Spatial Representation of the Results of Correlation Analysis

Chhaya Daudia and V.S. Phadke

Abstract

Quantitative techniques have been in vogue in geographical studies over the past fifty years. They are based on the tools developed by Mathematics or Statistics. Geographers use these to serve their ends and hence some modifications become inevitable to suit the intended purpose. Being interested in spatial dimension of the phenomena studied, it is necessary for geographers to extend the application of the tools developed by other disciplines to arrive at a spatial pattern which could then be interpreted. Some of them such as measures of central tendency or regression have been suitably amended to satisfy the needs of spatial analysis. Others such as correlation analysis continue to be discussed in an aspatial manner. This leaves our explanation incomplete. The present paper makes a modest attempt to suggest ways in which correlation analysis could be extended to arrive at spatial explanation.

Introduction

Quantitative techniques have been introduced to geography since the fifties of the twentieth century. Basically, they have their origin in the tools generated by Mathematics or Statistics. Geographers have modified these tools to suit to their requirement and applied them to analyse geographical data. However, while doing so very few techniques like measures of central tendency, regression, factor analysis have been extended to serve the needs of spatial representation (Sviatlovsky and Eells, 1937; Robinson and Bryson, 1957; Taaffe et al, 1963; Berry, 1966 ; Murdie, 1969). Taaffe (1970) provides a short but valuable insight into some of the techniques used in spatial interpretation. Yeates (1974) and Taylor (1977) have also proved a useful guide to geographers in their analysis of data

conceptually and functionally. While some of the techniques mentioned above have been suitably amended, many continue to be used in an aspatial manner. The resulting interpretation is more statistical and mathematical rather than geographical. Since geography is a spatial discipline, the results obtained using different techniques should be amenable to representation through maps. Geographers should focus on the latter and hence concentrate on arriving at a methodology that would help them to interpret the results spatially. They are working in this direction and one such attempt has discussed the methodology for interpreting the results of chi-square analysis spatially (Srivant and Phadke, 2001).

Correlation analysis is one of the techniques which is at present being interpreted aspatially. It is attempted using

four different methods, viz. scattergram which is graphical, and, statistical techniques such as concurrent deviation, Spearman's rank correlation and Karl Pearson's co-efficient of correlation which become increasingly sophisticated. While interpreting the results of correlation, one generally looks into three aspects, viz., trend (positive or negative), magnitude of the co-efficient value (0 to 1 on either side) and level of significance (reliability of the results obtained). Some researchers also go further and attempt to provide explanation for the departure of the value of coefficient from the ideal value of unity. Such a departure exists because all the units do not show proportionate changes in both the variables and that there are abnormally behaving units. The explanation, however, tends to be conceptual and functional when the spatial units showing such departure are not identified. An attempt should, therefore, focus on identifying such units. The thrust of the present paper is on such identification which also shows how such an identification and subsequent mapping improves interpretation by making it more geographical in nature.

Data Base

The paper is essentially of methodological type and hence data is used only as a means to illustrate the method. Moreover, it is also kept in mind that the demonstration of the method be easy and workable, if such techniques are to be made reachable to students, apart from researchers. The age of data and number of spatial units also become secondary considerations in such an exercise. Two variables, viz. Level of Urbanisation and Level of Female Literacy for different states of India for the year 1991 have, therefore, been selected. There are only 24 spatial units that make the exercise manageable. Once the methodology is grasped, it could be applied to more complex data to arrive at spatial interpretation.

Methodology

It is hypothesised that there is some correlation between level of urbanisation and level of female literacy. The nature and extent of this relation is estimated by following the well-trodden path using graphical and quantitative tools already in vogue. Subsequent to this, an attempt has been made to identify the abnormally behaving units along with the extent of abnormality shown by them. This information has then been used for mapping to arrive at a spatial pattern.

