Social Development Data and Societal Modelling: A Study in Indian Context

Rabi N. Subudhi

Abstract As we progress ahead and measure our growth in every field, science and technology or even societal development, every time, our focus is skewed and mostly myopic on what we achieved and what more to be done and how fast to be done. We seldom study what we really missed and skewed our development is. A retrospection and introspection, many a times, help us making a course correction. A development, or any economic growth, is no development if it is not for all and leaves a section of society. This paper looks at different social development models and important development indices which are in use today globally and analyses social development data in the Indian context. It refers to Mahalanobis D-square statistics and other Indian studies and attempts to suggest newer social development indices, using secondary data. It discusses issues relating to 'social-big-data', as available from secondary sources, published by various agencies. Paper presents a conceptual model for measuring the social wellbeing of individuals and families.

Keywords Social development · Development models · Development indices · HDI · Social data analysis

1 Introduction

As technology advances, gathering and retrieval of a large amount of data are now comparatively much easier. We have so many instances of handling very huge databases in real time, in many fields. Millions of share holders' data in the stock market or thousands of passengers' reservation data, railways or airlines, or the daily transactions of a big retail chain are some such examples, where very huge databases are handled, and used for decision making. Apart from business transaction data or administrative data, we also use large scientific experimental data for our research purpose. This data could be both quantitative as well as qualitative data. Here, in this paper, we shall confine our discussion to social development-related data.

R. N. Subudhi (✉)
School of Management, KIIT University Bhubaneswar, Bhubaneswar, India
e-mail: rabisubudhi@gmail.com

S. S. Rautaray et al. (eds.), *Data Science in Societal Applications*, Studies in Big Data 114, https://doi.org/10.1007/978-981-19-5154-1_4

It has been of enormous research interest to study and analyse the complexity and diversity of different demographic strata of any society or our humanity. How and why a region or a community or a religion or a caste/ creed is not so ‘developed’, as compared to other strata! A construct like ‘social development’ or ‘social acceptance’ is quantified and measured through several ‘representative’ operational indicators (or variables), differently by different scholars. But the commonality amongst all such known measures, as outlined in next section, is the use of a large amount of socio-economic and demographic data. We may call it social big data. One such example can be given on the record single-day COVID vaccination data (on 18.9.21, coinciding with the birthday of Indian PM, Sri Narendra Modi, when 25 million or 2.5 crore persons were vaccinated in India, according to Indian Express, of the same day), comprising of demographic details of each of the persons vaccinated, captured from Aadhar database.

Here, in this paper, we shall review some notable methodologies and approaches to measure human/social development, as available from literature, and then suggest and develop one model to measure the regional diversity, for different districts of Odisha (in India), using relevant secondary data.

2 Review of Literature

One of the most popular methodologies to measure human development is HDI (Human Development Index). This HDI is calculated by taking several socio-economic factors, and the formula is differently computed by different scholars.

The Human Development Report (1990) emphasised three measures of human development, which are mentioned below, as summarised by Mahabub-ul-Haq:

(a) *Human Development Index (HDI):* giving importance to ‘healthy life’, literacy, living standard and GDP.
(b) *Gender (related) development index (GDI):* giving importance to female health and life-expectancy, female—literacy and their income.
(c) *Human Poverty Index (HPI):* consisting of factors like ‘vulnerability to death’, ‘standard of living’, ‘percentage of malnourished children under five’.

For calculation of HDI, in his study, Tripathy [1] considered the indicators, like: Percentage of young school students (within the age group of 6–14 years), Percentage of male literacy, Percentage of female literacy, Number of hospital beds, for analysing district-wise human development in Odisha. Tripathy [1] found that human development is significantly affected by Life Expectancy Index (LEI), Education Index (EI), and Gross Domestic Product Index (GDPI) but out of them the high human development districts of Orissa are affected more by the GDPI. In the medium human development districts, EI is more weighted than LEI and GDPI. In case of the low human development districts of the state, EI has also more importance. Most indices have a negative relation with the HDI. There is huge number of errors in low human development districts as compared to high and medium human development districts.

The important factor is the health index, which influences human development in Orissa more as compared to other two factors like income and EI.

Foster et al. [2] in their study have taken State rankings, Changes in country rankings and Losses due to inequality as constructs and examined a new approach for integrating 'the distributive dimension into the HDI'. A new class of human development indices developed by them, in which, the traditional index and a family of indices both are included, which are sensitive to the distribution of human development. The general mean has been used by this class of indices to summarise achievements within each dimension of development, as well as to aggregate across dimensions.

