Jour. Ind. Soc. Ag. Statistics
51(2&3), 1998 : 187-202

Identifying the "Poor" Using Binary Classifiers

J. Roy and C.H. Sastry
Indian Statistical Institute, Calcutta

SUMMARY

To identify the "poor" to be supported through poverty eradication programs is an important problem for all Governments. The calorie consumption approach due to Dandekar and Rath [1] is difficult to apply because an elaborate survey of household consumption is needed. In a meeting of the Governing Council of the National Sample Survey Organisation of India, the Chairman B.S. Minhas suggested the use of simple questions, each with only two possible answers: "yes", or, "no" — where an "yes" answer would indicate some form of poverty. Rudra *et al* [2], carried out an exploratory survey of 160 households in 8 villages in West Bengal using 17 questions with "yes-no" answers. No objective methodology was developed in that paper for identifying the "poor" using the answers to these questions.

In this paper a methodology is developed for using this type of data for the above purpose. The concepts of "true identifier" and "fallible classifiers" of poverty, — or, for that matter of any particular trait — are introduced. When the true identifier is not available but instead of it a few fallible classifiers are available, a "best surrogate identifier" is defined based on all the fallible classifiers available. Computational methods are developed for obtaining this best surrogate identifier, as well as for reducing the number of original fallible classifiers in a logical way. The methodology is applied on data mentioned at [2] to demonstrate that only four of the 17 fallible classifiers are important for identifying the poor.

Keywords: Fallible classifiers, Binary vectors, Coefficient of agreement, Cluster analytic techniques.

1. Introduction

1.1 Household poverty is a multi-faceted characteristic, and inadequacy of energy derived from food is only one of them. One can think of inadequacy of a number of other items like housing, clothing, sanitation, education etc. -each of which is an indicator of poverty. In a survey of households, each household can be asked whether it had an inadequate provision of these primary necessities of life. An "yes"-answer to such a question would indicate that the household was poor in some sense. But an "yes"-answer to a single question by itself would not be necessarily definitive. Let an "yes"-answer be coded as 1 and a "no" as 0. the responses from a collection of households to a question of this type coded in the above manner can be represented as a binary statistical variable. If m such questions are used, on each of a collection of n households, the responses to the j-th question can be represented as the binary variable

$$X(j) = \{x(j, 1), x(j, 2), \dots, x(j, n)\}$$

where, for i = 1, 2, ..., n and j = 1, 2, ..., m, x (j, i) = 1 (or, 0) according as the response of the i-th household to the j-th question is "yes", (or, "no"). This will be called the j-th fallible classifier of poverty-fallible because by itself it does not determine poverty definitively.

1.2 There is not a priori reason to prefer any amongst the fallible classifiers to others: each is supposed to be equally valid and reliable.

1.3 In the ideal situation, when there is absolute agreement about the "poor", or, "not-poor" — status of each household, the collection of households can be described by a binary statistical variable

$$Y = \{y(1), y(2), \dots, y(n)\}$$

where for i = 1, 2, ..., n, y(i) = 1, (or, 0) according as the i-th household is, (or, is not) poor. Y will be called the true identifier of poverty, but it is seldom available.

1.4 The problem is to devise a "simple" procedure for deciding whether a particular household is or, not "poor", using as few as feasible of the fallible classifiers. The class of simple procedures to be considered is based on the Guttman score — the total number of "yes"-answers obtained from a household to all the fallible classifiers used. A household is classified as "poor" if this score is not below a cut-off value. The specific problem is to determine an "optimum" subset of fallible classifiers and an "optimum" cut-off score for that subset.

2. *Notations and Definitions*

2.1 Let $J = (j_1, j_2, \dots, j_k)$ be a given subset of the integers 1, 2, ..., m and $S(J) = X(j_1) + X(j_2) + \dots + X(j_k)$. When J = (1, 2, ..., m), the notation S will be used for S(J). In general the n components of S(J) will be denoted by {s(J,1), s(J, 2), ..., s(J,n)} and those of S by {s(1), s(2), ..., s(n)}.

2.2 For a given subset J and a given integer l, $0 \leq l \leq |J|$, our main interest will be in the use of a binary vector U(J, l)—to be called a simple composite fallible classifier—defined as:

$$U(J, l) = \{u(J, l, 1)\}, u(J, l, 2), ..., u(J, l, n)\}$$

where, u(J, l, i) = 1 (or, 0) according as s(J, i) $\geq l$ (or, $< l$) for i = 1, 2, ...n.

