Analysis on Education survey
71st Round Data for Chhattisgarh and India
(Sample Survey End Semester Project)

Srijan Chattopadhyay (BS2126)

Fall, 2023

Introduction

The surveys on social consumption relating to education, conducted by the National Sample Survey Office (NSSO) are the primary sources of data on various indicators on education scenario of the country; like literacy rates, attendance ratios, incentives received by the students, expenditure incurred for the purpose of education etc .. These are used for planning, policy formulation, and decision support and as input for further analytical studies by various Government organizations, academicians, researchers and scholars.
National Sample Survey Office (NSSO) conducted a nation-wide survey on 'Social Consumption: Education' as a part of its 7lstround Qanuary-June 2014) survey in the whole Indian Territory. The previous NSS rounds, during which the survey on similar subject was taken up, are the 35th round Quly 1980-June 1981), 42nd round Quly 1986-June 1987), 52nd round Quly 1995-June 1996) and the 64th round Quly 2007-June 2008). All the State Governments and Union Territories (except Andaman ts: Nicobar Islands, Chandigarh, Dadra ts: Nagar Haveli and Lakshadweep) also participated in the survey programme with at least on equal matching sample size basis.

Objective of the Survey

1. The foundation of the society is based on education. It brings economic and social prosperity. One can appositely say, a strong nation can be built by ensuring that each citizen of that nation is educated. In most of the countries, the government spends substantial amount on the creation as well as the functioning of the educational infrastructure. But to avail such facilities individuals too have to incur expenditure in the form of course fees, examination fees, cost of books and stationery etc. While information on the expenditure incurred by governments is available from budget documents, there is an increasing role of non- government organisations as well as individuals in the education sector. The generation of information on education and the expenditure by individuals through a specialised survey, therefore, has special significance in the contemporary context.
2. NSSO had conducted an all-India household survey on education during the period January – June 2014. The purpose of the survey was to collect information on participation of persons aged 5-29 years in pursuit of education in the country; the extent of use of educational infrastructure, facilities and incentives provided by the government and private sectors and its impact on current attendance status of population in the educational institutions; private expenditure incurred by households on education and the extent of educational wastage in terms of dropping-out and discontinuance, and its causes.
3. Increasing use of information technology in every sphere of day-to-day life at present seems to be one giant leap for the country. So it stands to reason that more importance has to be placed on computer literacy as education and computer now go with hand-in-hand. This survey also captured

some information on various facets of ability to operate computer along with possession of computer in the household and access to internet facility for a 14 plus aged member of a household.

4. NSSO conducted first all-India survey on social consumption in its 35th round (July1980 - June 1981) to obtain data on nature & quantum of benefits received by the people from public expenditure on various services such as health services (including mass immunisation and family welfare programmes), educational services, public distribution, etc. Subsequently, the subject of social consumption was covered by the NSSO in its 42nd round (July 1986 - June 1987) and 52nd round (July 1995 - June 1996) and 64th round (July 2007- June 2008). These surveys were conducted on the same lines as the 35th round survey with some modifications in subject coverage. The surveys covered both qualitative and quantitative aspects of educational services received by households. Qualitative aspects included literacy, educational level attained, current attendance/enrolment, drop-out, reason for drop-out, etc. Quantitative aspects of educational services included expenditure incurred by households in availing themselves of these services with its break-up into tuition fees, transport costs, etc. In NSS 47th round (July-December 1991), data on qualitative aspects of educational services were collected. This apart, person-wise information on literacy and educational attainment was collected for each surveyed household in all the household surveys of NSSO.

5. These indicators on education are required for planning, policy formulation and decision making at various levels within the government and outside. The results of the survey are likely to be used by the Ministry of Human Resource Development, Department of Electronics and Information Technology etc. apart from private data users and policy makers.

Sampling Design

Outline of sample design:

A stratified multi-stage design has been adopted for the 71st round survey. The first stage units (FSU) are the census villages (Panchayat wards in case of Kerala) in the rural sector and Urban Frame Survey (UFS) blocks in the urban sector. The ultimate stage units (USU) are households in both the sectors. In case of large FSUs, one intermediate stage of sampling is the selection of two hamlet-groups (hgs)/ sub-blocks (sbs) from each rural/ urban FSU.

Sampling Frame for First Stage Units:

For the rural sector, the list of 2011 census villages (henceforth the term 'village' would mean Panchayat wards for Kerala) constitutes the sampling frame. In case of Kerala, due to the non-availability of Panchayat wards based on census 2011, the available list of Panchayat wards based on census 2001 is used as the rural frame. For the urban sector, the latest updated list of UFS blocks (phase 2007-12) is considered as the sampling frame.

Stratification:

Stratum has been formed at district level. Within each district of a State/UT, generally speaking, two basic strata have been formed: (i) rural stratum comprising of all rural areas of the district and (ii) urban stratum comprising of all the urban areas of the district. However, within the urban areas of a district, if there are one or more towns with population 1 lakh or more as per Census 2011, each of them formed a separate basic stratum and the remaining urban areas of the district has been considered as another basic stratum.

Sub-stratification:

Rural sector:

If 'r' be the sample size allocated for a rural stratum, the number of sub-strata formed was 'r/2'. The villages within a district as per frame have been first arranged in ascending order of population[1]. Then sub-strata

1 to 'r/2' have been demarcated in such a way that each sub-stratum comprised a group of villages of the arranged frame and had more or less equal population.

Urban sector:

If 'u' be the sample size allocated for an urban stratum, the number of sub-strata formed was 'u/2'. For all strata, if u/2 >1, implying formation of 2 or more sub-strata, all the UFS blocks within the stratum have been first arranged in ascending order of total number of households in the UFS Blocks as per UFS phase 2007-12. Then sub-strata 1 to 'u/2' have been demarcated in such a way that each sub-stratum had more or less equal number of households.

Total sample size (FSUs):

8300 FSUs have been allocated for the central sample at all-India level. For the state sample, there are 9274 FSUs allocated for all-India. State wise allocation of sample FSUs is given in Table 1.

Allocation of total sample to States and UTs:

The total number of sample FSUs heve been allocated to the States and UTs in proportion to population as per Census 2011 subject to a minimum sample allocation to each State/ UT. While doing so, the resource availability in terms of number of field investigators has been kept in view.

Allocation of State/ UT level sample to rural and urban sectors:

State/UT level sample size has been allocated between two sectors in proportion to population as per Census 2011 with double weightage to urban sector subject to the restriction that urban sample size for bigger states like Maharashtra, Tamil Nadu etc. do not exceed the rural sample size. A minimum of 16 FSUs (minimum 8 each for rural and urban sector separately) is allocated to each State/ UT.

Allocation to strata:

Within each sector of a State/ UT, the respective sample size has been allocated to the different strata in proportion to the population as per Census 2011. Stratum level allocation has been adjusted to multiples of 2 with a minimum sample size of 2^{21}[21]. For special strata in the rural areas of Nagaland and A & N Islands, 4 FSUs has been allocated to each.

Allocation to sub-strata:

Allocation for each sub-stratum has been 2 in both rural and urban sectors.

Selection of FSUs:

For the rural sector, from each stratum/sub-stratum, required number of sample villages have been selected by Probability Proportional to Size With Replacement (PPSWR), size being the population of the village as per Census 2011. For the urban sector, from each stratum/sub-stratum, FSUs have been selected by Probability Proportional to Size With Replacement (PPSWR), size being the number of households of the UFS Blocks. Both rural and urban samples have been drawn in the form of two independent sub-samples and equal number of samples has been allocated among the two sub rounds.

Selection of hamlet-groups/ sub-blocks - important steps

1. Criterion for hamlet-group/ sub-block formation: After identification of the boundaries of the FSU, it is to be determined whether listing will be done in the whole sample FSU or not. In case the approximate present population of the selected FSU is found to be 1200 or more, it will be divided into a suitable number (say, D) of 'hamlet-groups' in the rural sector and 'sub-blocks' in the urban sector by more or less equalising the population as stated below.

approximate present population of the sample FSU	no. of hg's/sb's to be formed
less than 1200 (no hamlet-groups/sub-blocks)	1
1200 to 1799	3
1800 to 2399	4
2400 to 2999	5
3000 to 3599	6
.....and so on	-

For rural areas of Himachal Pradesh, Sikkim, Uttarakhand (except four districts Dehradun, Nainital, Hardwar and Udham Singh Nagar), Poonch, Rajouri, Udhampur, Reasi, Doda, Kistwar, Ramban, Leh (Ladakh), Kargil districts of Jammu and Kashmir and Idukki district of Kerala, the number of hamlet-groups will be formed as follows:

approximate present population of the sample village	no. of hg's to be formed
less than 600 (no hamlet-groups)	1
600 to 899	3
900 to 1199	4
1200 to 1499	5
1500 to 1799	6
.....and so on	-

Formation and selection of hamlet-groups/ sub-blocks:

In case hamlet-groups/ sub-blocks are to be formed in the sample FSU, the same should be done by more or less equalizing population. Note that while doing so, it is to be ensured that the hamlet-groups/ sub-blocks formed are clearly identifiable in terms of physical landmarks. Two hamlet-groups (hg)/ sub-blocks (sb) will be selected from a large FSU wherever hamlet-groups/ sub-blocks have been formed in the following manner – one hg/ sb with maximum percentage share of population will always be selected and termed as hg/ sb 1; one more hg/ sb will be selected from the remaining hg's/ sb's by simple random sampling (SRS) and termed as hg/ sb 2. Listing and selection of the households will be done independently in the two selected hamlet-groups/ sub-blocks. The FSUs without hg/ sb formation will be treated as sample hg/ sb number 1.

