factors like socioeconomic status, smoking cessation and overweight reduction plays a critical role in losing the Indigenous LE gap [5].

Scotland, Northern Island and England are forming new government policies to reduce life expectancy gap in result of various studies indicating health inequalities which are main determinants of life expectancy and are directly associated with income inequalities and a huge gap is witnessed in life expectancy of people living in deprived areas [6].

Numerous empirical studies have been already done in past exploring association of socioeconomic risk factors with life expectancy in European countries [7, 8]. Our study aims to explore several characteristics influencing life expectancy rate in Asian countries.

Increased per capita real income and higher expenditure on health has results in to a longer life expectancy according to a study made in Bangladesh [9] which shows that an average 8 days can be increased in a life expectancy by increasing one unit of per person health expenditure per capita (HEPC) and an increase of 33 days is expected by one US Dollar (USD) increment in GDP per capita.

## MATERIAL AND METHODS

### Data Collection

The sample/data consists of 43 Asian countries and other countries were dropped from the analysis due to

unavailability of data of some variables in these countries. Data of life expectancy at birth (Y), percent urban ($X_1$): percent of population living in urban areas, percent of married women (15-49 years) using contraception (all method) ($X_2$): percent of currently married or women of reproductive age who are currently using any methods of contraception, population per square kilometer ($X_3$), gross national income per capita (US$) ($X_4$), infant mortality rate ($X_5$), percent of dependent population ($X_6$): sum of percentages of population aged less than 15 years and more than 65 years, births per 1000 population ($X_7$) and deaths per 1000 population ($X_8$) were obtained from World Population Data Sheet, 2014 [10].

The dependent variable is the life expectancy at birth and this is divided into two categories: if the country have below 71 years life expectancy than group one and if the country have equal to or above 71 year life expectancy than group two. Two groups of dependent variable are made on the basis of overall Asia life expectancy in given year. In 2013 the overall world life expectancy is 71 year and overall Asia life expectancy is also same as world.

## METHODS

Discriminant Analysis finds a set of prediction equations based on independent variables that are used to classify individuals into groups [11, 12]. In many ways, discriminant analysis parallels multiple regression analysis. This method formulates linear equation which has been the most recognizable and the simplest interpretable measure of effect [13]. The main difference between these two techniques is that regression analysis deals with a continuous dependent variable, while discriminant analysis must have a discrete dependent variable. The mathematics of discriminant analysis is related very closely to the one-way MANOVA. In fact, the roles of the variables are simply reversed. The classification (factor) variable in the MANOVA becomes the dependent variable in discriminant analysis. The dependent variables in the MANOVA become the independent variables in the discriminant analysis.

The data for k group is taken and per group observations are represented by $N_k$. Total number of observations is represented by N. Measurements of P variables are represented by each observation. $X_{ki}$ represent the ith observation. The vector of means of variables across groups is represented by M; the vector of means of observation of group $k^{th}$ is represented by $M_k$. Define three sums of squares and cross products matrices, $S_T$, $S_W$ and $S_A$ as follows

$$S_T = \sum_{k=1}^{K}\sum_{i=1}^{N_K}(X_{ki} - M)(X_{ki} - M)'$$

$$S_W = \sum_{k=1}^{K}\sum_{i=1}^{N_K}(X_{ki} - M_k)(X_{ki} - M_k)'$$

$S_A = S_T - S_W$

Next, define two degrees of freedom values, $df_1$ and $df_2$:

$df_1 = K-1$, $df_2 = N-K$

A weighted average of the values of independent variables represents discriminant function. The selection of weights is made which separates the observation into groups. The average which comes from one group is high while the low value of the average comes from the other group. The problem reduces to one of finding the weights which, when applied to the data, best discriminate among groups according to some criterion. The solution reduces to finding the eigenvectors of $S_W^{-1}S_A$. The canonical coefficients are the elements of these eigenvectors.

A goodness-of-fit parameter, Wilks' lambda, is defined as follows:

$$\Lambda = \frac{|S_W|}{|S_T|} = \prod_{j=1}^{m}\frac{1}{1+\lambda_j}$$

where $\lambda_j$ is the $j_{th}$ eigenvalue corresponding to the eigenvector described above and m is the minimum of K- 1 and p.

