

ANÁLISE DE VIZINHANÇA: UMA MEDIDA DE PADRÃO DE DISTRIBUIÇÃO ESPACIAL

LÚCIA HELENA DE OLIVEIRA GERARDI (*) e
MIGUEL CEZAR SANCHEZ (*)

O Homem, desde seu aparecimento sobre a superfície da Terra, tem constatado que as coisas se acham distribuídas de forma desigual sobre esta superfície e tem percebido que há lugares onde certos eventos ocorrem de forma concentrada ou agrupada e outros onde há rarefação ou dispersão de itens.¹ A partir desta constatação, o Homem sentiu necessidade de expressar de forma mais objetiva os diferentes graus de concentração ou dispersão procurando, para isso, técnicas de medida que resultassem numa quantificação da avaliação visual feita sobre estas distribuições.

Um fato perceptível à primeira vista é que a falta de uma medida restringe as referências ao grau de concentração ou dispersão de uma distribuição apenas a termos relativos. Assim, uma distribuição só poderá ser chamada de concentrada quando a ela se compara uma outra menos concentrada. Vem de longo tempo a discussão de o que é concentrado e o que é disperso. Páginas e páginas já foram gastas com esse assunto, não chegando, os vários autores, a nenhuma conclusão de aceitação geral, o que demonstra o alto grau de subjetividade implícito nesses termos. Os termos concentração e dispersão, embora sejam utilizados muitas vezes como opostos, fazem parte do mesmo *continuum* onde a posição demarcada como mais concentrada significa também menos dispersa e aquela mais dispersa significa menos concentrada. A figura 1 demonstra essa relação. Considere o ponto A, por exemplo:

* Departamento de Geografia e Planejamento Regional — FFCL de Rio Claro, UNESP.

1. Neste ponto convém justificar o emprego indiscriminado dos termos: coisas, eventos e itens que, para os propósitos deste trabalho são sinônimos, significando todo fenômeno que possa ser espacialmente representado.

em relação ao ponto B ele é mais concentrado ou menos disperso. O ponto B, por sua vez é mais concentrado que C e mais disperso que A. Já o ponto C é menos concentrado que A e B e mais disperso que ambos. Sendo assim, mantido o mesmo referencial, é indiferente considerar uma distribuição mais ou menos dispersa ou mais ou menos concentrada.

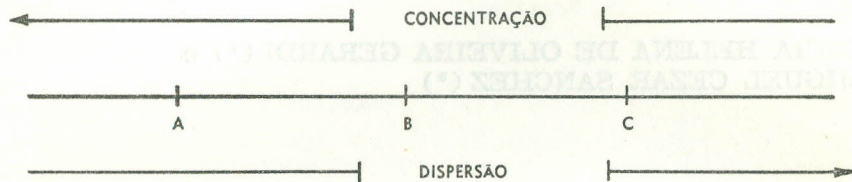


Fig. 1. Esquema do *continuum* concentração — dispersão.

Para superar a subjetividade e a ambigüidade dos conceitos de concentração e dispersão, nem só geógrafos têm se preocupado com a busca de medidas. Várias tentativas foram realizadas, principalmente por naturalistas, que se preocuparam com distribuições de plantas e animais, cuja descrição é extremamente importante para o desenvolvimento de sistemas de classificação mais específicos. Goodall (1952) realizou uma importante revisão da bibliografia a esse respeito, com relação a populações vegetais.

Dentre as várias tentativas, a que parece ter dado resultados mais satisfatórios foi aquela orientada no sentido de estabelecer a relação entre um indivíduo e seus vizinhos mais próximos, uma vez que o espaçamento entre os elementos de uma distribuição é visual e quantitativamente importante no estabelecimento do grau de concentração ou dispersão da distribuição.

Essa linha de análise aparece e é desenvolvida entre botânicos, sendo Dice (1952) o primeiro a usar a distância entre os vizinhos mais próximos como base para quantificar o padrão de distribuição de uma população vegetal, idéia cuja paternidade é também advogada por Skellam (1952). O desenvolvimento dessa linha de preocupação nos leva a P. J. Clark e F. C. Evans (1954), que propõem um índice de vizinhança baseado na distância de um indivíduo ao seu vizinho mais próximo como uma medida das relações espaciais em populações. Entre geógrafos, Dacey (1960 e 1962), Curry (1964), Getis (1964) e King (1968), são os que procuraram aplicar a técnica, relacionada a padrões de distribuição de povoamento e uso da terra.

