Multidimensional poverty index

An improvement in the existing framework

Bachelor of Statistics (Hons.)

INDIAN STATISTICAL INSTITUTE, KOLKATA

Mrinal Saini (BS1920)
Anurag Pal (BS1923)
Nalla Ritwik Narayana (BS1926)
Mridul Maheshwari (BS1929)
Gokavarapu Siva Lokesh (BS1930)

Under the supervision of

Prof. Bimal Kumar Roy
Prof. Sourabh Ghosh

Contents

Section 1

Introduction

1.1 What is Multidimensional Poverty

Poverty is often defined by one-dimensional measures - usually based on income. But no single indicator can capture the multiple dimensions of poverty.

Multidimensional poverty encompasses the various deprivations experienced by poor people in their daily lives – such as poor health, lack of education, inadequate living standards, disempowerment, poor quality of work, among others.

A multidimensional measure of poverty can incorporate a range of indicators that capture the complexity of this phenomena in order to inform policies aimed at reducing poverty and deprivation in a country.

1.2 Aim

To devise a method to check the reliability of the NITI Aayog's Method of Multidimensional Poverty measurement. We will also develop a few of our own models of Multidimensional Poverty measurement and compare their reliabilities.

1.3 Background

Multidimensional Poverty Index developed by Sabina Alkire and James Foster [1] was adopted by United Nations Development Programme in their Human Development Report. [2]

The AF Methodology is a general framework for measuring multidimensional poverty that identifies people as poor or not poor based on dual cutoff counting method. The first order cutoff within each component is applied to determine which person is "deprived" in that indicator. The information across all indicators is then aggregated to arrive at a deprivation score for each individual. The second order cutoff

is then applied to identify the individuals who are multidimensionally poor. NITI aayog is applying AF methods for determining poverty.
Because of the following features of Alkire-Foster Method, we will not try to give a new method of Multidimensional Poverty Measurement instead we will try to improve upon it.

- **Provides incentives for leaving no one behind and reaching the furthest behind first:** By reflecting the intensity of poverty (detailing the multiple deprivations that a family has at the same time), the national MPI has an advantage over headcount poverty measures since efforts to reduce the proportion of simultaneous hardships faced by the poor will reduce the MPI even if they have not yet moved out of poverty
- **Information to shape policy:** it allows robust disaggregation by groups (such as between urban and rural areas, subnational regions, gender, age groups,). One can also unpack the numbers to analyse the composition of poverty by dimensions and indicators – nationally, and at the level of States and districts, which allows for more efficient policy design, policy coordination and focus, and assignment of resources.

1.4 The Data Source

We did sample survey to collect data of 134 households across 3 Gram Panchayats in Giridih District.
We obtained the National Family Health Survey (NHFS) data by obtaining special access from Demographic and Health Survey(DHS-USAid). [3]

Section 2

Survey

2.1 Sampling

Our objective in designing the survey was to get a sample which is representative of population. To get a representaive sample was not difficult (taking n households from total population could have given us the sample) but to systematically build and do inference we adopted two stage sampling design. Our total population was all households in Giridih.

2.1.1 Stage-1:

Panchayats were classified in three stratas. The classification was based on distance from any block headquarters (not necessarily Giridih).

1. Strata I $<$ 5 Km
2. 5 Km $<$ Strata II $<$ 10 Km
3. Strata III $>$ 10 Km

One Panchayat was chosen from each class.

Once a Panchayat was chosen in each class, a village or two were selected in the panchayat to do the sampling.
(Choice of panchayat and village was done keeping in mind that the selected panchayats are in different directions from Giridih Headquarters and also availability of a key resource person in the village was an important deciding factor)

2.1.2 Stage-2:

For a particular choice of village we were provided by a list of households by the village headman (Mukhiya). Each village was divided into hamlets where a particular community members resided. Number of households to be sampled from each hamlet was proportion to their size in total population of village.

Selection of Households:

We were divided into four teams and each hamlet was divided into four parts (based on street numbers). Each team collected approximately equal number of samples. Households were selected at the discretion of the team.

Sample Size Determination

Given the time period in which we had to conclude our sampling we had an idea of how many households we could survey. Dividing them equally to the three strata in stage one and proportionally to each hamlet we got how many households we need to sample from each hamlet.
This design (in our judgement) was successful in giving a representative sample.