Formation of second stage strata and allocation of households:

SSS	composition of SSS within a sample FSU	number of households to be surveyed	
		FSU without hg/sb formation	FSU with hg/sb formation (for each hg/sb)
Schedule 25.0: Social Consumption: Health			
SSS 1	households having at least one child of age less than 1 year	2	1
SSS 2	from the remaining, households with at least one member (including deceased former member) hospitalised during last 365 days	4	2
SSS 3	other households	2	1
Schedule 25.2: Social Consumption: Education			
SSS 1	households with at least one student receiving technical/professional education	2	1
SSS 2	from the remaining, households having at least one student receiving general education	4	2
SSS 3	other households	2	1

Selection of households:

From each SSS, for both the schedules, the sample households are selected by SRSWOR

Estimation Procedure

Notations

- s = subscript for s-th stratum
- t = subscript for t-th sub-stratum
- m = subscript for sub-sample ($m = 1, 2$)
- i = subscript for i-th FSU [village (panchayat ward)/ block]
- d = subscript for a hamlet-group/ sub-block ($d = 1, 2$)
- j = subscript for j-th second stage stratum in an FSU/ hg/sb [$j = 1, 2$ or 3]
- k = subscript for k-th sample household under a particular second stage stratum within an FSU/ hg/sb

- D = total number of hg's/ sb's formed in the sample FSU
- $D^* = 0$ if $D = 1 = (D\check{}1)$ for FSUs with $D > 1$
- Z = total size of a rural/urban sub-stratum (= sum of sizes for all the FSUs of a sub-stratum)
- z = size of sample village/UFS block used for selection.
- n = number of sample FSUs surveyed including 'uninhabitated' and 'zero cases' but excluding casualty for a particular sub-sample and sub-stratum
- H = total number of households listed in a second-stage stratum of an FSU / hamlet-group or sub-block of sample FSU
- h = number of households surveyed in a second-stage stratum of an FSU / hamlet-group or sub-block of sample FSU
- x, y = observed value of characteristics x, y under estimation
- $\hat{X}, \hat{Y}$ = estimate of population total X, Y for the characteristics x, y

Under the above symbols,

$y_{stmidjk}$ = observed value of the characteristic y for the k-th household in the j-th second stage stratum of the d-th hg/ sb ($d = 1, 2$) of the i-th FSU belonging to the m-th sub-sample for the t-th sub-stratum of s-th stratum.

Formulae for Estimation of Aggregates

Schedules 25.0 & 25.2:

1. For jth second stage stratum

$$\hat{Y}_j = \frac{Z}{n_j} \sum_{i=1}^{n_j} \frac{1}{z_i} \left[\frac{H_{i1,j}}{h_{i1,j}} \sum_{k=1}^{h_{i1,j}} y_{i1,jk} + D_i^* \frac{H_{i2,j}}{h_{i2,j}} \sum_{k=1}^{h_{i2,j}} y_{i2,jk} \right]$$

2. For all second-stage strata combined

$$\hat{Y} = \sum_j \hat{Y}_j$$

Overall Estimate for Aggregates for a sub-stratum

Overall estimate for aggregates for a sub-stratum ($\hat{Y}_{st}$) based on two subsamples in a sub-stratum is obtained as

$$\hat{Y}_{st} = \frac{1}{2} \sum_{m=1}^{2} \hat{Y}_{stm}$$

Overall Estimate for Aggregates for a stratum

Overall estimate for a stratum $\hat{Y}_s$ will be obtained as $\hat{Y}_s = \sum_t \hat{Y}_{st}$

Overall Estimate of Aggregates at State/UT/all-India level

The overall estimate $\hat{Y}$ at the State/ UT/ all-India level is obtained by summing the stratum estimates $\hat{Y}_s$ over all strata belonging to the State/ UT/ all-India

Estimates of Ratios

Let $\hat{Y}$ and $\hat{X}$ be the overall estimates of the aggregates Y and X for two characteristics y and x respectively at the State/ UT/ all-India level. Then the combined ratio estimate of $R = \frac{Y}{X}$ is obtained as $\hat{R} = \frac{\hat{Y}}{\hat{X}}$

Estimation of Errors

The estimated variances of the above estimates will be as follows:

For aggregate $\hat{Y}$:

$$\text{Var}(\hat{Y}) = \sum_s \text{Var}(\hat{Y}_s) = \sum_s \sum_t \text{Var}(\hat{Y}_{st}) \tag{1}$$

where $\text{Var}(\hat{Y}_{st})$ is given by

$$\text{Var}(\hat{Y}_{st}) = \frac{1}{4}\left(\hat{Y}_{st}^{(1)} - \hat{Y}_{st}^{(2)}\right)^2 \tag{2}$$

where $\hat{Y}_{st}^{(1)}$ and $\hat{Y}_{st}^{(2)}$ are the estimates for sub-sample 1 and sub-sample 2 respectively for stratum s and sub-stratum t.

For ratio $\hat{R}$:

$$\text{MSE}(\hat{R}) = \sum_s \sum_t \left[\frac{(\hat{Y}_{st}^{(1)} - \hat{Y}_{st}^{(2)})}{\hat{X}^2} + \frac{(\hat{X}_{st}^{(1)} - \hat{X}_{st}^{(2)})}{\hat{X}^2}(\hat{R} - \hat{R}_{st})\right]^2 \tag{3}$$

Estimates of Relative Standard Error (RSE):

$$\text{RSE}(\hat{Y}) = \frac{\sqrt{\text{Var}(\hat{Y})}}{\hat{Y}} \times 100 \tag{4}$$

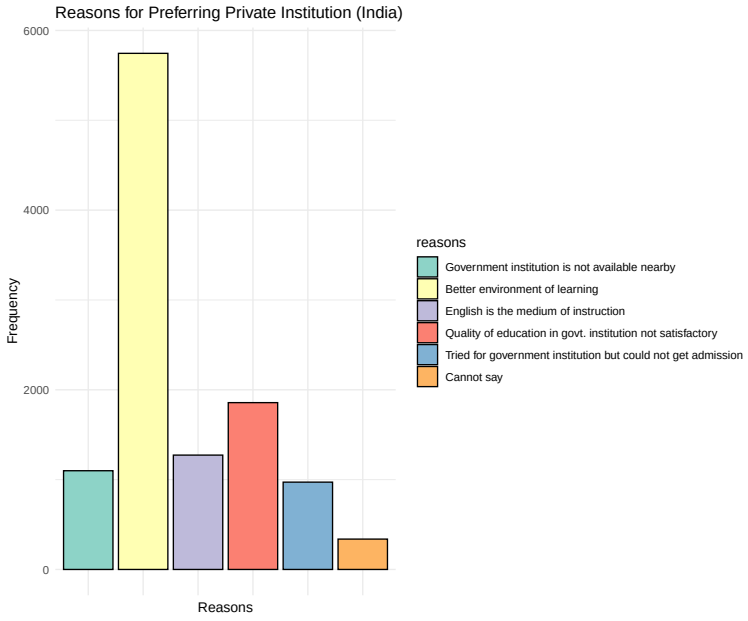
$$\text{RSE}(\hat{R}) = \frac{\sqrt{\text{MSE}(\hat{R})}}{\hat{R}} \times 100 \tag{5}$$

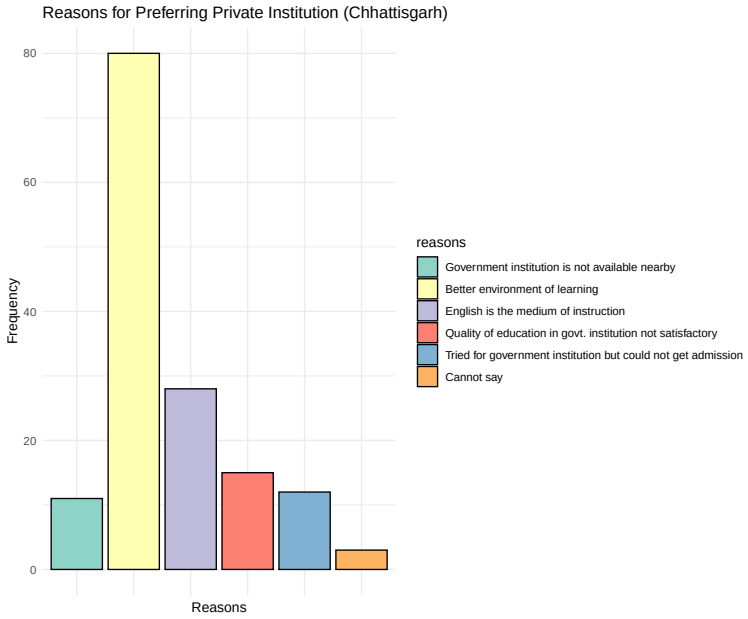
Goal of this project

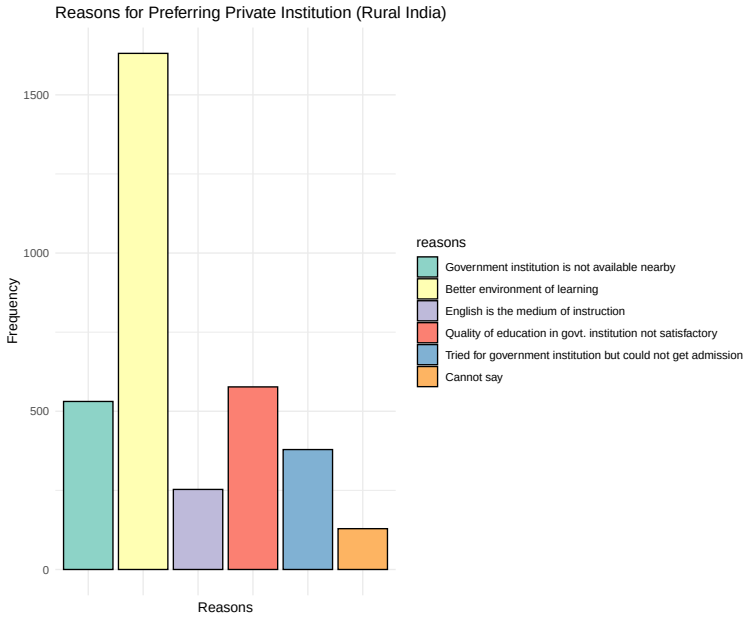
Main objective of this project is to look into the fact if the people of different sectors (Rural and Urban) and different genders (Male and Female) takes the basic education in the language same as their speaking language at home or others for specifically the state of Chhattisgarh and the Whole India. Also we have seen the distributions of persons taking basic education course in private institutions for various reasons.

