Probability and Statistics – II (2009): Group Assignment

Group 12, Section A

Group Members (alphabetical order):

Aarthy Sridhar, Kamlesh Chauhan, Mohit Garg, Prateek Sharma, R Tamizhmarai, Simit Batra

9mohitg@iimahd.ernet.in

Indian Institute of Management, Ahmedabad, India

1. Introduction

We identified the following five phenomena around us and analyzed those using different statistical techniques.

- *Ragging in India*

  We approached CURE (Coalition to Uproot Ragging from Education), an NGO working against ragging and obtained data on cases regarding various aspects of ragging and analyzed the data for proportions using *two-phase clustered sampling*.

- *Relation between literacy rate and expenditure on education*

  We analyzed the relationship between government expenditure on education and its impact on the literacy rates in the country using data on literacy rates as per the census data and data on the total expenditure on education as a percentage of the total government expenditure. *Regression* analysis was used.

- *Relation between HDI and importance to religion*

  We analyzed the data pertaining to the human development indices and religious importance values of different countries using *correlation coefficients and Goodman-Kruskal coefficient*.

- *Relation between inflow and outflow of the people from states*

  We analyzed data pertaining to inflow and outflow of migrants among states in a particular year using *Goodman-Kruskal coefficient*.

- *Validity of assumptions of stock market mathematical models*

  It has always been difficult to fit the variation in the prices in an established mathematical model. However, a lot of mathematical models assume the returns to be derived from a probability distribution like Normal or Lognormal. We apply Chi-Square goodness of fit tests to determine the validity of the assumption.

Table of Contents

1. Ragging in India

In order to study the ragging cases in India, we divide India geographically into clusters, with each state forming a cluster. Out of the total population of 35 States and Union Territories, we consider the major 27. Out of each of these 27 states, we pick out a sample of ragging cases from major newspapers. Thus, we have employed *two-phase clustered sampling.*

The data was obtained courtesy, Coalition to Uproot Ragging from Education (CURE)[1] and a processed version is presented in Exhibit 1

Methodology

The notations used are as follows:

N = number of clusters in population = 35

n = number of clusters selected in a simple random sample = 27

$\mathbf{M_i}$ = number of elements in cluster i, i = 1,2,3, ….. ,N

$\mathbf{m_i}$ = number of elements selected from cluster i

M = ΣM_i = number of elements in population

$\mathbf{y_{ij}}$ = value of jth unit in ith cluster

Let $\mathbf{p_i}$ = proportion of sampled elements from cluster i

$$p_i = \frac{1}{m_i} \sum_{i=1}^{n} y_{ij}$$

The estimate of population proportion = **p**

$$p = \frac{\sum_{i=1}^{n} M_i p_i}{\sum_{i=1}^{n} M_i}$$

[1] CURE is India's foremost anti-ragging non-profit organisation – http://www.noragging.com

$$p = \frac{\sum_{i=1}^{n} \mathrm{M_i p_i}}{M}$$

$$p = \sum_{i=1}^{n} \frac{M_i}{M} p_i$$

We assumed,

$$\frac{M_i}{M} = \frac{m_i}{m}$$

$$Var(p) = \frac{N^2}{M^2} \frac{N-n}{N} \frac{{s_p}^2}{n} + \frac{N}{nM^2} \sum_{i=1}^{n} {M_i}^2 \left(\frac{M_i - m_i}{M_i}\right) \left(\frac{p_i q_i}{m_i}\right)$$

$$s_p^2 = \frac{1}{n-1} \sum_{i=1}^{n} (M_i p_i - M_i p)^2$$

$$\frac{s_p^2}{M^2} = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{M_i}{M}\right)^2 (p_i - p)^2$$

$$\boldsymbol{Var(p) = N \frac{N-n}{n} \frac{{s_p}^2}{M^2} + \frac{N}{n} \sum_{i=1}^{n} \left(\frac{M_i}{M}\right)^2 \left(\frac{M_i - m_i}{M_i}\right) \left(\frac{p_i q_i}{m_i}\right)}$$

Let **σ** = Standard Deviation of p

$$\boldsymbol{\sigma = \sqrt{Var(p)}}$$

Confidence interval of p is

$$\boldsymbol{p_i \pm 1.96\sigma}$$

We found the proportion of total cases in India based on the following parameters:

1. Cases involving physical abuse
2. Cases involving sexual abuse

3. Cases involving verbal abuse
4. Cases involving injuries
5. Cases involving mental torture
6. Cases resulting in attempted suicide
7. Cases resulting in deaths
8. Cases resulting in student leaving college
9. Cases involving violent behavior
10. Cases resulting in punishments
11. Cases in Medical college
12. Cases in Engineering college
13. Cases in Other colleges

Results and Analysis

For each of these parameters, we calculated the population proportion and population standard deviation from the states data using two-phase clustered sampling. Thereafter, the confidence interval for 95% confidence was calculated and a graph plotting the proportions of different parameters was created (Table 1 and Table 2).

