

## A STATISTICAL STUDY OF FACTORS AFFECTING URBAN FEMALE LITERACY IN MAHARASHTRA

Anjum Ara M K Ahmad

Associate Professor, Dept. of Mathematics & Statistics  
Rizvi College of Arts, Science & Commerce, Mumbai

### ABSTRACT

The study will provide an appraisal of the factors that have a vital influence on the female literacy in the State of Maharashtra in urban areas. Eleven factors are used in the study of the variables affecting the female literacy rate in the urban data of Maharashtra. Principal Component Analysis method is used in the study. It is observed that four Principal Components have Eigen values more than one with total variance explained as 87.878 percent. A Varimax rotation method with Kaiser Normalization is used so that variable loads highly on a single factor and has small to moderate loadings on the remaining. Depending on the loading of a variable in Rotated Component Matrix the variables are categorized into four components.

**KEYWORDS:** Urban Female Literacy, Maharashtra, Principal Component Analysis

### 1. INTRODUCTION

The effects of female schooling are particularly important for policy-makers because promoting girls' education is a central development strategy [5]. Female literacy is promoted as a policy to increase household productivity by reducing fertility and improving child health, as well as a strategy to build the labour force [6]

Economic growth was expected to increase human capital in developing countries through investments in education and expanded labour markets. While economic growth and female literacy have mostly contributed to reduced fertility and improved child health, they have not consistently resulted in expansion of labour market opportunities for women [4]

Literacy can affect people's lives, both women's and men, through several channels. It increases access to knowledge, information and new ideas. It enhances overall efficiency, opportunities in the labour market and social status [1].

The study will provide an appraisal of the factors that have a vital influence on the female literacy in the urban areas of State of Maharashtra.

### 2. OBJECTIVES OF THE STUDY

1. To identify the factors whose eigenvalues are more than one using Principal Component Analysis.
2. The KMO and Bartlett Test of Sphericity is used to find out whether the data is fit for factor analysis.
3. To categorise the factors into components depending on loading of a variable in Rotated Component Matrix.

### 3. RESEARCH METHODOLOGY

Data is collected relating to the factors affecting the female literacy in urban Maharashtra. The study has included the following data.

#### 3.1 COLLECTION OF DATA

In order to cover the above objective in mind, there was a need to collect data from various Govt. agencies about the various facilities available at the district level. The study also collected data of various factors affecting female literacy at the district level.

The data available in Census (2001) and Census (2011) for the State of Maharashtra are used in the study. The district level mean child ever born (MCB) and female work participation rates (FWPR) using the Census (2011) data is calculated and included them as dependent variables. Also part of Census (2011), district level information is available on the literacy rates, urbanization; these are used as independent variables. The NFHS-I (1992-93) and NFHS-II (1998-99) reported that fertility was higher among rural, less educated women and Muslim in the period 1990-92 and 1996-98. As data is available on religious heterogeneity (i.e. the proportion of population belonging to the various religions) is given for Census (2001) only, this data is also used in the study.

The NFHS-I (1992-93) and NFHS-II (1998-99) gives the number and the type of medical facilities available at district level. The latest DLHS (2007-08) data is used to develop a medical composite index. This index is one of the variables in the study.

As data is available in the Census (2011) on percentage of population owning different assets like radio, transistor, television, telephone, bicycle, scooter, motor cycle, moped, car-jeep-van. Also the Indian Government has for long been using electronic and other mass media to promote family planning, child health and women empowerment hence the ownership of television and radio in a large segment of the population can possibly have a salutary impact on the female literacy rate. An attempt is made to study the effect of these assets on the dependent variable.

According to National Human Development Report 2001 the available Census data permits analysis of two aspects of quality of housing and shelter namely, living space or the number of rooms available to a household and the quality of construction of the residence i.e. whether a household resides in a pucca or a kutchha construction. The Census (2011) also presents data on quality of houses based on the material used for construction of walls and roof separately. Such information can be used to identify whether the house can be classified as kutchha, pucca or semi-pucca. There are also large inter-district variations in the availability of electricity, access to toilet facilities and safe drinking water at the household level, both in urban and in rural areas. An attempt will be made to use these factors in the Study.