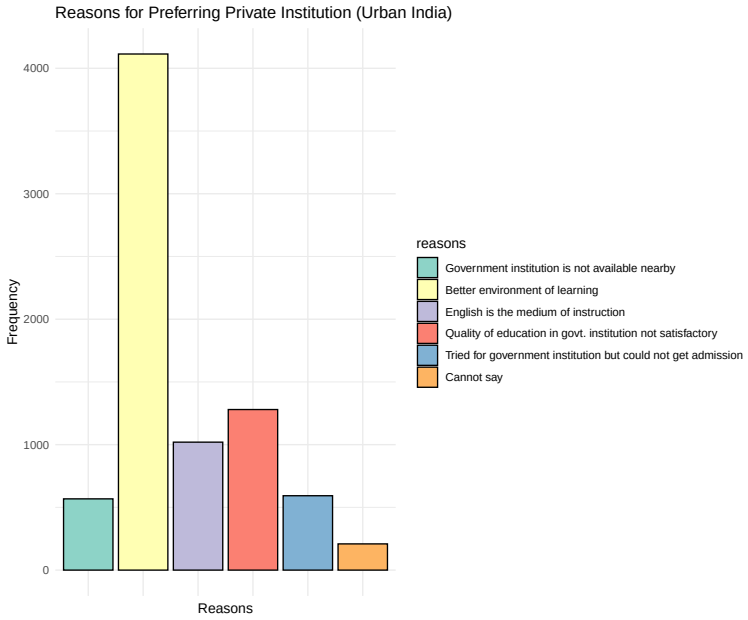
Results

Reasons for Private tutions statistics:






Reasons for Preferring Private Institution (Rural India)
Frequency
1500
1000
500
0
Reasons
reasons
Government institution is not available nearby
Better environment of learning
English is the medium of instruction
Quality of education in govt. institution not satisfactory
Tried for government institution but could not get admission
Cannot say


Reasons for Preferring Private Institution (Urban India)
Frequency
4000
3000
2000
1000
0
Reasons
reasons
Government institution is not available nearby
Better environment of learning
English is the medium of instruction
Quality of education in govt. institution not satisfactory
Tried for government institution but could not get admission
Cannot say

Estimated Percentages of languages:

All the estimated matrices for whole India are presented here:

- Estimated percentages in the matrix form for the whole India is in this table
- Estimated percentages in the matrix form for the whole India (Urban only) is in this table
- Estimated percentages in the matrix form for the whole India (Rural only) is in this table
- Estimated percentages in the matrix form for the whole India (Male only) is in this table
- Estimated percentages in the matrix form for the whole India (Female only) is in this table
- Estimated percentages in the matrix form for the whole India (Rural and Female only) is in this table
- Estimated percentages in the matrix form for the whole India (Rural and Male only) is in this table
- Estimated percentages in the matrix form for the whole India (Urban and Male only) is in this table
- Estimated percentages in the matrix form for the whole India (Urban and Female only) is in this table

Now, all the estimated matrices for Chhattisgarh are presented here:

- Estimated percentages in the matrix form for Chattisgarh is in this table
- Estimated percentages in the matrix form for Chattisgarh (Urban only) is in this table
- Estimated percentages in the matrix form for Chattisgarh (Rural only) is in this table
- Estimated percentages in the matrix form for Chattisgarh (Male only) is in this table
- Estimated percentages in the matrix form for Chattisgarh (Female only) is in this table
- Estimated percentages in the matrix form for Chattisgarh (Rural and Female only) is in this table
- Estimated percentages in the matrix form for Chattisgarh (Rural and Male only) is in this table
- Estimated percentages in the matrix form for Chattisgarh (Urban and Male only) is in this table
- Estimated percentages in the matrix form for Chattisgarh (Urban and Female only) is in this table

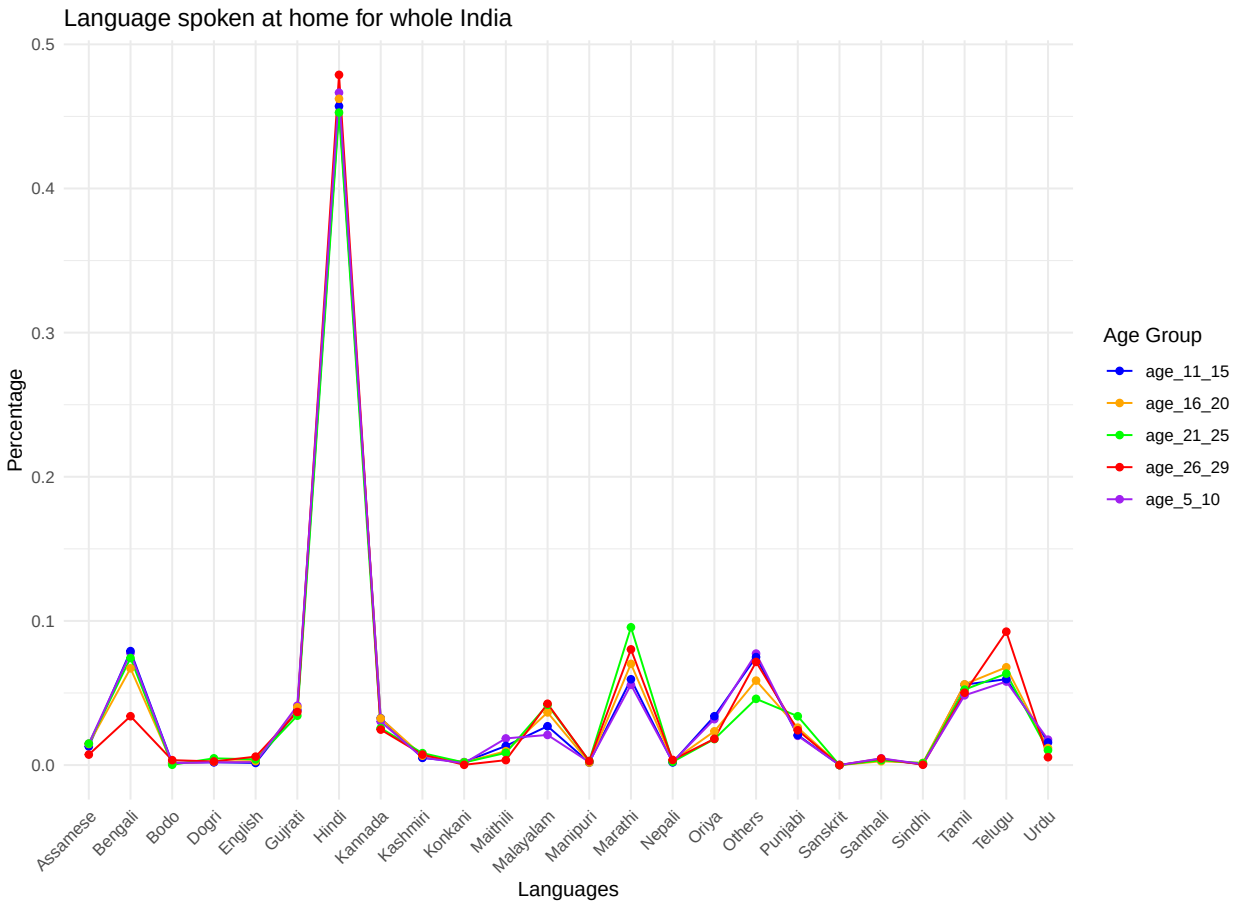
MSE of the estimates:

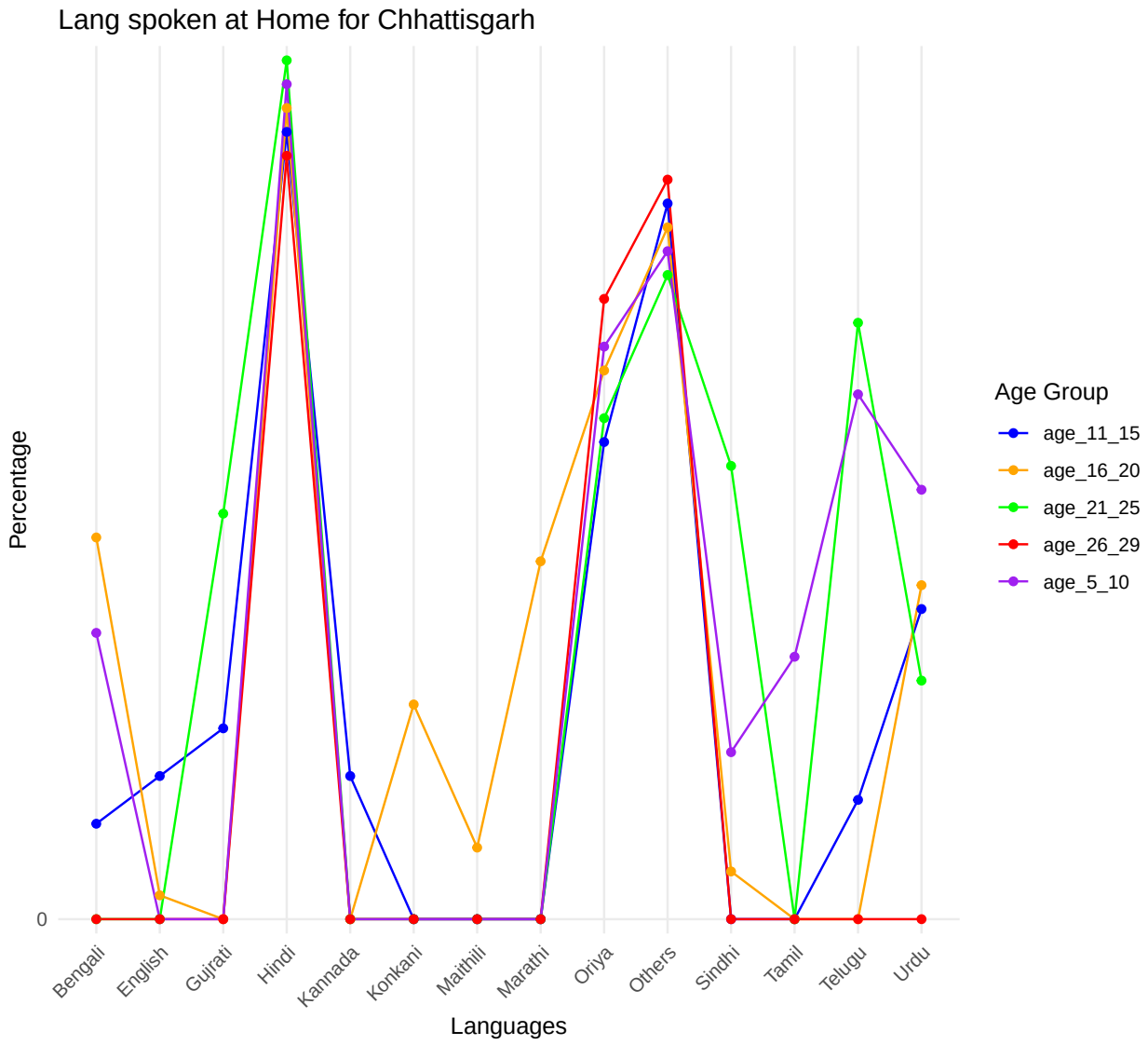
Due to computational complexities, all the MSEs are not presented. As Chattisgarh is relatively smaller than whole India, it took less time than India to compute. So, all the variance tables for Chattisgarh and some of whole India are presented.