2.3 The vector (1, 1, ..., 1) will be denoted by E.

2.4 A measure of agreement between two classifiers U and V will be defined as

$$r(U, V) = (UV' + (E-U)(E-V)')/n$$

This is analogous to the concept of the correlation coefficient between two statistical variables.

2.5 As a measure of agreement between a true identifier Y and a subset of fallible classifiers {X (j)}, for j in J, it is proposed to use R (J, Y) = max r (U (J, l)), where the maximum is with respect to l in the range 1, 2 ..., |J|. This will be called the coefficient of composite agreement and is analogous to the concept of multiple correlation in statistics.

3. Surrogate for True Identifier

3.1 In the absence of the true classifier Y the following surrogate is proposed from the class of binary vectors U(S, l): $0 \leq l \leq N$, by using the maximin principle. Let l^* be the value of l which maximises the minimum value of r (X (j), U(S, l)) with respect to l for l = 1, 2, ..., m. Then U* = U (S, l^*) is the proposed surrogate for Y. Whenever the true classifier Y is not available, U* obtained this way will be used in its place and when there is no possibility of confusion, denoted by Y itself. It should be noted that by definition, for the surrogate, R(S, U*) = 1 whereas for the true identifier when it exists, this is smaller than unity generally.

4. Choosing a Simple Composite Classifier

4.1 Two different procedures are described below for choosing an appropriate subset of linked variables for the purpose of constructing a simple composite classifier-one based on the concept of multiple agreement and the other based on representatives of possibly overlapping identified clusters of a linked classifiers.

4.2 The first procedure is similar to the forward step-wise procedure in multiple regression. In what follows, the symbol Y is used for the true identifier when it exists, or, for the best surrogate identifier when the true identifier does not exist. In the first step, one selects the single fallible classifier which has the highest coefficient of agreement with Y. Let this fallible classifier be denoted by $X^{(1)}$. The coefficient of agreement of $X^{(1)}$ with Y is noted. In the next step one more fallible classifier is selected, which along with $X^{(1)}$ has the highest measure of composite agreement with Y.

Let this fallible classifier be denoted by $X^{(2)}$. The coefficient of composite agreement of $X^{(1)}$ and $X^{(2)}$ with Y is then noted. One continues in this way and introduces an additional fallible classifier in each step until the coefficient of composite agreement ceases to increase, or, when it attains a preassigned high value.

4.3 The second procedure is heuristic, based on cluster analytic techniques. It does not use the coefficient of composite agreement of the chosen fallible classifiers with the true identifier of its surrogate. Instead, it groups the fallible classifiers into a number of "perfect" clusters or "cliques" and selects a representative from each clique. Given a lower limit L of the coefficient of agreement, a subset of fallible classifiers is said to be a "clique" if the coefficient of agreement between any two fallible classifiers belonging to the subset is at least L and if no other fallible classifier can be included in the subset without breaking this condition. A representative of a cluster of fallible classifiers is the one whose minimum agreement with other fallible classifiers is the highest.

4.4 In this paper the forward step-wise procedure is used in developing a procedure for identifying the "poor" in rural West Bengal. The heavy computations were carried on a PC using a program developed by the second author.

5. Data Description

Response to 17 questions were obtained from 160 families in an investigation on rural poverty conducted by Professor N. Bhattacharya, Ms. Snigdha Chakraborty and Dr. Krishna Majumdar, Indian Statistical Institute, Calcutta, in 8 villages of Jamboni block, Midnapur district, West Bengal, during December 1990 and May 1991. The yes/no answers have been coded as 0 or 1 in such a way that a response of 1 indicates poverty. The details are shown below.

Variable	Description	Response	
		Yes	No
1	Meat/Fish/Egg eaten last month	0	1
2	Per family bed room (≤ 1)	1	0
3	Room height (≤ 1.68 mtrs)	1	0
4	Living place protected from rain/storm	0	1
5	Posess woolen clothing	0	1
6	Per person woolen cloth (< 1)	1	0
7	Per adult lady saree (< 2)	1	0
8	Bed lacked mattresses	1	0
9	Household lacked blankets/quilts	1	0
10	Dining plates per adult (< 1)	1	0
11	Household has children (6-14 yrs.) not going to school due to economic constraints	1	0
12	During last 3 years whether any female in the household given special food before and after delivery	0	1
13	Had food throughout the year	0	1
14	Children (0-4 yrs) did not get milk daily	1	0
15	Begging as a profession	1	0
16	Food items borrowed/received as gift last month	1	0
17	Food items collected from nature/others property	1	0

6. *Method of Analysis*

Let X be an n × m matrix containing data from n households and m fallible classifiers. The (m + 1)th column contains the total of "1"s in the household.