2.2 Questionnaire

We had following objectives before framing the Questionnaire

1. Privacy of respondents is ensured.

2. Sensitive information (such as debt on the households, possession of jewellery, etc.,) was not included

3. Factors that determine poverty status are to be included whether they can be quantified or not.

4. All Factors which are used in existing framework of NITI Aayog Multidimensional poverty Index are to be included (to ensure that a comparative study can be done).

5. Questionnaire is concise so that dropout rate of respondents is low.

We studied the following sample survey questionnaire

- National Family Health Survey 4 & 5 [4]
- Consumer Expenditure Survey [5]
- Socio-Economic Caste Census (Used in Pradhan Mantri Jan Aawas Yojana) [6]
- We have attached the Questionnaire with the report

2.3 Data Summary

Total number of household from which data was collected = 134
Total number of individuals for whom data was collected = 928
Average household size = 6.92

Stage-1:

Panchayats selected:

Strata	Panchayat	Households
I	Motileda	46
II	Bengabad	41
III	Udnabad	47

Stage-2:

1. Motileda:

   Hamlet 1 : 25 Households

   Hamlet 2 : 21 Households

2. Bengabad:

   Hamlet 1 : 19 Households

   Hamlet 2 : 22 Households

3. Udnabad:

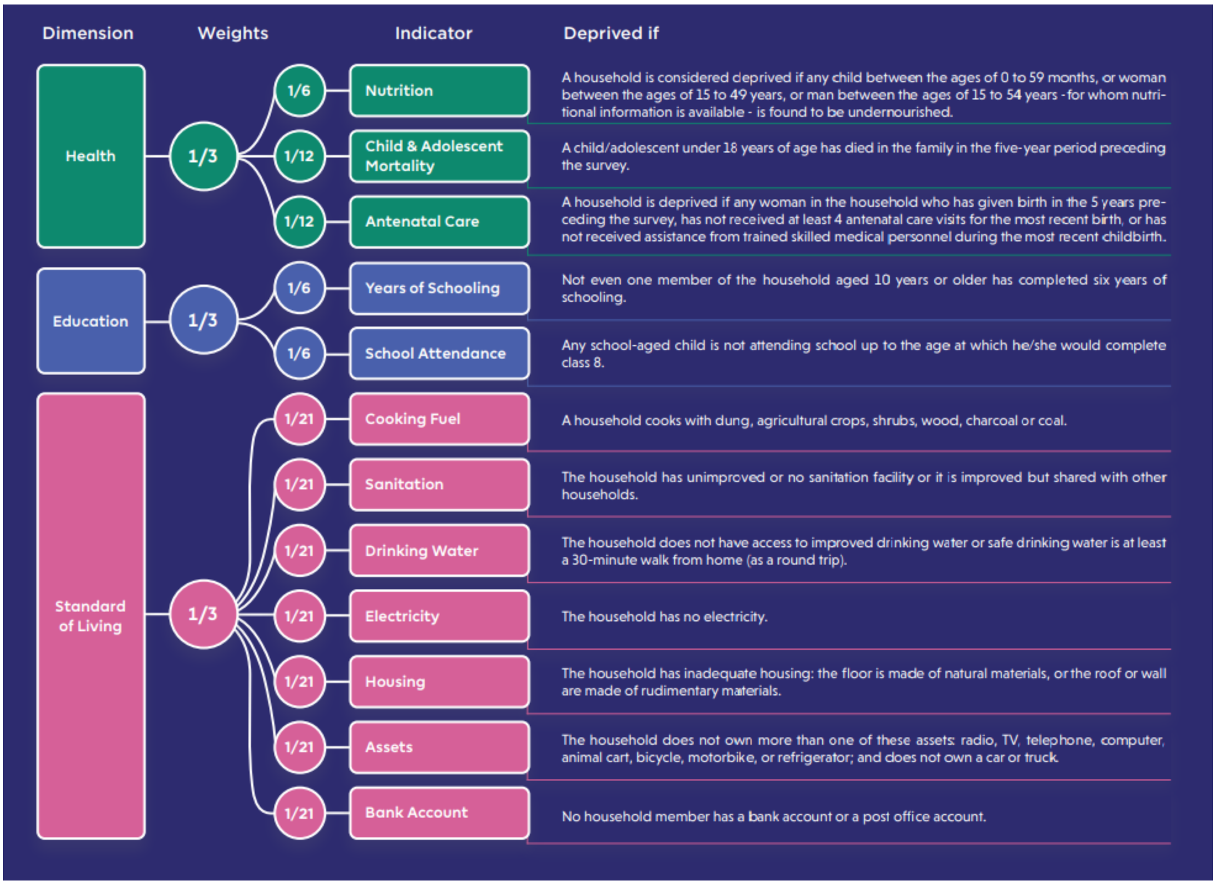
   Hamlet 1 : 25 Households

   Hamlet 2 : 22 Households

Section 3

Methodology

3.1 Existing Indicator weights for NITI Aayog's MPI



3.1.1 Calculations of MPI

Deprivation Status

If the achievement of an individual i in indicator j is denoted by $x_{ij'}$ the first order cutoff for indicator j is denoted by $z_{j'}$ and the status of the individual is denoted as $\mathrm{g}_{ij'}^0$, then $\mathrm{g}_{ij}^0 = 1$ if $x_{ij} < z_j$& $\mathrm{g}_{ij}^0 = 0$ otherwise for all $i = 1, 2 \cdots \mathrm{n}\& j = 1, 2 \cdots \mathrm{d}$

Poverty Cut-off

The identification function for multidimensional poverty is denoted by ρ. The function ρ is dependent on the deprivation status of an individual (x_i) given the cutoffs within an indicator (z) as well as on the cutoffs across indicators (k) and is therefore represented by: $\rho_k(x_i; z) = 1$ if $c_i \geq k$ and $\rho_k(x_i; z) = 0$ otherwise

Therefore, the function ρ considers an individual i as multidimensionally poor when her deprivation score (c_i) is greater than of equal to the second-order cutoff (k).

Calculating deprivation Score

