Journal of Statistics, Computing and Interdisciplinary Research
ISSN (P): 2707-7101
J. Stat. Comp. & Interdis. Rese.
Volume 2(1) (2020), 1-8
http://scir.wum.edu.pk/index.php/ojs

Factors Identification for Human Development Index

Anam Javaid[1] and Atif Akbar[2]
[1]Department of Mathematical sciences, University Sains Malaysia
[2]Department of Statistics, Bahauddin Zakariya UniversityMultan, Pakistan
Coresponding Auhors: anamjavaid7860@yahoo.com

Abstract

Human development index matters a lot for the economic condition of a country. It can be calculated with the help of the education index, health index, and GDP. So, by looking only at the value of the human development index (HDI), the economy of the country can be judged. The contribution of this research project is the identification of factors related to human development. The factors include the availability of the basic necessities, facilities related to education, health, and income. For this purpose, different important factors have been observed for the Multan district by taking MICS (2007-2008) survey data. Logistic regression is applied for the purpose of analysis. Significant factors are noted in each regression analysis based on the dependent factor.

Key words: Human developmentindex, Economic condition, Education health.

Introduction

Human development index able us to measure the development level of a country. The standard of living and the quality of life can be judge by it. Many researchers are working for the human development index (HDI) so that the quality of life for that country can be analyzed as Frey and Schneider (1979) analyzed the policy instrument that government used for making policy and they also indicated that which type of policy proposal accepted by the government. Akbar and Altaf (1995) analyzed the effect of different utility variables like electricity, gas, and water in the rent of the dwelling unit in the Karachi squatter arrangement. Peykarjou et al. (2011) studied the relationship between economic growth and health in Organization Islamic Conference members by using data from 2001-2009. Ahmad and Riaz (2012) estimated poverty in Pakistan after using data from 1974 -2009. Ajmair and Akhtar (2012) examined factors that effects on household consumption by using data of district Bhimber, AJK (Azad Jammu Kashmir). This present study is an attempt to finding determents of different variables likewise used in previous studies. This study is based on MICS (Multiple Indicator Cluster Survey) data of 2007-2008 for the Multan district.

One of the research was conducted by Luby and Rahbar (1999) as they examined the health facilities available to under five-year children. Later on, Glick and Sahn (2000) examined the multiple schooling indicators. Heltberg (2004) found the factors related to the fuel in the household. Javaid, A et al. (2018) reviewed the factors related to the human development index.

Many reserachers used different statistical techniques for finding the factors related to the human development index. This research project identified the significant factors related to HDI.

Methodology

Logistic Regression analysis

Rencher and Schaalje (2007) explained that in some regression situations, the response variable *y* has only two possible outcomes, for example, high blood pressure or low blood pressure, developing cancer of the esophagus or not developing it, whether a crime will be solved or not solved, and whether a bee specimen is a "killer" (Africanized) bee or a domestic honey bee. In such cases, the outcome y can be coded as 0 or 1 and we wish to predict the outcome (or the probability of the outcome) on the basis of one or more *x's*.To illustrate a linear model in which *y* is binary, consider the model with one *x*:

$$y_i = \beta_0 + \beta_1 x_1 + \epsilon_i; \qquad y_i = 0, 1, 2, \ldots, n. \tag{1}$$

Since y_i is 0 or 1, the mean $E(y_i)$ for each x_i becomes the proportion of observations at x_i for which $y_i = 1$. This can be expressed as

$$\begin{aligned} E(y_i) &= P(y_i = 1) = p_i, \\ 1 - E(y_i) &= P(y_i = 0) = 1 - p_i . \end{aligned} \tag{2}$$

The distribution $P\ (y_i = 0) = 1 - p_i$and $P(y_i = 1) = p_i$ in (3) is known as the Bernoulli distribution. By (1) and (2), we have

$$E\ (y_i) \quad = \quad p_i = \beta_0 + \beta_1 x_i. \qquad (3)$$

For the variance of y_i, we obtain

$$Var(y_i) \quad = \quad E[y_i - E(y_i)]^2$$
$$= \quad p_i (1 - p_i)\ (4)$$

By (3) and (4), we obtain

$$Var\ (y_i) \quad = \quad (\beta_0 + \beta_1 x_i)\ (1 - \beta_0 - \beta_1 x_i),$$

and the variance of eachy_i depends on the value of x_i. Thus the fundamental assumption of constant variance is violated, and the usual least-squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ computed. To obtain optimal estimators of b_0 and b_1, we could use generalized least squares estimators

$$\hat{\beta} \quad = \quad (X'V^{-1}X)^{-1}X'V^{-1}y$$

But there is an additional challenge in fitting the linear model(4). Since$E(y_i) = p_i$ is a probability, it is limited by $0 \leq p_i \leq 1$. If we fit (4) by generalized least squares to obtain

$$\hat{p}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i,$$

then$\hat{p}_i$ may be less than 0 or greater than 1 for some values of x_i. A model for $E\ (y_i)$ that is bounded between 0 and 1 and reaches 0 and 1 asymptotically (instead of linearly) would be more suitable. A popular choice is the logistic regression model.

$$p_i = E\ (yi) = \frac{e^{\beta_0 + \beta_1 x_1}}{1 + e^{\beta_0 + \beta_1 x_1}} = \frac{1}{1 + e^{-\beta_0 - \beta_1 x_1}} \qquad (5)$$

The model in *(5)* can be linearized by the simple transformation

$$ln(pi/(1 - pi)) \quad = \quad \beta_0 + \beta_1 x_1 \qquad (6)$$

Sometimes called the logit transformation.

