_____

A Statistical Assessment of the Role of Community's Learning and Living Environment to Determine Literacy Rate

Tooba Khan[1] and Nighat Zahra[2]

Abstract

A community (tehsil/town) level study has been carried out by analyzing data extracted from a household survey (tehsil-based Multiple Indicator Cluster Survey (MICS) 2007-08) of Punjab, Pakistan. The objectives of this study are to identify those community characteristics that significantly affect literacy rate and to identify differences and similarities among communities on the basis of factors significantly related to literacy rate. Average literacy rate in the province of Punjab is 59.32% with a standard deviation of 13.48. Statistical techniques of Multiple Linear Regression, Principal Component Analysis, Principal Components Regression and Cluster Analysis have been used. Characteristics identified to be significantly related to literacy rate are pre-school attendance, one-year overage for grade, primary school net attendance ratio, many years overage for grade, access to public primary school at nearest distance, access to public middle school at nearest distance, child disability, child labor, availability of health facilities at nearest distance, rudimentary floor of house, rudimentary roof of house, rudimentary walls of house, utilities at home, possessions, improved drinking water and adequate sanitation, large family size and receiving cash donations.

Keywords

Community, Environment, Literacy rate, Multicollinearity, Principal components regression

_____

[1] Department of Statistics, GC University, Lahore, Pakistan.
 Email: toobakhan1333@gmail.com
[2] Department of Statistics, GC University, Lahore, Pakistan.
 Email: nighat.zahra@hotmail.com

_____

1. Introduction

Literacy is often conventionally defined as ability to read and write at a specified age. UNESCO proclaimed 8th September as International Literacy Day and the objective of celebrating this day is to highlight and emphasize the significance of literacy among individuals, communities and societies. The period 2003-2012 has been declared as United Nations Literacy Decade (UNLD) by the General Assembly of the United Nations. This decade aims to promote literacy environment for all, to enhance levels of literacy and to empower individuals and communities in both developed and developing countries. During this decade, initiatives also include targeting most marginal or poorest social groups to combat poverty. Pakistan is at a serious risk of being unable to achieve Education For All (EFA) literacy goal by the year 2015 (UNESCO, 2006). Currently, according to Pakistan Social and Living Standards Measurement Survey (Government of Pakistan, 2013), literacy rate of population aged 10-years and above, in the province of Punjab is 60% which was the same in 2010-11. The MICS 2007-08 (Government of Punjab, 2009) reported the highest literacy rate in Samanabad Town (88.4%) of Lahore district and the lowest in Rojhan tehsil (16.7%) of Rajanpur district. In today's modern, challenging, rapidly changing and globalized world, literacy is not only seen as a phenomenon at individual level but is also considered as a contextual and societal phenomenon (UNESCO, 2006). Home and community factors play a pivotal role to determine literacy rate which may be considered to be a touchstone of educational outcomes.

There are complex and multifaceted factors which affect school attendance. Individual, household and community characteristics are all very important determinants of school attendance (Baschieri and Falkingham, 2009). Participation of youth in schooling and work is influenced by community where they live. Moreover, for household decisions, community-level factors are very important. In addition to poverty alone, barriers and opportunities should also be focused in the surrounding community to enhance education (Durrant, 1998). Learning- and teaching-focused policies should be combined with student-, family- and community-focused policies to lessen the influence of poverty on educational achievement (Egan, 2013). Increasing trend in attainment difference was explored between attenders and non-attenders of pre-school, and finally both groups are on their starkly distinct ways (Berlinski et al., 2008). Primary and secondary school enrolment, income and life expectancy (indicating importance of health for education) were found to be significant determinants of literacy rate (Fadiya, 2010). Effects of community level and household level factors on

education were investigated. Family composition, child characteristics, household welfare (material of walls, number of rooms in house, ownership of dwelling house, poverty status and total land managed by household) and many other variables turned out to be important determinants of primary school enrolment. Distance to school and to source of water and cognitive ability were also other important determinants of enrolment (Kabubo-Mariara and Mwabu, 2007). Results revealed that partly due to unavailability of schools at nearest distance and of adequate quality, and partly due to direct or indirect influence of lack of resources, children were not attending school (Kakwani et al., 2006). A part of the data revealed that students who entered at the age of 6-years in grade 1 were achieving significantly higher than 5-years-old entrants (Davis et al., 1980). It was concluded that there is more child labor, reduced schooling and more overage students in classrooms (age-grade distortion) due to having a large number of younger siblings (Patrinos and Psacharopoulos, 1997). As compared to non-working children, educational attainment of working children declined by 2-years of schooling. Child labor and grade repetition were found to be closely associated (Psacharopoulos, 1997). After making adjustments for various other variables associated with educational attainment, children's percentage entering lower or upper secondary school and family size were found to be strongly inversely associated with each other (Knodel and Wongsith, 1991).