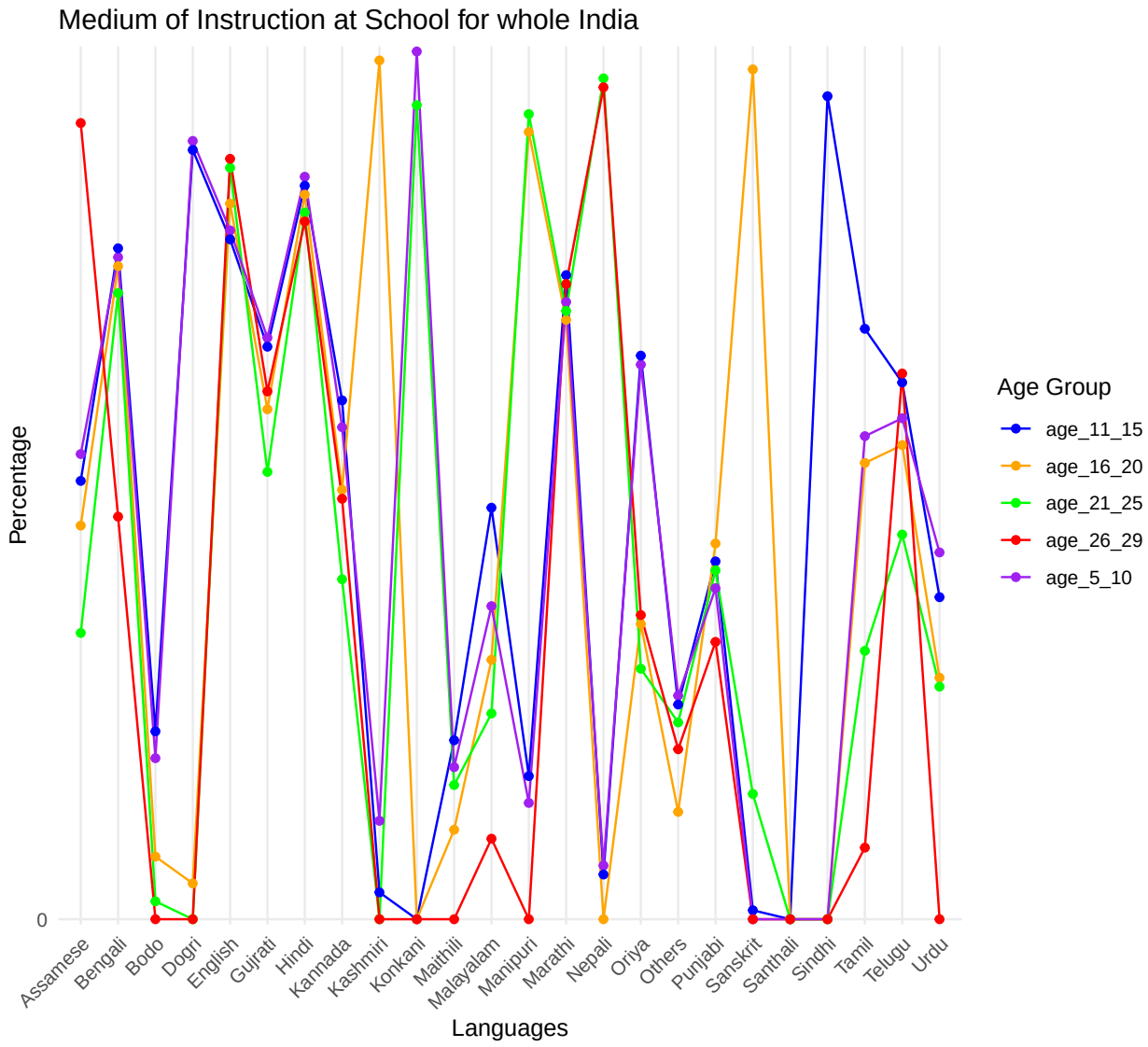
- MSE of the estimates of Chhattisgarh is in this table
- MSE of the estimates of Chhattisgarh (Urban Only) is in this table
- MSE of the estimates of Chhattisgarh (Urban and Male Only) is in this table
- MSE of the estimates of Chhattisgarh (Urban and Female Only) is in this table
- MSE of the estimates of Chhattisgarh (Rural Only) is in this table
- MSE of the estimates of Chhattisgarh (Rural and Female Only) is in this table
- MSE of the estimates of Chhattisgarh (Rural and Male Only) is in this table
- MSE of the estimates of Chhattisgarh (Male Only) is in this table
- MSE of the estimates of Chhattisgarh (Female Only) is in this table
- MSE of the estimates of whole India (Rural and Female Only) is in this table
- MSE of the estimates of whole India (Rural and Male Only) is in this table

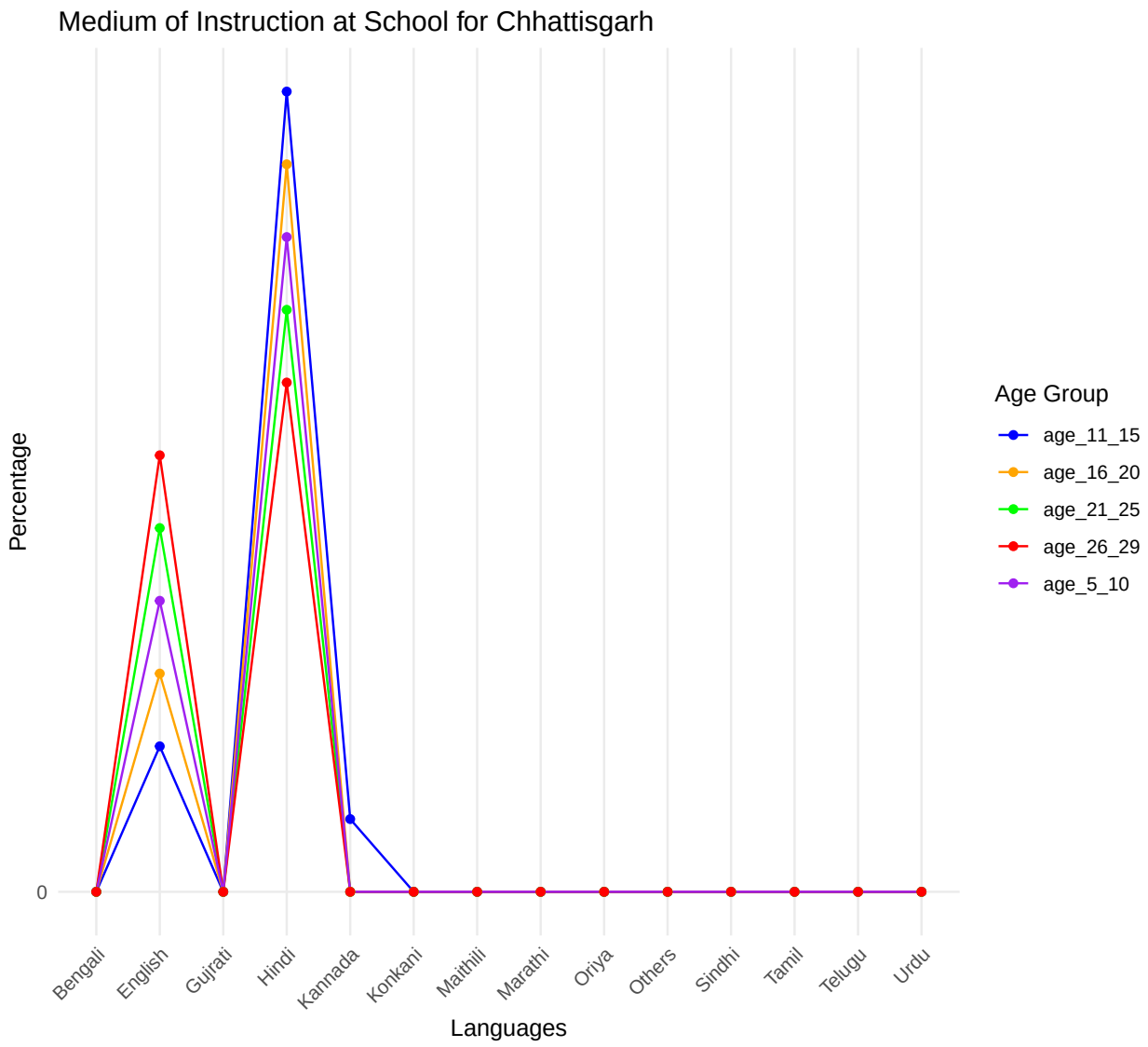
Distributions of Languages over age-groups School and Home wise

Now we present the visual representation of the age distribution of languages:




Lang spoken at Home for Chhattisgarh
Percentage
0
Bengali
English
Gujrati
Hindi
Kannada
Konkani
Maithili
Marathi
Oriya
Others
Sindhi
Tamil
Telugu
Urdu
Languages
Age Group
age_11_15
age_16_20
age_21_25
age_26_29
age_5_10


Medium of Instruction at School for whole India
Percentage
0
Assamese
Bengali
Bodo
Dogri
English
Gujrati
Hindi
Kannada
Kashmiri
Konkani
Maithili
Malayalam
Manipuri
Marathi
Nepali
Oriya
Others
Punjabi
Sanskrit
Santhali
Sindhi
Tamil
Telugu
Urdu
Languages
Age Group
age_11_15
age_16_20
age_21_25
age_26_29
age_5_10


Medium of Instruction at School for Chhattisgarh
Percentage
0
Bengali
English
Gujrati
Hindi
Kannada
Konkani
Maithili
Marathi
Oriya
Others
Sindhi
Tamil
Telugu
Urdu
Languages
Age Group
age_11_15
age_16_20
age_21_25
age_26_29
age_5_10

All the tables are presented next.....