Sharma [3] in his study examined the usefulness of HDI as a sensible measure of human development.

Acharya and Wall [4] have also contributed to the construction of HDI taking many unique indicators like environmental quality and human rights.

3 Mahalanobis Distance Statistics

Mahalanobis distance statistics was given by P.C. Mahalanobis in the year 1936, which was applied to measure the social development (distance) of different tribes (as compared to mainstream humanity). It is a distance between a point/vector (P) and a distribution (D). It is a multidimensional generalisation of the idea which measures the distance of a point from the mean of distribution by how many standard deviations. It has been applied for multivariate outlier detection, classifying the highly imbalanced (social) data sets and for subsequent classification.

Earlier, the distance between two points was measured by Euclidean distance. Mahalanobis distance statistics (or, D-square statistics) has a different approach, which can be illustrated by the following example:

Suppose there are two points in a 2D surface. It means it has two numerical columns, which are P and Q, in the given datasets. Therefore the distance between the two points, i.e. (P1, Q1) and (P2, Q2), can be stated as

$$\mathrm{d(P, Q)} = \sqrt{(P1 - Q1)^2 + (P2 - Q2)^2}$$

This measure can be extended as per the number of dimensions, like

$$\mathrm{d(P, Q)} = \sqrt{(P1 - Q1)^2 + (P2 - Q2)^2 + (P3 - Q3)^2 + \cdots + (Pn - Qn)^2}$$

It has two assumptions. First, the dimensions should be equally weighted and second, these should be independent of each other. It only measures the distance between two points. It doesn't consider the variation amongst the rest of the points.

The formula for calculating Mahalanobis distance is

$$\mathrm{D}^2 = (x - m)^T \cdot C^{-1} \cdot (x - m)$$

where,

D^2	Square of the Mahalanobis distance.
x	Vector of observations (row in a dataset).
m	Vector of mean values of independent variables (mean of each column).
C^{-1}	Inverse covariance matrix of independent variables (mean of each column).

In a normal distribution, the probability of distribution is concave exactly at that region, where the Mahalanobis distance is less than 1.

Nowadays, many software and statistical packages, like ‘R’, Python, Julia, etc., are used to compute Mahalanobis distance, conveniently, using large social data.

4 Role of Social Statistical Information and Development Models

Development models in social science research have seen many empirical studies, investigating diversity amongst demographic divisions of populations. These studies majorly used social development indicators across multiple developmental periods and within community, caste, religion, genetic-lineage, and examined relationships of related constructs and behavioural outcomes. Tests of social development model hypotheses have demonstrated significant utility in understanding the causal factors and relates to socio-psychological behaviours. Development models, aimed at measuring the degree of diversity (or distance from the bench-marked standard), use different types of information (secondary data), called ‘development indicators’ (or constructs).

5 Sources of Social Statistics

Giving priority to social issues and human problems, the main data sources for this type of study are usually secondary databases published by census organisation and other agencies collecting various vital statistics. For the construction of indices like HDI, poverty index, social acceptance index or happiness index, we need to have relevant data on other socio-anthropological indicators.

Such vital statistics, including housing, agriculture, health, education, etc. can be obtained, in India, from sources like local bodies, hospitals, dispensaries and local self govt., like:

https://www.censusindia.gov.in/

http://wwwmospi.nic.in/

https://www.mygov.in/

https://www.lgdirectory.gov.in/

Administrative units, like revenue department (for land holding), judiciary (for legal support system), disaster management and mitigation units (for distress rehabilitation data), crime-control-research units (for crime against the oppressed and marginalised or minority class), etc., (like: https://www.ncrb.gov.in/en/crime-in-india-table-addtional-table-and-chapter-contents) are also very useful for getting relevant secondary data.

6 Social Support System and Regional Safety Status: A Case Study of Odisha Districts

We can take a very large number of units for index-based comparisons, like all districts or blocks of India, or all (96,37,820) households of Odisha, for which data are readily available. But for space constraints, and as an illustration of index construction exercise, we shall take 'safety and security of citizens', only for the districts of Odisha.

For ranking districts of Odisha, on the basis of 'social safety and security', secondary data on the rate of latest crimes against children and ladies were taken and compared with the population density of each district. This latest crime Report-2020 (released in September 2021, by NCRB, at: https://www.ncrb.gov.in) was considered.