The intended objectives of this study are to
- Study the role of community's learning and living environment to determine literacy rate.
- Build a suitable model for identification of those community characteristics that significantly affect literacy rate.
- Identify differences and similarities among communities (tehsils/towns) on the basis of factors significantly related to literacy rate.
- Suggest the government of Punjab about various actions that must be taken for further advancement of literacy rate in the province.

2. Materials and methods

Household survey data of Punjab, Pakistan has been collected for this study. For analysis, certain variables of interest have been extracted from tehsil-based Multiple Indicator Cluster Survey (MICS) Punjab 2007-08 (Government of Punjab, 2009). MICS 2007-08 was conducted in the province of Punjab, Pakistan by the Bureau of Statistics (BOS), Planning and Development Department,

_____

Government of the Punjab, Pakistan in the period from December 2007 to April 2008. Total sample size for this survey was 91,280 households. The Punjab province is administratively divided into 9 divisions, 35 districts and 143 tehsils/towns and the survey provided information at the provincial, divisional, district and tehsil level. The sample was selected in two stages. In the first stage, within each of the sampling domains, by using Probability Proportional to Size sampling, enumeration areas (villages in rural areas and enumeration blocks in urban areas) were selected. In the second stage, within each randomly selected enumeration area, listing of households was carried out and Systematic Random samples (16 households in each rural area and 12 households in each urban area) were selected.

Diagnostics for developing and evaluating the Multiple Linear Regression Model have been performed. Principal Components Regression has been used to combat the problems resulting due to Multicollinearity. So, Principal Component Analysis has been performed to extract the components representing the structure of the independent variables to identify which of the independent variables make up which components so that these new components can be further analyzed. Hierarchical Cluster Analysis has been carried out by using Euclidean distance method for determining distance between clusters and the furthest neighbor (or complete linkage) method for clustering.

3. Results

Data of 143 tehsils/towns of Punjab, Pakistan has been statistically analyzed by using SPSS version 18.0 for Windows and E-Views 6, and conclusions have been derived.

Multiple Linear Regression has been performed by simultaneously entering all of the independent variables. The overall Multiple Linear Regression model has an F-ratio of 74.525, Mean Square Error of 16.775 and a significance level of 0.000 (Table 1). Hence, the overall model is significant.

Coefficient of Multiple Determination, R^2, indicates that 92% of the variability in literacy rate has been explained by its linear relationship with regressors pre-school attendance, primary school net attendance ratio, rudimentary floor of house and possessions. Adjusted R^2 that is a most useful measure for assessing success of the built model is 0.908 indicating that the Multiple Linear Regression model has accounted for 90.8% of the variance in literacy rate (Table 2).

_____

Out of 19 regressors, four regressors only (pre-school attendance, primary school net attendance ratio, rudimentary floor of house, and possessions) have turned out to be statistically significant, for their p-values are small (Table 3).

For testing whether or not the Regression model is mis-specified, Ramsey's RESET test has been carried out. The F-statistic is 0.392546 which is statistically insignificant, for the p-value is 0.5321 suggesting that the Multiple Linear Regression model is not mis-specified. Residuals obtained from Multiple Linear Regression model have been tested for normality. The Jarque-Bera statistic is 2.382726, and under the normality assumption, the probability of obtaining such a statistic is approximately 30%. There are no Outliers and influential cases in the data as none of the leverage values exceeds 0.5 indicating that no observation can be regarded as very high leverage (minimum leverage value = 0.027, maximum leverage value = 0.425) and all of the Cook's distance measures are below the cutoff of unity (minimum Cook's distance = 0.000, maximum Cook's distance = 0.381). Park Test (t = 1.014, p = 0.312), Glejser Test (t = 1.324, p = 0.188) and Koenker-Bassett (KB) Test (t = 0.936, p = 0.351) have been carried out to detect the presence of heteroscedasticity. The p-values of the three tests point to the conclusion that the residuals are homoscedastic that is, they have constant variance. As overall F-value of the Multiple Linear Regression model is significant, R^2 is very high, but a few t-ratios are statistically significant, these are all pointers to the presence of Multicollinearity among the explanatory variables. Values of Tolerance and Variance Inflation Factor have been presented in Table 4. Values of the Variance Inflation Factor of primary school net attendance ratio, access to private primary school at nearest distance and utilities at home are greater than 10 indicating that these variables are highly collinear with the remaining regressors (Table 4). Values of Tolerance closer to zero indicate higher degree of Collinearity of that regressor with the rest of the regressors. Examination of the off-diagonal elements of the correlation matrix also reveals significant Bi-variate correlations among most of the explanatory variables suggesting near linear dependence between pairs of regressors (Table 5).

Multicollinearity diagnostics have put forward the fact that Multicollinearity exists among the independent variables. Therefore, the Ordinary Least Square (OLS) Estimates, which have been obtained, are Best Linear Unbiased Estimates (BLUE) but with inflated standard errors.

