

Disparity in Consumption of Highly Nutritious Food among Rural Households in India: Data Mining Approach

Mayank Jain

Madras School of Economics
Chennai, Tamil Nadu, India

Pradip Kumar Bala

Associate Professor, Indian Institute of Management
Ranchi, Jharkhand, India

ABSTRACT

There are various approaches to measure the differences in household income. Income measured from the monetary earnings point of view is called the nominal income. Income measured from the consumption point of view i.e. the basket of consumption goods a household buys is called the real income. In this paper we analyze the income based on the consumption of Highly Nutritious Food. Using tools of data mining (C&R tree and C5.0), we predict the trends in expenditure of rural households of India in the consumption of Highly Nutritious Food (Pulses, Cereal Products, Meat, egg, milk and milk products) which are not available through the Public Distribution System (PDS). We study the differences in expenditure on Highly Nutritious Food (HNF) across caste, religion, states and demonstrate that multiple factors influence the expenditure on HNF. We also categorize the per capita consumption of HNF as Low, Average and High. The outcome shows disparity in consumption of Highly Nutritious Food and argues that it can be a way to look at income disparity.

Keywords

Highly Nutritious Food (HNF), Per Capita Consumption of Nutritious Food (PCCNF), Rural Households, Caste, Religion, Data Mining, C&R Tree, C5.0

1. INTRODUCTION

There has always been evidences of high inequalities in the Indian population. This inequality persists in the nominal income, consumption, social status, welfare etc. In this paper we study inequality based on the consumption of Highly Nutritious Food (Pulses, Cereal Products, Meat, egg, milk and milk products) which are not available through the Public Distribution System (PDS). With the rising food prices, food expenditures are increasingly dominating the household budget. The households are forced to reduce their consumption basket and depend solely on the basic food for their nutritional requirement. This issue is more severe and pronounced when we study the differences in expenditure on HNF among the rural households.

This paper proposes to review the trends in expenditure on HNF's using various input and outcome measures. This is an alternative approach to study the differences in income among households. Households with low incomes would tend to consume less of HNF and depend more on food available through PDS. This paper also identifies the differences in income based on consumption of HNF among households with different sources of income (like agricultural, non

agricultural labour, artisan, service, business etc). It tries to investigate whether holding a ration card creates any incentive to consume more of HNF. There is also a difference in consumption of HNF among different regions of India (North, South, East, West, Central and North East) as discovered from the data. There are various other interesting findings which will be discussed subsequently.

Findings of the research incorporated in this article are based on the household data from India Human Development Survey, 2004-05 (IHDS). The IHDS is a nationally representative survey of 41,554 households organised by researchers from the University of Maryland and the National Council of Applied Economic Research. It is a multi-topic multi-purpose survey containing information about a variety of dimensions of social and economic well being of the households. These data are in public domain and at an all-India level, poverty, education, household structure and employment levels recorded in this survey are comparable to those recorded by Census and the National Sample Survey albeit with some exceptions associated with the survey design (Desai et al 2010)

2. LITERATURE REVIEW

The problem of differences in income has been addressed by various authors, but very less literature is available about the differences in consumption of HNF. Most of the study in this area is based on earnings point of view and on per capita calorie consumption from the expenditure point of view. Few of them are "Income inequality in village India: The role of caste" by Swaminathan and Rawal (2011), which examines the role of caste in understanding inequality in incomes in rural India. In the paper "Regional Heterogeneity in Food Consumption and Nutrition Intake in India" Srivastava et al, the authors raise the issue of difference in consumption of cereal crops across Indian states. In their article "Food and Nutrition in India: Facts and Interpretation" Deaton and Dreze (2009) analyze per capita consumption from the calories perspective.

3. UNDERSTANDING THE PROBLEM

The Indian social system is characterised by various social groups or Caste. The *New Shorter Oxford Dictionary* defines caste as "a hindu hereditary class of socially equal persons, united in religion and usually following similar occupations, distinguished from other caste in the hierarchy by its relative degree of purity or pollution" [Ed Lesley Brown, Clarendon Press, Oxford, 1993]. In the data the caste is categorized as Brahmin, Scheduled Caste (SC), Scheduled Tribe (ST), Other Backward Class (OBC), Others.