Esse artigo tem por objetivo básico tornar acessível a técnica proposta por Clark e Evans, aplicando-a a três distribuições e procurar explorar seu potencial relacionando os índices obtidos ao grau de ajustamento de superfícies de tendência.

O ÍNDICE DE VIZINHANÇA

Supondo que uma distribuição espacial aleatória de pontos tem, probabilisticamente, chances extremamente reduzidas de ser totalmente concentrada ou totalmente dispersa, o conceito de aleatoriedade é o ponto de referência para medida de espaçamento entre pontos ou índice de Vizinhança (IV), o qual representa a maneira e a magnitude pelas quais distribuições dadas se afastam daquela considerada aleatória.

Segundo o proposto por Clark e Evans, a aleatoriedade é um conceito eminentemente espacial, ao considerar que "qualquer ponto tenha a mesma chance que qualquer outro ponto de ocorrer em qualquer subárea, assim como que qualquer subárea tenha a mesma chance que qualquer outra subárea de receber um ponto e que a localização de qualquer ponto não seja influenciada por aquela de qualquer outro ponto" (Clark e Evans, 1954, p. 446).

A aleatoriedade, entendida como um conceito espacial, é influenciada pelos limites do espaço objeto de investigação. Assim, na figura 2, considerando a distribuição limitada no espaço A, pode-se falar em aleatoriedade. Porém, se os limites do espaço forem transferidos para B, percebe-se claramente uma distribuição concentrada de pontos. O bom senso do pesquisador deve prevalecer na escolha destes limites do espaço estudado. Seria ideal que estes limites fossem naturais, para distribuição de eventos que independem da ação do homem. Porém, como em Geografia grande parte dos eventos resulta da ação intencional do homem organizando o espaço, e raramente essas ações se circunscrevem a limites naturais, grande cuidado é necessário no estabelecimento das fronteiras do espaço a ser estudado.

O cálculo do índice de vizinhança tem como base medidas lineares não orientadas, tomadas entre cada indivíduo da população e seu vizinho mais próximo.² O conjunto dessas medi-

2. Quando um ponto tem dois outros como seus vizinhos mais próximos, com igual distância, apenas um deles deve ser escolhido para tomada de medida. Por outro lado, quando um ponto é o vizinho mais próximo de outros dois, duas medidas devem ser tomadas a partir desse ponto. A unidade de

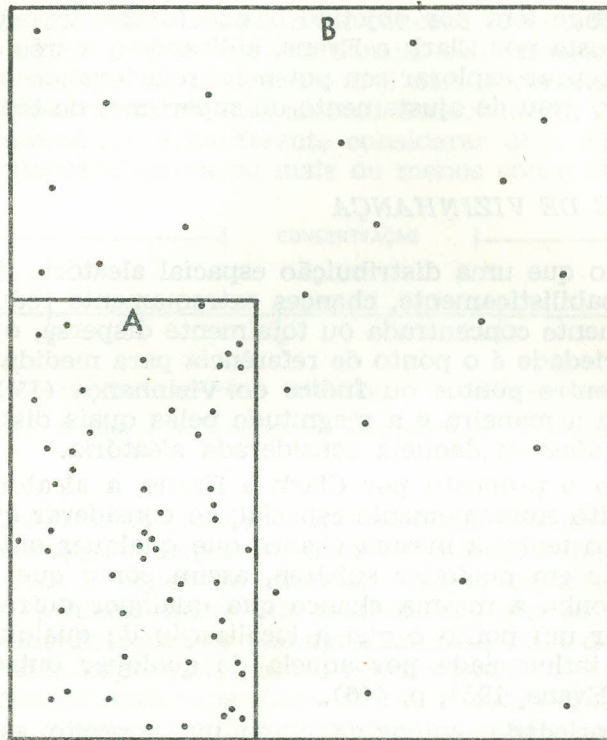


Figura 2. Influência dos limites do espaço na consideração do padrão de distribuição de pontos.

das tomadas para uma população dada (abrangendo todo o universo ou uma amostra selecionada aleatoriamente) pode ser transformada em uma medida de distância média chamada

Distância Média Observada (\bar{D}_{obs}) ou $\bar{D}_{obs} = \frac{\sum d}{N}$ onde:

d = distância entre cada indivíduo e seu vizinho mais próximo e N = número total de pontos na área considerada.

Uma vez que a aleatoriedade é o ponto de referência, calcula-se a distância média entre os vizinhos mais próximos supondo que os pontos estivessem distribuídos de forma aleatória. Para o cálculo da Distância Média Aleatória (\bar{D}_{ale}), usa-se:

medida pode ser estabelecida a critério do pesquisador, porém deve ser a mesma tanto na tomada das distâncias como no cálculo da área (km de distância/km² de área ou cm de distância/cm² de área).