Languages(Age)	**5-10**	**11-15**	**16-20**	**21-25**	**26-29**
Gujrati	0.0412	0.0381	0.0401	0.0343	0.0368
Hindi	0.4665	0.4571	0.4622	0.4527	0.4788
Urdu	0.0176	0.0157	0.0118	0.0103	0.0054
Malayalam	0.0209	0.0269	0.0365	0.0418	0.0425
Sindhi	0.0006	0.0006	0.0015	0.0013	0.0002
English	0.0020	0.0015	0.0027	0.0040	0.0058
Bengali	0.0777	0.0791	0.0671	0.0743	0.0339
Tamil	0.0484	0.0558	0.0560	0.0523	0.0502
Oriya	0.0318	0.0339	0.0234	0.0180	0.0182
Assamese	0.0148	0.0129	0.0146	0.0149	0.0073
Maithili	0.0185	0.0133	0.0096	0.0085	0.0034
Santhali	0.0047	0.0038	0.0026	0.0033	0.0045
Others	0.0774	0.0749	0.0586	0.0459	0.0716
Marathi	0.0557	0.0595	0.0702	0.0956	0.0803
Punjabi	0.0203	0.0208	0.0260	0.0339	0.0241
Nepali	0.0021	0.0018	0.0030	0.0023	0.0036
Telugu	0.0579	0.0596	0.0678	0.0634	0.0926
Kannada	0.0302	0.0325	0.0324	0.0256	0.0246
Manipuri	0.0020	0.0017	0.0017	0.0028	0.0028
Konkani	0.0013	0.0020	0.0020	0.0018	0.0002
Bodo	0.0013	0.0015	0.0011	0.0002	0.0034
Kashmiri	0.0052	0.0050	0.0065	0.0083	0.0073
Sanskrit	0.0000	0.0000	0.0000	0.0000	0.0000
Dogri	0.0019	0.0020	0.0025	0.0047	0.0025

Table 1: Language Spoken at Home for whole India (Age Distribution)

Languages	5-10	11-15	16-20	21-25	26-29
Hindi	0.7447	0.6996	0.7160	0.7453	0.6279
Gujrati	0.0000	0.0010	0.0000	0.0038	0.0000
Sindhi	0.0009	0.0000	0.0002	0.0057	0.0000
Bengali	0.0016	0.0003	0.0026	0.0000	0.0000
Telugu	0.0099	0.0004	0.0000	0.0330	0.0000
Oriya	0.0243	0.0066	0.0135	0.0088	0.0426
Marathi	0.0000	0.0000	0.0026	0.0000	0.0000
Others	0.2132	0.2883	0.2608	0.2020	0.3296
Urdu	0.0040	0.0019	0.0025	0.0013	0.0000
Kannada	0.0000	0.0009	0.0000	0.0000	0.0000
English	0.0000	0.0009	0.0001	0.0000	0.0000
Konkani	0.0000	0.0000	0.0013	0.0000	0.0000
Maithili	0.0000	0.0000	0.0003	0.0000	0.0000
Tamil	0.0014	0.0000	0.0000	0.0000	0.0000

Table 2: Language Spoken at Home for Chhattisgarh (Age Distribution)

Languages(Age)	5-10	11-15	16-20	21-25	26-29
Gujrati	0.0391	0.0364	0.0301	0.0191	0.0308
Hindi	0.4469	0.4453	0.4259	0.3349	0.3304
Urdu	0.0113	0.0086	0.0031	0.0019	0.0000
Malayalam	0.0084	0.0145	0.0036	0.0010	0.0003
Sindhi	0.0000	0.0001	0.0000	0.0000	0.0000
English	0.2318	0.2002	0.3410	0.5038	0.5128
Bengali	0.0662	0.0716	0.0563	0.0479	0.0140
Tamil	0.0254	0.0403	0.0215	0.0043	0.0003
Oriya	0.0323	0.0348	0.0080	0.0033	0.0080
Assamese	0.0237	0.0183	0.0138	0.0064	0.0000
Maithili	0.0005	0.0006	0.0003	0.0004	0.0000
Santhali	0.0000	0.0000	0.0000	0.0000	0.0000
Others	0.0014	0.0012	0.0003	0.0007	0.0006
Marathi	0.0458	0.0537	0.0411	0.0417	0.0500
Punjabi	0.0090	0.0108	0.0113	0.0108	0.0048
Nepali	0.0003	0.0002	0.0000	0.0001	0.0001
Telugu	0.0291	0.0317	0.0246	0.0128	0.0317
Kannada	0.0273	0.0303	0.0183	0.0101	0.0164
Manipuri	0.0004	0.0005	0.0000	0.0000	0.0000
Konkani	0.0001	0.0000	0.0000	0.0000	0.0000
Bodo	0.0006	0.0006	0.0003	0.0001	0.0000
Kashmiri	0.0003	0.0002	0.0001	0.0000	0.0000
Sanskrit	0.0000	0.0001	0.0001	0.0004	0.0000
Dogri	0.0000	0.0000	0.0002	0.0000	0.0000

Table 3: Medium of Instruction at School for Whole India (Age Distribution)

Languages	5-10	11-15	16-20	21-25	26-29
Hindi	0.9032	0.9253	0.9151	0.7077	0.6686
Gujrati	0.0000	0.0000	0.0000	0.0000	0.0000
Sindhi	0.0000	0.0000	0.0000	0.0000	0.0000
Bengali	0.0000	0.0000	0.0000	0.0000	0.0000
Telugu	0.0000	0.0000	0.0000	0.0000	0.0000
Oriya	0.0000	0.0000	0.0000	0.0000	0.0000
Marathi	0.0000	0.0000	0.0000	0.0000	0.0000
Others	0.0000	0.0000	0.0000	0.0000	0.0000
Urdu	0.0000	0.0000	0.0000	0.0000	0.0000
Kannada	0.0000	0.0009	0.0000	0.0000	0.0000
English	0.0968	0.0737	0.0849	0.2923	0.3314
Konkani	0.0000	0.0000	0.0000	0.0000	0.0000
Maithili	0.0000	0.0000	0.0000	0.0000	0.0000
Tamil	0.0000	0.0000	0.0000	0.0000	0.0000

Table 4: Medium of Instruction at School for Chhattisgarh (Age Distribution)

Languages(Age)	5-10	11-15	16-20	21-25	26-29
Gujrati	0.0377	0.0346	0.0323	0.0240	0.0493
Hindi	0.4731	0.4648	0.4822	0.4618	0.4628
Urdu	0.0120	0.0106	0.0083	0.0065	0.0020
Malayalam	0.0149	0.0212	0.0295	0.0424	0.0381
Sindhi	0.0000	0.0001	0.0000	0.0000	0.0000
English	0.0016	0.0006	0.0011	0.0010	0.0011
Bengali	0.0846	0.0841	0.0722	0.0744	0.0203
Tamil	0.0336	0.0406	0.0406	0.0399	0.0268
Oriya	0.0365	0.0395	0.0258	0.0177	0.0106
Assamese	0.0174	0.0152	0.0186	0.0154	0.0117
Maithili	0.0231	0.0171	0.0127	0.0131	0.0022
Santhali	0.0061	0.0050	0.0039	0.0054	0.0085
Others	0.0918	0.0907	0.0727	0.0625	0.1214
Marathi	0.0506	0.0553	0.0651	0.0900	0.0548
Punjabi	0.0198	0.0202	0.0221	0.0357	0.0213
Nepali	0.0023	0.0019	0.0029	0.0024	0.0006
Telugu	0.0520	0.0541	0.0631	0.0640	0.1178
Kannada	0.0299	0.0308	0.0315	0.0221	0.0326
Manipuri	0.0017	0.0015	0.0015	0.0027	0.0022
Konkani	0.0012	0.0019	0.0011	0.0012	0.0000
Bodo	0.0017	0.0021	0.0016	0.0004	0.0054
Kashmiri	0.0058	0.0059	0.0080	0.0110	0.0087
Sanskrit	0.0000	0.0000	0.0000	0.0004	0.0000
Dogri	0.0023	0.0024	0.0030	0.0064	0.0017

Table 5: Language Spoken at Home for Rural India (Age Distribution)

Languages	5-10	11-15	16-20	21-25	26-29
Hindi	0.7573	0.6934	0.7153	0.7514	0.2343
Gujrati	0.0000	0.0000	0.0000	0.0000	0.0000
Sindhi	0.0000	0.0000	0.0000	0.0000	0.0000
Bengali	0.0000	0.0000	0.0000	0.0000	0.0000
Telugu	0.0000	0.0000	0.0000	0.0000	0.0000
Oriya	0.0242	0.0046	0.0125	0.0072	0.0000
Marathi	0.0000	0.0000	0.0000	0.0000	0.0000
Others	0.2168	0.3020	0.2721	0.2414	0.7657
Urdu	0.0000	0.0000	0.0000	0.0000	0.0000
Kannada	0.0000	0.0000	0.0000	0.0000	0.0000
English	0.0000	0.0000	0.0000	0.0000	0.0000
Konkani	0.0000	0.0000	0.0000	0.0000	0.0000
Maithili	0.0000	0.0000	0.0000	0.0000	0.0000
Tamil	0.0017	0.0000	0.0000	0.0000	0.0000

Table 6: Language Spoken at Home for Rural Chhattisgarh (Age Distribution)

Languages	5-10	11-15	16-20	21-25	26-29
Gujrati	0.0386	0.0353	0.0281	0.0191	0.0476
Hindi	0.5140	0.5092	0.5034	0.4273	0.4285
Urdu	0.0081	0.0055	0.0021	0.0023	0.0000
Malayalam	0.0070	0.0134	0.0036	0.0008	0.0006
Sindhi	0.0000	0.0000	0.0000	0.0000	0.0000
English	0.1396	0.1156	0.2341	0.3730	0.3589
Bengali	0.0745	0.0780	0.0636	0.0552	0.0213
Tamil	0.0238	0.0359	0.0207	0.0056	0.0005
Oriya	0.0388	0.0422	0.0105	0.0052	0.0128
Assamese	0.0294	0.0233	0.0196	0.0102	0.0001
Maithili	0.0007	0.0008	0.0004	0.0008	0.0000
Santhali	0.0000	0.0000	0.0000	0.0000	0.0000
Others	0.0017	0.0013	0.0004	0.0011	0.0001
Marathi	0.0473	0.0555	0.0461	0.0513	0.0470
Punjabi	0.0105	0.0128	0.0120	0.0156	0.0018
Nepali	0.0003	0.0003	0.0000	0.0002	0.0001
Telugu	0.0339	0.0364	0.0319	0.0178	0.0583
Kannada	0.0304	0.0332	0.0225	0.0138	0.0224
Manipuri	0.0004	0.0003	0.0000	0.0001	0.0000
Konkani	0.0001	0.0002	0.0001	0.0001	0.0000
Bodo	0.0007	0.0008	0.0004	0.0002	0.0000
Kashmiri	0.0004	0.0003	0.0001	0.0000	0.0000
Sanskrit	0.0000	0.0000	0.0000	0.0006	0.0000
Dogri	0.0000	0.0000	0.0003	0.0000	0.0000

Table 7: Medium of Instruction at School for Rural India (Age Distribution)

Languages	5-10	11-15	16-20	21-25	26-29
Hindi	0.9663	0.9632	0.9758	0.8994	0.7985
Gujrati	0.0000	0.0000	0.0000	0.0000	0.0000
Sindhi	0.0000	0.0000	0.0000	0.0000	0.0000
Bengali	0.0000	0.0000	0.0000	0.0000	0.0000
Telugu	0.0000	0.0000	0.0000	0.0000	0.0000
Oriya	0.0000	0.0000	0.0000	0.0000	0.0000
Marathi	0.0000	0.0000	0.0000	0.0000	0.0000
Others	0.0000	0.0000	0.0000	0.0000	0.0000
Urdu	0.0000	0.0000	0.0000	0.0000	0.0000
Kannada	0.0000	0.0000	0.0000	0.0000	0.0000
English	0.0337	0.0368	0.0242	0.1006	0.2015
Konkani	0.0000	0.0000	0.0000	0.0000	0.0000
Maithili	0.0000	0.0000	0.0000	0.0000	0.0000
Tamil	0.0000	0.0000	0.0000	0.0000	0.0000