			Confidence Interval	
Parameters	**Proportion**	**Variance**	**Lower Limit**	**Upper Limit**
Physical	0.3727	0.001454	0.29796417	0.44744265
Sexual	0.1969	0.000691	0.14533716	0.248363626
Verbal	0.1076	0.000392	0.06878262	0.146440481
Injuries	0.2231	0.000814	0.16715991	0.279034316
Mental	0.0945	0.000382	0.05620141	0.132774964
Attemp. suicide	0.0446	0.000209	0.01625504	0.072983803
Deaths	0.1286	0.000518	0.08399259	0.173225257
Left coll	0.0630	0.000264	0.03112005	0.094864206
Violent behavior	0.0499	0.000212	0.02136108	0.078376454

Punishments	0.2835	0.000976	0.22222577	0.344703368
Medical	0.1627	0.000604	0.11454282	0.210916495
Engineering	0.2992	0.001213	0.23093981	0.367485386
Other	0.2913	0.000883	0.23310296	0.349574204

Table 1: Proportion Interval Estimates

The following is a graph of the proportion of ragging cases vs. the different parameters.

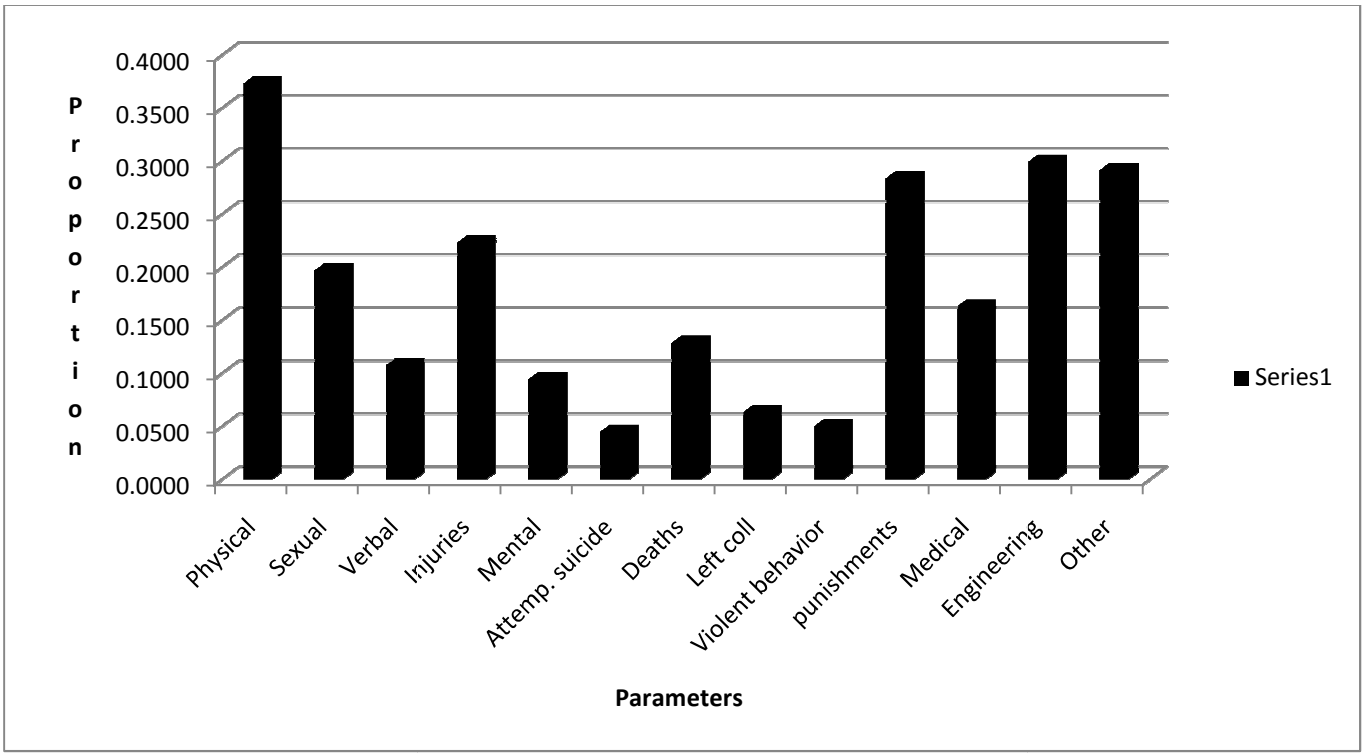

Table 2: Proportion Point Estimates

Conclusion

The conclusions from the analysis are as follows:

- It is clear from the analysis that more than $1/3^{rd}$ cases of ragging involve physical abuse while close to $1/5^{th}$ are sexual in nature.
- Another interesting statistic is that engineering and medical colleges together account for over 45% cases in the country. This strengthens the belief that professional colleges are badly hit due to the menace of ragging.