Results of Analysis

1. Scattergram Method: Data related to level of urbanisation (X) and level of female literacy (Y) were plotted on a scattergram and it was found to show a positive trend as could be judged from the scatter of points rising upward towards right (Fig. 1). The trend line was superimposed using semi-average method; this was preferred to that by inspection to ensure greater accuracy. It is observed that all the points of the scatter do not lie on the trend line, suggesting departure from the expectation. The points falling above the trend line indicate that female literacy is higher than expected in relation to urban population and vice versa. The simplest way to bring out abnormally behaving states is to divide them into two classes, viz. positive and negative departure,

Table 1

Identification of Residuals by Various Methods

Sr. No.	STATE	U. P (%)	F. L (%)	dx	dy	dx. dy	Rx	Ry	Rx-Ry	zx	zy	zy-zx
1	A.P.	26.89	32.72	-	-	-	10	19	-9	0.30	-0.84	-1.14
2	Arunachal P	12.80	29.69	-	+	-	21	20	1	-1.09	-1.03	0.06
3	Assam	11.10	43.03	-	+	-	22	16	6	-1.26	-0.18	1.08
4	Bihar	13.14	22.89	+	-	-	20	23	-3	-1.06	-1.46	-0.40
5	Goa	41.01	67.09	+	+	+	2	3	-1	1.69	1.36	-0.33
6	Gujarat	34.49	48.64	-	-	+	4	10	-6	1.05	0.18	-0.83
7	Haryana	24.63	40.47	-	-	+	12	17	-5	0.07	-0.34	-0.41
8	H. P.	8.69	52.13	-	+	-	24	6	18	-1.50	0.40	1.90
9	Karnataka	30.92	44.34	+	-	-	6	15	-9	0.69	-0.09	-0.78
10	Kerala	26.39	86.17	-	+	-	11	1	10	0.25	2.58	2.33
11	M. P.	23.18	28.85	-	-	+	13	21	-8	-0.07	-1.08	-1.01
12	Maharashtra	38.69	52.32	+	+	+	3	5	-2	1.46	0.42	-1.04
13	Manipur	27.52	47.60	-	-	+	8	11	-3	0.36	0.12	-0.24
14	Meghalaya	18.60	44.85	-	-	+	16	14	2	-0.52	-0.06	0.46
15	Mizoram	46.10	78.60	+	+	+	1	2	-1	2.19	2.10	-0.09
16	Nagaland	17.21	54.75	-	-	+	17	4	13	-0.66	0.57	1.23
17	Orissa	13.38	34.68	-	-	+	19	18	1	-1.03	-0.71	0.32
18	Punjab	39.55	50.41	+	+	+	7	8	-1	0.56	0.29	-0.26
19	Rajasthan	22.88	20.44	-	-	+	14	24	-10	-0.10	-1.62	-1.52
20	Sikkim	9.10	46.69	-	+	-	23	12	11	-1.46	0.06	1.52
21	T. N.	34.15	51.33	+	+	+	5	7	-2	1.01	0.35	-0.66
22	Tripura	15.30	49.65	-	-	+	18	9	9	-0.85	0.25	1.10
23	U. P.	19.84	25.31	+	-	-	15	22	-7	-0.39	-1.31	-0.92
24	W. B.	27.48	46.56	+	+	+	9	13	-4	0.35	0.05	-0.31

which could be mapped by distinct shades. In order to bring in further refinement within each group, a subdivision has been attempted by taking up a difference of 10% female literacy on either side. Using this diagram, states were identified in each of the categories. These have been treated in the same way as regression residuals and mapped appropriately (Fig 2a).