Table 8: Medium of Instruction at School for Rural Chhattisgarh (Age Distribution)

Languages(Age)	5-10	11-15	16-20	21-25	26-29
Gujrati	0.0516	0.0475	0.0556	0.0460	0.0229
Hindi	0.4466	0.4363	0.4226	0.4422	0.4968
Urdu	0.0343	0.0295	0.0186	0.0147	0.0093
Malayalam	0.0387	0.0424	0.0505	0.0411	0.0475
Sindhi	0.0023	0.0024	0.0045	0.0028	0.0005
English	0.0031	0.0037	0.0060	0.0074	0.0112
Bengali	0.0572	0.0653	0.0570	0.0741	0.0491
Tamil	0.0922	0.0969	0.0865	0.0666	0.0762
Oriya	0.0178	0.0187	0.0186	0.0184	0.0267
Assamese	0.0074	0.0067	0.0067	0.0143	0.0023
Maithili	0.0046	0.0029	0.0036	0.0031	0.0048
Santhali	0.0003	0.0004	0.0001	0.0009	0.0000
Others	0.0344	0.0323	0.0307	0.0270	0.0160
Marathi	0.0706	0.0710	0.0803	0.1020	0.1087
Punjabi	0.0218	0.0224	0.0337	0.0318	0.0274
Nepali	0.0017	0.0016	0.0031	0.0021	0.0069
Telugu	0.0755	0.0744	0.0772	0.0628	0.0644
Kannada	0.0308	0.0371	0.0341	0.0297	0.0156
Manipuri	0.0031	0.0025	0.0022	0.0028	0.0034
Konkani	0.0018	0.0025	0.0036	0.0024	0.0004
Bodo	0.0000	0.0000	0.0001	0.0000	0.0011
Kashmiri	0.0032	0.0026	0.0034	0.0052	0.0056
Sanskrit	0.0001	0.0000	0.0000	0.0006	0.0000
Dogri	0.0009	0.0009	0.0013	0.0027	0.0035

Table 9: Language Spoken at Home for Urban India (Age Distribution)

Languages	5-10	11-15	16-20	21-25	26-29
Hindi	0.7573	0.6934	0.7153	0.7514	0.2343
Gujrati	0.0000	0.0000	0.0000	0.0000	0.0000
Sindhi	0.0000	0.0000	0.0000	0.0000	0.0000
Bengali	0.0000	0.0000	0.0000	0.0000	0.0000
Telugu	0.0000	0.0000	0.0000	0.0000	0.0000
Oriya	0.0242	0.0046	0.0125	0.0072	0.0000
Marathi	0.0000	0.0000	0.0000	0.0000	0.0000
Others	0.2168	0.3020	0.2721	0.2414	0.7657
Urdu	0.0000	0.0000	0.0000	0.0000	0.0000
Kannada	0.0000	0.0000	0.0000	0.0000	0.0000
English	0.0000	0.0000	0.0000	0.0000	0.0000
Konkani	0.0000	0.0000	0.0000	0.0000	0.0000
Maithili	0.0000	0.0000	0.0000	0.0000	0.0000
Tamil	0.0017	0.0000	0.0000	0.0000	0.0000

Table 10: Language Spoken at Home for Urban Chhattisgarh (Age Distribution)

Languages	5-10	11-15	16-20	21-25	26-29
Gujrati	0.0409	0.0393	0.0342	0.0192	0.0122
Hindi	0.2472	0.2730	0.2723	0.2288	0.2210
Urdu	0.0206	0.0169	0.0051	0.0016	0.0000
Malayalam	0.0125	0.0173	0.0037	0.0011	0.0000
Sindhi	0.0000	0.0002	0.0000	0.0000	0.0000
English	0.5063	0.4286	0.5532	0.6543	0.6844
Bengali	0.0415	0.0544	0.0420	0.0396	0.0058
Tamil	0.0303	0.0521	0.0229	0.0027	0.0000
Oriya	0.0129	0.0148	0.0030	0.0012	0.0026
Assamese	0.0068	0.0048	0.0023	0.0021	0.0000
Maithili	0.0000	0.0001	0.0002	0.0000	0.0000
Santhali	0.0000	0.0000	0.0000	0.0009	0.0000
Others	0.0007	0.0010	0.0002	0.0002	0.0011
Marathi	0.0415	0.0489	0.0312	0.0307	0.0533
Punjabi	0.0048	0.0054	0.0097	0.0053	0.0080
Nepali	0.0000	0.0000	0.0000	0.0000	0.0000
Telugu	0.0150	0.0191	0.0102	0.0071	0.0021
Kannada	0.0182	0.0226	0.0098	0.0058	0.0097
Manipuri	0.0003	0.0012	0.0000	0.0000	0.0000
Konkani	0.0002	0.0000	0.0000	0.0001	0.0000
Bodo	0.0003	0.0000	0.0000	0.0000	0.0000
Kashmiri	0.0001	0.0000	0.0000	0.0000	0.0000
Sanskrit	0.0000	0.0003	0.0001	0.0002	0.0000
Dogri	0.0000	0.0000	0.0000	0.0000	0.0000

Table 11: Medium of Instruction at School for Urban India (Age Distribution)

Languages	5-10	11-15	16-20	21-25	26-29
Hindi	0.9663	0.9632	0.9758	0.8994	0.7985
Gujrati	0.0000	0.0000	0.0000	0.0000	0.0000
Sindhi	0.0000	0.0000	0.0000	0.0000	0.0000
Bengali	0.0000	0.0000	0.0000	0.0000	0.0000
Telugu	0.0000	0.0000	0.0000	0.0000	0.0000
Oriya	0.0000	0.0000	0.0000	0.0000	0.0000
Marathi	0.0000	0.0000	0.0000	0.0000	0.0000
Others	0.0000	0.0000	0.0000	0.0000	0.0000
Urdu	0.0000	0.0000	0.0000	0.0000	0.0000
Kannada	0.0000	0.0000	0.0000	0.0000	0.0000
English	0.0337	0.0368	0.0242	0.1006	0.2015
Konkani	0.0000	0.0000	0.0000	0.0000	0.0000
Maithili	0.0000	0.0000	0.0000	0.0000	0.0000
Tamil	0.0000	0.0000	0.0000	0.0000	0.0000

Table 12: Medium of Instruction at School for Urban Chhattisgarh (Age Distribution)

R Codes and Outputs

First let's import the required libraries.

```
library(haven)
library(dplyr)
library(rstatix)
library(ggplot2)
```

Then we have imported the data, subsetted the data with respect to the state Chattisgarh. Again subsetted with respect to age, as we require this for age distribution.

```
data1 = read_dta(file.choose())
df1 = read_dta(file.choose())
df2 = df1
df1 = df2[,c("HH_ID","psrl_no","sex")]
result <- merge(data1, df1, by = c("HH_ID", "psrl_no"))
data1 = result
data1$lang_home = as_factor(data1$lang_home)
data1$medium_instruction = as_factor(data1$medium_instruction)
data1$sector = as_factor(data1$sector)
data1$indicator = ifelse(as.numeric(data1$medium_instruction) == as.numeric(data1$lang_home),
                         1,
                         0)
data1$sex = as_factor(data1$sex)
data1 <- data1 %>% arrange(HH_ID)

dat_cha = subset(data1, state_cd == "22")

estimate = function(data) {
  sum(data$wgt_combined)
}

data1R = subset(data1, sector == "Rural")
data1U = subset(data1, sector == "Urban")
dat_chaR = subset(dat_cha, sector == "Rural")
```

```
dat_chaU = subset(dat_cha, sector == "Urban")
data1M = subset(data1, sex == "Male")
data1F = subset(data1, sex == "Female")
dat_chaM = subset(dat_cha, sex == "Male")
dat_chaF = subset(dat_cha, sex == "Female")
data1RM = subset(data1R, sex == "Male")
data1RF = subset(data1R, sex == "Female")
data1UM = subset(data1U, sex == "Male")
data1UF = subset(data1U, sex == "Female")
dat_chaRM = subset(dat_chaR, sex == "Male")
dat_chaRF = subset(dat_chaR, sex == "Female")
dat_chaUM = subset(dat_chaU, sex == "Male")
dat_chaUF = subset(dat_chaU, sex == "Female")
```