- Another startling statistic is that 10% to 24% of the cases either lead to deaths or attempted suicides. This has severe implications for the national education system since this implies that ragging as a phenomenon hampers the educational environment.

This has implications for policy decisions at the national level as well as college level. Moreover, law enforcement authorities need to be aware of this data to be able to come up with effective mechanisms to deal with ragging.

2. Literacy Rates vs. Education Expenditure

The education level of a country's population is a parameter in the socio-economic development index. In order to improve the education level, the government spends a large amount of money each year under several programmes and schemes to fulfill various target objectives. One such objective is the national literacy rate. The literacy rate is defined as "The total percentage of the population of an area at a particular time aged seven years or above who can read and write with understanding. Here the denominator is the population aged seven years or more" [(1)].

Methodology

In order for the government to determine and allocate expenditure on education, it is important to understand its impact on the various target objectives. In this study, we attempt to analyze the relationship between government expenditure on education and its impact on the literacy rates in the country. We collected data on the male, female and total literacy rates as per the census data from 1951 to 2001 (Exhibit 2). We also collected data on the total expenditure on education as a percentage of the total government expenditure per year from 1951-52 to 2004-05 (Exhibit 3).

Since the change in the literacy rate over a given time period would depend on the expenditure on education over the period, we used linear regression as a means of estimating the relationship between expenditure and literacy rates. The literacy data was available only for five censuses and hence, we could not estimate a regression for more than three independent variables. However, the data on education expenditure was available for each of the ten years between two consecutive censuses. Thus, we reduced the number of independent variables from ten to three.

We followed a simple reduction technique in which we took the independent variables as the arithmetic average of first four years, next three years and the last three years of data (Table 3).

Results and Analysis

We used the *linest(…)* [2] function in Microsoft Excel to estimate the regression coefficients. We did not use the constant term in the regression since we the change in literacy rate should be zero (ignoring contribution by population increase) if the expenditure is zero. The estimation results are provided in Table 4.

Year	Literacy Rates			Change in Literacy Rates			Expenditure % Total Budget		
	Male	*Female*	*Person*	*ΔMale*	*ΔFemale*	*ΔPerson*	*Years 1 to 4*	*Years 5 to 7*	*Years 8 to 10*
1951	27.16	8.86	18.33						
1961	40.4	15.34	28.31	13.24	6.48	9.98	8.7525	10.91	11.54
1971	45.95	21.97	34.45	5.55	6.63	6.14	9.9425	9.976667	9.716667
1981	56.37	29.75	43.56	10.42	7.78	9.11	10.0175	10.62333	10.98667
1991	64.13	39.29	52.21	7.76	9.54	8.65	10.9225	12.50667	13.36333
2001	79.56	54.28	65.38	15.43	14.99	13.17	13.045	13.14	14.34

Table 3: Literacy Rates and Education Expenditure

		ΔMale	*ΔFemale*	*ΔPerson*
Coefficients	Years 1 to 4	0.0156 ± 2.2124%	2.1670 ± 0.2405%	0.3886 ± 0.9883%
	Years 5 to 7	-3.3437 ± 5.2694%	-5.6116 ± 0.5729%	-2.4115 ± 2.3540%
	Years 8 to 10	4.0415 ± 4.2613%	4.2040 ± 0.4633%	2.7394 ± 1.9036%
Goodness of Fit (r^2)		0.9454	0.9991	0.9858
Standard Error (s_y)		4.09037	0.4447	1.8273
Fischer F-Statistic		11.54566 (90% < Confidence < 95%)	778.2823 (Confidence > 99%)	46.1122 (98.5% < Confidence < 99%)
Degrees of Freedom		2	2	2
Regression ss		579.5167	461.7918	461.8858

Residual ss	33.46231	0.3956	6.6777

Table 4: Regression Statistics

The results for the change in male literacy rate (ΔMale) do not seem to be linearly dependent on the independent variables. The Fischer F-Statistic gives a low measure of confidence – somewhere in between 90% and 95%. The change in female literacy rate (ΔFemale), however, has an excellent match with the regression variables with the Fischer F-Statistic giving a measure of confidence of over 99%. The overall literacy rate (ΔPerson) is a reasonable fit to the expenditure on education as a percentage of total expenditure. The Fischer F-Statistic measures the confidence of this fit to be in between 98.5% and 99%.