The canonical correlation between the $j_{th}$ discriminant function and the independent variables is related to these eigenvalues as follows: $r_{cj} = \sqrt{\frac{\lambda_j}{1+\lambda_j}}$

The overall covariance matrix, T is given by: $T = \left(\frac{1}{N-1}\right) S_T$

The within group covariance matrix, W is given by: $W = \left(\frac{1}{N-K}\right) S_W$

The among group (or between group) covariance matrix, A is given by: $A = \left(\frac{1}{K-1}\right) S_A$

The linear discriminant functions are defined as: $LDF_k = W^{-1}M_k$

The standardized canonical coefficients are given by: $v_{ij}\sqrt{w_{ij}}$

where $v_{ij}$ are the elements of V and $w_{ij}$ are the elements of W.

The correlations between the independent variables and the canonical variates are given by:

$$Corr_{jk} = \frac{1}{\sqrt{w_{ij}}}\sum_{i=1}^{p} v_{ik}w_{ji}$$

## RESULTS AND DISCUSSION

The purpose of the present study was to examine the relationship between life expectancy at birth under two categories (life expectancy above and equal to 71 or below 71 years) and various associated factors of life expectancy. Out of 43 countries included in our analysis we have 18 countries showing life expectancy below 71 years and 25 countries have above 71 years.

Mean scores on each variable is listed in group means table no. 1. Mean of 1st three factors was high in group 2 with mean values 71.600, 58.080 and 878.480 respectively. Mean values of $X_4$, $X_5$, $X_6$ and $X_7$ was high in group 1 with mean values 6582.778, 46.833, 38.056 and 26.889 respectively. However mean value of factor $X_8$ was almost equal in both groups on the average.

In order to access the importance of each predictor standardized coefficients are presented in table 2. Univariate

analysis is conducted to present the list of independent variables influencing the dependent variable. We can say five variables carry positive signs indicating their significance contribution in discriminating the countries with equal to and above 71 years LE. On the other hand insignificant variable is $X_3$. Variables 1, 4, 5, 6 and 7 are highly significant in our analysis. The values used for calculating the cutting scores for below 71 years LE and equal to or above 71 years LE are 1.857 and -1.337 respectively.

Table 3 provides the information about the overall significance of the model. Clustering of model under two categories was significant. Unexplained error is 27% (Wilks' lambda: 0.277). Out of total variation 72% of the variability is explained by this model (canon $corr^2$: 0.723). Strong association is detected (canonical correlation coefficient: 0.850) between two groups namely set of all independent variables and two group of LE scores of all countries

.

Table 1: Group means and correlation with life expectancy

| Variables | Group 1 | Group 2 | Overall | Correlation with LE |
|---|---|---|---|---|
| $X_1$ | 38.722 | 71.600 | 57.837 | 0.748 |
| $X_2$ | 46.444 | 58.080 | 53.209 | 0.432 |
| $X_3$ | 156.500 | 878.480 | 576.256 | 0.487 |
| $X_4$ | 6582.778 | 31821.200 | 21256.279 | 0.592 |
| $X_5$ | 46.833 | 13.600 | 27.512 | -0.779 |
| $X_6$ | 38.056 | 29.320 | 32.9767 | -0.670 |
| $X_7$ | 26.889 | 16.160 | 20.651 | -0.842 |
| $X_8$ | 06.889 | 05.120 | 5.861 | -0.360 |
| Count | 18 | 25 | 43 | |

Table 2: Summary of interpretive measures for discriminant analysis

| Variables | Standardized Coefficients | Wilks Lambda | Discriminant Loading(Rank) | $df_1$ | $df_2$ | Univariate F Ratio | Sig. |
|---|---|---|---|---|---|---|---|
| $X_1$ | -0.187 | 0.609 | -0.497(8) | 1 | 41 | 26.359 | 0.000 |
| $X_2$ | 0.153 | 0.882 | -0.227(6) | 1 | 41 | 5.495 | 0.024 |
| $X_3$ | 0.026 | 0.946 | -0.147(5) | 1 | 41 | 2.318 | 0.136* |
| $X_4$ | -0.034 | 0.771 | -0.337(7) | 1 | 41 | 12.162 | 0.001 |
| $X_5$ | 0.791 | 0.345 | 0.855(1) | 1 | 41 | 78.013 | 0.000 |
| $X_6$ | -0.536 | 0.636 | 0.469(3) | 1 | 41 | 23.448 | 0.000 |
| $X_7$ | 0.755 | 0.529 | 0.585(2) | 1 | 41 | 36.500 | 0.000 |
| $X_8$ | 0.268 | 0.855 | 0.256(4) | 1 | 41 | 06.977 | 0.012 |
| Group centroid group 1 | | | 1.857 | | | | |
| Group centroid group 2 | | | -1.337 | | | | |