The counting vector for individual i up to the j^{th} indicator (denoted by c_i), also known as deprivation score, is their status in each indicator $\left(\mathrm{g}_{ij}^0\right)$ multiplied by the weight (w_j) assigned to that indicator. The deprivation score (or weighted deprivation) of individual i can thus be denoted as: $c_i = w_1 g_{i1}^0 + w_2 g_{i2}^0 + \ldots + w_j g_{ij}^0$ or $c_i = \sum_{j=1}^{d} w_j g_{ij}^0$ where $\sum_{j=1}^{d} w_j = 1$

Censored Deprivation Score

Censored Deprivation Scores Censored scores are denoted as $c_i(k)$ to differentiate them from deprivation scores ci. Thus, if $c_i < k$, then $c_i(k) = 0$ and if $c_i \geq k$ then $\mathrm{c}_i(k) = \mathrm{c}_i$. Thus, $\mathrm{c}_i(k)$ is the deprivation score of the multidimensionally poor.

Headcount Ratio

$\mathrm{H} = \frac{\mathrm{q}}{n}$ where q is the total number of multidimensionally poor individuals identified (i.e., the total number of individuals for whom $\rho_k(xi; z) = 1$) and n is the total population.

Intensity

$$\mathrm{A} = \frac{1}{\mathrm{q}} \sum_{i=1}^{\mathrm{q}} \mathrm{c}_i(k)$$

Where $c_i(k)$ is the censored deprivation score (i.e. deprivation score of multidimensionally poor individuals) up to the i^{th} individual and q is the number of multidimensionally poor individuals.

Multidimensional Poverty Index

$\mathrm{M}_0 = \mathrm{HxA}$ or $\mathrm{HxA} = \frac{\mathrm{q}}{n} \times \frac{1}{\mathrm{q}} \sum_{i=1}^{\mathrm{q}} \mathrm{c}_i(k) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{c}_i(k) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{\mathrm{d}} \mathrm{w}_j\ \mathrm{g}_{ij}^{0}(k)$ The MPI, therefore, is the share of weighted deprivations faced by multidimensionally poor individuals divided by the total population. The MPI is therefore known as the adjusted headcount ratio.

From NITI Ayog framework we got the following observations

Intensity=0.4402
Headcount Ratio=0.2424
MPI=0.1067

Section 4

Measure for reliability of model

Based on our judgement, we have classified each household as poor or non-poor. For the same households, we applied NITI Aayog's two stage cutoff criteria to decide whether they are multidimensionally poor or not poor. Now we wish to check whether the poverty status given by NITI Aayog is in agreement with our judgement or not. For this we would calculate Cohen's Kappa for both the vectors of poverty status.