Using data from Table 1, it was first studied to find if any association existed between the occurrence of crime and sex ratio, and with density. As expected, it was found (from Table 2) that there existed a positive correlation between population density and crime incidences (higher in more populous districts). It is interesting to note here that there is a negative correlation between the sex ratio and rate of crime. That is, the higher the sex ratio, less the crime (Table 2).

Table 1 Crime statistics of districts of Odisha

1	2	3	4	5	6	7
Crime against SC-ST	Crime against women	Crime against children	Districts of Odisha	Sex ratio	Density	Literacy
130	1246	228	Angul	943	200	77.53
131	1613	411	Balasore	957	610	79.79
98	497	106	Baragarh	977	254	74.62
88	884	336	Bhadrak	981	601	82.78
168	973	185	Bolangir	987	251	64.72
35	211	46	Boudh	991	142	71.61
144	656	184	Cuttack	940	667	85.5

(continued)

Table 1 (continued)

1	2	3	4	5	6	7
34	200	37	Deogarh	975	106	72.57
98	1091	217	Dhenkanal	947	268	78.76
13	268	93	Gajapati	1043	134	53.49
93	1002	132	Ganjam	983	430	71.09
107	1056	212	Jagatsinghpur	968	682	86.59
198	1411	341	Jajpur	973	630	80.13
66	436	118	Jharsuguda	953	274	78.86
80	842	187	Kalahandi	1003	199	59.22
26	352	115	Kandhamal	1037	91	64.13
112	1413	217	Kendrapara	1007	545	85.15
77	765	191	Keonjhar	988	217	68.24
38	554	126	Khurda	929	800	86.88
33	647	170	Koraput	1032	157	49.21
18	446	148	Malkangiri	1020	106	48.54
144	1490	448	Mayurbhanj	1006	242	63.17
46	508	156	Nayagarh	915	248	80.42
60	673	207	Nabarangpur	1019	231	46.43
60	240	82	Nuapada	1021	158	57.35
147	1544	274	Puri	963	488	84.67
42	547	185	Rayagada	1051	137	49.76
59	781		Sambalpur	976	157	76.22
65	173	170	Sonepur	960	261	74.42
54	465	38	Sundargarh	973	216	73.34

Note Estimated data for 2020, as available from the following sources
Source for column (1, 2, 3): https://www.ncrb.gov.in/crime-in-india-table-contents
Source for column (5, 6, 7): https://www.census2011.co.in/census/state/districtlist/orissa.html

Table 2 Correlation table

Correlation	
r(CAW-literacy)	0.37
r(CAW-Density)	0.513
r(CAW-Sex Ratio)	−0.21
r(CSCST-Literacy)	0.552
r(SCST-Density)	0.63
r(SCST-Sex Ratio)	−0.42

Note CAW = Crime against women, CSCST = Crime against SC-ST persons;
Correlation values computed by author using data from Table 1

Table 3 (a): Worst status states; (b): Better performing districts; (c): Average

(a) *Index > 1*	Worst status
4.454	Mayurbhanj
3.816	Angul
2.503	Sambalpur
2.049	Malkangiri
1.789	Rayagada
1.649	Koraput
1.617	Kalahandi
1.578	Kandhamal
1.327	Dhenkanal
1.059	Bolangir
(b) *Index < 1*	*Best status*
−2.746	Sonepur
−2.704	Khurda
−2.294	Cuttack
−1.744	Boudh
−1.524	Bhadrak
−1.516	Nuapada
−1.318	Deogarh
−1.179	Baragarh
(c) *1 < Index < 1*	
Ganjam	
Nayagarh	
Gajapati	
Jajpur	
Sundargarh	
Kendrapara	
Balasore	
Puri	
Nabarangpur	
Keonjhar	

7 Development of Social Security Status of Districts

Using crime data for each district, and other population statistics, an relative index is presented here, by adding all crimes and then dividing by population density, by the following formula:

$$\text{SIID (Social Insecurity Index of Districts)} = [(\text{CAW} + \text{CAC} + \text{CA SCST})/\ \text{PD}] - (\mu),$$

where CAW = Crime against women, CAC = Crime against Children, CASCST = Crime against SC-ST, and PD = Population density, μ = State average ratio = Total Crime/State PD.

The following table, presented in three blocks, gives 'best (with negative score)', 'worst (with positive scores, arising because of the higher number of crimes in the denominator)' and average (or status-quo) status of districts of Odisha. The same methodology can be applied to find the national average and then all districts of India.