* insignificance

**Table 3: Eigenvalues and multivariate test**

| Function | Eigenvalue | % of Variance | Cumulative % | Canonical Correlation | (Canonical Correlation)$^2$ |
|---|---|---|---|---|---|
| 1 | 2.604[a] | 100 | 100 | 0.850 | 0.723 |
| Wilks Lambda Test | | | | | |
| Test of Function (s) | | Wilks Lambda | Chi-square | df | Sig. |
| 1 | | 0.277 | 47.439 | 8 | 0.000 |

**a.** First 1 canonical discriminant functions were used in the analysis.

Table 4: Fisher's linear discriminant functions

| Variables | Group 1 | Group 2 |
|---|---|---|
| Constant | -79.325 | -71.598 |
| $X_1$ | 0.050 | 0.079 |
| $X_2$ | 0.803 | 0.773 |
| $X_3$ | -0.002 | -0.002 |
| $X_4$ | 0.001 | 0.001 |
| $X_5$ | 0.037 | -0.170 |
| $X_6$ | 0.777 | 1.070 |
| $X_7$ | 2.181 | 1.762 |
| $X_8$ | 3.544 | 3.149 |

**Table 5: Classification count table for grouping**

| | Predicted Group Membership | | |
|---|---|---|---|
| Actual Group | 1 | 2 | Total |
| 1 | 17 (94.4) | 01 (5.60) | 18 (100) |
| 2 | 01 (4.00) | 24 (96.0) | 25 (100) |
| Total | 18 | 25 | 43 |

Table 4 shows that countries individual characteristics determine their LE. This table represents two classification functions, one for each of the two groups. Each function is represented vertically. Two equations are given below

Group 1: $-79.325+0.050X_1+0.803X_2-0.002X_3+0.001X_4+0.037X_5+0.777X_6+2.181X_7+3.544X_8$ (1)

Group 2: $-71.598+0.079X_1+0.773X_2-0.002X_3+0.001X_4-0.170X_5+1.070X_6+1.762X_7+3.149X_8$ (2)

95.34% of original grouped cases correctly classified

In table 5 the accuracy of discriminant analysis is presented. Out of forty three countries, eighteen countries with below LE, seventeen (94.4%) of them classified correctly and one (5.6%) incorrectly classified as LE equal to and above 71 years. Twenty five countries with LE above average, twenty four (96.0%) of them classified correctly and one (4.0%) were incorrectly classified as LE below average. Hence, out of 43 cases, 41 were classified correctly, which means the accuracy of the application of the discriminant analysis is 95.34%.

Table 5 also helps in increasing the accuracy of assigning clusters. Assignment of row 12 to Indonesia having equaled to or above 71 years LE was actually misclassified. Similarly misclassification of cluster is detected in row 30 to Philippines. However according to discriminant analysis it would have been a better fit if this row was assigned to group of country having equal to or above 71 years LE.

## CONCLUSION

The purpose of this study was to demonstrate that how discriminant analysis predicts country above or below average LE on the basis of associated factors of LE. Predictive equation is found for classifying new individuals or interpreting the predictive equation to better understand the relationships that may exist among the variables. Better accuracy is achieved by reclassification of variables that were misclassified [14, 15]. Successful discrimination is made between countries with LE below or above average. Clustering LE of countries under two categories is statistically significant on the basis of discriminant analysis. Analyzing set of eight variables shows seven of them significantly help in discriminating between countries with below or above average LE. Therefore it is suggested to reclassified country that was misclassified under below average LE.

## REFERENCES


[1] Lei, H., Li, L., Liu, X. and Mao, A. (2009). Quantitative study on socioeconomic determinants of life expectancy in Beijing, China. Journal of Evidence-Based Medicine, 2, 92-98.