To obtain Regression coefficients with great precision or accuracy, correlated independent variables have been transformed to uncorrelated components so that

_____

Regression may be run on these new variables called Component Scores. Four components have been extracted from a set of 19 original variables, Pearson's Correlation Coefficients have been computed between literacy rate and each of the Principal Components where first and third Principal Component have been found to be significantly linearly correlated with literacy rate at 1% level of significance. Standardized literacy rate has been regressed on first and third Principal Component. The Principal Components Regression Model has an F-value of 362.673, Mean Square Error of 0.164 and a significance level of 0.000. Therefore, the overall Principal Components Regression Model is significant (Table 6).

Coefficient of Multiple Determination, R^2, indicates that 83.7% of the variability in literacy rate has been explained by its linear relationship with the two Principal Components. It is clear from adjusted R^2 that the Principal Components Regression model has accounted for 83.5% of the variance in literacy rate (Table 7). Standardized Principal Components Regression model is,
$$Z(LR) = 0.842(PC1) + 0.357(PC3)$$

As constant in the Multiple Linear Regression model is the expected value of literacy rate (a measure of educational outcomes) when all of the explanatory variables in the model are assumed to have zero values, but in this study, each variable under analysis has values for the entire community (tehsil/town), therefore, it is totally unreasonable to assume zero value for any explanatory variable included in the model. So, it has no practical interpretation.

Results show that the variables having positive influence on literacy rate are: pre-school attendance, one-year overage for grade, primary school net attendance ratio, access to public primary school at nearest distance, access to public middle school at nearest distance, child disability, availability of health facilities at nearest distance, utilities at home, possessions, and improved drinking water and adequate sanitation whereas those having negative influence on literacy rate are: many years overage for grade, child labor, rudimentary floor of house, rudimentary roof of house, rudimentary walls of house, large family size and receiving cash donations (Table 8).

In order to evidence differences among tehsils/towns of Punjab province on the basis of first and third Principal Components which have been found to be significantly correlated with literacy rate, Hierarchical Cluster Analysis has been carried out. Cluster Analysis has provided a solution with eleven Clusters which

_____

have been presented in Table 9. Case index of each tehsil/town has been displayed and the district to which each tehsil/town belongs has also been shown. Serial number of each district has been presented in the Appendix. Tehsils/towns lying in each cluster are homogeneous with respect to "prosperity and development" (PC1) and "economical characteristics" (PC3).

Literacy rate of each tehsil/town has been plotted in Figure 1 and all of the plotted points have been labeled with case index of each tehsil/town. As there is homogeneity among tehsils/towns in each cluster with respect to determinants of literacy rate, similarly, literacy rates of tehsils/towns falling in each cluster are close enough reaffirming the pattern that emerged from Cluster Analysis.

Mean literacy rate of cluster 4 is lowest (44.4%) whereas it is highest (80.6%) in cluster 9. With respect to mean literacy rate, clusters have been arranged in descending order as given below:

Clusters: High $\underline{9, 2, 8, 11, 10, 5, 3, 6, 1, 7, 4}$ Low

4. Discussion

Significant relation of pre-school attendance with literacy rate implies that communities (tehsils/towns) where there is increased percentage of 3-4 years old children currently attending pre-school education, literacy rates in those tehsils/towns are also high as due to learning environment, there is motivation and curiosity for learning that induce school readiness among children. Studies of Berlinski et al. (2008) and Daniels (1995) support this conclusion.

Influence of 1-year overage for grade is positive and significant demonstrating that literacy rate increases as percentage of students who are 1-year overage for grade (primary school entry age = 6-years) increases revealing the fact that there is more tendency of sending children to primary schools at the age of 6-years instead of 5-years which is an official entry age into grade 1 in Pakistan. The work of Davis et al. (1980) supports this conclusion.

Positive significant relation of primary school net attendance ratio with literacy rate illustrates that rise in primary school net attendance ratio produces increase in literacy rate that can be considered to be a measure of educational outcomes. The work of Fadiya (2010) in Nigeria also supports this conclusion.

_____

Those tehsils/towns have lower literacy rates where there is increased percentage of those students who are many years overage for grade due to dropping out, grade repetition etc. The work of Patrinos and Psacharopoulos (1997) in Peru supports this conclusion.

Tehsils/towns where higher percentage of households has access to public primary and public middle schools at nearest distance, and health facilities are also available at nearest distance, their literacy rates are high as well. Studies of Kabubo-Mariara and Mwabu (2007) in Kenya, Kakwani et al. (2006) in African countries, and Knodel and Wongsith (1991) in Thailand support this conclusion.