4.1 Cohen's Kappa

Cohen's Kappa (κ) is used to measure inter-rater reliability for categorical items. In other words, Cohen's Kappa is a measure for the agreement between two independent raters that are rating the same thing take away the effect of how often the raters may agree by chance. [7]

4.1.1 Evaluating Cohen's Kappa

The definition of κ is

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where p_o is the relative observed agreement between the raters, p_e is the probability of chance agreement between the raters.
If the raters are in complete agreement then $\kappa = 1$, and a score of 0 mean that there is random agreement between the raters. The value of κ can be negative, which can occur by chance if there is no relationship between the ratings of the two raters, or it may reflect a real tendency of the raters to give differing ratings.

For a 2×2 contingency table:

B \ A	**Yes**	**No**
Yes	a	b
No	c	d

$$p_o = \frac{\text{total agreements}}{\text{total observations}} = \frac{a+d}{a+b+c+d}$$

To calculate p_e we first calculate:

$$\begin{aligned} p_y &= \text{probability that both raters would say Yes at random} \\ &= \frac{a+b}{a+b+c+d} \times \frac{a+c}{a+b+c+d} \end{aligned}$$

$$\begin{aligned} p_n &= \text{probability that both raters would say No at random} \\ &= \frac{b+d}{a+b+c+d} \times \frac{c+d}{a+b+c+d} \end{aligned}$$

So the overall random agreement probability p_e is the probability that they agreed on either Yes or No i.e.

$$p_e = p_{Yes} + p_{No}$$

4.1.2 Interpreting Cohen's Kappa

Magnitude guidelines for Cohen's Kappa:

Value of κ \ **Guideline**	**Landis & Koch**	**Fleiss**
< 0.00	Poor	Poor
0.01 to 0.20	Slight	
0.21 to 0.40	Fair	
0.41 to 0.60	Moderate	Fair to Good
0.61 to 0.75	Substantial	
0.76 to 0.80		Excellent
0.80 to 1.00	Almost Perfect	

Section 5

Logistic Regression

Since AF approach is using a linear classifier, and we wish to stay on the same framework we would only suggest improvements in choice of indicators/weights.

5.0.1 Basic setup of logistic regression

We are given a data set containing N points. Each point i consists of a set of m input variables $x_{1,i}...x_{m,i}$ (also called independent variables, predictor variables), and a binary outcome variable Y_i (also known as a dependent variable), i.e. it can assume only the two possible values say 0 and 1 or in our case poor or non-poor. The goal of logistic regression is to use the dataset to create a predictive model of the outcome variable.
In simple binary logistic regression model, we assumed a linear relationship between the predictor variables and the log-odds (also called logit) of the event that $Y_i = 1$. And probability conditioned on the explanatory variables, follows a Bernoulli distribution with parameters p_i, the probability of the outcome of 1 for trial i. Also the expected value of each Y_i is equal to the probability of success pi, which is a general property of the Bernoulli distribution.

$$Logit(p) = \log_b \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m$$

$$Y_i \mid x_{1,i}, \ldots, x_{m,i} \sim \text{Bernoulli}(p_i)$$

5.0.2 Constrained Logistic

Multidimensional poverty index has a weightage scheme which is analogous to a linear classification model. The weights of the indicators signifies their contribution to the deprivation experienced. Having negative coefficient on an indicator in logistic regression would imply that deprivation in that indicator makes a individual relatively well-off when compared to non deprivation in that particular indicator, so we have a constraint on our coefficient in our logistic regression to be non-negative.