#### 3.2 ANALYSIS OF DATA

Various Statistical Techniques like Correlation, Factor Analysis, and Principal Component Analysis is used in the study to identify the important factors affecting female literacy in Maharashtra. The Economic Wealth Index and the Medical Index is created using Principal Component Analysis. Various softwares like Excel, SPSS are applied in the Study.

##### 3.2.1 PRINCIPAL COMPONENTS ANALYSIS

PCA is a multivariate statistical technique used to reduce the number of variables in a data set into a smaller number of 'dimensions'. In mathematical terms, from an initial set of  $n$  correlated

variables, PCA creates uncorrelated indices or components, where each component is a linear weighted combination of the initial variables. For example, from a set of variables  $X_1$  through to  $X_n$ ,

$$PC_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1n}X_n$$

$$\vdots$$

$$PC_m = a_{m1}X_1 + a_{m2}X_2 + \dots + a_{mn}X_n$$

where  $a_{mn}$  represents the weight for the  $m^{\text{th}}$  principal component and the  $n^{\text{th}}$  variable, i.e. in brief if there are  $n$  correlated variables  $X_1, \dots, X_n$ , each principal component (PC) is the sum of each variable multiplied by its weight (the weight for each variable is different in each principal component).

The weights for each principal component are given by the eigenvectors of the correlation matrix, or if the original data are standardized, the co-variance matrix. The variance ( $\lambda$ ) for each principal component is given by the Eigen value of the corresponding eigenvector. The components are ordered so that the first component ( $PC_1$ ) explains the largest possible amount of variation in the original data. The second component ( $PC_2$ ) is completely uncorrelated with the first component, and explains additional but less variation than the first component. Subsequent components are uncorrelated with previous components; therefore, each component captures an additional dimension in the data, while explaining smaller and smaller proportions of the variation of the original variables. The higher the degree of correlation among the original variables in the data, the fewer components required to capture common information.

Before applying the PCA following tests are necessary:

**a) Kaiser-Meyer-Olkin Measure of Sampling Adequacy:** This measure varies between 0 and 1, and values closer to 1 are better. A value of 0.5 is a suggested minimum.

**b) Bartlett's Test of Sphericity:** This tests the null hypothesis stating that the correlation matrix is an identity matrix. An identity matrix is one in which all of the diagonal elements are 1 and all off diagonal elements are 0. This null hypothesis should be rejected.

Taken together, these tests provide a minimum standard which should be passed before principal components analysis (or factor analysis) should be conducted.

#### 4.1 FACTOR ANALYSIS IN URBAN MAHARASHTRA

Eleven factors as given below are considered to study the Factors affecting the female literacy rate in Urban Maharashtra. The respective Mean and Std. Deviation of the Factors are given below.

**Table 4.1: Descriptive Statistics of Factors in Urban Maharashtra**

Sr.No.	Variables	Factors	Mean	Std. Deviation
1	MLR	Male Literacy Rate	89.7380	4.99967
2	MWPR	Male Work Participation Rate	54.8410	3.15581
3	FWPR	Female Work Participation	30.6201	12.47140
4	SEXRATIO	Sex Ratio	947.96	42.791
5	HINDUS	Percentage of Hindus	78.8300	10.47183
6	MUSLIM	Percentage of Muslims	11.7626	8.48844
7	SC	Scheduled Caste	12.7575	4.75579
8	ST	Scheduled Tribe	9.5849	13.46229
9	MNB	Mean Number of Births	2.9682	.32880
10	EWI	Economic Wealth Index	-.0420	.75942
11	IMR	Infant Mortality Rate	51.71	15.067

Source: Calculated with help of SPSS, data taken from Census

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.			.615
Bartlett's Test of Sphericity	Approx. Chi-Square		1219.425
	df		55
	Sig.		.000

Source: Derived using SPSS

The Factor Analysis is run in the urban data of all the districts of Maharashtra for the selected factors The KMO and Bartlett Test of Sphericity (Table 4.1) has given the Measure of Sampling Adequacy as 0.615. Also the Bartlett's Test of Sphericity is significant hence concluding that the data is fit for principal component analysis.