Positive impact of child disability on literacy rate has appeared highlighting the fact that those tehsils/towns where there is higher percentage of those children who are 2-9 years old with at least one reported disability (difficulties in understanding, learning, speaking, seeing, hearing, moving, sitting, standing or walking and any kind of mental disability) have higher literacy rates as well. The work of Eisenberg et al. (1998) supports this conclusion. Previous research findings suggest that children with disabilities may also positively affect the personality of their siblings. In this case, siblings of the children with disabilities are far more responsible and due to an increased opportunity of care giving, helping and teaching their younger handicapped brothers and sisters, they choose those professions which are linked to provision of services and education. Severe retardation surely adversely affects all of the family members.

Reduced literacy rate has been observed in those tehsils/towns where child labor is high. The work of Psacharopoulos (1997) in Latin America supports this conclusion.

Where families are residing in tehsils/towns having houses with rudimentary floor, rudimentary roofs and rudimentary walls, being low income earners, lower literacy rates have been found there. The findings of Kabubo-Mariara and Mwabu (2007) in Kenya also support this conclusion.

Literacy rates of those tehsils/towns are higher where larger percentage of household members have more than three utilities (Electricity, gas, telephone, mobile phone, computer, internet, radio, television, cable T.V., washing machine, stitching machine, turbine or donkey pump, water filter, iron, fridge/freezer, microwave/cooking range, fan/cooler and air conditioner) and also where there is high percentage of household members in possession of at least one of the

_____

bicycle, scooter/motorcycle, car or any other vehicle, watch and animal drawn cart. The findings of Anh et al. (1998) in Vietnam support this conclusion. Advanced literacy rates have been observed in those tehsils/towns where improved sources of drinking water have been provided to households and all of them are using improved sanitary means for disposing of excreta. Hence, due to safe and clean environment, individuals are less likely to suffer from diseases. He work of Durrant (1998) in Pakistan supports this conclusion.

As percentage of households with number of household members 10 or more increases, children in these houses do not find sufficient area for learning practices as houses are often overcrowded due to large family sizes, so literacy rates decline in those tehsils/towns as those households cannot afford schooling expenses of all of their wards. Studies of Knodel and Wongsith (1991) in Thailand support this conclusion.

This study has also described that with an increment in the percentage of households who received cash donations (Zakat or in other forms) during the year preceding the survey year, literacy rate declines. It means that households who are receiving cash donations are ultra-poor and they make use of the received amount only on their basic necessities of life (food, clothing etc.) and do not receive enough amounts that may be spent on schooling of their children. It also appears that the percentage of households who received cash donations is low as small number of people give cash donations or pay Zakat.

On the basis of Principal Components significantly related with literacy rate, tehsils/towns have been divided into eleven clusters by means of cluster analysis. These eleven clusters have been ordered according to their learning and living environment and the clusters which have been assigned lower ranks are more inclined towards social exclusion.

5. Conclusion

The present study or analysis, which has been carried out in the United Nations Literacy Decade (UNLD), has delineated poverty as a significant determinant of literacy rate. Pre-school attendance, one-year overage for grade, primary school net attendance ratio, many years overage for grade, access to public primary school at nearest distance, access to public middle school at nearest distance, child disability, child labor, availability of health facilities at nearest distance,

_____

rudimentary floor of house, rudimentary roof of house, rudimentary walls of house, utilities at home, possessions, improved drinking water and adequate sanitation, large family size and receiving cash donations are the characteristics that have been identified to be significantly related to literacy rate.

6. Recommendations

The government of Punjab must make sure that Zakat is distributed regularly by all the rich and well-heeled people to the deserving people. If government makes policies in such a way that abundant wealth of well-to-do people can be distributed among the needy then substantial differences regarding development of communities can be observed. This money may also help the impoverished people to alleviate their poverty and they may also utilize this received amount on schooling of their children. Therefore, the use of this Islamic institution should be made in an intelligent way, so that the country can be changed into a true welfare state.

Although, in around two-thirds of the 204 countries and territories, official primary school entry age is 6-years, nonetheless, living in a developing country like Pakistan, awareness needs to be imparted among parents about the fact that one-year late entry into grade 1 will also result in one-year late entry into the job market.

In the MICS which will be conducted in the future, questions regarding children's satisfaction with schools or teachers, teachers behavior, fellow students behavior etc. should also be considered so that these can also be taken into account and their influences examined on literacy rate.

Various techniques can be used to combat the problems resulting due to Multicollinearity. The proposed general approaches include collection of additional data, re-specifying the model and using other estimation methods such as Ridge Regression and Principal Components Regression. Moreover, by considering Principal Axis Factoring as the extraction method for Factor Analysis with VARIMAX rotation, Factor Scores can also be derived and used in Multiple Regression. Besides, other approaches such as Robust Regression and Regression Tree Method may also be used in further studies.