Taking K, which varies from 0 to m, as the cut-off score, the number of households, a, satisfying the condition

$$(X(i, j) = 1 \text{ and } X(i, m+1) \geq K) \text{ or}$$
$$(X(i, j) = 0 \text{ and } X(i, m+1) < K)$$

is obtained for each fallible classifier.

Minimum of such A's over all the fallibles, for a fixed K, when maximised over all K's, gives an optimum cut-off score Ck. This is obtained as 9 in this computation. Now, build a surrogate classifier X(i, m + 2) by the criterion

$$X(i, m+2) = 1 \text{ if } X(i, m+1) \geq Ck$$
$$= 0 \text{ otherwise}$$

for each household.

Starting with one fallible classifier which gives maximum agreement with the surrogate classifier, we examine stepwise, L-tuple, where L = 2, 3, ..., m,

Agreement of the Guttman score for different cut-off values with individual classifiers

Cutoff score	Fallible classifiers																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	163	769	575	250	919	1000	281	956	531	838	163	281	781	475	13	300	313
1	163	769	575	250	919	1000	281	956	531	838	163	281	781	475	13	300	313
2	163	769	575	250	919	1000	281	956	531	838	163	281	781	475	13	300	313
3	163	769	575	250	919	1000	281	956	531	838	163	281	781	475	13	300	313
4	169	775	581	256	913	994	288	950	538	844	169	288	788	481	19	306	319
5	194	800	581	281	900	969	313	925	550	856	194	313	813	506	44	319	344
6	219	800	606	306	875	944	338	913	575	844	219	325	813	531	69	344	356
7	281	800	644	381	813	869	363	850	613	819	294	388	825	581	144	394	394
8	344	750	731	481	738	756	438	750	625	731	381	475	825	619	256	456	431
9	500	669	763	638	556	538	544	556	631	550	550	619	656	638	475	600	525
10	650	481	613	713	369	313	644	344	631	438	750	719	494	638	700	688	600
11	738	369	538	725	231	150	706	194	569	300	850	756	356	600	863	675	675
12	800	281	463	750	144	63	744	106	519	225	850	731	281	550	938	713	700
13	825	244	438	750	94	13	719	56	481	175	825	731	231	538	988	700	700
14	838	231	425	750	81	0	719	44	469	163	838	719	219	525	988	700	688
15	838	231	425	750	81	0	719	44	469	163	838	719	219	525	988	700	688
16	838	231	425	750	81	0	719	44	469	163	838	719	219	525	988	700	688
17	838	231	425	750	81	0	719	44	469	163	838	719	219	525	988	700	688

MaxMin of Agreement = 475 Cut-off Value = 9

Agreement of the individual classifier with the surrogate

500	669	<u>763</u>	638	556	538	544	556	631	550	550	619	656	638	475	600	525

Max. Agreement Attained : 762.5 with Variable = 3

Stepwise selection of Additional Fallible Classifiers

	Step 2 : Additional Variable (s)															
Cutoff score	2	1	[4]	5	6	7	8	9	10	11	12	13	14	15	16	17
0	538	538	538	538	538	538	538	538	538	538	538	538	538	538	538	538
1	663	744	[781]	563	538	713	563	713	600	763	756	650	719	763	750	688
2	769	519	619	756	763	594	756	681	713	550	625	769	681	475	613	600

Cut-off Value = 1 Maximum Agreement = 781.25
Chosen Additional Variable = 4

	Step 3 : Additional Variable (s)														
Cutoff score	1	2	5	6	7	8	9	10	11	12	[13]	14	15	16	17
0	538	538	538	538	538	538	538	538	538	538	538	538	538	538	538
1	750	675	575	538	713	563	713	594	775	769	644	738	781	763	694
2	669	781	769	781	725	775	725	756	681	738	[813]	769	625	731	706
3	481	613	613	619	506	619	594	600	494	513	600	531	469	506	525

Cut-off Value = 2 Maximum Agreement = 812.5
Chosen Additional Variable = 13

fallible classifiers and select that additional fallible which contributes to maximum agreement with the surrogate classifier. The agreements shown in the results are A/n*1000.

Results of Guttman score agreements, the optimum cut-off value and the first few iterations of stepwise selection of additional fallible classifiers are shown in tables before.

7. *Conclusion*

The tabulated results show that the contribution to the agreement improves until upto step no. 4, stabilises until step no. 7, dips in step no. 8 and gradually increases.