5.1 Estimate the parameter

In general setup, the Log-likelihood

$$\begin{aligned}\ell\left(\beta_0, \beta\right) &= \sum_{i=1}^{n} Y_i \log p\left(x_i\right)+\left(1-Y_i\right) \log 1-p\left(x_i\right) \\ &= \sum_{i=1}^{n}\left(Y_i \sum_{j=0}^{p} \beta_j x_{ij}-\log \left(1+e^{\sum_{j=0}^{p} \beta_j x_{ij}}\right)\right)\end{aligned}$$

To estimate the parameters $\beta_0, \ldots, \beta_p$, we may compute the MLE. For the function $f(x)=\log \left(1+e^x\right), f^{\prime}(x)=\frac{e^x}{1+e^x}=1-\frac{1}{1+e^x}$. Then setting the partial derivatives of the log-likelihood equal to 0 yields the equations (for $m=0, \ldots, p$)

$$0=\frac{\partial l}{\partial \beta_m}=\sum_{i=1}^{n} x_{im}\left(Y_i-\frac{e^{\sum_{j=0}^{p} \beta_j x_{ij}}}{1+e^{\sum_{j=0}^{p} \beta_j x_{ij}}}\right).$$

We are not going to be able to set this to zero and solve exactly. (No closed-form solution.) We can however approximately solve it numerically.

5.1.1 Newton raphson method

Suppose that the objective f is a function of multiple arguments, $f\left(\beta_1, \beta_2, \ldots \beta_p\right)$. Let's bundle the parameters into a single vector, w. Then the Newton update is

$$\beta^{(n+1)}=\beta^{(n)}-H^{-1}\left(\beta^{(n)}\right) \nabla f\left(\beta^{(n)}\right)$$

where ∇f is the gradient of f, its vector of partial derivatives $\left[\partial f / \partial \beta_1, \partial f / \partial \beta_2, \ldots \partial f / \partial \beta_p\right]$, and H is the Hessian of f, its matrix of second partial derivatives, $H_{ij}=\partial^2 f / \partial \beta_i \partial \beta_j$. But it is actually really slow

These equations may be solved numerically (e.g. by Newton-Raphson) to obtain the MLEs $\hat{\beta}_0, \ldots, \hat{\beta}_p$. To estimate the conversion probability for a new impression with covariates $\tilde{x}_1, \ldots, \tilde{x}_p$, we may use the plugin estimate

$$\hat{p}=\frac{e^{\hat{\beta}_0+\hat{\beta}_1 \tilde{x}_1+\ldots+\hat{\beta}_p \tilde{x}_p}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 \tilde{x}_1+\ldots+\hat{\beta}_p \tilde{x}_p}}.$$

5.1.2 Estimation in Constrained Model

The Likelihood function for a constrained model remains the same except for the fact that the parameter space in which we find the MLE reduces under the constraint that all coefficients are non negative.

5.2 K-Fold Cross-validation

To get the Logistic Regression coefficients for our data, we randomly split the data into k groups, or folds, of approximately equal sizes. At each step, 1 of the k folds is used as a testing set and Logistic model is fit on the remaining $k-1$ folds. The accuracy is then calculated on the test fold. This procedure is repeated k times, and each time a different group of observations is treated as the testing set. We select the set of coefficients for which the testing accuracy is the highest. We used $k = 5$ for getting the value of the coefficients i.e. we split our data into 5 folds with approximately 27 ($= 134/5$) observations in each.

Section 6

Results

6.1 NITI Vs Judgement

NITI \ Judgement	**Non Poor**	**Poor**
Non Poor	69	32
Poor	4	29

Cohen's Kappa $= 0.437$

Interpretation

Value of Kappa suggests that there is fair-moderate agreement between our judgement and NITI Aayog's method.

6.2 Model with weights that we found using constrained logistic regression model

6.2.1 Current Model

Weights	
Indicators	**Indicator Weights**
Nutrition	0.0276
Child & Adolescent Mortality	0.0512
Antenatal Care	0.0391
Years of schooling	0.0000
School attendance	0.0206
Cooking fuel	0.0799
Sanitation	0.0308
Drinking water	0.0000
Electricity	0.4703
Housing	0.0583
Assets	0.1208
Bank account	0.1015

Cutoff : 0.1327

6.2.2 Confusion Matrix

Judgement \ Model	**Non Poor**	**Poor**
Non Poor	63	10
Poor	9	52

Accuracy:0.86
Precision:0.88
Cohen's Kappa:0.715

Remark

The weight vector for this model has almost 47% of weight on electricity. This means any person with electrified house but deprivation in all other indicators will still get a score of 0.53 only. This is inconsistent with our purpose of framing a index based on multiple deprivation indicators. To solve this we combine electricity and assets with an "OR" function.