Tehsils/towns which appear in the clusters that have been ranked as low must be paid more attention. Effective policies and strategies must be made for the

development of these tehsils/towns. In today's technology-driven world, for empowering role of literacy and setting education as the first priority, strong efforts (efficient literacy policies, strategies and programs) are required in order to transform vicious cycles of poverty and illiteracy into virtuous cycles of health, education and development.

Table 1: Analysis of Variance (ANOVA) (MLR)

Source of Variation	Sum of Squares (SS)	d.f.	Mean Square (MS)	Computed *F*	p-value
Regression	23753.361	19	1250.177	74.525	.000
Residual	2063.358	123	16.775		
Total	25816.719	142			

Constructed by the authors

Table 2: Model summary (MLR)

R	R Square	Adjusted R Square	Standard Error of the Estimate
.959	.920	.908	4.0958

Constructed by the authors

Table 3: Coefficients (MLR)

Variable	Unstandardized Coefficient B	Unstandardized Coefficient Standard Error	Standardized Coefficient Beta	*t*-statistic	p-value
Constant	-11.766	14.583		-.807	.421
PRE_SA	.349	.076	.201	4.625	.000
OVERAGE_1Y	-.069	.066	-.067	-1.050	.296
NAR	.523	.099	.472	5.261	.000
OVERAGE_MY	.119	.081	.056	1.468	.145
PRIM_SD_G	-.164	.095	-.093	-1.723	.087
MID_SD_G	.057	.045	.073	1.257	.211
PRIM_SD_P	.092	.052	.146	1.790	.076
MID_SD_P	-.065	.044	-.118	-1.487	.140
CH_DIS	.169	.165	.030	1.028	.306
CH_LABOUR	-.067	.119	-.019	-.559	.577
HEALTH_F	.057	.034	.077	1.697	.092
R_FLOOR	-.188	.045	-.311	-4.147	.000
R_ROOF	.049	.056	.053	.877	.382
R_WALLS	.045	.051	.061	.884	.379
UTILITY	.114	.110	.102	1.035	.303
POSSESS	.385	.168	.134	2.297	.023
ISW_SED	-.048	.049	-.063	-.981	.329
F_SIZE_L	.221	.122	.061	1.811	.073
ZAKAT	-.823	.417	-.059	-1.971	.051

Constructed by the authors

_____

Table 4: Collinearity statistics

Variable	Tolerance	VIF
PRE_SA	.346	2.893
OVERAGE_1Y	.158	6.342
NAR	.081	12.410
OVERAGE_MY	.453	2.209
PRIM_SD_G	.221	4.520
MID_SD_G	.191	5.230
PRIM_SD_P	.097	10.283
MID_SD_P	.103	9.710
CH_DIS	.752	1.329
CH_LABOUR	.545	1.833
HEALTH_F	.313	3.196
R_FLOOR	.116	8.636
R_ROOF	.176	5.681
R_WALLS	.137	7.310
UTILITY	.067	14.858
POSSESS	.192	5.216
ISW_SED	.156	6.423
F_SIZE_L	.581	1.721
ZAKAT	.717	1.395