[2] Yavari, K and Mehrnoosh, M. (2006). Determinants Life Expectancy: A Cross-Country Analysis. Iranian Economic Review, 11(15), 131-142.

[3] Sen, A. (1999). Development as Freedom. Alfred A. Knopf, New York.

[4] Razzak, H., Ali, M. and Ali, M. (2015a). Principal Component Analysis of Socioeconomic Factors and Their Association with Life Expectancy at Birth in Asia. International Journal of Multidisciplinary Academic Research, 3(1), 14-23.

[5] Zhao, Y., Wright, J., Begg, S. and Guthridge, S. (2013). Decomposing Indigenous life expectancy gap by risk factors: A life table analysis. Population Health Metrics, 11(1), http://www.pophealthmetrics.com/content/11/1/1, 10/01/2015.

[6] Health inequalities and the social determinants of health, (2012). Royal College of Nursing, Policy and international department 20 Cavendish Square London, Policy Briefing No. 01/12.

[7] Iacobuta, A.O. and Cuza, A.I. (2012). Socio-economic determinants of life expectancy: a cross-country analysis.
Journal of Academy of Business and Economics, 12(2), 76-82.

[8] Wang, H., Schumacher, A.E., Levitz, C.E., Mokdad, A.H. and Murray, C.J.L. (2013). Left behind: widening disparities for males and females in US country life expectancy, 1985-2010. Population Health Matrices, 11(8), 1-15.

[9] Mahumud, R.A., Rawal, L.B., Hossain, G., Hossain, R. and Isam, N. (2013). Impact of Life Expectancy on Economics Growth and Health care expenditures: A Case of Bangladesh. Universal Journal of Public Health, 1(4), 180-186.

[10] http://www.prb.org (Population Reference Bureau Data Sheet, 2014).

[11] Fisher, R.A. (1936). The use of Multiple Measurements in Taxonomic Problems. Annals of Eugenies, pp:179-188.

[12] Johnson, R.A. and Wichern D.W. (1992). Applied Multivariate Statistical Analysis. Englewood cliffs new jersey.

[13] Hedges, L.V. (1988). The meta-analysis of test validity studies: Some new approaches. In H. Wainer and H. I. Braun (Eds.). Test Validity, Erlbaum, pp:191-212.

[14] Ramayah, T., Jantan, M. and Chandramohan, K. (2004). Retrenchment Strategy in Human Resource Management: The Case of Voluntary Separation Scheme (VSS). Asian Academy Management Journal, 9(2), 35-62.

[15] Razzak, H, Ali, M. and Ali, M. (2015b). Application of Discriminant Analysis to Predict the Institute's Annual Performance in Sargodha Board. World Applied Sciences Journal, 33(2), 213-219.