The emerging spatial pattern brings out that Kerala is at the highest level on the positive side and Rajasthan is at the highest level on the negative side. This matches with our perception about these two states. Kerala's position is primarily because of social factors that have enabled it to achieve the highest level of female literacy. Similarly, Rajasthan's position on the lowest
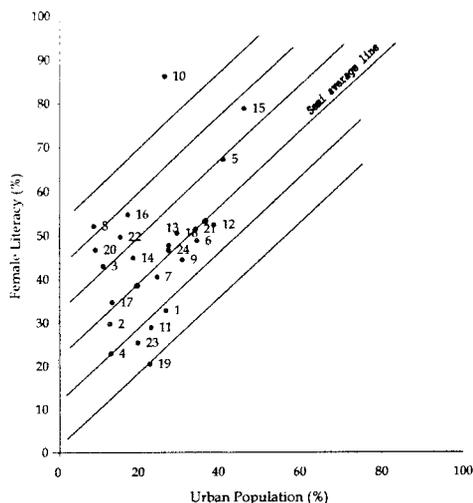
Fig. 1

**Urban Population and Female Literary in the
States of India 1991**

rung is because of poor attention provided towards females in general and female literacy in particular. Other residuals could be interpreted in a similar manner. In general, hill areas show positive residuals because of low level of their urbanisation and/or missionary influence. Economically developed and highly urbanised states stand out by their negative residuals. The advantage of the technique is identification of departing units without going into the rigour of regression analysis.

2.Concurrent Deviation Method: The calculation of co-efficient of concurrent deviation is a routine exercise carried out by all in order to get a quick idea about the extent of relationship. In order to identify departing units, following considerations become essential. There are two types of deviations, viz. concurrent and non-concurrent. Concurrent deviations do not

require any explanation as they are as per our expectations. Therefore, our interest lies in nonconcurrent deviations. Seven states have been identified to have such non-concurrent deviations. They are classified into two groups, based on whether non-concurrence is the result of positive or negative deviation in female literacy (Table I); it would indicate whether female literacy is higher or lower than expected for a given level of urbanisation. Further subdivision is not possible because of the limitation of the method, viz. it takes into consideration only nature of deviation in relation to the previous value and not the extent. These results have been mapped to understand the spatial pattern, if any (Fig. 2b).

The map reveals that Kerala, Himachal Pradesh, Sikkim and Assam fall in one group, showing positive departure of female literacy in relation to level of urbanisation. This appears to bring together widely differing states on the female literacy front. While Kerala is expected to be on the positive side because of the factors indicated above, others have been roped in not because of their commendable performance on the female literacy front but because of their low level of urbanisation. Uttar Pradesh, Bihar and Karnataka show negative departure where social backwardness is the major factor for the first two; Karnataka gets included simply because of its high level of urbanisation. As is known, this method is crude and its discussion here serves only academic interests.

3. Spearman's Rank Correlation Method: Calculation of the co-efficient of rank correlation does not require any discussion. The basic steps are indicated in Table 1. Both

the variables have been ranked in descending order of their magnitude. In order to identify departing units, following considerations become essential. If the correlation were perfect, ranks of both the variables would be identical in the case of direct relationship. The departure in the magnitude of the co-efficient from unity suggests the existence of abnormally behaving units. The difference in corresponding ranks provides the extent of departure shown by each unit. Since, here, researchers are interested in knowing how far the female literacy departs from expectation for a given level of urbanisation, the ranks of the latter should have been subtracted from the corresponding ranks of the former. However, since the ranks are assigned from the highest to the lowest values, ranks of female literacy have been substracted from the corresponding ranks of level of urbanisation to give a correct pattern. Differences in ranks have been classified and mapped using choropleth techniques; two different sets of shades have been used for positive and negative departures (Fig 2c).

The map brings out the fact that the states of Himachal Pradesh, Sikkim, Nagaland, Kerala, Assam, Tirupura, Meghalaya, Arunachal Pradesh and Orissa show positive departure suggesting high female literacy vis-a-vis the expectation. Surprisingly, Himachal Pradesh stands at the highest level while Kerala slides to the third highest level. This is contrary to our general perception, but it is mainly due to limitation of the method that focuses on position rather than magnitude. It should also be kept in mind that one is not looking into female literacy as such but female literacy in relation to level of urbanisation. For low level of

urbanisation that Himachal Pradesh has, low level of female literacy also stands high. Similar explanation is valid for higher positive departure with Sikkim and Nagaland compared to Kerala, and, same level in the case of Assam and Tripura. Arunachal Pradesh, Orissa and Meghalaya show expected pattern of departure. Negative departures are to be explained in the same manner.