The parameters β_0 and β_1 in (5) and (6) are typically estimated by the method of maximum likelihood. For a random sample $y_1, y_2, \ldots, y_n$ from the Bernoulli distribution with

$P(y_i = 0) = 1 - p_i$and $P(y_i = 1) = p_i$, the likelihood function becomes

$$L(\beta_0, \beta_1) = f(y_1, y_2, \ldots, y_n, \beta_0, \beta_1) = \prod_{i=1}^{n} f(y_i;\ \beta_0, \beta_1) = \prod_{i=1}^{n} p_i^{y_i}(1 - p_i)^{1-y_i} (7)$$

Taking the logarithm of both sides of *(7)* and using *(6)*, we obtain

$$\ln L(\beta_0, \beta_1) = \sum_{i=1}^{n} y_i(\beta_0 + \beta_1 x_1) - \sum_{i=1}^{n} \ln(1 + e^{\beta 0 + \beta 1 x1}) \quad (8)$$

Differentiating *(8)* with respect to β_0 and β_1 and setting the results equal to zero gives

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \frac{1}{1+ e^{-\beta_0-\beta_1 x_1}} \quad (9)$$

$$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} \frac{x_i}{1+e^{-\beta_0-\beta_1 x_1}} \quad (10)$$

These equations can be solved iteratively for $\hat{\beta}_0$and$\hat{\beta}_1$.

Results and Discussion

Data used in analysis was taken from MICS(2007-2008) with 51 variablescontaining 2892 observations. The results are analysed using the different dependent factors of HDI with the remaining independent factors. The coding are observed as for the factors related to HDI. The codings are observed in Table 1.

Table 1:*Variable names with their Notations*

S.No.	Variable Name	Notations	S.No.	Variable Name	Notations
1	Gender	A1	27	Bicycle	F20
2	Type of dwelling	A2	28	Motorcycle or scooter	F21
3	House Ownership	A3	29	Car or truck	F22
4	Gas	F2	30	Animal-driven cart	F23
5	Main material of roof	A5	31	House area (Marla)	C1
6	Mai material of wall	A6	32	House value (RS.)	C2
7	Television	F4	33	Own agriculture land	C3
8	Electricity	F1	34	Own livestock?	C4
9	Number of rooms for sleeping	A4	35	Received any pension last year?	C5
10	Radio/Tape recorder	F3	36	Cable TV	F5
11	Purchase goods from government utility stores	C6	37	Consider government utility stores as beneficial to a common man	C7

12	Type of fuel used for cooking	A7	38	Primary-Government. Boys	E1
13	Middle-Government. Boys	E5	39	Primary-Government. Girls	E2
14	Mobile Phone	F7	40	Primary-Private .Boys	E3
15	Computer	F8	41	Primary-Private. Girls	E4
16	Internet Connection	F9	42	Telephone	F6
17	Secondary-Government. Boys	E9	43	Middle-Government. Girls	E6
18	Air- Conditioner	F11	44	Middle-Private. Boys	E7
19	Washing Machine/Dryer	F12	45	Middle-Private. Girls	E8
20	Air cooler or fan	F13	46	Refrigerator/Freezer	F10
21	Cooking range/Microwave	F14	47	Secondary-Government. Girls	E10
22	Secondary-Private. Girls	E12	48	Secondary-Private. Boys	E11
23	Iron	F16	49	Stitching Machine	F15
24	Distance to nearest health facility (in minutes)	H2	50	Type of nearest health facility (government. or Private)	H1
25	Donkey Pump or turbine	F18	51	Water Filter	F17
26	Watch	F19			

In this research, the factors related to education are kept as the dependent factors. Using the logistic regression analysis. Results are noted in terms of the significant factors only.

$$ln\left(\frac{E1}{1-E1}\right) = 0.8808\ E_2 - 0.2651\ E_6 - 0.4532E_9 - 0.3653E_{11} + 0.4085E_{12}$$

$$ln\left(\frac{E2}{1-E2}\right) = -1.1912A_1 - 1.9286E_1 - 0.4685E_5 - 0.7496E_7 - 0.8257E_{10} - 1.2034E_{12}$$

$ln(\frac{E3}{1-E3})$ = $-0.6580E_4$

$ln\left(\frac{E4}{1-E4}\right)$ = $-0.6501E_3$

$ln(\frac{E5}{1-E5})$ = $-0.2408E_2 -1.2849E_6 -0.3878E_7 + 0.5595E_8 - 1.2964E_9 + 0.8859E_{10} + 0.6695E_{11} - 0.7418E_{12}$