**Appendix**

| Sr. No. | Country | Group | Y | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 01 | Afghanistan | 1 | 61 | 24 | 21 | 48 | 2000 | 74 | 48 | 35 | 8 |
| 02 | Armenia | 2 | 74 | 63 | 55 | 101 | 8140 | 10 | 30 | 14 | 9 |
| 03 | Azerbaijan | 2 | 74 | 53 | 51 | 110 | 16180 | 11 | 28 | 18 | 6 |
| 04 | Bahrain | 2 | 76 | 100 | 62 | 1901 | 36140 | 8 | 23 | 15 | 2 |
| 05 | Bangladesh | 1 | 70 | 26 | 61 | 1101 | 2810 | 33 | 34 | 20 | 6 |
| 06 | Bhutan | 1 | 68 | 36 | 66 | 16 | 7210 | 47 | 35 | 22 | 6 |
| 07 | Cambodia | 1 | 63 | 20 | 51 | 82 | 2890 | 45 | 37 | 24 | 6 |
| 08 | China | 2 | 75 | 54 | 85 | 143 | 11850 | 15 | 26 | 12 | 7 |
| 09 | China Hong Kong SAR | 2 | 84 | 100 | 80 | 6589 | 54260 | 16 | 26 | 8 | 6 |
| 10 | Georgia | 2 | 75 | 54 | 53 | 69 | 7040 | 11 | 31 | 11 | 11 |
| 11 | India | 1 | 66 | 31 | 55 | 394 | 5350 | 44 | 36 | 22 | 7 |
| **12** | **Indonesia** | **2** | **71** | **50** | **62** | **132** | **9260** | **32** | **34** | **20** | **6** |
| 13 | Iran | 2 | 74 | 71 | 82 | 47 | 15600 | 16 | 29 | 19 | 5 |
| 14 | Iraq | 1 | 69 | 71 | 53 | 80 | 15220 | 29 | 43 | 31 | 5 |
| 15 | Japan | 2 | 83 | 91 | 54 | 336 | 37630 | 19 | 39 | 8 | 10 |
| 16 | Jordan | 2 | 73 | 83 | 61 | 85 | 11660 | 17 | 38 | 28 | 4 |
| 17 | Kazakhstan | 1 | 70 | 55 | 51 | 6 | 20570 | 28 | 32 | 23 | 8 |
| 18 | Korea South | 2 | 81 | 81 | 80 | 507 | 33440 | 29 | 27 | 9 | 5 |
| 19 | Kuwait | 2 | 74 | 98 | 52 | 205 | 88170 | 8 | 25 | 18 | 2 |
| 20 | Kyrgyzstan | 1 | 70 | 34 | 36 | 29 | 3070 | 27 | 35 | 28 | 7 |
| 21 | Laos | 1 | 68 | 34 | 50 | 29 | 4570 | 68 | 39 | 26 | 6 |
| 22 | Lebanon | 2 | 80 | 87 | 58 | 478 | 17390 | 9 | 29 | 13 | 4 |
| 23 | Malaysia | 2 | 75 | 71 | 49 | 91 | 22460 | 7 | 32 | 17 | 5 |
| 24 | Maldives | 2 | 74 | 41 | 35 | 1241 | 9890 | 9 | 31 | 23 | 3 |
| 25 | Mongolia | 1 | 67 | 68 | 55 | 2 | 8810 | 26 | 32 | 28 | 6 |
| 26 | Nepal | 1 | 68 | 17 | 50 | 184 | 2260 | 46 | 39 | 22 | 7 |
| 27 | Oman | 2 | 76 | 75 | 32 | 13 | 52170 | 9 | 25 | 20 | 2 |
| 28 | Pakistan | 1 | 65 | 35 | 35 | 244 | 4920 | 74 | 42 | 28 | 8 |
| 29 | Palestinian Territory | 2 | 73 | 83 | 53 | 731 | 4900 | 20 | 43 | 31 | 4 |
| **30** | **Philippines** | **1** | **69** | **63** | **55** | **334** | **7820** | **23** | **38** | **24** | **6** |
| 31 | Qatar | 2 | 78 | 100 | 38 | 206 | 123860 | 7 | 15 | 11 | 1 |
| 32 | Saudi Arabia | 2 | 74 | 81 | 24 | 14 | 53780 | 16 | 33 | 22 | 4 |
| 33 | Singapore | 2 | 83 | 100 | 62 | 8034 | 76850 | 20 | 27 | 9 | 5 |
| 34 | Sri Lanka | 2 | 74 | 15 | 68 | 315 | 9470 | 9 | 34 | 18 | 6 |
| 35 | Tajikistan | 1 | 67 | 26 | 28 | 58 | 2500 | 57 | 39 | 34 | 7 |
| 36 | Thailand | 2 | 75 | 47 | 79 | 129 | 13510 | 11 | 28 | 12 | 8 |
| 37 | Timor-Leste | 1 | 62 | 30 | 22 | 81 | 6410 | 63 | 47 | 37 | 10 |
| 38 | Turkey | 2 | 75 | 77 | 73 | 99 | 18760 | 10 | 33 | 16 | 5 |
| 39 | Turkmenistan | 1 | 65 | 47 | 48 | 11 | 12920 | 47 | 32 | 22 | 8 |
| 40 | United Arab Emirates | 2 | 77 | 83 | 28 | 113 | 58090 | 6 | 16 | 15 | 1 |
| 41 | Uzbekistan | 1 | 68 | 51 | 65 | 69 | 5340 | 44 | 32 | 22 | 5 |
| 42 | Vietnam | 2 | 73 | 32 | 76 | 273 | 5030 | 15 | 31 | 17 | 7 |
| 43 | Yemen | 1 | 63 | 29 | 34 | 49 | 3820 | 68 | 45 | 36 | 8 |