Conclusion

From our analysis, we could conclude that

- Literacy rate is positively correlated with expenditure in education as a percentage of total expenditure.
- Due to the excellent goodness of fit with the change in female literacy, it is likely that during the years we analyzed, the education expenditure was used to fund schemes focusing specifically on female literacy.

3. Relation between Human Development Index and Importance to religion

Religion has always been a topic of intense debate and discussion around the world. In recent times, its importance has grown manifold and it wields greater power and influence on world politics and the social well being of nations. It is common belief that religion is good for the general welfare of the society and for the social well-being of a nation. But, it becomes necessary to statistically verify this claim. Hence, we decided to study the correlation between the importance that a country and its people accord to religion and the human development index of the country.

Methodology

The following methodologies were adopted to analyze the data pertaining to the human development indices and religious importance values of different countries.

1. Correlation Coefficient:

   We calculated the correlation coefficient to estimate the magnitude and nature of correlation that existed between the two factors under study, namely, importance given to religion (r) and human development index (h). Additionally, we also calculated the correlation coefficient between importance given to religion (r) and average life expectancy in the country (l).
2. Goodman-Kruskal: We also estimated the Goodman-Kruskal coefficient γ using the following formula:

   $\gamma = (\Pi c - \Pi d) / (\Pi c + \Pi d)$

   Πc = Sum of concordant pairs

   Πd = Sum of discordant pairs

The data pertaining to HDI and life expectancy was sourced from the UNDP website's annual report. The countries were matched with the Gallup 2009 world survey on religious importance in various countries (Exhibit 4). Overall, 134 countries were studied. The human development index, as defined by the UNDP, captures three chief dimensions: health, education and GDP. This normalized measure is good implication on whether the country is developing, developed or underdeveloped. Importance to religion was the survey findings from Gallup 2009 world poll and measured the proportion of belief among the population.

Results and Analysis

The correlation coefficients and Goodman-Kruskal γ were found and tabulated in Table 5.

Factors	Correlation Coefficient	Goodman-Kruskal
HDI & Religion	-0.680900676	-0.872463768

Life & Religion	-0.577919351	-0.641783567

Table 5: Correlation and Goodman-Kruskal Measures

Conclusion

The negative correlation coefficients in both the cases (with HDI and with life expectancy) prove contrary to common beliefs. The implication is that, as the country and its people attach more importance to religion, the social wellbeing of the society as a whole reduces. This could be due to the detrimental effect that religion seems to have on the functioning of the social machinery. The Scandinavian countries like Sweden, and Norway, which are known for its liberal and irreligious society, scored high on the HDI ranking with high GDPs and life expectancies (Norway:2, Sweden: 7).

It can also be inferred that a poorer country and its people tended to be more religious. For instance, Niger had a high religious index of 97% and had a correspondingly low HDI of 0.374, ranking in the underdeveloped countries of the world.

4. Inter State Migration

Inter-state migration in India has been a continuous topic of discussion. There is a general perception that there is a mismatch in the number of people moving into and out of each state. In order to verify the above assumption, we collected data pertaining to inflow and outflow of migrants in a particular year. Using the data collected, we intend to check whether such an association exists between the number of people moving out and into state.

Methodology

Inflow and outflow data for 16 states was obtained to establish a relationship. The following methodologies were adopted to analyze the data pertaining to interstate migration across different states:

- Goodman-Kruskal:

  We calculated the Goodman-Kruskal coefficient γ to determine the association between inflow and outflow of migrants across states in India. Two factors used in the study were movement of population into and out of the state under consideration. The estimate was found using the following formula:

  $\gamma = (\Pi c - \Pi d) / (\Pi c + \Pi d)$

  Πc = Sum of concordant pairs

  Πd = Sum of discordant pairs

Results and Analysis

For the study, 15 major Indian states were selected while the others were classified under other/Union Territory category. The inflow and outflow data of migrants from each state to other 15 states was sourced from collected (Exhibit 5). Thereafter, total population moving out and into the state was tabulated (Table 6). The concordant and discordant pairs were then established using Excel.

Outflow/Inflow	0-10	10-20	20-30	30-40	40-50	50-60
0-10	3	2	1	0	0	0
10-20	0	5	1	0	0	0
20-30	0	0	0	0	0	1
30-40	1	0	0	0	0	0
40-50	0	1	0	0	0	0
50-60	0	0	0	1	0	0

Table 6: Outflow/Inflow table (all figures in '00s)

The Goodman-Kruskal estimate γ was found and tabulated as below (Table 7).