Caste system dominates a larger part of decision making in the rural household scenario. A person of the Brahmin caste enjoys the supreme position and is generally well-off. The *Others* are the people who are not Brahmins but club together with Brahmins to form the General Category. The Scheduled Caste are people who were historically disadvantaged and were considered untouchables. Scheduled Tribes are the ethnic tribal groups. Other Backward Class are socially and educationally backward communities.

India is a secular country. People following various religions, Hindu, Muslim, Sikh, Christians, Jains, Buddhists, Tribals etc all live in harmony. The vast landscape of India is divided into 28 states and seven union territories. No socio-economic study on India can be complete without taking in account the influence of Caste, Religion and Region in their decision making.

This study is important and unique in various aspects. Firstly, No study has been done to assess the differences in consumption expenditure across household based on highly nutritious food consumption. Most studies till now have focussed on only calorie intake and minimum calorie requirement. Here, we argue that calorie intake is necessary, but getting calories only from rice and wheat makes consumption monotonous. Even the poorest of the poor household consumes rice. But its consumption of HNF is limited due to its inability to purchase them. As income rise the consumption decisions also change and the expenditure on HNF's increase.

Having focussed only on the rural households, the point to investigate is whether the rise in income is treated as the same by people of all caste, religion, and states. Prima facie it appeared that there is a difference in expenditure decisions across them. What might be the reasons contributing to these? Do the people living in North India spend the same on HNF as compared to South India? Do the OBC's increase their expenditure on HNF in the same proportion as the Brahmins? Do the Hindu's and Muslim's record the same increase in their consumption of HNF's? Why not?

4. RESEARCH METHOD

Household data from IHDS contained recorded data of 41,554 households from 1503 villages and 971 urban neighbourhoods across India. This data contained 924 variables. Out of these, we used 30 variables for our research. States were categorized in North, East, West, South, Central and North East.

North included states of Delhi, Haryana, Himachal Pradesh, Jammu & Kashmir, Punjab, Rajasthan, Uttar Pradesh, Uttaranchal (presently Uttarakhand), South included states of Andhra Pradesh, Karnataka, Kerala, Pondicherry, Tamil Nadu. East were the states of Bihar, Jharkhand, Orissa, Sikkim, West Bengal. West included states of Dadra & Nagar Haveli, Daman and Diu, Goa, Gujarat, Maharashtra. Central included states of Chhattisgarh and Madhya Pradesh. North East included the states of Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Tripura. The states were divided into these 6 zones so as to capture the vast geography of India in a condensed form.

From the whole data only the data pertaining to Rural India was retained. These rural household were 26735 in number. Two new variables were created, 1) "Nutritious food" in which the expenditure on 6 highly nutritious food items (Pulses, Cereal Products, Meat, egg, milk and milk products) were taken and summed together. 2) PCCNF which recorded the per capita monthly consumption on nutritious food by

dividing *Nutritious Food* by the number of persons in that particular household. The names of the remaining variables were changed so that it is more clearly understood.

Using data mining to predict the trends in the household data was an interesting idea. Data Mining was applied on this data set using SPSS Clementine 12.0. PCCNF was taken as the output variable and was treated with various individual input variables (religion, caste, ration card, no. of meals per day, state zone etc) and combination of input variables using C&R Tree. "The Classification and Regression (C&R) Tree node is a tree-based classification and prediction method. This method uses recursive partitioning to split the training records into segments with similar output field values. The C&R tree node starts by examining the input fields to find the best split, measured by the reduction in an impurity index that results from the split. The split defines two subgroups, each of which is subsequently split into two more subgroups, and so on, until one of the stopping criteria is triggered. All splits are binary (only two subgroups)." *Clementine Help*.

Later on, the per capita consumption of Nutritious food was categorized into 3.