	Initial	Extraction
Male Literacy Rate	1.000	.865
Male Work Participation Rate	1.000	.872
Female Work Participation	1.000	.906
Sex Ratio	1.000	.894
Percentage of Hindus	1.000	.835
Percentage of Muslims	1.000	.946
Scheduled Caste	1.000	.884
Scheduled Tribe	1.000	.787
Mean Number of Births	1.000	.928
Economic Wealth Index	1.000	.931
Infant Mortality Rate	1.000	.818

Extraction Method: Principal Component Analysis.

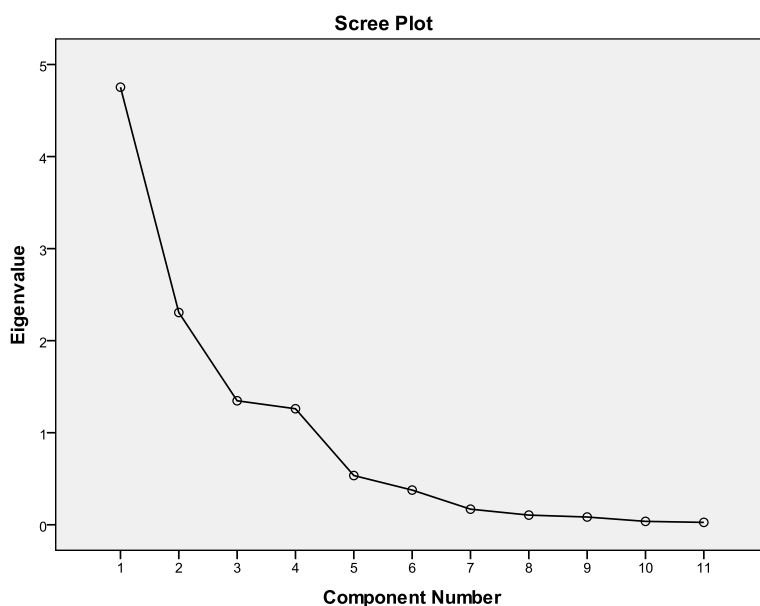
Source: Derived using SPSS

From the Communalities Table (Table 4.3), it is observed that all the factors are important.

Comp	Initial Eigen values			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cum %	Total	% of Variance	Cum %	Total	% of Variance	Cum %
1	4.754	43.214	43.214	4.754	43.214	43.214	3.339	30.352	30.352
2	2.306	20.961	64.175	2.306	20.961	64.175	3.153	28.666	59.018
3	1.347	12.246	76.421	1.347	12.246	76.421	1.833	16.662	75.681
4	1.260	11.457	87.878	1.260	11.457	87.878	1.342	12.197	87.878
5	.535	4.862	92.740						
6	.377	3.427	96.167						
7	.170	1.541	97.709						
8	.105	.956	98.664						
9	.084	.765	99.429						
10	.037	.336	99.765						
11	.026	.235	100.000						

Extraction Method: Principal Component Analysis.

Source: Derived using SPSS

**Figure 4.1: Scree Plot of Urban Data**

From Table 4.4 it is observed that Four Principal Components are extracted with Eigen values more than one. These four principal components explain 87.878 percent variance. Also the Scree Plot (Figure 4.1) shows that four principal components are sufficient for the data.