Factors	Goodman-Kruskal Estimate

Inflow and Outflow of Migrants	0.4492754

Table 7: Goodman-Kruskal Measure

Conclusion

The positive correlation between number of people moving out of state and number of people coming into state was established. This proves contrary to common belief regarding interstate migration. The implication is that, as more people migrate into a state, the number of people moving out of that state also increases. In comparison to the positive correlation established, Bihar and Maharashtra did not follow the above mentioned trend. In Bihar, total outflow exceeded inflow and in Maharashtra inflow was found to be more than outflow.

It can also be inferred that people from different states identify different needs and their decision to migrate is not just governed by economic factors. However this migration could be occurring among people belonging to different strata and not only the same socio-economic status.

5. Validating Stock Market Model Assumptions

Stock prices in the financial markets fluctuate with high unpredictability. It has always been difficult to fit the variation in the prices in an established mathematical model. However, a lot of mathematical models assume the returns to be derived from a probability distribution like Normal or Lognormal. We intend to identify the relationship between the distribution of the closing stock prices of Sensex and popular standard distribution methods:

1. Standard Normal Distribution
2. Lognormal Distribution

Methodology

Daily closing prices of BSE Sensex for the following durations were collected [3]. Due to the large number of samples in the raw data, we have omitted it from the exhibits.

- 2nd January, 2008 to 30th December, 2008 (to be referred to as data1)
- 2nd January, 2007 to 30th December, 2008 (to be referred to as data2)
- 2nd January, 2003 to 30th December, 2008 (to be referred to as data3)
- 2nd January, 1998 to 30th December, 2008 (to be referred to as data4)

In order to establish a relationship between sensex variations and Normal distribution or Lognormal Distribution, we can measure goodness of fit for each of the distributions and the sensex using chi-square test. If a sample of the sensex closing price does not fit the Normal or Lognormal distributions, it will be concluded that there is no certainty about the relationship between the sensex variation and the mathematical model under consideration.

Results and Analysis

Case 1:

In the first analysis, we investigate whether sensex follows Normal Distribution or not. The hypothesis is formulated as:

Let H_0: The population has a normal distribution

Let H_a: The population does not have a normal distribution

- We select a random sample such as data1 and calculate the sample mean and sample standard deviation.
- Intervals are then defined based on the confidence interval and the abscissa values of the Normal distribution such that the area between the successive abscissa is 5.
- We record the observed frequency (f_i) of the data1 values in each interval defined.
- Now, we calculate the expected number of occurrences e_i for each interval of values as defined above.

- The value of the test statistic (chisqv) is given by:

  chisqv = $\sum_{i=1}^{k} \frac{(fi-ei)^2}{ei}$

  Now the hypothesis will be rejected using p-value approach such that,

  Reject H_0 if p-value $\leq \alpha$

  Where α is the level of confidence and there are k – 2 - 1 degrees of freedom

This process is repeated for all the four data sets namely, data1, data2, data3 and data4 and observations are recorded.

Case 2:

We also investigated whether returns earned from the Sensex prices follow lognormal distribution or not. Data elements are defined as Ln(Sensex value for the previous day/current sensex value). Using these data values, the steps followed in case 1 are repeated as it is for this case as well.

The results of the tests are presented in Table 8.

	Normal Distribution		Lognormal Distribution	
Data 1	chisqv	129.4	chidist	39.6
	p-value	4.25846E-10	p-value	0.699389559
	result	reject H0	result	do not reject H0
	α	0.05	α	0.05
Data 2	chisqv	268.4	chisqv	146.2
	p-value	1.84047E-18	p-value	0.000582
	result	reject H0	result	reject H0
	α	0.05	α	0.05
Data 3	chisqv	1249.6	Chisqv	532.6
	p-value	7.5E-118	p-value	5.2775E-16
	result	reject H0	result	reject H0
	α	0.05	α	0.05
Data 4	chisqv	5968.4	chisqv	881.8
	p-value	0	p-value	7.24503E-19
	result	reject H0	result	reject H0
	α	0.05	α	0.05

Table 8: Chi-Square test results of stock market data

Conclusions

Since the hypothesis is being rejected for most of the data sets, we can conclude that the sensex closing prices as well sensex log return values do not follow either Lognormal or Normal Distribution.