PCCNF in the range of INR

[0-100):	Low
[100-200):	Average
>200:	High

This is the discretisation of range variable. We categorize this as per our own discretion in order to find a better and clear picture of the results with numeric values. On the discreet and categorized values, we use the C5.0 modelling and note the results. "The C5.0 node builds either a decision tree or a rule set. The model works by splitting the sample based on the field that provides the maximum information gain at each level. The target field must be categorical. Multiple splits into more than two subgroups are allowed." *Clementine Help*

5. EMPIRICAL RESULTS USING C&R TREE

5.1 Across Caste

The predicted (average) Per Capita Consumption of Nutritious Food of rural households across India in Rupees at 2005's current prices was INR 116.46. The predicted average consumption per capita of a caste and their proportion in total population is given in Table 1.

Brahmins and others consume more than the national average consuming. All other caste consume less than the national average of 116.46. Brahmins are followed by OBC and SCs. The worst performer is the ST, who consume at an extreme low of 77.45 per person.

5.2 Across Religion

Classifying and regressing on the basis of religion, we see that the minority communities (those who are less than 5 % of the total population, except the tribals are the ones consuming most of the HNF. The Sikhs consume as high as 195.54 per head. We can accredit it to the culture of the community, and their very high consumption of Milk, Milk products and meat.

The Christians and Jains are well off communities, both with good education and the latter forming the entrepreneurial class. Hindus and Muslim combined, who form 92.58 % of the total population consumed slightly less than the national average of 116.46. The variations in these communities due to

caste has already been discussed in 5.1. Tribals show the least consumption of 59.39 per capita. Table 2 shows the % of all the religions and their per capita consumption.

Table 1: Consumption of HNF across Caste

Caste	% of total population	PCCNF (in INR)
Brahmins & Other	26.719	158.94
OBC	40.035	108.91
SC	22.264	98.27
ST	10.982	77.45

Table 2: Consumption of HNF across Religion

Religion	% of total population	PCCNF (in INR)
Sikh	2.71	195.54
Christians & Jain	3.26	139.70
Hindus & Muslim	92.58	114.22
Tribals	1.45	59.39

5.3 Across Regions/States

Across regions, the households of North India spent on an average INR 174.63 per capita. We further sub divide North India and club Punjab & Haryana together and keep rest of the North Indian States together we notice an interesting result. Punjab and Haryana have a per capita consumption of HNF at INR 222.23 which is almost double to that of the national average. While the rest of North India consumes INR 155.09 worth of highly nutritious food per capita per month. All the remaining part of rural India consumes below the national average. South Indian households have a per capita consumption of INR 108.06 while the West India has a predicted consumption of INR 98.95. The Central and East Indian States show a PCCNF of INR 56.51 and INR 70.90 respectively. The North Eastern rural households, though 5% of the sample spent INR 128.95 which is above the national average of INR 116.46.

North Indian States Delhi, Haryana, Himachal Pradesh, Jammu and Kashmir have a mean PCCNF of above INR 200 each. 4 North-Eastern States Arunachal Pradesh, Mizoram, Manipur and Nagaland have PCCNF greater than INR 150. Assam and Tripura spend on an average INR 121. Meghalaya has the average consumption of INR 66.77 along with Madhya Pradesh. The tribal states of Chhattisgarh, Jharkhand and Orissa have per capita consumptions of nutritious food below INR 50. Gujarat, Andhra Pradesh and Uttar Pradesh, Karnataka, Tamil Nadu, Dadra and Nagar Haveli consume in the range of INR 101 – 107. Bihar and Maharashtra showed predicted consumption of INR 81.7 and INR 87.94 respectively. West Bengal had an average consumption of INR 96.68. Kerala and Uttaranchal consumed INR 130.74 and INR 138.26.

Table 3: Consumption of HNF across Regions

Region	North	South	East	West	Central	North-East
PCCNF	174.63	108.06	70.90	98.95	56.51	128.95

5.4 Across Income

Consumption of HNF also shows high correlation with income. The following table shows income range generated by SPSS Clementine and the expenditure incurred by household with those income range in PCCNF.