Above small illustration, just for Odisha districts, used secondary data of 77 × 35 matrix size. We can suggest a similar exercise for other social development (secondary) data, with a very large set of indicators. Following illustrative conceptual model is suggested for possible construction of SWI (Social Wellbeing Index). Odisha State Government administration should look into the state of affairs at distorts like, Mayurbhanj, Angul, Sambalpur, and Malkangiri. Things are looking really good for the districts of Sonpur, Khurdha, Cuttack, Boudh, and Bhadrak.

8 Conceptual Model for Constructing Family Wellbeing Model

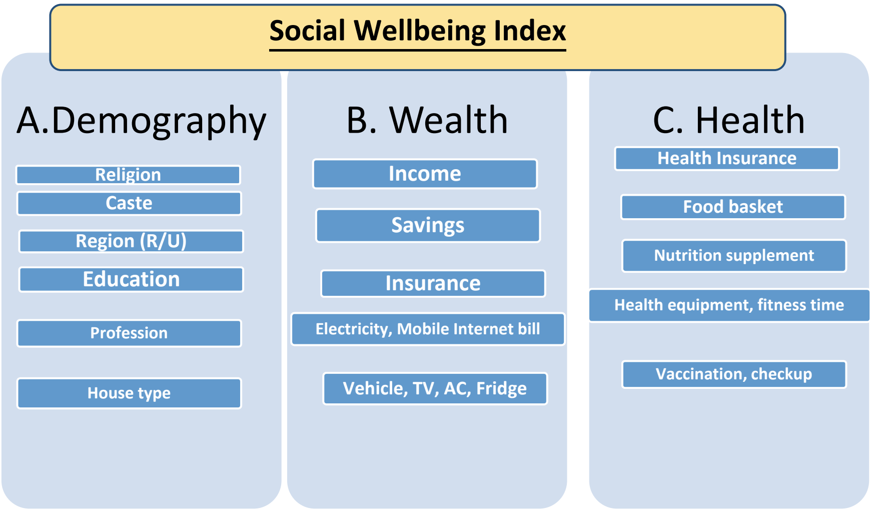
To understand and compare the relative status of family wellbeing, at the household level, we may consider following socio-economic and demographic factors and can get data from various sources, as mentioned above. This can further be expanded and have more operational variables/indicators.

Input Variables/Indicators

Religion|Caste|Region (rural/Urban)|Income|Education|Profession|House-type|Vehicles|Insurance|Food basket budget| Savings|Water source|Health/vaccination|Electricity Bill|LPG Bill|News magazine Bill|Mobile-internet bill| TV, Fridge, AC availability|.

Output/Criterion Variable

Social Wellbeing Index (and perception scores).



[Author's own suggestion]

9 Conclusion

Studying and constructing social development indices are of great social importance, having practical use value for policymakers, planners and administrators. It can indirectly help measuring the impacts of intervention programmes. As found from this illustrative study, we can observe better gender ratio can have a positive role in checking crimes and creating a better social security system. In a similar way, as a possible future research suggestion, from the above conceptual model on the construction of 'social wellbeing index', we can possibly find the relative importance of smaller, yet newer indicators, like vaccination and health insurance plans. We suggest that, more and more such measures should be initiated, we relatively newer social indicators, using various scales.

There are many other virgin areas, where indices similar to HDI could be applied.

Acknowledgements The author acknowledges the help and contribution of his research scholar, Mr. Smruti Malhar Mohapatro, for the collection of data and other secondary information, from the above cited sources, identified by the author.

References

1. Tripathy, U.: Estimation of human development index in Orissa: district-wise analysis. IUP J. Manag. Econ. **8**(4) (2010)
2. Foster, J.E., Lopez-Calva, L.F., Szekely, M.: Measuring the distribution of human development: methodology and an application to Mexico. J. Hum. Dev. **6**(1), 5–25 (2005)
3. Sharma, S.D.: Making the human development index (HDI) gender-sensitive. Gend. Dev. **5**(1), 60–61 (1997)
4. Acharya, A., Wall, H.J.: An evaluation of the United Nations' human development index. J. Econ. Soc. Meas. **20**(1), 51–65 (1994)
5. Tripathy, U.: Estimation of human development index in Orissa: district-wise analysis. IUP J. Manag. Econ. **VIII**(4) (2010)

Web-Source

6. https://www.odisha.data.gov.in/catalog/households-marital-status-sex-and-age-head-household-india-and-states#web_catalog_tabs_block_10