4. Karl Pearson's Product Moment Method: This is a more sophisticated method compared to rank correlation as it takes into consideration each and every value. On calculation of 'r' value, one comes to know the extent of its departure from the ideal magnitude of unity, but it is necessary to know the spatial units which are responsible for such a departure. This is done by adopting a method similar to the one used in rank correlation. The values of both the variables are standadised and then a difference in these values is worked out. In the present case, differences between Z-score values of female literacy and those of urban population have been derived and classified into nine classes at an interval of 0.5 Z-score difference and mapped using choropleth technique (Fig. 2d).

The map provides a satisfactory pattern in that Kerala emerges as the highest positive residual while Rajasthan is at the bottom with respect to performance of female literacy for given levels of their urbanisation. The reason for a spatial view conforming to our expectations lies in the sophisticated nature of the method that takes into consideration each and every value.
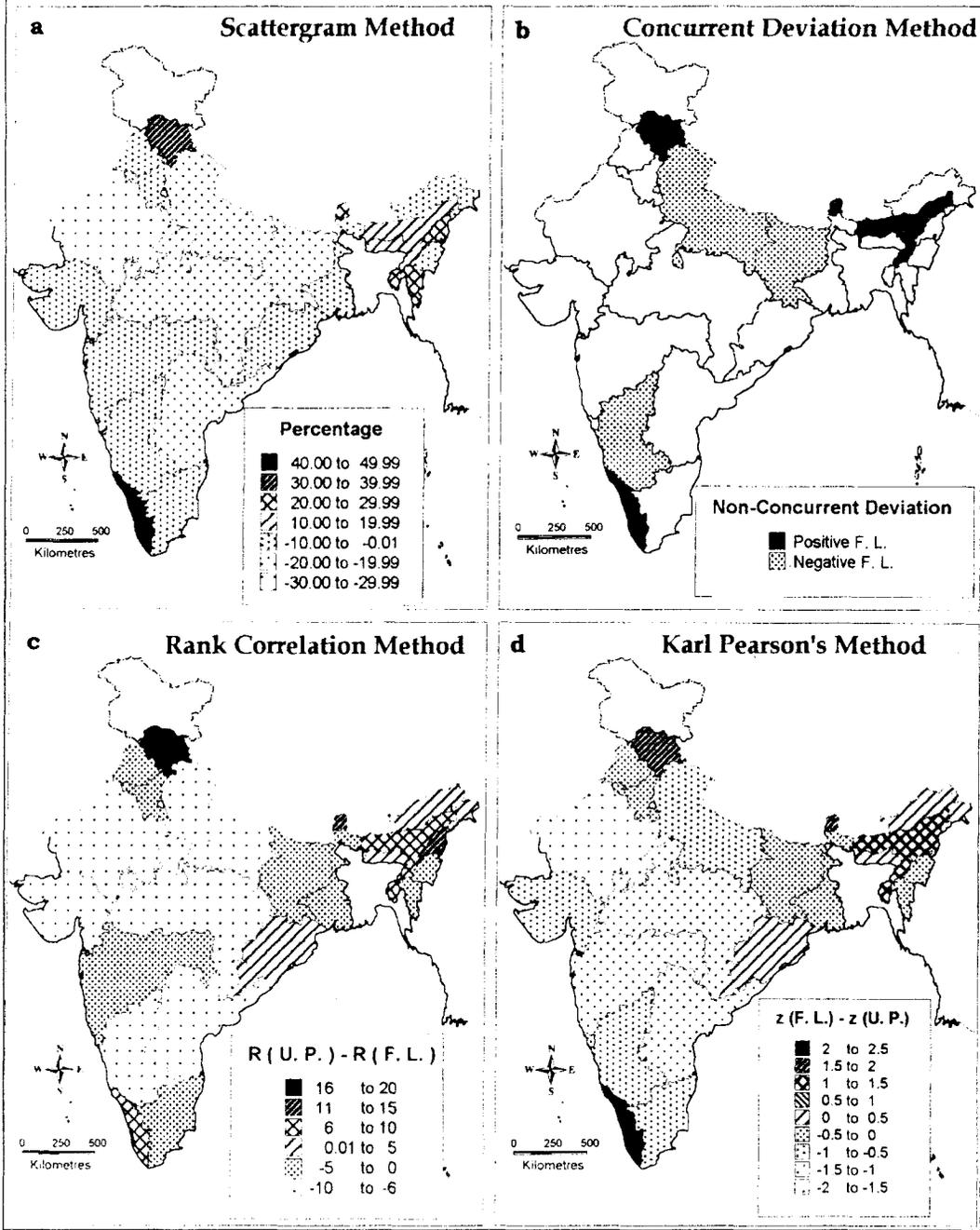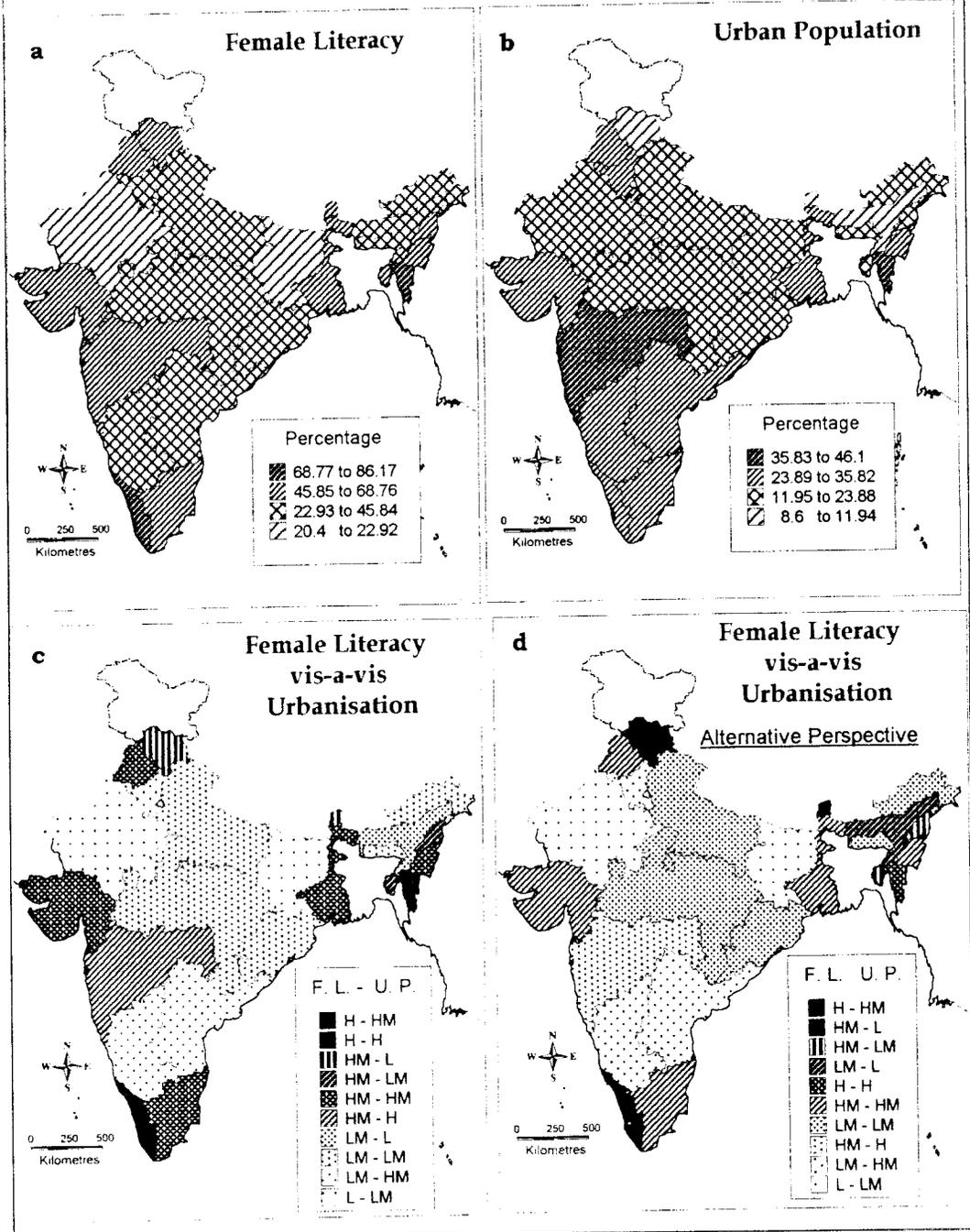
INDIA
Female Literacy Residuals
1991