Then we are just seeing the data:

```
head(data1)
```

```
##       HH_ID psrl_no centre   fsu round sch sample sector nssregion state_cd
## 1 110001101      03    008 11000    71 252      1  Urban       242       24
## 2 110001101      04    009 11000    71 252      1  Urban       242       24
## 3 110001102      09    014 11000    71 252      1  Urban       242       24
## 4 110001201      05    012 11000    71 252      1  Urban       242       24
## 5 110001202      06    009 11000    71 252      1  Urban       242       24
## 6 110001203      03    007 11000    71 252      1  Urban       242       24
##   district state_district stratum sstratum sround ssample  fod hamlet_grp sss
## 1       07           2407     035       13      1       1 2410          1   1
## 2       07           2407     035       13      1       1 2410          1   1
## 3       07           2407     035       13      1       1 2410          1   1
## 4       07           2407     035       13      1       1 2410          1   2
## 5       07           2407     035       13      1       1 2410          1   2
## 6       07           2407     035       13      1       1 2410          1   2
##   hhsno level age age_entry levl_curr_attend course_cd acad_session
## 1    01    05  20         5               13         6            6
## 2    01    05  19         5               15         4            6
## 3    02    05  18         5               13        14            6
## 4    01    05   7         5                7         1           12
## 5    02    05   6         5                7         1           12
## 6    03    05  19         5               11         5            6
##   present_class inst_type nature_inst reason_private medium_instruction
## 1             2         2          NA              5            English
## 2             2         2          NA              5            English
## 3             2         3          NA              5            Gujrati
## 4             2         2          NA              2            Gujrati
## 5             2         2          NA              3            English
## 6             2         2          NA              5            English
##   lang_home course_type edu_free tut_fee_wav amt_wav reason_wav scholarship
## 1   Gujrati           1        2           3      NA         NA           2
## 2   Gujrati           1        2           3      NA         NA           2
## 3   Gujrati           1        2           3      NA         NA           2
## 4   Gujrati           1        2           3      NA         NA           2
## 5   Gujrati           1        2           3      NA         NA           2
## 6   Gujrati           1        2           3      NA         NA           2
##   schlor_amt schlor_type scholr_agency textbooks stationery midday_meal
## 1         NA          NA            NA         6          6           2
```

```
## 2         NA          NA            NA         6          6           2
## 3         NA          NA            NA         6          6           2
## 4         NA          NA            NA         6          6           2
## 5         NA          NA            NA         6          6           2
## 6         NA          NA            NA         6          6           2
##   meal_agency mode_transport concession dist_inst changed_edu_inst
## 1          NA              3          1         5                1
## 2          NA              3          1         5                1
## 3          NA              3          1         5                1
## 4          NA              1         NA         1                1
## 5          NA              2         NA         5                1
## 6          NA              3          1         5                1
##   private_coaching purpose_priv_coach nss nsc    mlt  wgt_ss wgt_combined
## 1                2                 NA   1   2 870573 8705.73     4352.865
## 2                2                 NA   1   2 870573 8705.73     4352.865
## 3                2                 NA   1   2 870573 8705.73     4352.865
## 4                2                 NA   1   2 856531 8565.31     4282.655
## 5                2                 NA   1   2 856531 8565.31     4282.655
## 6                2                 NA   1   2 856531 8565.31     4282.655
##      sex indicator
## 1 Female         0
## 2 Female         0
## 3 Female         1
## 4 Female         1
## 5 Female         0
## 6   Male         0
```

```
head(dat_cha)
```

```
##          HH_ID psrl_no centre   fsu round sch sample sector nssregion state_cd
## 2214 112711101      05    016 11271    71 252      1  Urban       222       22
## 2215 112711101      06    017 11271    71 252      1  Urban       222       22
## 2216 112711101      09    018 11271    71 252      1  Urban       222       22
## 2217 112711101      10    019 11271    71 252      1  Urban       222       22
## 2218 112711201      03    008 11271    71 252      1  Urban       222       22
## 2219 112711201      04    009 11271    71 252      1  Urban       222       22
##      district state_district stratum sstratum sround ssample  fod hamlet_grp
## 2214       11           2211     026       02      2       2 2210          1
## 2215       11           2211     026       02      2       2 2210          1
## 2216       11           2211     026       02      2       2 2210          1
## 2217       11           2211     026       02      2       2 2210          1
## 2218       11           2211     026       02      2       2 2210          1
## 2219       11           2211     026       02      2       2 2210          1
##      sss hhsno level age age_entry levl_curr_attend course_cd acad_session
## 2214   1    01    05  21         5               15         4           12
## 2215   1    01    05  18         5               11         3           12
## 2216   1    01    05  16         5               10         1           12
## 2217   1    01    05  18         5               13        12           12
## 2218   2    01    05  13         5                8         1           12
## 2219   2    01    05  11         5                8         1           12
##      present_class inst_type nature_inst reason_private medium_instruction
## 2214             2         3          NA              1              Hindi
## 2215             2         2          NA              2            English
## 2216             2         2          NA              2            English
## 2217             2         4          NA             NA              Hindi
```

```
## 2218            2         3           1              3            English
## 2219            2         3           1              3            English
##      lang_home course_type edu_free tut_fee_wav amt_wav reason_wav scholarship
## 2214     Hindi           2        2           3      NA         NA           2
## 2215     Hindi           1        2           3      NA         NA           2
## 2216     Hindi           1        2           3      NA         NA           2
## 2217     Hindi           2        2           3      NA         NA           2
## 2218     Hindi           1        2           3      NA         NA           2
## 2219     Hindi           1        2           3      NA         NA           2
##      schlor_amt schlor_type scholr_agency textbooks stationery midday_meal
## 2214         NA          NA            NA         1          1           2
## 2215         NA          NA            NA         1          1           2
## 2216         NA          NA            NA         1          1           2
## 2217         NA          NA            NA         1          6           2
## 2218         NA          NA            NA         1          1           2
## 2219         NA          NA            NA         1          1           2
##      meal_agency mode_transport concession dist_inst changed_edu_inst
## 2214          NA              1         NA         1                1
## 2215          NA              1         NA         1                1
## 2216          NA              1         NA         1                1
## 2217          NA              1         NA         1                1
## 2218          NA              9         NA         2                1
## 2219          NA              9         NA         2                1
##      private_coaching purpose_priv_coach nss nsc    mlt  wgt_ss wgt_combined
## 2214                2                 NA   1   2 339553 3395.53     1697.765
## 2215                1                  1   1   2 339553 3395.53     1697.765
## 2216                1                  1   1   2 339553 3395.53     1697.765
## 2217                2                 NA   1   2 339553 3395.53     1697.765
## 2218                1                  1   1   2 463027 4630.27     2315.135
## 2219                1                  1   1   2 463027 4630.27     2315.135
##        sex indicator
## 2214   Male         1
## 2215   Male         0
## 2216   Male         0
## 2217 Female         1
## 2218   Male         0
## 2219   Male         0
```

Then we are extracting all the languages used in Chattisgarh and India

```
alllang_cha = union(unique(dat_cha$lang_home),
                    unique(dat_cha$medium_instruction))

alllang_ind = union(unique(data1$lang_home),
                    unique(data1$medium_instruction))

alllang_cha
```

```
##  [1] "Hindi"    "Gujrati"  "Sindhi"   "Bengali"  "Telugu"   "Oriya"
##  [7] "Marathi"  "Others"   "Urdu"     "Kannada"  "English"  "Konkani"
## [13] "Maithili" "Tamil"
```

```
alllang_ind
```

```
##  [1] "Gujrati"   "Hindi"     "Urdu"      "Malayalam" "Sindhi"    "English"
##  [7] "Bengali"   "Tamil"     "Oriya"     "Assamese"  "Maithili"  "Santhali"
```

```
## [13] "Others"    "Marathi"   "Punjabi"   "Nepali"    "Telugu"    "Kannada"
## [19] "Manipuri"  "Konkani"   "Bodo"      "Kashmiri"  "Sanskrit"  "Dogri"
```

The following function will return the estimate matrix

```
estimatelang = function(data, alllang){
  A = matrix(0,nrow=length(alllang),ncol=length(alllang))
  for(i in 1:length(alllang)){
    for(j in 1:length(alllang)){
      A[i,j] = estimate(subset(data, lang_home == alllang[i]
                           & medium_instruction == alllang[j]))
    }
  }
  colnames(A)<- rownames(A) <- alllang
  return(A)
}
```

The following function will rerurn the variance matrix

```
variancelang1lang2 = function(data, lang1, lang2){
  R = (estimate(subset(data, lang_home == lang1
                           & medium_instruction == lang2))/estimate(data))
  data_split <- data %>%
    df_split_by(stratum) #stratum
  var = 0
  for(s in 1:length(data_split$data)){
    vars = 0
    data_s = data_split$data[[s]]
    data_s_split = data_s %>% df_split_by(sstratum) #substratum
    for(t in 1:length(data_s_split$data)){
      varst = 0
      data_st = data_s_split$data[[t]]
      data_st_split = data_st %>% df_split_by(ssample) #sub sample
      yest1 = estimate(subset(data_st_split$data[[1]], lang_home
                           == lang1 & medium_instruction == lang2))
      xest1 = estimate(data_st_split$data[[1]])
      if(length(data_st_split$data) < 2) yest2 = xest2 = 0
      else{
        yest2 = estimate(subset(data_st_split$data[[2]], lang_home
                           == lang1 & medium_instruction == lang2))
        xest2 = estimate(data_st_split$data[[2]])
      }
      yest = yest1 - yest2
      xest = xest1 - xest2
      varst = yest^2 +R^2*xest^2-2*R*yest*xest
      vars = vars + varst
    }
    var = var + vars
  }
  return(var/(4*estimate(data)^2))
}


variancelang = function(data, alllang){
  A = matrix(0,nrow=length(alllang),ncol=length(alllang))
  for(i in 1:length(alllang)){
```

```
    for(j in 1:length(alllang)){
      A[i,j] = variancelang1lang2(data, alllang[i], alllang[j])
    }
    print(i)
  }
  colnames(A)<- rownames(A) <- alllang
  return(A)
}
```

Now, we will be doing age wise analysis.

```
# Rural + Urban, Country and State
dat_1 = subset(data1, age %in% c(5:10))
dat_2 = subset(data1, age %in% c(11:15))
dat_3 = subset(data1, age %in% c(16:20))
dat_4 = subset(data1, age %in% c(21:25))
dat_5 = subset(data1, age %in% c(26:29))

data1 <- data1 %>% arrange(HH_ID)

dat_cha = subset(data1, state_cd == "22")

dat_cha_1 = subset(dat_cha, age %in% c(5:10))
dat_cha_2 = subset(dat_cha, age %in% c(11:15))
dat_cha_3 = subset(dat_cha, age %in% c(16:20))
dat_cha_4 = subset(dat_cha, age %in% c(21:25))
dat_cha_5 = subset(dat_cha, age %in% c(26:29))


# Rural, Country and State
data1R = subset(data1, sector == "Rural")
dat_1R = subset(data1R, age %in% c(5:10))
dat_2R = subset(data1R, age %in% c(11:15))
dat_3R = subset(data1R, age %in% c(16:20))
dat_4R = subset(data1R, age %in% c(21:25))
dat_5R = subset(data1R, age %in% c(26:29))

dat_chaR = subset(dat_cha, sector == "Rural")

dat_cha_1R = subset(dat_chaR, age %in% c(5:10))
dat_cha_2R = subset(dat_chaR, age %in% c(11:15))
dat_cha_3R = subset(dat_chaR, age %in% c(16:20))
dat_cha_4R = subset(dat_chaR, age %in% c(21:25))
dat_cha_5R = subset(dat_chaR, age %in% c(26:29))

# Urban, Country and State
data1U = subset(data1, sector == "Urban")
dat_1U = subset(data1U, age %in% c(5:10))
dat_2U = subset(data1U, age %in% c(11:15))
dat_3U = subset(data1U, age %in% c(16:20))
dat_4U = subset(data1U, age %in% c(21:25))
dat_5U = subset(data1U, age %in% c(26:29))

dat_chaU = subset(dat_cha, sector == "Urban")
```

```
dat_cha_1U = subset(dat_chaR, age %in% c(5:10))
dat_cha_2U = subset(dat_chaR, age %in% c(11:15))
dat_cha_3U = subset(dat_chaR, age %in% c(16:20))
dat_cha_4U = subset(dat_chaR, age %in% c(21:25))
dat_cha_5U = subset(dat_chaR, age %in% c(26:29))
```