The sequence of fallible variables and their corresponding maximum agreements are:

Step No. :	1	2	3	4	5	6	7	8	9	10
Variables:	3	4	13	12	15	6	8	11	16	1
Cut-off value:	9	1	2	2	2	3	4	4	5	5
Agreement:	762	781	812	850	850	850	850	837	850	881
Step No. :	11	12	13	14	15	16	17			
Variables:	7	2	17	14	9	5	10			
Cut-off value:	5	6	6	7	7	8	9			
Agreement:	881	887	906	918	962	081	1000			

Thus, we may conclude that the contribution of fallible classifiers 3 (smaller room heights), 4 (unprotected living place from rain and storms), 13 (lack of food throughout the year) and 12 (no special food before/after delivery) are important and adequate to identify the "poor" in the concerned rural areas.

ACKNOWLEDGEMENTS

Our thanks are due to Professor Nikhilesh Bhattacharya, Dr. Snigdha Chakraborty and Dr. Krishna Majumdar, of Indian Statistical Institute, Calcutta for providing us with the data from their survey on rural poverty for the illustrative example discussed above.

REFERENCES

[1] Dandekar, V.M. and Rath, Nilakantha, 1971. Poverty in India. *The Indian School of Political Economy*, Pune.

[2] Rudra, Ashok, Chakraborty, Snigdha, Majumdar, Krishna and Bhattacharya, Nikhilesh, 1993. Criteria for identification of rural poor-preliminary results based on a survey in West Bengal. An article in the book in honour of Professor Dhiresh Bhattacharya, Bangiya Arthanithi Parishad.

Appendix

Values of poverty indicators for each household

S. No.	Poverty Indicators 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	TOT	FREQ
1.	0	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	13	
2.	0	1	1	0	1	1	1	1	1	1	0	1	1	1	0	1	1	13	2
3.	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	12	
4.	0	1	0	0	1	1	1	1	1	1	1	1	1	1	0	1	0	12	
5.	0	1	1	1	1	1	1	1	0	1	1	0	1	1	0	0	1	12	
6.	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	0	12	
7.	0	1	1	1	1	1	0	1	1	1	0	1	1	1	0	0	1	12	
8.	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1	12	
9.	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	1	0	12	
10.	0	1	1	0	1	1	1	1	1	1	1	0	1	0	0	1	1	12	8
11.	1	1	1	0	1	1	0	1	0	1	1	1	0	1	0	1	0	11	
12.	0	1	1	1	1	1	0	1	1	1	0	1	1	1	0	0	0	11	
13.	0	1	0	0	1	1	1	1	1	1	1	0	1	1	0	1	0	11	
14.	0	1	1	0	1	1	1	1	1	1	1	0	1	0	1	0	0	11	
15.	0	1	1	1	1	1	1	1	1	0	0	1	1	1	0	0	0	11	
16.	0	1	1	0	1	1	0	1	1	1	0	1	1	1	0	1	0	11	
17.	0	1	1	1	1	1	0	1	1	1	1	0	1	0	0	0	1	11	
18.	1	1	1	0	1	1	0	1	0	1	0	1	1	1	0	0	1	11	
19.	0	1	1	0	1	1	0	1	0	1	1	1	1	1	0	0	1	11	
20.	0	1	1	1	1	1	0	1	1	1	0	1	1	1	0	0	0	11	
21.	0	1	1	0	1	1	1	1	1	1	0	1	1	1	0	0	0	11	
22.	0	1	1	0	1	1	0	1	1	1	0	1	1	1	0	1	0	11	
23.	0	1	1	1	1	1	0	1	1	1	1	0	1	0	0	0	1	11	
24.	0	1	1	0	1	1	0	1	1	1	1	0	1	1	0	0	1	11	14
25.	0	1	0	0	1	1	0	1	1	1	0	1	1	1	0	1	0	10	
26.	1	0	1	0	1	1	1	1	1	1	0	0	1	0	0	1	0	10	
27.	0	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	0	10	
28.	0	1	0	0	1	1	0	1	1	1	0	1	1	1	0	1	0	10	
29.	0	0	1	0	1	1	1	1	1	1	1	0	1	0	0	1	0	10	
30.	0	1	1	1	0	1	1	1	1	1	0	0	1	0	0	1	0	10	
31.	1	1	1	0	1	1	1	1	0	1	0	0	1	1	0	0	0	10	
32.	0	1	1	1	1	1	0	1	1	0	0	1	1	1	0	0	0	10	
33.	0	1	1	1	1	1	1	1	0	1	0	0	1	0	0	1	0	10	
34.	1	0	0	1	0	1	1	1	1	1	0	0	1	1	0	0	1	10	
35.	0	1	1	1	1	1	0	1	1	1	0	0	0	0	0	1	1	10	
36.	1	0	1	0	1	1	0	1	1	1	0	0	1	1	0	0	1	10	
37.	0	1	0	0	1	1	0	1	1	1	0	1	1	1	0	1	0	10	