Exhibits

Exhibit 1: Ragging Data [(4)]

STATE (Cluster)	**No. of cases (mi)**	**Physical**	**Sexual**	**Verbal**	**Injuries**	**Mental**	**Attemp. suicide**	**Deaths**	**Left coll**	**Violent behavior**	**punishments**	**Medical**	**Engineering**	**Other**
ANDHRA PRADESH	**29**	8	7	4	4	2	3	5	1	1	9	4	11	8
AP	**18**	4	3	1	2	3	2	6	1	1	5	2	8	5
ASSAM	**2**	1	1	0	1	0	0	0	0	0	1	0	1	1
BIHAR	**7**	2	0	0	4	0	1	1	0	1	1	2	0	2
CHATTISGARH	**6**	1	1	0	0	0	0	1	0	0	3	1	1	2
DELHI	**10**	2	0	3	1	1	0	0	1	1	4	1	0	3
GOA	**3**	0	2	1	0	1	0	0	1	0	0	0	0	1
GUJARAT	**14**	4	3	1	1	1	0	1	1	1	4	4	0	3
HARYANA	**8**	1	1	0	1	0	0	0	0	0	1	0	1	4
HIMACHAL	**14**	9	0	2	4	1	0	2	0	0	4	3	0	3
JAMMU & KASHMIR	**4**	1	0	0	2	0	0	0	0	0	0	2	0	1
JHARKHAND	**8**	6	1	1	1	0	2	1	0	0	3	1	4	1
KARNATAKA	**14**	6	2	1	2	0	1	3	1	2	5	3	6	1
KERALA	**26**	5	6	2	6	3	4	1	4	2	8	3	7	9
MADHYA PRADESH	**14**	4	2	0	3	1	0	1	1	0	5	1	5	4
MAHARASHTRA	**23**	10	5	1	6	6	0	5	4	1	5	6	5	9
MANIPUR	**1**	0	0	0	0	0	0	1	0	0	0	1	0	0
MP	**7**	6	0	0	2	0	0	2	0	2	3	0	2	4
NEW DELHI	**6**	3	4	1	2	0	0	0	0	0	3	0	1	5
ORISSA	**12**	5	2	3	4	1	1	2	1	0	2	0	4	6
PUNJAB	**36**	5	11	5	5	4	0	3	2	1	17	7	11	11
RAJASTHAN	**8**	5	3	0	2	1	0	1	1	0	0	1	1	1

TAMIL NADU	**14**	3	4	2	1	2	0	5	0	0	4	4	5	3
TRIPURA	**4**	3	0	1	2	0	0	0	1	0	0	1	2	0
UP	**45**	21	6	5	15	2	1	5	1	1	7	11	14	12
UTTARANCHAL	**4**	2	2	1	1	1	1	0	1	0	1	0	1	3
WEST BENGAL	**44**	25	9	6	13	6	1	3	2	5	13	4	24	9
Grand Total	**381**	**142**	**75**	**41**	**85**	**36**	**17**	**49**	**24**	**19**	**108**	**62**	**114**	**111**

Exhibit 2: Literacy Rate in India [5]

Year	**All India Literacy**		
	Male	*Female*	*Person*
1901	9.8	0.6	5.3
1911	10.6	1.1	5.9
1921	12.2	1.8	7.2
1931	15.6	2.9	9.5
1941	24.9	7.3	16.1
1951	27.16	8.86	18.33
1961	40.4	15.34	28.31
1971	45.95	21.97	34.45
1981	56.37	29.75	43.56
1991	64.13	39.29	52.21
2001	79.56	54.28	65.38

Exhibit 3: Annual expenditure on education as a percentage of total expenditure [6]

In Crore)							
Year	**% of Education expenditure to all Sectors expenditure**		**Year**	**% of Education expenditure to all Sectors expenditure**		**Year**	**% of Education expenditure to all Sectors expenditure**
1951-52	7.92		1974-75	10.74		1997-98	12.75

1952-53	8.43		1975-76	10.3		1998-99	14
1953-54	8.82		1976-77	9.96		1999-00	14.6
1954-55	9.84		1977-78	11.61		2000-01	14.42
1955-56	10.65		1978-79	11.46		2001-02	12.89
1956-57	11.47		1979-80	10.83		2002-03	12.6
1957-58	10.61		1980-81	10.67		2003-04	11.98
1958-59	10.9		1981-82	10.3		2004-05	12.13
1959-60	11.73		1982-83	12.52			
1960-61	11.99		1983-84	10.07			
1961-62	11.7		1984-85	10.8			
1962-63	9.47		1985-86	12.99			
1963-64	9		1986-87	11.78			
1964-65	9.6		1987-88	12.75			
1965-66	9.82		1988-89	13.08			
1966-67	9.56		1989-90	13.64			
1967-68	10.55		1990-91	13.37			
1968-69	9.38		1991-92	13.14			
1969-70	9.61		1992-93	13.15			
1970-71	10.16		1993-94	12.94			
1971-72	9.53		1994-95	12.95			
1972-73	9.7		1995-96	13.34			
1973-74	10.1		1996-97	13.33			