Now, let's have the estimates:

```
(estimate_data_1 = sum(dat_1$indicator*dat_1$wgt_combined))
```

```
## [1] 67982696
```

```
(estimate_data_2 = sum(dat_2$indicator*dat_2$wgt_combined))
```

```
## [1] 66802151
```

```
(estimate_data_3 = sum(dat_3$indicator*dat_3$wgt_combined))
```

```
## [1] 32887341
```

```
(estimate_data_4 = sum(dat_4$indicator*dat_4$wgt_combined))
```

```
## [1] 5557194
```

```
(estimate_data_5 = sum(dat_5$indicator*dat_5$wgt_combined))
```

```
## [1] 471477.2
```

```
(estimate_dat_cha_1 = sum(dat_cha_1$indicator*dat_cha_1$wgt_combined))
```

```
## [1] 1527056
```

```
(estimate_dat_cha_2 = sum(dat_cha_2$indicator*dat_cha_2$wgt_combined))
```

```
## [1] 1530055
```

```
(estimate_dat_cha_3 = sum(dat_cha_3$indicator*dat_cha_3$wgt_combined))
```

```
## [1] 740719.2
```

```
(estimate_dat_cha_4 = sum(dat_cha_4$indicator*dat_cha_4$wgt_combined))
```

```
## [1] 107415.4
```

```
(estimate_dat_cha_5 = sum(dat_cha_5$indicator*dat_cha_5$wgt_combined))
```

```
## [1] 4031.05
```

```
#Estimate and Variance of the Percentages

estimatelang = function(dat_cha) {
  sum(dat_cha$wgt_combined)
}


variancelang = function(dat_cha) {
  dat_cha_split <- dat_cha %>%
    df_split_by(stratum) #stratum
  e = 0
  for (s in 1:length(dat_cha_split$data)) {
    es = 0
```

```
    dat_cha_s = dat_cha_split$data[[s]]
    dat_cha_s_split = dat_cha_s %>% df_split_by(sstratum) #substratum
    for (t in 1:length(dat_cha_s_split$data)) {
      est = 0
      dat_cha_st = dat_cha_s_split$data[[t]]
      dat_cha_st_split = dat_cha_st %>% df_split_by(ssample) #sub sample
      for (m in 1:length(dat_cha_st_split$data)) {
        estm = 0
        dat_cha_stm = dat_cha_st_split$data[[m]]
        dat_cha_stm_split = dat_cha_stm %>% df_split_by(sss) #second stage stratum
        for (j in 1:length(dat_cha_stm_split$data)) {
          estmj = 0
          dat_cha_stmj = dat_cha_stm_split$data[[j]]
          dat_cha_stmj_split = dat_cha_stmj %>% df_split_by(fsu) #fsu
          for (i in 1:length(dat_cha_stmj_split$data)) {
            estmji = 0
            dat_cha_stmji = dat_cha_stmj_split$data[[i]]
            dat_cha_stmji_split = dat_cha_stmji %>% df_split_by(hamlet_grp) #hamlet_grp
            for (d in 1:length(dat_cha_stmji_split$data)) {
              estmjid = 0
              dat_cha_stmjid = dat_cha_stmji_split$data[[d]]
              estmjid = sum(dat_cha_stmjid$wgt_ss) #all households
              estmji = estmji + estmjid
            }
            estmj = estmj + estmji
          }
          estm = estm + estmj
        }
        est = (est - estm) ^ 2
      }
      es = es + est
    }
    e = e + es
  }
  return(e / 4)
}


#Extracting all the languages

alllang_cha = union(unique(dat_cha$lang_home),
                    unique(dat_cha$medium_instruction))

alllang_ind = union(unique(data1$lang_home),
                    unique(data1$medium_instruction))


#Function for giving the percentages language wise
langhomestatistics = function(data, alllang) {
  vect = numeric(length(alllang))
  k = 0
  for (lang in alllang) {
    k = k + 1
```

```
    datlang = subset(data, lang_home == lang)
    if (dim(datlang)[1] == 0)
      vect[k] = 0
    else
      vect[k] = estimatelang(datlang)
  }
  return (vect/sum(vect))
}

langschoolstatistics = function(data, alllang) {
  vect = numeric(length(alllang))
  k = 0
  for (lang in alllang) {
    k = k + 1
    datlang = subset(data, medium_instruction == lang)
    if (dim(datlang)[1] == 0)
      vect[k] = 0
    else
      vect[k] = estimatelang(datlang)
  }
  return (vect/sum(vect))
}


#Making the data frames
#Rural + Urban (Country)

langhome_ind = as.data.frame(cbind(
  alllang_ind,
  langhomestatistics(dat_1, alllang_ind),
  langhomestatistics(dat_2, alllang_ind),
  langhomestatistics(dat_3, alllang_ind),
  langhomestatistics(dat_4, alllang_ind),
  langhomestatistics(dat_5, alllang_ind)
))

langschool_ind = as.data.frame(cbind(
  alllang_ind,
  langschoolstatistics(dat_1, alllang_ind),
  langschoolstatistics(dat_2, alllang_ind),
  langschoolstatistics(dat_3, alllang_ind),
  langschoolstatistics(dat_4, alllang_ind),
  langschoolstatistics(dat_5, alllang_ind)
))

colnames(langhome_ind) = colnames(langschool_ind) = c("languages","5-10","11-15",
                                                      "16-20","21-25","26-29")

#Rural (Country)

langhome_indR = as.data.frame(cbind(
  alllang_ind,
  langhomestatistics(dat_1R, alllang_ind),
```

```
  langhomestatistics(dat_2R, alllang_ind),
  langhomestatistics(dat_3R, alllang_ind),
  langhomestatistics(dat_4R, alllang_ind),
  langhomestatistics(dat_5R, alllang_ind)
))

langschool_indR = as.data.frame(cbind(
  alllang_ind,
  langschoolstatistics(dat_1R, alllang_ind),
  langschoolstatistics(dat_2R, alllang_ind),
  langschoolstatistics(dat_3R, alllang_ind),
  langschoolstatistics(dat_4R, alllang_ind),
  langschoolstatistics(dat_5R, alllang_ind)
))

colnames(langhome_indR) = colnames(langschool_indR) = c("languages","5-10","11-15",
                                                        "16-20","21-25","26-29")

#Urban (Country)

langhome_indU = as.data.frame(cbind(
  alllang_ind,
  langhomestatistics(dat_1U, alllang_ind),
  langhomestatistics(dat_2U, alllang_ind),
  langhomestatistics(dat_3U, alllang_ind),
  langhomestatistics(dat_4U, alllang_ind),
  langhomestatistics(dat_5U, alllang_ind)
))

langschool_indU = as.data.frame(cbind(
  alllang_ind,
  langschoolstatistics(dat_1U, alllang_ind),
  langschoolstatistics(dat_2U, alllang_ind),
  langschoolstatistics(dat_3U, alllang_ind),
  langschoolstatistics(dat_4U, alllang_ind),
  langschoolstatistics(dat_5U, alllang_ind)
))

colnames(langhome_indU) = colnames(langschool_indU) = c("languages","5-10","11-15",
                                                        "16-20","21-25","26-29")

#Rural + Urban (State)
langhome_cha = as.data.frame(cbind(
  alllang_cha,
  langhomestatistics(dat_cha_1, alllang_cha),
  langhomestatistics(dat_cha_2, alllang_cha),
  langhomestatistics(dat_cha_3, alllang_cha),
  langhomestatistics(dat_cha_4, alllang_cha),
  langhomestatistics(dat_cha_5, alllang_cha)
))

langschool_cha = as.data.frame(cbind(
  alllang_cha,
```

```
  langschoolstatistics(dat_cha_1, alllang_cha),
  langschoolstatistics(dat_cha_2, alllang_cha),
  langschoolstatistics(dat_cha_3, alllang_cha),
  langschoolstatistics(dat_cha_4, alllang_cha),
  langschoolstatistics(dat_cha_5, alllang_cha)
))

colnames(langhome_cha) = colnames(langschool_cha) = c("languages","5-10","11-15",
                                                      "16-20","21-25","26-29")

#Rural (State)

langhome_chaR = as.data.frame(cbind(
  alllang_cha,
  langhomestatistics(dat_cha_1R, alllang_cha),
  langhomestatistics(dat_cha_2R, alllang_cha),
  langhomestatistics(dat_cha_3R, alllang_cha),
  langhomestatistics(dat_cha_4R, alllang_cha),
  langhomestatistics(dat_cha_5R, alllang_cha)
))

langschool_chaR = as.data.frame(cbind(
  alllang_cha,
  langschoolstatistics(dat_cha_1R, alllang_cha),
  langschoolstatistics(dat_cha_2R, alllang_cha),
  langschoolstatistics(dat_cha_3R, alllang_cha),
  langschoolstatistics(dat_cha_4R, alllang_cha),
  langschoolstatistics(dat_cha_5R, alllang_cha)
))

colnames(langhome_chaR) = colnames(langschool_chaR) = c("languages","5-10","11-15",
                                                        "16-20","21-25","26-29")
#Urban (Country)

langhome_chaU = as.data.frame(cbind(
  alllang_cha,
  langhomestatistics(dat_cha_1U, alllang_cha),
  langhomestatistics(dat_cha_2U, alllang_cha),
  langhomestatistics(dat_cha_3U, alllang_cha),
  langhomestatistics(dat_cha_4U, alllang_cha),
  langhomestatistics(dat_cha_5U, alllang_cha)
))

langschool_chaU = as.data.frame(cbind(
  alllang_cha,
  langschoolstatistics(dat_cha_1U, alllang_cha),
  langschoolstatistics(dat_cha_2U, alllang_cha),
  langschoolstatistics(dat_cha_3U, alllang_cha),
  langschoolstatistics(dat_cha_4U, alllang_cha),
  langschoolstatistics(dat_cha_5U, alllang_cha)
))

colnames(langhome_chaU) = colnames(langschool_chaU) = c("languages","5-10","11-15",
```

```
                                                    "16-20","21-25","26-29")
```

Conclusion

- For Chhattisgarh, inspite of heterogeneity in home spoken language, only Hindi and English (and, Kannada) are used in school.
- Quite a few Hindi spoken people learn their native language in school, fraction of english medium students is more in Urban areas
- SE is very less, that means that the ratios are accurate enough.
- For the whole India, lot of heterogeneity in medium of instructions are also observed, which means more fractions learn their native language in school.
- The main reason for private tuition is "Better Environment of Learning", be it Chhattisgarh, or whole country or rural or urban, this never changes.
- The most spoken languages in the country are Hindi, Bengali, Marathi, Tamil, Telegu etc..
- Medium of Instructions is evenly distributed among the languages in spite of lot of Hindi spoken people.

References

- https://microdata.gov.in/nada43/index.php/catalog/136