Constructed by the authors

Table 5: Correlation matrix

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}
X_1	1	.485**	.579**	.134	.413**	.652**	.426**	.452**	.058	-.269**	.511**	-.766**	-.705**	-.637**	.646**	.558**	.691**	-.065	.206*
X_2	.485**	1	.891**	.283**	.439**	.566**	.491**	.461**	-.016	-.429**	.420**	-.649**	-.615**	-.549**	.593**	.640**	.517**	-.356**	.058
X_3	.579**	.891**	1	.424**	.641**	.701**	.620**	.578**	-.042	-.438**	.532**	-.748**	-.742**	-.722**	.789**	.763**	.670**	-.349**	.158
X_4	.134	.283**	.424**	1	.518**	.264**	.458**	.392**	-.100	-.221**	.295**	-.161	-.348**	-.455**	.498**	.473**	.311**	.040	.308**
X_5	.413**	.439**	.641**	.518**	1	.654**	.665**	.617**	-.148	-.233**	.596**	-.470**	-.614**	-.684**	.821**	.665**	.631**	.017	.248**
X_6	.652**	.566**	.701**	.264**	.654**	1	.702**	.704**	-.138	-.444**	.693**	-.778**	-.708**	-.767**	.795**	.734**	.818**	-.147	.206*
X_7	.426**	.491**	.620**	.458**	.665**	.702**	1	.934**	-.140	-.496**	.626**	-.601**	-.646**	-.740**	.731**	.758**	.607**	-.223**	.238**
X_8	.452**	.461**	.578**	.392**	.617**	.704**	.934**	1	-.163	-.470**	.629**	-.616**	-.649**	-.751**	.704**	.729**	.583**	-.191*	.204*
X_9	.058	-.016	-.042	-.100	-.148	-.138	-.140	-.163	1	.058	-.139	-.027	-.028	.063	-.116	-.091	-.140	.126	.218**
X_{10}	-.269**	-.429**	-.438**	-.221**	-.233**	-.444**	-.496**	-.470**	.058	1	-.260**	.454**	.359**	.352**	-.316**	-.506**	-.277**	.268**	-.117
X_{11}	.511**	.420**	.532**	.295**	.596**	.693**	.626**	.629**	-.139	-.260**	1	-.604**	-.638**	-.762**	.727**	.692**	.720**	.049	.194*
X_{12}	-.766**	-.649**	-.748**	-.161	-.470**	-.778**	-.601**	-.616**	-.027	.454**	-.604**	1	.826**	.789**	-.762**	-.733**	-.775**	.257**	-.158
X_{13}	-.705**	-.615**	-.742**	-.348**	-.614**	-.708**	-.646**	-.649**	-.028	.359**	-.638**	.826**	1	.789**	-.835**	-.711**	-.709**	.145	-.297**
X_{14}	-.637**	-.549**	-.722**	-.455**	-.684**	-.767**	-.740**	-.751**	.063	.352**	-.762**	.789**	.789**	1	-.878**	-.796**	-.796**	.092	-.254**
X_{15}	.646**	.593**	.789**	.498**	.821**	.795**	.731**	.704**	-.116	-.316**	.727**	-.762**	-.835**	-.878**	1	.827**	.846**	-.105	.257**
X_{16}	.558**	.640**	.763**	.473**	.665**	.734**	.758**	.729**	-.091	-.506**	.692**	-.733**	-.711**	-.796**	.827**	1	.735**	-.213*	.221**
X_{17}	.691**	.517**	.670**	.311**	.631**	.818**	.607**	.583**	-.140	-.277**	.720**	-.775**	-.709**	-.796**	.846**	.735**	1	-.029	.135
X_{18}	-.065	-.356**	-.349**	.040	.017	-.147	-.223**	-.191*	.126	.268**	.049	.257**	.145	.092	-.105	-.213*	-.029	1	-.036
X_{19}	.206*	.058	.158	.308**	.248**	.206*	.238**	.204*	.218**	-.117	.194*	-.158	-.297**	-.254**	.257**	.221**	.135	-.036	1

**. Correlation is significant at the 1% level of significance (2-tailed).*. Correlation is significant at the 5% level of significance (2-tailed)

Table 6: Analysis of Variance (ANOVA) (PCR)

Source of Variation	Sum of Squares (SS)	d.f.	Mean Square (MS)	Computed *F*	p-value
Regression	118.889	2	59.445	362.673	.000
Residual	23.111	141	.164		
Total	142.000	143			

Constructed by the authors

Table 7: Model summary (PCR)

R	R Square	Adjusted R Square	Standard Error of the Estimate
.915	.837	.835	.40485408

Constructed by the authors

Table 8: Coefficients of original variables (PCR)

Variable	Coefficient	Standard Error	*t*-statistic	p-value
Constant	32.825	1.4419	22.764	.000
PRE_SA	.285	.0075	38.146	.000
OVERAGE_1Y	.129	.0065	19.732	.000
NAR	.115	.0098	11.729	.000
OVERAGE_MY	-.314	.0080	-39.319	.000
PRIM_SD_G	.028	.0094	2.991	.003
MID_SD_G	.080	.0045	18.023	.000
PRIM_SD_P	-.004	.0051	-.817	.416
MID_SD_P	.004	.0044	.963	.337
CH_DIS	.157	.0163	9.606	.000
CH_LABOUR	-.117	.0118	-9.937	.000
HEALTH_F	.039	.0033	11.796	.000
R_FLOOR	-.105	.0045	-23.366	.000
R_ROOF	-.094	.0055	-16.944	.000
R_WALLS	-.047	.0051	-9.220	.000
UTILITY	.065	.0108	5.950	.000
POSSESS	.152	.0166	9.186	.000
ISW_SED	.085	.0048	17.751	.000
F_SIZE_L	-.271	.0121	-22.413	.000
ZAKAT	-1.342	.0413	-32.520	.000