Exhibit 4: HDI, Life Expectancy and Importance to Religion for each country [7; 8]

Country	HDI	Life Expectancy	Importance to religion (%)
Albania	0.801	76.2	33
Algeria	0.733	71.7	91
Angola	0.446	41.7	88
Argentina	0.869	74.8	63
Armenia	0.775	71.7	69
Australia	0.962	80.9	32
Austria	0.948	79.4	55
Azerbaijan	0.746	67.1	51
Bahrain	0.866	75.2	97

Bangladesh	0.547	63.1	100
Belarus	0.804	68.7	28
Belgium	0.946	78.8	37
Belize	0.778	75.9	61
Benin	0.437	55.4	92
Bolivia	0.695	64.7	87
Bosn&herz	0.803	74.5	66
Botswana	0.654	48.1	77
Brazil	0.8	71.7	88
Bulgaria	0.824	72.7	34
BurkinaFaso	0.37	51.4	87
Burundi	0.413	48.5	92
Cambodia	0.598	58	85
Cameroon	0.532	49.8	96
Canada	0.961	80.3	45
CentralAfricanRep	0.384	43.7	94
Chad	0.388	50.4	94
Chile	0.867	78.3	70
Colombia	0.791	72.3	87
Comoros	0.561	64.1	95
Congo	0.411	45.8	98
CôtedIvoire	0.432	47.4	88
Croatia	0.85	75.3	67
Cuba	0.838	77.7	34
Cyprus	0.903	79	75
Czech	0.891	75.9	21
Denmark	0.949	77.9	18
Djibouti	0.516	53.9	98
Dominican Rep	0.779	71.5	86
Ecuador	0.772	74.7	82
Egypt	0.708	70.7	99
ElSalvador	0.735	71.3	85
Estonia	0.86	71.2	14
Finland	0.952	78.9	29
France	0.952	80.2	25
Georgia	0.754	70.7	80
Germany	0.935	79.1	40
Ghana	0.553	59.1	93
Greece	0.926	68.2	70
Guatemala	0.689	69.7	86
Guinea	0.642	45.8	97

Guyana	0.75	65.2	87
Haiti	0.529	59.5	75
Honduras	0.7	69.4	89
Hungary	0.874	72.9	39
Iceland	0.968	81.5	38
India	0.619	63.7	79
Indonesia	0.728	69.7	99
Iran	0.759	70.2	83
Ireland	0.959	78.4	58
Israel	0.932	80.3	50
Italy	0.941	80.3	72
Jamaica	0.736	72.2	70
Japan	0.953	82.3	25
Jordan	0.773	71.9	96
Kazakhstan	0.794	65.9	49
Kenya	0.521	52.1	94
Korea	0.921	77.9	45
Kuwait	0.891	77.3	94
Kyrgyzstan	0.696	65.6	68
Lao	0.601	63.2	96
Latvia	0.855	72	36
Lebanon	0.772	71.5	91
Lithuania	0.862	72.5	39
Luxembourg	0.944	78.4	39
Madagascar	0.533	58.4	93
Malawi	0.437	46.3	98
Malaysia	0.811	73.7	95
Mali	0.38	53.1	94
Mauritania	0.55	63.2	98
Mexico	0.829	75.6	70
Moldova	0.708	68.4	70
Morocco	0.646	70.4	99
Mozambique	0.384	42.8	87
Myanmar	0.583	60.8	96
Namibia	0.65	51.6	92
Nepal	0.534	62.6	93
Netherlands	0.953	79.2	33
New Zealand	0.943	79.8	33
Nicaragua	0.71	71.9	82
Niger	0.374	55.8	97
Nigeria	0.47	46.5	94

Norway	0.968	79.8	20
Pakistan	0.551	64.6	96
Panama	0.812	75.1	87
Paraguay	0.755	71.3	89
Peru	0.773	70.7	85
Philippines	0.771	71	95
Poland	0.87	75.2	75
Portugal	0.897	77.7	72
Qatar	0.875	75	95
Romania	0.813	71.9	84
Russian	0.802	65	30
Rwanda	0.452	45.2	90
Saudi	0.812	72.2	95
Senegal	0.499	62.3	98
SierraLeone	0.336	41.8	98
Singapore	0.922	79.4	51
Slovakia	0.863	74.2	47
Slovenia	0.917	77.4	39
SouthAfrica	0.674	50.8	85
Spain	0.949	80.5	38
SriLanka	0.743	71.6	99
Sudan	0.526	57.4	95
Sweden	0.956	80.5	17
Switzerland	0.955	81.3	42
Syrian	0.724	73.6	88
Tajikistan	0.673	66.3	79
Tanzania	0.467	51	96
Thailand	0.781	69.6	94
Togo	0.512	57.8	80
Trinidad&	0.814	69.2	92
Tunisia	0.766	73.5	91
Turkey	0.775	71.4	89
UAE	0.868	78.3	93
Uganda	0.505	49.7	92
UK	0.946	79	29
Ukraine	0.788	67.7	37
Uruguay	0.852	75.9	41
USA	0.951	77.9	67
Venezuela	0.792	73.2	79
VietNam	0.733	73.7	35
Yemen	0.508	61.5	94

