Jour. Ind. Soc. Ag. Statistics
53(3), 2000 : 232–242

Era of Information : Relevance of Basics in Statistics[1]

Padam Singh
Additional Director General,
Indian Council of Medical Research, New Delhi

1. There has been impressive advances in the development of statistical theory as well as its applications. Many new disciplines have come up. The advent of computers has helped the discipline in a great deal.

Present era is the era of information. Statisticians have to show their relevance and should play a crucial role in generating appropriate and relevant information for the society. In this, one has to make use of the latest advancements in the theory of statistics as well as the computers. But, in the process the 'basics' in statistics should not be lost sight of. While using more advanced methods there is more emphasis on high level theories and sophisticated softwares, but the feel for the data gets missed. It is the application of the elementary tools of statistics through which the information generated could be understood better by a common man then through the parameters generated using sophisticated models and advanced softwares as currently in vogue. I would like to take some examples to illustrate my points on simple applications of statistics and drawing appropriate information through the same. These examples are not exhaustive but only used as illustrations.

It is the information in the simple language of statistics which is understood more and better by the commoners. I dedicate this lecture in the memory of Dr. V.G. Panse who was a great visionary and had a very good intuition on application of statistics in day to day life in general and Agriculture and Animal Sciences in particular.

2. Norms and Reference Standards

Norms and reference standards have good deal of application in day to day life. For example, the calorie requirement norm of 2425 K cal per consumption unit is usually referred to in studying the trends and assessing the improvements in dietary intake. When these norms were developed the latest

1 Dr. V.G. Panse Memorial Lecture delivered at the 54th Annual Conference of the Indian Society of Agricultural Statistics at Narendra Deva University of Agriculture & Technology, Narendra Nagar, Kumarganj, Faizabad, U.P. on 29th November, 2000.

methodology and available data at that point of time were utilised. Now more recent data have become available which incorporates significant improvements/changes in the life styles warranting relook into these norms. It is also relevant in the context that the per capita per cu consumption level is not improving whereas nutrition status measured through anthropometry has shown considerable improvement, which does not appear to be consistent.

Table 1. Nutrient intake and related parameters

	1975-79	1988-89	1996-97
Calorie intake	2349	2283	2108
Children (%)			
Severe malnutrition	15.0	8.7	6.2
Moderate malnutrition	47.5	43.8	44.3
Adults chronic energy deficiency (%)			
Male	55.6	49.0	45.5
Female	51.8	49.3	48.2

Sever malnutrition < 60% of NCHS standards
Moderate malnutrition 60 – 75% of NCHS standards
Chronic energy deficiency BMI value < 18.5
Source : NNMB – 1999

These suggest relook into the norm. The data collected by IRMS, Delhi for Rajasthan indicate that per cu consumption norm should be around 2150. This has implications on the magnitude as well as the trend in the proportion of these below RDI.

3. Use of Averages

While teaching averages a classic example is quoted of implication of a wrong decision on the basis of average height of members of a family vis-a-vis the average depth of a river. But the important message of that classical example is often forgotten.

Let me take the example of assessing the proportion of persons with intake less than the recommended dietary intake (RDI). The data at household level are analyzed in terms of nutrient intake per consumption unit and then used in determining the number of persons below the recommended dietary intake. If the per cu intake for an household is less than the RDI norm (2425 K cal) all members of household are taken as below the norm or vice versa. This analysis ignores intra family variation in the nutrient intake. This might exaggerate the problem for an economy where the average intake is below the

Table 2. Estimation of average calorie requirement per cu - Rajasthan

	Urban			Rural		
Age, Sex Occupational Categories	Sample (%) (N_i)	RDA (R_i)	cu (C_i)	Sample (%) (N_i)	RDA (R_i)	cu (C_i)
<1 year Male	1.4	760	0.3	2.0	760	0.3
<1 year Female	1.4	760	0.3	1.9	760	0.3
1-3 year Male	3.3	1240	0.5	3.7	1240	0.5
1-3 year Female	2.9	1240	0.5	3.4	1240	0.5
4-6 year Male	4.7	1690	0.7	5.2	1690	0.7
4-6 year Female	3.6	1690	0.7	4.5	1690	0.7
7-9 year Male	3.9	1950	0.9	4.3	1950	0.9
7-9 year Female	3.8	1950	0.9	3.9	1950	0.9
10-12 year Male	4.4	2190	1.0	4.5	2190	1.0
10-12 year Female	3.8	1970	0.9	4.1	1970	0.9
13-15 year Male	3.7	2450	1.1	3.3	2450	1.1
13-15 year Female	3.4	2060	1.0	3.0	2060	1.0
16-17 year Male	2.0	2640	1.2	1.6	2640	1.2
16-17 year Female	1.6	2060	0.9	1.1	2060	0.9
>18 year Male						
Sedentary	25.2	2425	1.0	16.0	2425	1.0
Moderate	3.6	2850	1.2	11.0	2850	1.2
Heavy	0.1	3800	1.6	0.3	3800	1.6
>18 year Female						
Sedentary	26.7	1875	0.8	24.1	1875	0.8
Moderate	0.4	2225	0.9	1.9	2225	0.9
Heavy	0.0	2925	1.3	0.0	2925	1.3