Table 4: Consumption of HNF across generated Income range

Income Range	% of total population	PCCNF (in INR)
0-27553.96	55.96	87.09
27553.96-38219.78	12.65	114.20
38219.78-46502.74	6.17	130.61
46502.74-71183.87	10.80	148.53
71183.87-102548	6.88	176.05
102548-158270.05	4.47	213.71
>158270.05	3.07	244.66

Majority of households had their annual income between INR 0 to INR 27553.96. These low incomes were clearly reflected in the expenditure of HNF's where the households spent an average of INR 87.09 per capita per month. The households who with annual income between INR 27553.96 to INR 38219.78 (12.65 % of households) had a PCCNF of INR 114.20 which is still below the National Average. Households with annual income above 38219.78 to 46502.74 (6.17 %) spent more than the national average. Their average PCCNF was 130.61. Thus, we see that 68.61%, i.e., more than 2/3rd rural households of India consumed nutritious food below the national average. This suggests that National Average is affected by the extreme values of the high consumption of the higher income households. Households with higher annual income subsequently showed proportionately higher consumptions of HNF's. The highest average consumption per head of INR 244.66 was shown by rural households with income greater than INR 1,58,270. Though they are just 3.07 % of the total sample.

5.5 Across Sources of Income

Comparing from the point of view of main income source of rural households, agricultural and non agricultural labourers are the worse off in terms of consumption of HNF. Household with Agricultural labourers (19.71% of the total households) spent INR 79.47 on HNF while those with Non Agricultural labourers (16.52 %) spent INR 94.03 per capita. Artisans (4.43 %) spent on an average INR 111.45. Households with other professions showed a consumption above the national predicted value of expenditure on HNF (INR 116.46).

Households involved in some type of cultivation or other work (38.21 %) consumed INR 123.76 worth of HNF while those involved in petty trade (3.00 %) spent INR 125.07. When the main income source was allied agriculture (1.19 %) the expenditure on HNF per head was INR 159.63 whereas in case of Salary as the main source of income for the household (10.68 %) expenditure per head on HNF increased to INR 163.76. The most interesting finding is that the highest consumption was shown by household whose main source of income was rental income from property or pension (2.53%). They consumed as high as INR 198.25 worth of HNF. The reason for this needs to be studied.

Table 5: Consumption of HNF across Source of Income

Source of Income	% of population	PCCNF (In INR)
Agricultural Labourer	19.71	79.47
Non Agricultural Labourer	16.52	94.03
Artisan	4.43	111.45
Cultivation	38.2	123.79
Petty Trade	3.01	125
Business/Profession	3.74	131
Pension/Rent	2.53	198.25
Allied Agriculture	1.18	159.62
Salaried	10.68	163.76

5.6 Across Other Variables

Applying the data mining techniques on other variables yielded many interesting results. A Variable POOR was assigned a value YES if the household was poor according to the 2005 poverty estimates which differed across states. Taking POOR as input variable when data mining was performed, we found that 20.72 % of the total households which came in the category of poor; their average consumption of HNF was as low as INR 36.45. The households which were not poor (79.28 %) showed a higher average consumption of INR 137.35.

The poor household who had a ration card could the HNF performed marginally better. Their consumption was INR 37.98. While those without a ration card had to sacrifice with almost INR 8 per head of PCCNF to consume at INR 28.87. If the poor belonged to the ST community, their condition was all the more pathetic as they could consume only INR 27.84 worth of HNF's. The poor belonging to SCs and OBCs had a predicted average consumption of 37.18. Brahmins and Others of the poor category consumed better than the SCs and STs. Their consumption per capita was INR 47.78.