a   Scattergram Method

Percentage

- 40.00 to 49.99
- 30.00 to 39.99
- 20.00 to 29.99
- 10.00 to 19.99
- -10.00 to -0.01
- -20.00 to -19.99
- -30.00 to -29.99

0   250   500
Kilometres

b   Concurrent Deviation Method

Non-Concurrent Deviation

- Positive F. L.
- Negative F. L.

0   250   500
Kilometres

c   Rank Correlation Method

R (U. P.) - R (F. L.)

- 16   to 20
- 11   to 15
- 6   to 10
- 0.01 to 5
- -5   to 0
- -10   to -6

0   250   500
Kilometres

d   Karl Pearson's Method

z (F. L.) - z (U. P.)

- 2   to 2.5
- 1.5 to 2
- 1   to 1.5
- 0.5 to 1
- 0   to 0.5
- -0.5 to 0
- -1   to -0.5
- -1.5 to -1
- -2   to -1.5

0   250   500
Kilometres

Fig. 2

INDIA
1991

a Female Literacy

b Urban Population

Percentage
- 68.77 to 86.17
- 45.85 to 68.76
- 22.93 to 45.84
- 20.4 to 22.92

0 250 500
Kilometres

Percentage
- 35.83 to 46.1
- 23.89 to 35.82
- 11.95 to 23.88
- 8.6 to 11.94

0 250 500
Kilometres

c Female Literacy
vis-a-vis
Urbanisation

F. L. - U. P.
- H - HM
- H - H
- HM - L
- HM - LM
- HM - HM
- HM - H
- LM - L
- LM - LM
- LM - HM
- L - LM

0 250 500
Kilometres

d Female Literacy
vis-a-vis
Urbanisation

Alternative Perspective

F. L. U. P.
- H - HM
- HM - L
- HM - LM
- LM - L
- H - H
- HM - HM
- LM - LM
- HM - H
- LM - HM
- L - LM

0 250 500
Kilometres

Fig. 3

Alternative way of viewing the problem

Since the aim of the present exercise was oriented to find out the departures, rather than deriving the value of the degree of association between the variables concerned, the problem was also looked upon in a slighly different way. Since female literacy, here, is being considered in relation to urban population, it was thought that the performance of the states on this front could also be analysed using cartographic wisdom. For this purpose, classification of female literacy was attempted taking mean as the base. States were identified having values below 0.5 mean, between 0.5 mean and mean, mean and 1.5 mean, and, above 1.5 mean. Similar procedure was followed for level of urbanisation. Two separate maps were prepared using the above information (Fig. 3 a & b). Each level of female literacy could get combined with four different levels of urbanisation. Therefore, four such levels would would bring 16 possible combinations. They have been listed with female literacy ranging from high to low while corresponding level of urbanisation from low to high; this has been done because a state with high female literacy in relation to low urbanisation level is a better performer than the one with high female literacy and high urbanisation. In our analysis, only 10 out of 16 combinations actually existed, they have been mapped (Fig. 3c). This is also attempted in a slighly different manner taking into consideration the difference in the levels of female literacy and level of urbanisation (Fig. 3d). They are arranged in a sequence where combinations with higher female literacy come first, then follow those with equal level in both and lastly are the ones where female literacy level is lower than that of urbanisation.