$$\text{Average Calorie Norm per cu} = \frac{\Sigma N_i R_i C_i}{\Sigma N_i C_i}$$

Urban = 2125
Rural = 2161

average norm and might send the signals of complecency when it reaches the level just above the average norms.

Percent of Persons Below the Norm for

Households with per cu consumption around the RDA = 42.8

Households with per cu consumption below the norm = 69.2

Households with per cu consumption above the norm = 14.8

For Delhi the percentage of persons below the norm were estimated as 45% using household level averages whereas the same was only 35% using individual level data.

4. Inconsistency between Norms

Many times norms for several parameters are obtained independent of each other and the consistency between them is ignored or lost sight of.

For example, while assessing and classifying an individual as normal on the basis of Anthropometric measurements for adults certain norms are being used. For adults a person with the height of more than 145 cms is considered as normal. Using weight as the measurement an adult with more than 45 kg weight is classified as normal. Making use of both weight and height of adults, yet another index is used which measures Body Mass Index (BMI) as under :

BMI = weight (kg) divided by height (cm) square

The cut off point for classification of an individual as normal is the BMI value of 18.5 or more. It may be seen that this cut off of 18.5 or more is not consistent with the norm for height of 145 cms and to that of weight as 45 kg as the BMI based on these would be over 20.

5. Grouping of Data

The grouping of data and frequency distribution is taught at the very initial stages and is the foundation for the development of theory of statistics. There is an important issue which relates to the method of grouping and analysis of grouped data. Let me take the example of studying the age at which there is a maximum undernutrition in children upto 5 years of age. The general

Table 3. Percentage of underweight children in UP

Age (Months)	Under-Nutrition (%)
3-5	29
6-11	58
12-23	60
24-35	55
36-47	50
48-59	47

Based on SD classification
Source : Sheila Vir & A.K. Nigam

groupings considered in this situation are 3-5 months, 6-11 months, 12-23 months, 24-35 months, 36-47 months and 48-59 months.

An analysis of incidence of underweight based on data in terms of "age groups" is presented in (Table-3).

As per the "age group" analysis, it is observed that the maximum underweight occurs at the age of 12-23 months. Since the "age groups" are not equal in time intervals, the analysis as above may give erroneous information. Using the same data the analysis by "single months" was also undertaken (Table-4).

Table 4. Percentage of underweight children in UP

Age (Months)	Under-Nutrition (%)
3	16
4	23
5	42
6	43
7	52
8	62
9	61
10	63
11	69
12	61
13	55
14	60
15	61
16	59
17	56
18	62
19	61
20	45
21	57
22	61
23	59

The analysis reveals that the maximum under nutrition in fact occurs in children at the age of 8-11 months and the plateau begins at 12th month itself

and not at 24 months. Thus, the data using single months as age provides better insight than the grouped data. This also suggest that the grouping of months in a category has to have some basis, the underlying assumption being homogeneity which is generally over looked.

6. Quality of Life Approach Based on Verifiable Indicators as an Alternative to the Income/Expenditure Approach for Measurement and Identification of Poverty

Measurement of Poverty

The concept of poverty line is based on the average calorie norms of 2400 calories per capita per day for rural areas and 2100 calories per capita per day for urban areas. Since entire planning exercise is in monetory terms, the average calorie requirement norm has to be translated into monetory equivalent. The data of National Sample Survey Organisation (NSSO) on consumption expenditure for 1973-74 was used for working out the consumption expenditure per capita per month corresponding to the desired average calorie norm.

It was estimated that, on an average, consumption expenditure of Rs. 49.09 per capita per month (1973-74) corresponded to the calorie intake of 2400 per capita per day in rural areas and consumption expenditure of Rs. 56.64 per capita per month to the calorie intake of 2100 per day in urban areas.

This is referred to as the base poverty line. This poverty line is updated using CPI of agricultural workers for rural areas and CPI of industrial workers for urban areas.