Constructed by the authors

_____

Table 9: Clusters

Cluster 1			Cluster 2		
Case Index	**Tehsil/Town**	**District**	**Case Index**	**Tehsil/Town**	**District**
1	Ahmedpur East	1	2	Bahawalpur City	1
4	Hasilpur	1	31	Faislabad Town	8
7	Bahawalnagar	2	33	Iqbal Town	8
9	Fort Abbas	2	50	Gujrat	12
11	Minchinabad	2	51	Kharian	12
16	DG Khan	4	52	Sara-e-Alamgir	12
19	Karor Lal Esan	5	55	Mandi Bahauddin	14
21	Alipur	6	59	Shakargarh	15
22	Jatoi	6	60	Daska	16
25	Jampur	7	63	Sialkot	16
35	Tandlianwala Town	8	67	Allama Iqbal Town	17
36	Chinniot	9	79	Shah Kot	19
39	Ahmadpur Sial	9	108	Gujjar Khan Town	28
74	Kasur	18	125	Dina	31
92	Kabirwala	22	126	Jhelum	31
93	Mian Channu	22	**Cluster 4**		
96	Kehror Pacca	23	5	Khairpur Tamewali	1
103	Arifwala	26	17	Taunsa	4
130	Kot Momin	32	20	Layyah	5
140	Noorpur Thal	34	23	Kot Addu	6
141	Essa Khel	35	24	Muzaffargarh	6
143	Piplan	35	26	Rajanpur	7
Cluster 3			27	Rojhan	7
3	Bahawalpur Sadar	1	37	Jhang	9
28	Chak Jhumra Town	8	38	Shorkot	9
34	Sumundri Town	8	89	Jalalpur Pirwala Town	21
54	Pindi Bhattian	13	104	Pakpattan	26
56	Malakwal	14	106	Depalpur	27
57	Phalia	14	133	Shahpur	32
94	Jahanian	22	138	Mankera	33
95	Dunya Pur	23	**Cluster 6**		
98	Burewala	24	12	Khanpur	3
99	Mailsi	24	13	Liaqatpur	3
100	Vehari	24	14	RY Khan	3
131	Sahiwal	32	15	Sadiqabad	3
132	Sargodha	32	41	Kamalia	10
135	Bhakkar	33	47	Nowshera Virkan Town	11
142	Mianwali	35	75	Chunian	18
Cluster 5			76	Pattoki	18
6	Yazman	1	82	Muridke	20

Case Index	Tehsil/Town	District	Case Index	Tehsil/Town	District
8	Chishtian	2	**Cluster 8**		
10	Haroonabad	2	30	Jinnah Town	8
29	Jaranwala Town	8	32	Madina Town	8
40	Gojra	10	43	Aroop Town	11
42	TT Singh	10	44	Kamoke Town	11
53	Hafizabad	13	45	Khiali Shahpur Town	11
58	Narowal	15	46	Nandipur Town	11
77	Nankana Sahib	19	48	Qila Didar Singh Town	11
78	Safdarabad	19	49	Wazirabad Town	11
80	Sangla Hill	19	68	Nishtar Town	17
81	Ferozewala	20	69	Ravi Town	17
83	Sharaqpur Sharif	20	71	Wahga Town	17
84	Sheikhupura	20	**Cluster 9**		
91	Khanewal	22	61	Pasrur	16
97	Lodhran	23	64	Lahore Cantt	17
101	Chichawatni	25	65	Aziz Bhatti Town	17
102	Sahiwal	25	66	Data Ganj Bakhsh Town	17
105	Okara	27	70	Shalimar Town	17
107	Renala Khurd	27	72	Samanabad Town	17
129	Bhalwal	32	73	Gulberg Town	17
134	Sillanwali	32	112	Rawal Town	28
139	Khushab	34	115	Taxila Town	28
Cluster 7			122	Chakwal	30
18	Choubara	5	**Cluster 10**		
90	Shujabad Town	21	62	Sumbrial	16
136	Darya Khan	33	86	Mumtazabad Town	21
137	Kallur Kot	33	87	Shah Rukan-e-Alam Town	21
Cluster 11			88	Sher Shah Town	21
85	Boson Town	21	113	Potohar Town	28
109	Kahuta Town	28	114	Kallar Sayaddan Town	28
110	Kotli Sattian Town	28	116	Attock	29
111	Murree Town	28	117	Fateh Jang	29
120	Jand	29	118	Hasanabdal	29
121	Pindigheb	29	119	Hazro	29
			123	Choa Saidan Shah	30
			124	Talagang	30
			127	Pind Dadan Khan	31
			128	Sohawa	31