38.	0	1	1	1	1	1	0	1	1	1	0	0	1	1	0	0	0	10	
39.	0	1	1	1	1	1	0	1	1	1	0	0	1	0	0	1	0	10	
40.	0	1	1	0	1	1	0	1	0	0	0	1	1	1	0	1	1	10	
41.	0	1	1	1	1	1	0	1	0	1	1	0	1	1	0	0	0	10	
42.	0	1	1	1	1	1	0	1	1	1	0	0	1	0	0	0	1	10	
43.	0	1	1	1	1	1	0	0	0	1	0	1	1	1	0	0	1	10	
44.	1	1	1	0	1	1	0	1	0	1	1	1	1	0	0	0	0	10	
45.	0	1	0	0	1	1	0	1	1	1	0	1	1	1	0	1	0	10	
46.	0	1	1	0	1	1	1	1	0	1	0	1	1	1	0	0	0	10	
47.	1	1	1	0	1	1	0	1	1	1	0	0	1	0	0	1	0	10	
48.	0	1	1	0	1	1	0	1	0	1	0	1	1	1	0	1	0	10	
49.	0	1	0	0	1	1	0	1	1	1	1	0	1	1	0	1	0	10	
50.	0	1	0	1	1	1	0	1	1	1	1	0	1	0	0	0	1	10	26
51.	0	1	1	0	1	1	1	1	1	1	0	0	1	0	0	0	0	9	
52.	0	1	1	1	1	1	0	1	0	0	0	0	1	1	0	1	0	9	
53.	1	1	1	0	1	1	0	1	1	0	0	0	1	1	0	0	0	9	
54.	0	1	1	0	1	1	0	1	0	1	0	1	1	1	0	0	0	9	
55.	0	1	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	9	
56.	0	1	1	0	0	1	1	1	1	1	0	0	1	0	0	0	1	9	
57.	0	1	1	0	1	1	0	1	1	1	1	0	1	0	0	0	0	9	
58.	1	1	0	1	1	1	0	1	1	1	0	0	1	0	0	0	0	9	
59.	0	1	1	0	1	1	0	1	1	1	0	0	1	1	0	0	0	9	
60.	0	1	1	1	1	1	0	1	0	1	0	1	0	1	0	0	0	9	
61.	0	1	1	1	1	1	0	1	0	1	0	0	1	0	0	0	1	9	
62.	1	1	1	0	1	1	0	1	1	0	0	0	1	0	0	1	0	9	
63.	0	1	0	0	1	1	0	1	1	1	0	1	1	1	0	0	0	9	
64.	0	1	1	0	1	1	0	1	0	1	0	0	1	1	0	1	0	9	
65.	0	1	1	0	1	1	1	1	0	1	0	0	1	0	0	0	1	9	
66.	0	1	1	0	1	1	0	1	0	1	0	0	1	0	0	1	1	9	
67.	0	1	1	0	1	1	0	1	0	1	0	0	1	1	0	1	0	9	
68.	1	0	1	1	1	1	0	1	0	1	0	0	1	0	0	0	1	9	
69.	1	1	1	0	1	1	0	0	0	0	0	1	1	1	0	0	1	9	
70.	0	1	0	0	1	1	0	1	0	1	0	1	1	1	0	1	0	9	
71.	0	1	1	0	1	1	0	1	0	1	0	0	1	1	0	0	1	9	
72.	0	1	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0	9	
73.	0	1	1	0	1	1	0	1	0	1	0	1	1	1	0	0	0	9	
74.	0	0	1	0	1	1	0	1	1	0	0	1	0	1	0	1	1	9	
75.	0	1	1	0	1	1	1	1	0	1	0	0	0	1	0	1	0	9	
76.	0	1	1	0	1	1	1	1	1	0	0	0	1	0	0	1	0	9	
77.	0	1	1	0	1	1	1	1	1	0	1	0	1	0	0	0	0	9	