$\bar{D}_{ale} = \frac{1}{2\sqrt{(N/A)}}$ onde: 1 e 2 são constantes, N é o número total de pontos e A é a área considerada. Assim N/A representa a densidade de pontos na área específica.

A relação entre as distâncias médias observadas e aquelas calculadas considerando uma distribuição aleatória de pontos, resulta num índice de Vizinhança (IV), que representa a medida do grau de afastamento da distribuição estudada em relação à distribuição aleatória com respeito ao vizinho mais próximo.

Assim, $IV = \frac{\bar{D}_{obs}}{\bar{D}_{ale}}$.

Se a distribuição observada for perfeitamente aleatória, o valor \bar{D}_{obs} será exatamente igual ao valor \bar{D}_{ale} e IV será igual a um. Se todos os pontos da distribuição observada pudessem, teoricamente, estar localizados exatamente no mesmo local, a distância média entre eles seria zero e qualquer que fosse o valor de \bar{D}_{ale} , resultaria em IV igual a zero, significando a máxima concentração de pontos; por outro lado, sob condições de espaçamento máximo, numa distribuição hexagonal regular, onde cada indivíduo estaria equidistante de seis outros, o valor de IV seria 2,1491, conforme demonstram Clark e Evans (1954, p. 492). Assim, o índice IV tem uma amplitude limitada, com medidas exatas de um padrão de distribuição de pontos completamente concentrados (zero), aleatório (um), ou perfeitamente uniforme ou disperso (2,15). O índice de Vizinhança indica que, numa dada distribuição, a distância média observada é IV vezes maior que a distância média esperada numa distribuição aleatória dentro da mesma área e com o mesmo número de pontos. Assim, um IV igual a 0,60 significa que os vizinhos mais próximos estão em média 0,60 unidades (km, cm, milhas) mais distantes uns dos outros do que o esperado em condições de aleatoriedade.

Embora o índice IV possa ser utilizado de forma descritiva, ou seja, como medida do grau de concentração ou dispersão de uma distribuição de pontos, graças ao seu referencial aleatório ele pode ser usado de uma forma inferencial no sentido de indicar o afastamento de um padrão de distribuição de pontos daquele esperado em condições aleatórias. Como qualquer teste inferencial, IV utilizado desta forma deve ter sua significância calculada através da probabilidade de que a distribuição por ele representada tenha ocorrido por pura chance.

A transformação do índice IV na variável z estandarizada da curva normal é a base deste teste de significância (Hammond e McCullagh, 1974, p. 241-242): $z = \frac{\bar{D}_{obs} - \bar{D}_{ale}}{\sigma \bar{D}_{ale}}$ onde: \bar{D}_{ale} é

obtido pelo cálculo da fórmula: $\sigma \bar{D}_{ale} = \frac{0,26136}{\sqrt{\{N(N/A)\}}}$, sendo o numerador da fração uma constante, N o número total de pontos e A a área onde eles estão localizados.

Obtido o valor z , é suficiente compará-lo com uma tabela de probabilidades associadas a valores de z na distribuição normal para se obter a probabilidade de que a distribuição observada tenha ocorrido por pura chance.

APLICAÇÃO DE IV A DISTRIBUIÇÕES REAIS

Para testar os resultados de IV quando aplicado a distribuições reais de pontos tomou-se como base três distribuições estabelecidas de acordo com os seguintes critérios: a) *localização dos pontos*: o interesse subjacente no trabalho é estudar a distribuição do cultivo do café no Estado de São Paulo. Assim, este fator econômico se constitui na primeira restrição à localização dos pontos, isto é, foram selecionados os pontos em que se constatou a presença deste cultivo. Os dados disponíveis sobre área cultivada com café são publicados a nível de município e portanto a segunda restrição para a localização dos pontos foi a própria malha municipal do Estado. O posicionamento dos pontos dentro de cada município obedeceu à escolha do ponto central de cada unidade municipal, determinado pelo cruzamento de duas perpendiculares, que se constituiu na terceira restrição à localização dos pontos; b) *escolha de pontos*: como a quantidade de pontos com dados de produção de café era muito grande (252), optou-se pela escolha de um subconjunto de pontos adotando-se para isso vários critérios de escolha diferentes: 1. sorteio — com o objetivo de imprimir um caráter aleatório na obtenção dos pontos a serem considerados, procedeu-se, com o auxílio de uma tábua de números equiprováveis, a um sorteio de 100 pontos representando uma amostra de 39,68 por cento do total de pontos considerados, cuja distribuição espacial pode ser verificada na figura 3a; 2. limite arbitrário preestabelecido — representa uma escolha intencional de 73 pontos baseada no critério de pertinência a uma classe que representasse os municípios responsáveis por 0,5 por cento ou mais da área cultivada com