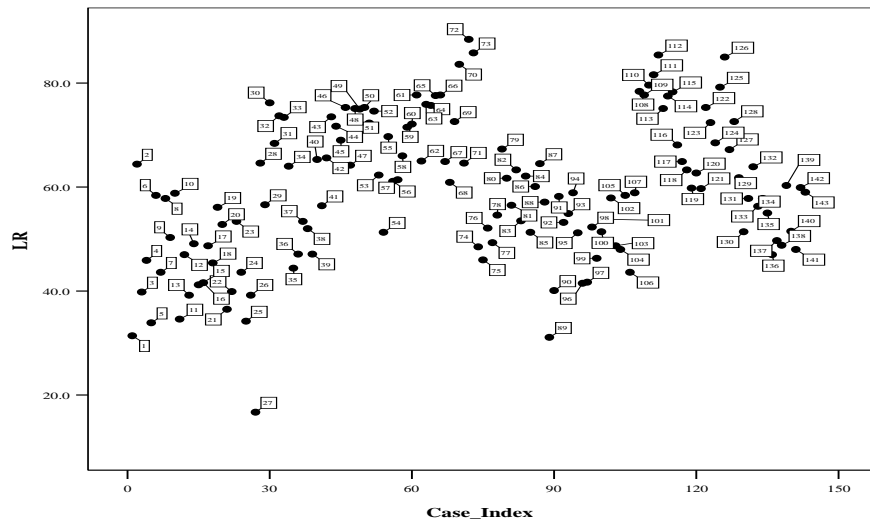Constructed by the authors

Figure 1: Literacy rate

References

1. Anh, T. S., Knodel, J., Lam, D. and Friedman, J. (1998). Family size and children's education in Vietnam. *Demography*, **35(1)**, 57-70.
2. Baschieri, A. and Falkingham, J. (2009). Staying in school: Assessing the role of access, availability, and economic opportunities: The case of Tajikistan. *Population, Space and Place*, **15(3)**, 205-224.
3. Berlinski, S., Galiani, S. and Manacorda, M. (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of Public Economics*, **92(5-6)**, 1416-1440.
4. Daniels, S. (1995). Can pre-school education affect children's achievement in primary school? *Oxford Review of Education*, **21(2)**, 163-178.
5. Davis, B.G., Trimble, C.S. and Vincent, D. R. (1980). Does age of entrance affect school achievement? *The Elementary School Journal*, **89(3)**, 133-143.
6. Durrant, V. L. (1998). Community influences on schooling and work activity of youth in Pakistan. *The Pakistan Development Review*, **37(4)** Part II, 915-937.
7. Egan, D. (2013). Poverty and low educational achievement in Wales: Student, family and community interventions. York: Joseph Rowntree Foundation. Retrieved from: http://www.jrf.org.uk/sites/files/jrf/wales-education-poverty-summary.pdf

---

8. Eisenberg, L., Baker, B. L. and Blacher, J. (1998). Siblings of children with mental retardation living at home or in residential placement. *Journal of Child Psychology and Psychiatry*, **39(3)**, 355-363.

9. Fadiya, B. B. (2010). Determinants of educational outcomes in Nigeria (1975-2008). *European Journal of Social Sciences*, **15(4)**, 501-511.

10. Government of Pakistan (2011). Pakistan Social and Living Standards Measurement Survey (PSLM 2010-11). Statistics Division, Federal Bureau of Statistics, Islamabad, Pakistan.

11. Government of Pakistan (2013). Pakistan Social and Living Standards Measurement Survey (PSLM 2011-12). Statistics Division, Pakistan Bureau of Statistics, Islamabad, Pakistan.

12. Government of Punjab (2009). Multiple Indicator Cluster Survey (MICS) PUNJAB 2007-08. Planning and Development Department, Bureau of Statistics.

13. Kabubo-Mariara, J. and Mwabu, D.K. (2007). Determinants of school enrolment and education attainment: Empirical evidence from Kenya. *South African Journal of Economics*, **75(3)**, 572-593.

14. Kakwani, N., Soares, F. and Son, H. H. (2006). Cash transfers for school-age children in African countries: Simulation of impacts on poverty and school attendance. *Development Policy Review*, **24(5)**, 553-569.

15. Knodel, J. and Wongsith, M. (1991). Family size and children's education in Thailand: Evidence from a national sample. *Demography*, **28(1)**, 119-131.

16. Patrinos, H. A. and Psacharopoulos, G. (1997). Family size, schooling and child labor in Peru: An empirical analysis. *Journal of Population Economics*, **10(4)**, 387-405.

17. Psacharopoulos, G. (1997). Child labor versus educational attainment: Some evidence from Latin America. *Journal of Population Economics*, **10(4)**, 377-386.

18. UNESCO (2006). Education For All (EFA) Global Monitoring Report 2006, Literacy for Life. Retrieved from:
http://unesdoc.unesco.org/images/0014/001416/141639e.pdf

_____

Appendix

Districts

Serial no.	District	Serial no.	District
1	Bahawalpur	19	Nankana Sahib
2	Bahawalnagar	20	Sheikhupura
3	Rahim Yar Khan	21	Multan
4	Dera Ghazi Khan	22	Khanewal
5	Layyah	23	Lodhran
6	Muzaffargarh	24	Vehari
7	Rajanpur	25	Sahiwal
8	Faisalabad	26	Pakpattan
9	Jhang	27	Okara
10	Toba Tek Singh	28	Rawalpindi
11	Gujranwala	29	Attock
12	Gujrat	30	Chakwal
13	Hafizabad	31	Jhelum
14	Mandi Bahauddin	32	Sargodha
15	Narowal	33	Bhakkar
16	Sialkot	34	Khushab
17	Lahore	35	Mianwali
18	Kasur		