Table 6: Consumption of HNF regressing along the variable POOR (people below official poverty line)

Variable	PCCNF (in INR)
Poor	36.45
Not Poor	137.35
Poor with Ration Card	37.98
Poor without Ration Card	28.87
Poor and Brahmin/Other	47.78
Poor and SC, OBC	37.18
Poor and ST	27.84

Poor was a variable recorded values YES or NO. Those considered POOR under current official poverty line were given value YES

Mining with multiple input variables, some interesting results were noted. Even among the POOR, the Brahmin poor had the highest per capita PCCNF at INR 47.8 while the poor STs consumed the lowest at INR 27.84. Agricultural labourers of the North India were the best in consumption, spending INR 124.88 in HNF while those of Central India were the worst, consuming only INR 33.46. Among the Brahmins, the North Indian Brahmins show the maximum consumption consuming HNF worth INR 202.51 while the Brahmins of central and east India were the poorest in consumption, consuming only INR 108.80. The STs of North India consumed the maximum HNF INR 160.88, but we may ignore this value owing to the very less percentage of ST community in North India. Among the regions of higher ST population, interestingly, the states of North East performed better than all others consuming INR 136.92 worth of HNF. While the tribals of Central India were the worst performers consuming only INR 32.52. Even though the agricultural labourers consumed the least, the very few Brahmin agricultural labourers had a higher consumption of more than INR 100. While agricultural labourers coming under ST category consumed the least, INR 60.77. Similar was the result for Non-Agricultural labourers.

6. RESULTS USING THE C5.0 MODELLING

In the C5.0 modelling, we take the discreet PCCNF as the output variable and take multiple input variables, like caste, religion, income, state zones, ration card etc are taken. We note various results which are in resonance with the results obtained through C&R tree. The noted results are as follows:

68.34% of the households living in central, east, south and west India had low consumptions of HNFs, while the North India predicted almost equal shares in almost all income groups.

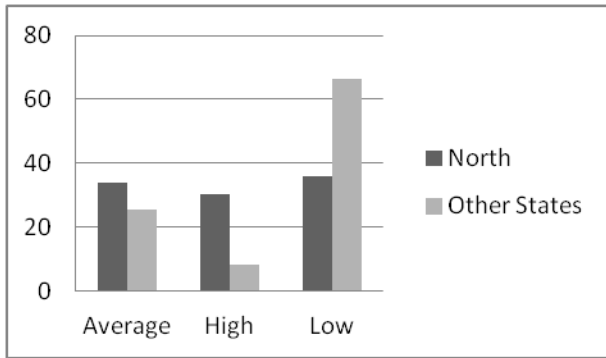


Figure 1 : Comparing North with Other States

Very few (around 23%) Brahmins in North India predict low consumption of HNF. As compared to this, around 50% of the all other communities (OBC, SC, ST) predicted low consumptions. This suggests the income disparity across caste.

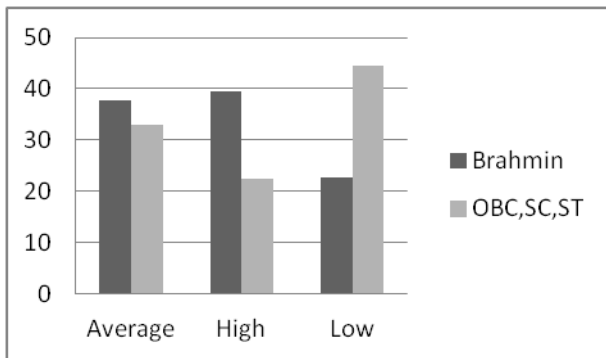


Figure 2 : Brahmins Vs OBS,SC,ST in North India

The maximum High consumption is shown by the others (general caste apart from Brahmins) in North India. It is 45.65 % of all others population

Taking the case of North East, the Christian community is better off with almost 63 % of them consuming more than the Low consumption. Most Hindu's have average consumption, but a substantial number of them are also in the low consumption categories. The tribals and muslims are the worse off with more than 50% of them consuming less than INR 100 per capita per month.

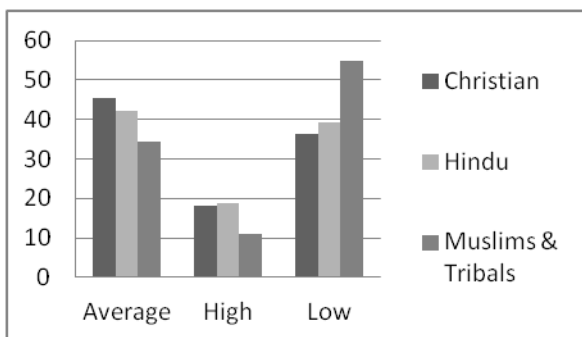


Figure 3 : Consumption across Religions in North East

The Sikh and Jain are the most well off communities with more than 40% of them showing High consumptions of HNFs.