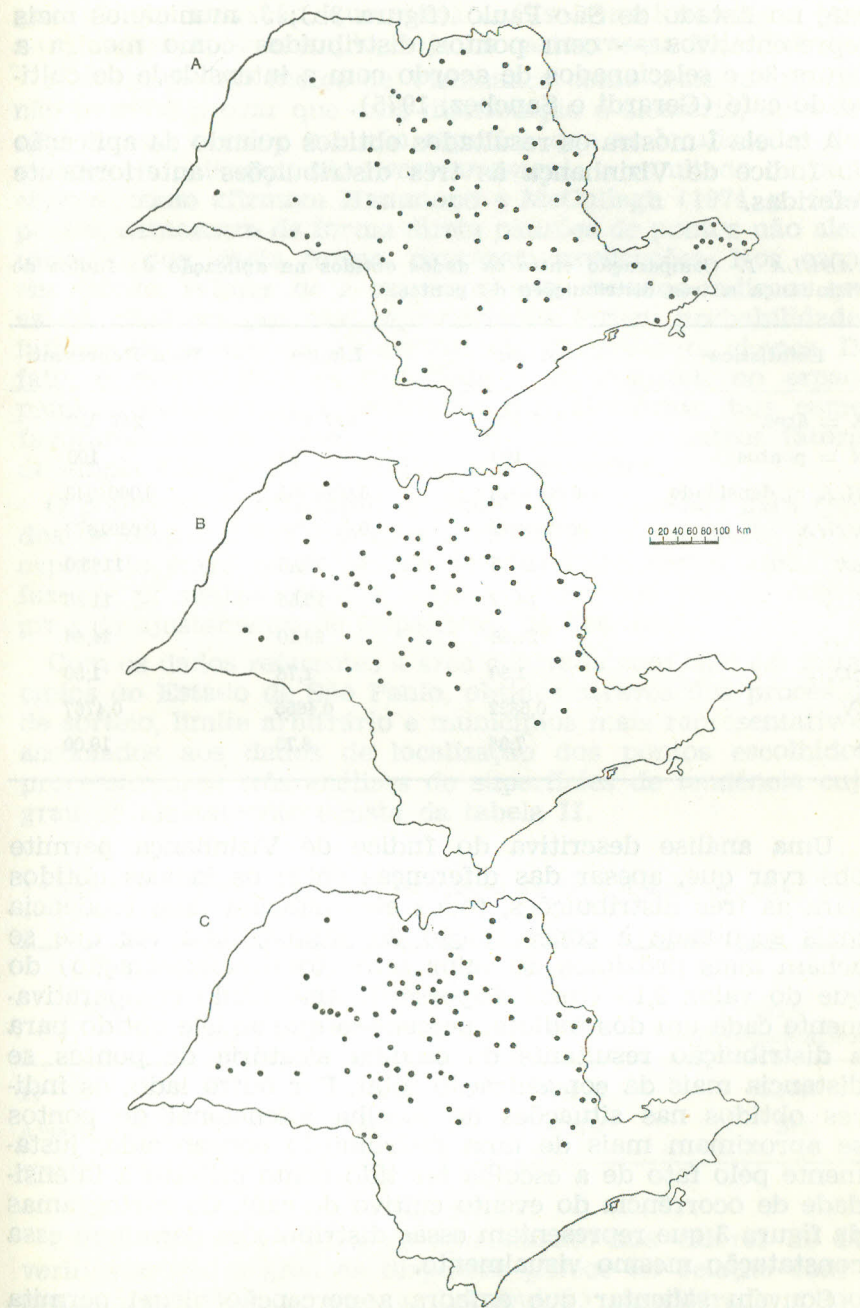


Figura 3. Distribuição de pontos no Estado de São Paulo, escolhidos segundo 3 critérios.

café no Estado de São Paulo (figura 3b); 3. municípios mais representativos — cem pontos distribuídos como mostra a figura 3c e selecionados de acordo com a intensidade de cultivo do café (Gerardi e Sanchez, 1975).

A tabela I mostra os resultados obtidos quando da aplicação do Índice de Vizinhança às três distribuições anteriormente referidas.