6.3 Model by merging electricity and assets

6.3.1 Current Model

Weights	
Indicators	**Indicator Weights**
Nutrition	0.0520
Child & Adolescent Mortality	0.0965
Antenatal Care	0.0740
Years of schooling	0.0000
School attendance	0.0386
Cooking fuel	0.1506
Sanitation	0.0583
Drinking water	0.0000
Housing	0.1099
Bank account	0.1914
Electricity and Assets	0.2287

Cutoff : 0.2503

6.3.2 Confusion Matrix

Judgement \ Model	**Non Poor**	**Poor**
Non Poor	63	10
Poor	9	52

Accuracy: 0.86
Precision:0.88
Cohen's Kappa:0.715

Remark

Kappa value have increased to above 0.70. This indicates that the model and our judgement are in agreement.
In the next model we try to incorporate a new indicator which is not used by NITI Aayog.

6.4 Weights after including chronic disease as a new indicator

6.4.1 Current Model

Weights	
Indicators	**Indicator Weights**
Nutrition	0.0343
Child & Adolescent Mortality	0.0633
Antenatal Care	0.0134
Years of schooling	0.0320
School attendance	0.0000
Cooking fuel	0.1215
Sanitation	0.0175
Drinking water	0.0000
Electricity	0.4083
Housing	0.0467
Assets	0.1208
Bank account	0.1341
Disease	0.0082

Cutoff : 0.1668

6.4.2 Confusion Matrix

Judgement \ Model	**Non Poor**	**Poor**
Non Poor	62	11
Poor	11	52

Accuracy:0.84
Precision:0.85
Cohen's Kappa:0.669

Remark

Cohen's kappa suggest this model has "good" agreement with our judgement. But still the same problem of high weight on electricity persists. Again merging Electricity and assets might prove to be helpful.

6.5 Merging electricity and assests and Weights after including disease

6.5.1 Current Model

Weights	
Indicators	**Indicator Weights**
Nutrition	0.0577
Child & Adolescent Mortality	0.1068
Antenatal Care	0.0230
Years of schooling	0.0537
School attendance	0.0000
Cooking fuel	0.2052
Sanitation	0.0296
Drinking water	0.0000
Housing	0.0787
Bank account	0.2264
Disease	0.0139
Electricity and Assets	0.2287

Cutoff : 0.2816

6.5.2 Confusion Matrix

Judgement \ Model	**Non Poor**	**Poor**
Non Poor	62	11
Poor	11	50

Accuracy: 0.86
Precision: 0.85
Cohen's Kappa: 0.669

6.6 Conclusion

The kappa value for NITI Aayog's model indicates moderate agreement between the existing framework of determining poverty status and the poverty status according to our judgement.
The model where electricity and assets are merged stands out to be a better alternative.

6.6.1 Way forward

For the case of data based policy making, we suggest that the Development Monitoring and Evaluation Office (DMEO), an attached office under NITI Aayog, to develop an index based on weights and cut-off calculated by logistic regression model.
For this purpose a batch of sample collection personnel should be trained in such a way that each of them has similar understanding of poverty of a household. In this way when these personnel goes for sampling they can collect the quantitative data for each household and at same time they can judge the household condition based on their understanding of poverty.
With the availability of both the quantitative data for each household and their poverty status, fitting a constrained logistic regression model will give us a improved choice of weight scheme for constructing MPI.

Bibliography

[1] Alkire & Foster (2011). *Counting and Multidimensional Poverty Measurement. Journal of Public Economics. 95 (7–8): 476–487.*

[2] National Multidimensional Poverty Index - NITI Aayog https://www.niti.gov.in/sites/default/files/2021-11/National_MPI_India-11242021.pdf

[3] India: Standard DHS, 2015-16 https://dhsprogram.com/methodology/survey/survey-display355.cfm

[4] Link to NFHS website
http://rchiips.org/nfhs/nfhs4.shtml

[5] Link to Household Consumer Expenditure page in National Data Archive
http://microdata.gov.in/nada43/index.php/catalog/CEXP

[6] Link to Questionnaire for Socio Economic and Caste Census-2011
https://rural.nic.in/sites/default/files/Questionnaire for Socio Economic and Caste Census-2011_1.pdf

[7] Cohen's kappa
https://en.wikipedia.org/wiki/Cohen's_kappa