Households across the country with annual income less than 38,325 have low consumptions of HNFs.

Among all income groups, the others show high consumption with Brahmins marginally behind them. All other communities are far behind the average consumptions of Brahmins and others.

An interesting result was that 48.53% Brahmins without a ration card were found to have low consumption of HNFs.

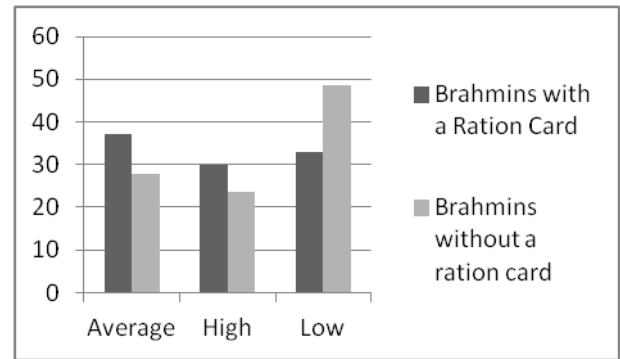


Figure 4 : Comparison of Brahmins with and without a Ration Card

42.94% of households belonging to general caste (apart from Brahmins) and involved in allied agriculture or obtaining income from rent and pensions are seen to have very high consumption of nutritious food.

7. CONCLUSION

Disparity in income is clearly seen in the empirical results and categorized results when we look at it from the point of view of per capita consumption of nutritious food. Interpreting this consumption as Income, the Brahmins and Others across India are better off than all other communities while the ST community is the poorest. A large income gap is also seen between the Brahmins and STs. The Sikh and Jain are the higher income earning religious communities, so their consumption of HNF is also higher compared to all other religions. The rural households of North India are better off than all others; this might be due higher agricultural productivity and production & consumption of milk and milk products in states of Punjab, Haryana and western Uttar Pradesh.

PCCNF highly correlates with increase in Income. Increases in Consumption of highly nutritious food can also be observed as increase in real income. It is important to note that agricultural labourers are the poorest in consumption of nutritious food and thus income. One of the reasons for this can be the disguised unemployment in agriculture. Renting a property and a salaried job generates higher income and thus higher consumption. Thus, summing up, we may argue that disparity in consumption of nutritious food provides a better picture of the income disparity of rural Indian households. Highly Nutritious Foods are the food items which a household has to buy at market prices, thus its basket is suited for estimating the real income.

8. REFERENCES

[1] Swaminathan, M and Rawal, V (2011): *Income inequality in village India: The role of caste* (ECINEQ 2011-207, August 2011)

- [2] Angus Deaton, Jean Dreze (2009): *Food and Nutrition in India: Facts and Interpretation* (Economic and Political Weekly, 14 February, 2009, Vol XLIV, NO 7, Pages 42-65)
- [3] Desai Sonalde, Reeve Vanneman “India Human Development Survey” 2005, IPSCR 226263) Desai Sonalde, Amaresh Dubey, B L Joshi, Mitali Sen, Abusaleh Shariff and Reeve Vanneman (2010) : *Human Development in India: Challenges for a society in Transition* (New Delhi: Oxford University Press)
- [4] *C&R tree*, Help , Clementine 12.0
- [5] Pujari, Arun K: *Data Mining Techniques* (Universities Press, Hyderabad, India, Second Edition)
- [6] Madhura Swaminathan, Vikas Rawal (2011): *Income inequality in village India: The role of caste* ECINEQ WP 2011-207
- [7] Shraddha Srivastava, Amarnath Tripathi, A R Prasad: *Regional Heterogeneity in Food Consumption and Nutrition Intake in India* (Banaras Hindu University)
- [8] Mahendra Dev, S (2005): *Calorie Norms and Poverty* (Economic and Political Weekly, February 19, 2005 Vol XL, NO.0.