The poverty line which was estimated for the base year assumes a relationship of consumption of expenditure (in rupees) and the average calorie requirement norm on an average. This assumption does not hold good.

For measurement of poverty the per capita income or expenditure is used which is difficult to assess. Also, there has been a recent controversy about

Table 5. Poverty line (Rs. per capita per month)

Year	Rural	Urban
1977-78	56.84	70.33
1983-84	89.50	115.65
1987-88	115.20	162.16
1993-94	205.84	281.35

Table 6. Cross tabulation of percentage of persons below poverty line and below calorie norm

	Rural			Urban		
	Below poverty line	Above poverty line	Total	Below poverty line	Above poverty line	Total
1977-78						
Below calorie norm	45.32	12.47	57.79	37.33	11.95	49.28
Above calorie norm	12.31	29.21	42.21	12.66	38.06	50.72
Total	57.63	42.37	100.00	49.94	40.01	100.00
1983-84						
Below calorie norm	37.75	28.29	66.64	26.31	34.37	60.68
Above calorie norm	3.63	29.73	33.37	2.47	36.85	39.32
Total	41.38	58.62	100.00	28.78	71.22	100.00
1987-88						
Below calorie norm	29.39	36.37	65.76	18.08	38.67	56.75
Above calorie norm	2.97	31.27	34.24	2.78	40.47	43.25
Total	32.36	67.64	100.00	20.86	79.14	100.00
1993-94						
Below calorie norm	31.20	26.20	57.40	29.00	35.87	64.80
Above calorie norm	6.10	36.50	42.60	3.50	31.70	35.20
Total	37.30	62.70	100.00	32.50	67.50	100.00

the reference period of 30 days verses 7 days on this subject. Therefore, on the measurement of quality of life it is worth while considering other alternatives than the present one on poverty.

This is in view of the fact that now, data have become available from National Family Health Surveys (1992-93-I) and (1998-1999-II) which are based on a sample of about 90,000 households (Table-7).

Using this data or similar data from other surveys quality of life indices, based on housing characteristics, viz. ownership of household assets/goods, livestock and agriculture could be constructed by assigning appropriate scores to individual variables. These indices could be used for measurement of quality of life as proxy for poverty and for identification of poor households. A household below the certain aggregate score could be taken as poor. Several alternatives could be considered in this regard. In one, only the housing characteristics can be considered and in another, assets can also be included.

Table 7. Percent distribution of households by housing characteristics according to residence, India, 1992-93 & 1998-99

Housing characteristic	Urban	Rural NFHS-2	Total	Urban	Rural NFHS-1	Total
Electricity						
Yes	91.3	48.1	60.1	82.8	38.7	50.9
No	8.7	51.9	39.9	17.2	61.3	49.1
Total percent	100.0	100.0	100.0	100.0	100.0	100.0
Source of drinking water						
Piped	74.5	25.0	38.7	69.5	19.3	33.1
Hand Pump	18.1	47.3	39.2	18.1	41.6	35.1
Well water	6.0	23.5	18.7	9.2	32.1	25.8
Surface water	0.4	3.5	2.6	1.0	5.1	3.9
Other	1.0	0.7	0.8	2.2	2.0	2.0
Total percent	100.0	100.0	100.0	100.0	100.0	100.0
Sanitation facility						
Flush toilet	63.9	8.8	24.0	60.1	6.9	21.6
Pit toilet/latrine	16.8	10.0	11.9	15.5	5.9	8.6
Other	0.0	0.1	0.1	0.3	0.1	0.1
No facility	19.3	81.1	64.0	24.1	87.1	69.7
Total percent	100.0	100.0	100.0	100.0	100.0	100.0
Main type of fuel used for cooking						
Wood	23.1	73.1	59.3	29.6	77.0	63.9
Crop residues	0.5	8.1	6.0			
Dung cakes	1.4	8.4	6.5	3.0	12.2	9.7
Coal/coke/lignite/ charcoal	4.9	1.7	2.6	8.6	2.3	4.1
Kerosene	21.5	2.7	7.9	22.5	1.9	7.6
Electricity	0.8	0.2	0.4	1.0	0.1	0.4
Liquid petroleum gas	46.9	5.1	16.7	33.4	1.9	10.6
Biogas	0.6	0.5	0.6			
Other	0.2	0.2	0.2	1.9	4.5	3.8
Total percent	100.0	100.0	100.0	100.0	100.0	100.0
Type of house						
Kachha	9.4	41.4	32.5	17.2	60.4	48.5
Semi-pucca	24.4	39.5	35.3	26.2	28.4	27.8
Pucca	66.0	19.0	32.0	56.6	11.2	23.7
Missing	0.2	0.2	0.2			
Total percent	100.0	100.0	100.0	100.0	100.0	100.0
Persons per room						
<3	68.6	60.2	62.5	63.4	58.6	59.9
3-4	19.5	24.4	23.1	21.8	25.4	24.4
5-6	8.3	10.7	10.0	9.9	10.9	10.6
7 +	3.5	4.5	4.2	4.7	5.0	5.0
Missing	0.1	0.1	0.1		0.1	0.1
Total percent	100.0	100.0	100.0	100.0	100.0	100.0
Mean number of persons per room	2.5	2.8	2.7	2.7	2.8	2.8
Number of households	25,243	65,953	91,196	24,424	64,138	88,562