78.	0	1	1	1	0	1	0	1	1	1	0	0	1	0	0	0	1	9	
79.	0	1	1	0	1	1	0	1	0	0	0	1	1	1	0	0	1	9	
80.	0	1	1	1	1	1	0	1	1	1	0	0	1	0	0	0	0	9	
81.	0	1	1	0	1	1	1	1	1	1	0	0	1	0	0	0	0	9	
82.	0	0	0	1	1	1	1	1	1	1	0	1	0	1	0	0	0	9	
83.	0	1	0	1	0	1	1	1	0	1	0	0	1	0	0	1	1	9	
84.	0	1	1	1	1	1	0	1	1	1	0	0	1	0	0	0	0	9	
85.	0	1	1	1	1	1	0	1	1	1	0	0	1	0	0	0	0	9	
86.	1	1	0	0	1	1	0	1	0	1	0	0	1	1	0	0	1	9	36
87.	0	0	0	0	1	1	0	1	1	1	0	1	1	1	0	0	0	8	
88.	1	1	0	0	1	1	0	1	1	1	0	0	1	0	0	0	0	8	
89.	0	1	1	0	1	1	0	1	1	1	0	0	1	0	0	0	0	8	
90.	0	1	0	0	1	1	1	1	0	1	0	0	1	0	0	1	0	8	
91.	0	1	1	0	1	1	0	1	1	0	0	0	1	0	0	0	1	8	
92.	0	1	1	0	1	1	0	1	1	1	0	0	1	0	0	0	0	8	
93.	0	1	1	0	1	1	0	1	0	1	0	0	1	0	0	1	0	8	
94.	0	1	0	0	1	1	0	1	1	1	0	0	1	0	0	0	1	8	
95.	0	1	1	1	1	1	0	0	0	1	0	0	1	1	0	0	0	8	
96.	0	1	1	0	1	1	0	1	1	1	0	0	1	0	0	0	0	8	
97.	0	1	1	0	1	1	1	1	0	1	0	0	1	0	0	0	0	8	
98.	0	1	1	0	1	1	0	1	1	1	0	0	1	0	0	0	0	8	
99.	1	1	0	0	1	1	1	1	0	1	0	0	1	0	0	0	0	8	
100.	1	0	0	0	1	1	0	1	0	1	1	1	1	0	0	0	0	8	
101.	0	1	0	0	0	1	1	1	1	1	0	0	0	1	0	0	1	8	
102.	0	0	0	0	1	1	1	1	0	1	1	0	1	0	0	1	0	8	
103.	0	0	0	0	1	1	1	0	1	1	0	0	1	1	0	0	1	8	
104.	0	0	0	0	1	1	0	1	1	1	1	0	1	0	0	0	1	8	
105.	0	0	0	0	1	1	1	1	0	1	0	0	1	1	0	1	0	8	
106.	0	0	0	0	1	1	1	1	0	1	0	1	1	1	0	0	0	8	
107.	0	1	0	0	1	1	0	1	1	1	0	0	1	1	0	0	0	8	
108.	0	0	0	0	1	1	1	1	0	1	0	0	1	1	0	1	0	8	
109.	0	1	1	1	0	1	0	1	0	1	0	0	1	1	0	0	0	8	
110.	1	0	1	0	1	1	0	1	0	1	0	0	1	1	0	0	0	8	
111.	0	1	1	0	1	1	0	1	1	1	0	0	0	1	0	0	0	8	
112.	0	1	1	1	1	1	0	1	0	0	0	0	1	0	0	0	1	8	
113.	0	1	0	0	1	1	0	1	0	1	0	1	1	1	0	0	0	8	
114.	0	1	1	0	1	1	0	1	1	1	0	0	1	0	0	0	0	8	
115.	1	1	1	0	0	1	0	1	0	1	1	0	1	0	0	0	0	8	
116.	0	1	0	1	1	1	0	1	0	1	0	0	0	1	0	0	1	8	
117.	0	0	0	0	1	1	0	1	0	1	0	1	1	1	0	0	1	8	
118.	0	1	0	0	1	1	0	1	1	0	0	1	1	1	0	0	0	8	