The last map provides a pattern which is nearly a replica of the one provided by the residuals in Karl Pearson's method. It is the simplest way to understand the performance of states on female literacy front in relation to level of urbanisation.

Conclusion

The aim of the present paper was to suggest a methodoloy to represent the results of correlation analysis for spatial explanation. This has been achieved. However, it should be noted that all the methods for doing this are not equally appealing and this is because of the limitations associated with individual method used to derive correlation. Co-efficient of concurrent deviation is the crudest of all the methods and gives only nominal classes of positive and negative departure without any possility of finer classification. Rank correlation does not give perceived representation because of its use of positions rather than actual magnitudes. Scattergram and Karl Pearson's methods come close; however, among the two, the latter is more accurate mathe-matically. In the former, there is some arbitrariness in drawing lines on either side of the trend line. Moreover, the trend line itself is poorly defined. Despite these limitations, it needs to be admitted that these two methods provide a quick way of identifying departing units which are not normally worked out in the correlation exercise. An alternative route based on cartographic analysis is also suggested which gives an idea of abnormally behaving units. The results of this method tally very closely with that of Karl Pearson's method and hence it can be tried without hesitation for quick identification of depar-ting units as an aid to spatial explanation.

The exercise illustrates the case of two positively correlated variables. When the relationship is negative, the technique would need a suitable amendment. There will not be any change while working with scattergram method. In other methods, signs of one of the variables need to be changed before carrying out the suggested operations. This should invariably be done for independent variable, although it is true that in correlation execise one does not make such a consideration, viz. - independent/ dependent variables.

Acknowledgement

References

Berry, B.J. L. (1966) : Essays on Commodities and the Spatial Structure of the India Economy, Research Paper 111, Department of Geography, University of Chicago.

Murdie, R. A. (1969) : Factorial Ecology of Metropolitan Toronto- 1951-61, Research Paper 116, Department of Geography, University of Chicago.

Robinson, A.H. and Bryson, R.A. (1957) : 'A Method for Describing Quantitatively the Correspondence of Geographical Distribution', Annals AAG, 47, 379-391.

Sviatlovsky, E.E. and Eells, W. C. (1937) : 'The Centrographical Method and Regional Analysis', Geographical Review, 27, 240-254.

Srivant, S. and Phadke, V. S. (2001) : 'Cartographic Representation of the Results of Chisquare Analysis for Spatial Explanation', Transactions, IIG, 23 (1 & 2), 47- 52.

Taaffe, E.J. (1970) : Geography, Prentice Hall Inc. Englewood Cliffs, N.J., 44-49.

Taaffe, E.J., Morrill, R. L. and Gould, P. R. (1963) : 'Transport Expansion in Underdeveloped Countries: A Comprehensive Analysis', Geographical Review, 53, 503-529.

Taylor, P.J.(1977) : Quantitative Methods in Geography-An Introduction to Spatial Analysis, Houghton Mifflin Com., Boston.

Yeates, M. (1974) : Introduction to Quantitative Techniques in Human Geography, McGraw Hill, New York.

Ms. Chhaya Daudia
Teacher, Taraben Master English Primary School, Juhu Tara Road, Santactuz (W), Mumbai 400 056.

Dr. V.S. Phadke
Professor and Head
Department of Geography
University of Mumbai
Vidyanagari, Mumbai 400 098.