Table 8. Scores for the variables used in the construction of quality of life indices

	Variable				Scores
1.	Separate room for cooking	Yes			=1
		No			=0
2.	Type of house	Pucca			=2
		Semi-pucca			=1
		Kachha			=0
3.	Source of light	Electricity			=2
		Kerosene or gas or oil			=1
		Others			=0
4.	Fuel for cooking	Electricity or gas or bio-gas			=2
		Coal or charcoal or kerosene			=1
		Others			=0
5.	Source of drinking water	Well or pipe or hand-pump (Own)			=2
		Well or pipe or hand-pump (Public)			=1
		Others			=0
6.	Toilet facility	Own flush toilet			=3
		Flush toilet (Public or shared) or own pit toilet			=2
		Shared pit toilet or public pit toilet			=1
		Others			=0
7.	Type of livestock owned	Bullock			=2
		Cow			=2
		Buffalo			=2
		Goat			=1
		Sheep			=1
		Camel			=2
8.	Ownership of goods	Sweing machine			=2
		Clock/Watch			=1
		Sofa set			=2
		Fan			=2
		Radio/Transistor			=2
		Refrigerator			=3
		Television			=3
		VCR/VCP			=3
		Bicycle			=2
		Motorcycle/Scooter			=3
		Car			=4
9.	Ownership land	Acres	Irrigated	Unirrigated	
		No land	0	0	
		Less than one	1	0	
		1-1.99	2	1	
		2 – 5	3	2	
		5 +	4	3	

The trends in poverty can be worked out using the data of NFHS-1 (1992-93) and 2 (1998-99).

Similar data are also available through NSSO and a good time trend could be generated using this data.

Total score on the basis of housing characteristics could be worked out for each household and compared with the minimum score required for that household to be classified as poor or non poor. One of the ways of assigning score is presented in Table-8.

There is also an important issue relating to what would be the minimum score to be used as cut off for the purpose. It could be on the basis of the minimum needs/provisions for a reasonably good quality of life. Alternatively to start with it could be the one which corresponds to the accepted magnitude of poverty. The one which is consistent with the magnitude of poverty at All India level for 1993-94 corresponds to score of 3 for rural areas and of 7 for urban areas, considering only the housing variable. But when one takes both housing as well as the assets the corresponding scores are 8 for rural and 14 for urban (Table-9).

Table-9. Nearest cut-off point corresponding to the Planning Commission's poverty ratio at All India level

Index	Rural	Urban	Combined
QLI	8	14	9
QLI-1	3	7	3

It may be noted that these scores could be by various combinations on availability of household charateristics and assets. The scores for different states are worked out and based on housing characteristics alone, the percentage of households below the desirable quality of life for rural are given in Table 10.

It could be seen that this alternative approach of quality of life index based on household characteristics and assets owned is simple workable and could be easily understood. Further this would also help in identification of the poor.

These are only some of illustrations. There could be more. It is imperative for statisticians to prove their mettle by coming out with analysis and interpretation of the data to generate appropriate information on the issues and problems dogging the country.

Table 10. Percentage of households below desired quality of life - Rural

Sl. No.	States/UTs	
1	Andhra Pradesh	43.5
2	Arunanchal Pradesh	29.5
3	Assam	49.1
4	Bihar	69.9
5	Goa	10.2
6	Gujarat	35.6
7	Haryana	18.6
8	Himachal Pradesh	15.9
9	Jammu region of J & K	25.7
10	Karnataka	37.0
11	Kerala	17.8
12	Madhya Pradesh	69.2
13	Maharashtra	42.7
14	Orissa	74.3
15	Punjab	5.2
16	Rajasthan	48.6
17	Tamil Nadu	42.1
18	Uttar Pradesh	58.0
19	West Bengal	72.0