Zambia	0.434	40.5	97
Zimbabwe	0.513	40.9	81

Exhibit 5: Inter-State Flow of Migrants among Major States in India [9]

(1999-2000)

States/UTs	Andhra Pradesh	Assam	Bihar	Gujarat	Haryana	Karnataka	Kerala	Madhya Pradesh	Maharashtra	Orissa	Punjab	Rajasthan	Tamil Nadu	Uttar Pradesh	West Bengal	Other States/Uts	All India
Andhra Pradesh	-	0	158	513	3	2944	241	530	3569	552	48	5	1521	278	308	389	11059
Assam	172	-	6	91	19	8	37	137	86	133	104	54	40	442	692	557	2578
Bihar	557	347	-	522	2608	51	50	1081	3389	2096	3389	548	66	9793	8313	1673	34483
Gujarat	127	12	154	-	47	108	252	545	6047	339	69	1608	0	311	57	319	9995
Haryana	0	0	58	0	-	5	1	154	399	45	3015	1769	14	1835	30	1342	8667
Karnataka	3364	0	22	46	16	-	1230	1044	8836	153	68	147	2117	177	135	1127	18482
Kerala	404	0	8	90	164	2119	-	203	1698	2	11	79	3118	114	181	307	8498
Madhya Pradesh	182	0	22	613	139	42	111	-	5364	590	199	2671	77	4656	131	452	15249
Maharashtra	1610	13	15	6746	205	2695	691	3722	-	476	174	850	900	2385	240	839	21561
Orissa	689	1	414	7	69	56	264	1131	1008	-	47	73	32	70	1420	137	5418
Punjab	78	1	129	19	3277	201	111	65	400	171	-	2553	227	2543	337	2736	12848
Rajasthan	614	82	106	2957	5043	261	174	1973	2965	49	907	-	408	2177	729	177	18622
Tamil Nadu	1732	165	60	68	26	3783	2903	180	3751	45	11	35	-	127	873	1836	15595
Uttar Pradesh	511	47	1294	3339	7612	157	154	9202	16174	68	4954	2796	270	-	1996	2746	51320
West Bengal	230	121	1130	122	19	132	20	785	1640	1563	238	366	258	2140	-	554	9318
Other States/Uts	1309	500	2035	2965	3946	1483	1596	1982	3839	794	3486	1341	2805	9885	3570	-	41536
India	**11579**	**1289**	**5611**	**18098**	**23193**	**14045**	**7835**	**22734**	**59165**	**7076**	**16720**	**14894**	**11853**	**36933**	**19012**	**15191**	**X**

Works Cited

1. *Wikipedia.* [Online] [Cited: September 07, 2009.] http://en.wikipedia.org/wiki/Literacy_in_India.

2. *Colby College.* [Online] [Cited: September 06, 2009.] http://www.colby.edu/chemistry/PChem/notes/linest.pdf.

3. *Yahoo Finance.* [Online] [Cited: September 09, 2009.] http://finance.yahoo.com/q/hp?s=%5EBSESN.

4. **CURE.** *Ragging in India.* s.l., India : Coalition to Uproot Ragging from Education (www.noragging.com), August 2009.

5. **Government, Tamil Nadu.** [Online] [Cited: September 07, 2009.] http://www.tn.gov.in/schooleducation/statistics/table4-lit.htm.

6. **Indiastat.** *Indiastat.com.* [Online] [Cited: September 07, 2009.] http://www.indiastat.com/table/economy/8/expenditureoneducationaspercentageoftotalbudgetandgrossdomesticproductstatedomesticproductgdpsdp/369168/9213/data.aspx.

7. *Gallup.* [Online] [Cited: September 09, 2009.] https://worldview.gallup.com.

8. **UNDP.** Human Development Report 2007/2008. [Online] [Cited: September 09, 2009.] http://hdr.undp.org/en/reports/global/.

9. **Indiastat.com.** *Indiastat.com.* [Online] [Cited: September 02, 2009.] http:// www. indiastat.com /urbanareas/31/migration /16181/stats. aspx.