119.	0	1	0	0	1	1	0	1	1	1	0	0	1	0	0	0	1	8	
120.	0	1	1	1	1	1	0	1	0	1	0	0	1	0	0	0	0	8	
121.	0	0	0	0	1	1	0	1	1	1	0	0	0	1	0	1	1	8	35
122.	1	0	0	0	1	1	0	1	0	1	0	1	0	1	0	0	0	7	
123.	0	1	0	0	1	1	0	1	1	1	0	0	0	0	0	0	1	7	
124.	1	0	0	0	1	1	0	1	1	1	0	0	0	1	0	0	0	7	
125.	0	1	0	0	1	1	1	1	0	1	0	0	1	0	0	0	0	7	
126.	0	1	0	0	1	1	0	1	0	1	1	0	1	0	0	0	0	7	
127.	1	0	1	0	1	1	0	1	0	1	0	0	0	0	0	0	1	7	
128.	0	1	0	0	1	1	0	1	1	1	0	0	1	0	0	0	0	7	
129.	0	1	0	0	0	1	0	1	0	1	1	0	1	1	0	0	0	7	
130.	0	1	1	0	1	1	1	1	0	1	0	0	0	0	0	0	0	7	
131.	0	1	0	0	1	1	0	1	1	1	0	0	1	0	0	0	0	7	
132.	0	0	0	0	1	1	0	1	0	1	0	0	1	0	0	1	1	7	
133.	0	1	0	0	0	1	0	1	1	1	0	0	1	0	0	0	1	7	
134.	0	1	0	1	1	1	0	1	1	0	0	0	0	1	0	0	0	7	
135.	0	1	0	0	1	1	1	1	0	0	0	0	1	0	0	1	0	7	
136.	0	1	0	0	0	1	0	0	1	1	0	1	0	1	0	0	1	7	
137.	0	0	0	0	1	1	0	1	1	1	0	0	0	0	0	1	1	7	
138.	0	1	0	0	1	1	0	1	0	1	0	0	0	1	0	1	0	7	
139.	1	1	0	0	1	1	0	1	0	1	0	0	1	0	0	0	0	7	18
140.	0	0	0	0	1	1	1	1	0	1	0	0	1	0	0	0	0	6	
141.	0	1	1	0	1	1	0	1	0	1	0	0	0	0	0	0	0	6	
142.	0	0	0	0	0	1	0	1	1	1	0	0	0	1	0	0	1	6	
143.	1	0	1	0	1	1	0	1	0	1	0	0	0	0	0	0	0	6	
144.	0	0	0	0	1	1	0	1	0	0	0	1	0	1	0	1	0	6	
145.	0	1	0	0	1	1	0	1	0	0	0	0	1	0	0	1	0	6	
146.	0	0	1	0	1	1	0	1	1	1	0	0	0	0	0	0	0	6	
147.	0	1	0	0	1	1	0	1	0	1	0	0	1	0	0	0	0	6	
148.	0	0	0	0	1	1	1	0	0	1	0	0	1	0	0	0	1	6	
149.	0	1	0	0	1	1	0	1	0	1	0	0	0	0	0	0	1	6	
150.	0	1	0	0	1	1	1	1	0	0	0	0	1	0	0	0	0	6	
151.	0	1	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	6	12
152.	0	1	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	5	
153.	0	1	0	0	1	1	0	1	0	0	0	0	1	0	0	0	0	5	
154.	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0	1	5	
155.	0	0	0	0	1	1	0	1	0	1	0	1	0	0	0	0	0	5	4
156.	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	4	
157.	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	1	0	4	
158.	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	4	
159.	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	4	4
160.	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	3	1
																		12	