$ln\left(\frac{E6}{1-E6}\right)$= $-0.4779E_1 + 0.3348E_2 - 1.2909E_5 + 1.3245E_9 -1.5580E_{10} + 0.2520E_{12} -0.2691H_1 + 0.2106H_2$

$ln\left(\frac{E7}{1-E7}\right)$ = $-0.6394E_8 + 0.2956E_{11} + 0.2936E_{12}$

$ln\left(\frac{E8}{1-E8}\right)$ = $-0.6389E_7 -0.2000E_9 + 0.1814E_{10} + 0.2691E_{11} -0.2738E_{12}$

$ln\left(\frac{E9}{1-E9}\right)$ = $0.4219F_2 -0.3705E_1 + 0.1807E_2 - 0.6265E_5 + 0.4315E_6 + 0.6455E_7 - 0.6837E_8 - 2.0560E_{10}$

$ln\left(\frac{E10}{1-E10}\right)$ = $-0.1555E_2 + 0.2378E_5 -0.4847E_6 -0.9142E_7 + 0.9218E_8 + 1.9146E_9$

$ln\left(\frac{E11}{1-E11}\right)$ = $- 0.5612E_7 + 0.5416E_8 - 0.5947E_{12}$

$ln\left(\frac{E12}{1-E12}\right)$ = $-0.2158F_2 + 0.1241E_1 -0.1419E_2 + 0.4999E_7 -0.5934E_{11}$

By observing the different education factors, it is observed that for the education analysis, the average distance of school, gender, family structure, biological child, number of sisters in a family, number of brother in a family, occupation father, education father, education mother, house wealth, rural area, average distance of school, pupil child ratio, pupil teacher's ratio, %age of men with a white collar job, % age of women with a white collar job, district development index, national GDP per capita, distance to nearest health facility, consider government utility store as beneficial to a common man, telephone, own agriculture land, animal driven cart and air conditioner are main significant factors for education. *i.e*. these all above variables have main effect on education.

For the health facilities, logistic regression is analysed using the health factors as dependent variable. The results are noted.

$ln\left(\frac{H1}{1-H1}\right)$ = $-4.6837H_2$

$ln\left(\frac{H2}{1-H2}\right)$ = $-1.6511 F_{18} – 7.6416 H_1$

In analysis for health facility,the important significant factors are distance to nearest health facility in minutes, education, age cohort, income, and respiratory problem.

For the house ownership, significant factor is identified using the logistic regression analysis.

$$ln\left(\frac{A3}{1-A3}\right) = -1.4080A_2$$

For house ownership, significant variables were type of dwelling with solid cement, dwelling with wall finish, index of quality of structure, number of floors, dwelling in neighborhoods, utility variables, wall, rooms and secondary government boys are main factors. *i.e.* for house ownership all above variables are important.

Conclusion

From the above result, important factors for different educational, health, and for other used dependent variable can be seen so its mean the above significant factors plays an important role in development of that dependent factor. The government policies can also be made for increment of HDI by focusing on these kind of factors. For the future work, multivariate analysis can be done for this kind of dataset.

Acknowledgments

The author would like to thank the Dr.AtifAkbar for the support and guidance in this research project.

Acknowledgments

The authors would like to thank the Editor for their guidline.

References

Ahmad, K. and Riaz, A. (2012). An Econometric Model of Poverty in Pakistan: ADRL Approach to Co-integration. *Asian Journal of Business and Management sciences*; 1(3),75-84.

Ajmair, M. and Akhtar, M. (2012).Household Consumption in Pakistan (A Case Study of District Bhimber, AJK).*European Journal of Scientific Research*. ISSN 1450-216X Vol.75 No.3, 448-457.

Akbar, M. and Altaf, A. (1995). A Rent Hedonic Model of the low Income Housing Market in Karachi. *Pakistan Economic and Social Review*; 33(1/2), 79-90.

Frey, B.S. and Schneider, F. (1979).An Econometric Model with an Endogenous Government Sector.*Springer* 34, 29-43.

Glick, P., and Sahn, D. (2000). Gender and education impacts on Employment and Earnings in a Developing Country: the Case of Guinea. *Economic Development and Cultural Change;* 45(4), 793–823.

Heltberg, R. (2004). Fuel switching: Evidence from Eight Developing Countries. *Energy Economics;* 26, 869-887.

Javaid, A., Akbar, A., & Nawaz, S. (2018). A Review on Human Development Index. *Pakistan Journal of Humanities and Social Sciences*, 6(3), 357-369.

Noorali, R. Luby, S. and Rahbar, M. H. (1999). Does Use of a Government Service Depend on Distance from the Health Facility?.*Health Policy and Planning*; 14(2), 191-197

Peykarjou.Gollu, R. B. and Gashti, H.P. (2011) Studying the Relationship Between Health and Economic Growth in OICMember states. 3(8), 1-14.

Rencher, A. C., &Schaalje, G. B. (2007). *Linear models in statistics*. John Wiley & Sons.