**Table 4.5: Rotated Component Matrix<sup>a</sup>**

	Component			
	1	2	3	4
Male Literacy Rate	<b>-.634</b>	-.244	.632	.063
Male Work Participation Rate	-.041	<b>.884</b>	-.045	-.294
Female Work Participation	<b>.608</b>	.712	-.143	.096
Sex Ratio	.064	.095	-.058	<b>.937</b>
Percentage of Hindus	.044	<b>.779</b>	-.328	.344
Percentage of Muslims	-.108	<b>-.926</b>	.105	-.258
Scheduled Caste	.395	-.060	<b>.843</b>	-.117
Scheduled Tribe	.427	.240	<b>-.739</b>	.003
Mean Number of Births	<b>.947</b>	-.174	-.006	-.013
Economic Wealth Index	<b>-.872</b>	-.116	-.109	-.382
Infant Mortality Rate	<b>.743</b>	.476	-.140	-.142
Extraction Method: Principal Component Analysis.				
Rotation Method: Varimax with Kaiser Normalization.				
a. Rotation converged in 5 iterations.				

Source: Calculated using SPSS

The four components are divided into four groups of factors based on the maximum loadings using the Rotated Component Matrix (Table 4.5). These groups are given headings as mentioned in Table 4.5.

**Table 4.6: Classification of Factors in Rural Maharashtra**

Component 1	Component 2	Component 3	Component 4
Fertility, Economic Condition, Infant Mortality and Female Work Participation	Religion & Male Work Participation	SC & ST	Sex Ratio
( percent of Variance is 43.214)	( percent of Variance is 20.961)	( percent of Variance is 12.246)	( percent of Variance is 11.457)
Mean Number of Births (0.947)	Percentage of Muslims (-0.926)	Scheduled Caste (0.843)	Sex Ratio (0.937)
Economic Wealth Index (-0.872)	Male Work Participation Rate (0.818)	Scheduled Tribe (-0.739)	
Infant Mortality Rate (0.743)	Percentage of Hindus ( 0.779)		
Male Literacy Rate (-0.634)			
Female Work Participation (0.608)			

Source: Derived from Table 4.5

#### 4. CONCLUSION

The Factor Analysis was run in the combine data of all the districts of Maharashtra. Eleven variables are used to study the effect of female literacy in Maharashtra. The KMO and Bartlett Test of Sphericity are significant and concluded that the data is fit for principal component analysis. In the Communalities Table it is observed that all the Factors are important. Four Principal Components have Eigen values more than one with total variance explained as 87.878 percent. A Varimax rotation method with Kaiser Normalization is used so that variable loads highly on a single factor and has small to moderate loadings on the remaining.

The variables are categorized into four components (Table 4.6) depending on the loading of a variable in Rotated Component Matrix. Factors Fertility, Economic Condition, Infant Mortality and Female Work Participation has maximum loading on the first component. Variables Religion & Male Work Participation has maximum loading on the second component. SC & ST has maximum loading on the third component and Sex Ratio has maximum loading on the fourth component. Hence these factors are classified into four categories Component 1, 2, 3 and 4.

**ACKNOWLEDGEMENT:** I acknowledge UGC for sanctioning my minor research project on which this paper is based.

## REFERENCES

- [1] Cleland, John and Jejeebhoy Shireen, "Maternal Schooling and Fertility: Evidence from Censuses and Surveys", in *Girl's Schooling, Autonomy and Fertility Change in South Asia*. Roger Jeffery and Alaka M. Basu (eds.). Thousand Oaks, C.A.: Sage Publications. : pp. 72-106, 1996.
- [2] Government of Maharashtra, "Economic Survey of Maharashtra, 2010-11," Directorate of Economics & Statistics, Planning Department, Mumbai, 2011.
- [3] Gujarati Damodar N. , "Basic Econometrics", Second Edition, 1988.
- [4] Lincove, Jane Arnold, "Growth, Girls' Education, and Female Labor: A Longitudinal Analysis", *The Journal of Developing Areas*, 41, 2: pp. 45-68, 2008.
- [5] Schultz, T. P., "Why Governments Should Invest More to Educate Girls", *World Development*, 30, 2: pp. 207-225, 2002.
- [6] Summers, L., "Investing in All the People: Educating Women in Developing Countries", EDI Working Papers (No. 45), The World Bank, Washington DC, 1994.