2. Program Listing

```
50  REM ANALYSIS OF 0.1 DATA ON RURAL POVERTY INDICATORS
60  DEFINT A-Z
70  OPTION BASE 1
80  DIM X(201, 22)
90  INPUT "DATA FILE name ="; A$
100 OPEN "I", # 1, A$
110 INPUT "No. of variables = "; M
120 INPUT "OUTPUT File name ="; mB$
130 OPEN "O", # 2, B$
140 I = 1
150 X (I, M + 1) = 0
160 FOR J = 1 TO M
170 IF EOF (1) THEN GOTO 220
180 INPUT # 1, X (I, J)
190 X(I, M + 1) = X (I, M + 1) + X (I, J)
200 NEXT J
210 I = I +1: GOTO 150
220 CLOSE # 1 : N = I - 1
230 PRINT # 2,"          AGREEMENT OF THE GUTTMAN SCORE FOR "
240 PRINT # 2,"             DIFFERENT CUT-OFF  VALUES WITH"
250 PRINT # 2,"                INDIVIDUAL CLASSIFIERS"
260 PRINT # 2," CUTOFF         FALLIBLE CLASSIFIERS"
270 PRINT # 2," SCORE";
280 FOR I = 1 TO M: PRINT # 2, USING "####";I;: NEXT I
290 PRINT # 2," ": X$ = STRING$ (10, 45): PRINT # 2, X$
300 AMAX = 0
310 FOR K = 0 TO M
320 AMIN = N+1
330 PRINT # 2, USING " ####" ;K; " ";
340 FOR J = 1 TO M
350 A = 0
360 FOR I = 1 TO N
```

```
370 IF (X (I, J) = 1 AND X (I, M + 1) >=K ) OR (X (I, J) = 0 AND X (I, M + 1)
    < K) THEN A = A+ 1
380 NEXT I
390 IF A < AMIN THEN AMIN = A
400 PRINT # 2, USING "####"; (A/N) * 1000;
410 NEXT J
420 IF AMIN > AMAX THEN AMAX = AMIN: CK = K
430 PRINT 2, " "
440 NEXT K
450 PRINT #2, X$: PRINT #2, " MaxMin of Agreement = "; (AMAX/N)*
    1000; "Cut-off Value = "; CK
460 PRINT # 2, " "
470 PRINT #2, "          AGREEMENT OF THE INDIVIDUAL CLASSIFIER
    WITH THE SURROGATE"
480 PRINT #2, " "
490 FOR I = 1 TO N
500 IF X (I, M +1) >= CK THEN X (I, M + 2) =1 ELSE X (I, M+ 2) = 0
510 X (I, M + 1) = 0
520 NEXT I
530 AMAX = 0 : PRINT #2, " ";
540 FOR J = 1 TO M
550 A = 0 : X (N + 1, J) = J
560 FOR I =1 TO N
570 IF (X (I, J) = 1 AND X (I, M + 2) = 1 ) OR (X (I, J)) =
    0 AND X (I, M + 2) = 0 ) THEN A = A + 1
580 NEXT I
590 IF A > AMAX THEN AMAX = A : CJ = J
600 PRINT # 2, USING "####"; (A/N)* 1000;
610 NEXT J : PRINT # 2, " "
620 PRINT # 2, " Max. Agreement Attained:"; (AMAX/N) *1000;" With
    variable = ";CJ
630 FOR I = 1 TO N + 1
640 EX = X(I, 1)
650 X (I, 1) = X (I, CJ)
```

```
660 X (I, CJ) = EX
670 X (I, M + 1) = X (I, 1)
680 NEXT I : PRINT #2, " "
690 PRINT # 2, " STEPWISE SELECTION OF ADDITIONAL FALLIBLE
    CLASSIFIERS"
700 PRINT # 2, " "
710 FOR K = 1 TO M - 1
720 PRINT # 2, "STEP " K + 1; ":" "Additional Variable (s) "
730 PRINT # 2 , "Cutoff";
740 FOR L = K + 1 TO M: PRINT #2, USING "####"; X (N + 1, L);: NEXT L:
    PRINT #2 ," "
750 PRINT # 2, "SCORE ——"
760 AMAX = 0
770 FOR KI = 0 TO K + 1
780 ALMAX = – (N + 1)
790 PRINT # 2, USING "####";KI; " ";
800 FOR KJ = K + 1 TO M
810 A = 0
820 FOR I = 1 TO N
830 SX = X (I, M + 1) + (I, KJ)
840 IF (SX>= KI AND X (I, M + 2) = 1) OR (SX < KI AND X (I, M + 2) = 0)
    THEN A = A + 1
850 NEXT I
860 IF A > ALMAX THEN ALMAX = A: CKJ = KJ
870 PRINT # 2, USING "####"; (A/N)*1000;
880 NEXT KJ
890 IF ALMAX > AMAX THEN AMAX = ALMAX: CKI = CKJ: TKI= KI
900 PRINT # 2, " "
910 NEXT KI
920 PRINT # 2, " "
930 PRINT #2," Cut-off Value = "; TKI, "Maximum Agreement= ";
    (AMAX/N)* 1000
940 PRINT #2, " Chosen Additional Variable =" ; X (N + 1, CKI)
950 PRINT #2, " "
```

```
960 FOR I = 1 TO N + 1
970 EX = X (I, K +1)
980 X ( I, K + 1) = = X (I, CKI)
990 X (I, CKI) = EX
1000 X (I, M + 1) = X (I, M + 1)+ X(I, K + 1)
1010 NEXT I : NEXT K
1020 PRINT # 2, "With Cut-off Value As "; CK
1030 PRINT # 2, "Predicted Indicator Variable Values are:"
1040 PRINT # 2, " "
1050 K = 0
1060 FOR I = 1 TO N
1070 PRINT #2, USING "##"; X (I, M + 2);
1080 K = K + 1
1090 IF K = 32 THEN PRINT # 2, " " : K = 0
1100 NEXT I
1110 PRINT #2, " "
1120 PRINT #2, " The Sequence of fallible Variables are:"
1130 PRINT # 2, " ": PRINT # 2; " ";
1140 FOR I = 1 TO M
1150 PRINT # 2, USING "####"; X (N + 1, I);
1160 NEXT I : PRINT # 2, " "
1170 CLOSE # 2
1180 END
```