TABELA I. Comparação entre os dados obtidos na aplicação do Índice de Vizinhança a três distribuições de pontos.

Estatística	Sorteio	Limite	Mais Represent.
A = área	247 320	247 320	247 320
N = pontos	100	73	100
N/A = densidade	0,0004043	0,0002951	0,0004043
$\sqrt{N/A}$	0,0201072	0,0171784	0,0201072
Σd	1447,5	989,5	1185,0
\bar{D}_{obs}	14,47	13,55	11,85
\bar{D}_{ale}	24,86	29,10	24,86
$\sigma \bar{D}_{ale}$	1,30	1,78	1,30
IV	0,5822	0,4655	0,4767
z	7,98	8,73	10,00

Uma análise descritiva do Índice de Vizinhança permite observar que, apesar das diferenças entre os índices obtidos para as três distribuições, todos eles indicam uma tendência mais acentuada à concentração de pontos, uma vez que se acham mais próximos do valor zero (total concentração) do que do valor 2,15 (total dispersão). Analisando comparativamente cada um dos índices, percebe-se que aquele obtido para a distribuição resultante de escolha aleatória de pontos se distancia mais da concentração total. Por outro lado, os índices obtidos nas situações de escolha intencional de pontos se aproximam mais de uma distribuição concentrada, justamente pelo fato de a escolha ter tido como critério a intensidade de ocorrência do evento cultivo de café. Os cartogramas da figura 3 que representam essas distribuições permitem essa constatação mesmo visualmente.

Convém salientar que embora a percepção visual permita estabelecer gradações gerais de concentração, o Índice de Vizinhança tem a vantagem de apontar variações mínimas desse

grau de concentração, que numa análise visual não seria possível avaliar (por exemplo, veja os cartogramas 3b e 3c).

A utilização do Índice de Vizinhança como teste inferencial não permite provar que uma distribuição é aleatória, uma vez que mesmo o valor um não significa que as localizações dos eventos na paisagem são *necessariamente* o resultado da pura chance, como afirmam Hammond e McCullagh (1974, p. 241), porém, demonstra de forma direta padrões de pontos não aleatórios e que, desta forma, requerem explicações. Nos casos em estudo, valores de z extremamente elevados indicam que as distribuições por eles representadas teriam probabilidades ínfimas de ocorrerem aleatoriamente ou por pura chance. De fato, o evento café se distribuiu e se localizou no espaço paulista obedecendo a princípios não aleatórios, tais como, favorabilidade de solos, *inputs* de mercado e outros fatores de ampla divulgação na bibliografia específica.

O potencial de utilização do Índice de Vizinhança para estudos de distribuições espaciais é muito grande e muito pouco explorado entre nós. Uma de suas aplicabilidades, ainda em fase de primeiros estudos, seria verificar sua relação com o grau de ajustamento de Superfícies de Tendência.

Com os dados referentes à área cultivada com café em municípios do Estado de São Paulo, obtidos através dos processos de sorteio, limite arbitrário e municípios mais representativos, associados aos dados de localização dos pontos escolhidos, processaram-se três análises de superfícies de tendência cujo grau de ajustamento consta da tabela II.

TABELA II. Comparação entre os valores de IV e o grau de ajustamento de superfícies de tendência.

Índices	Critérios		
	Sorteio	Limite	Mais Representativos
IV	0,5822	0,4655	0,4767
Ajustamento	54%	27%	20%

Comparando-se o grau de ajustamento aos valores de IV, verifica-se que o grau de dispersão parece ter relação com o ajustamento. Veja-se que a distribuição considerada menos agrupada é aquela em que se verifica um maior ajustamento, ou que distribuições com o mesmo número de pontos e relati-

vas ao mesmo fenômeno, porém, com graus de dispersão diferentes, resultaram em percentuais de ajustamento bem discrepantes (verifique na tabela II os critérios de sorteio e de municípios mais representativos).

Considerando o exemplo apresentado, pode-se deduzir que aplicações mais numerosas e estudos mais detalhados do Índice de Vizinhança devem ser realizados no âmbito da Geografia, uma vez que os padrões de distribuição espacial de fenômenos são de interesse fundamental na pesquisa geográfica.

BIBLIOGRAFIA

- Clark, J. P. e Evans, F. C. (1954), Distance to nearest neighbor as a measure of spatial relationships in populations, *Ecology*, vol. 35, nº 4, p. 445-453.
- Curry, L. (1964), The random spatial economy: an exploration in settlement theory, *Annals of the Association of American Geographers*, vol. 54, p. 138-146.
- Dacey, M. F. (1960), A note on the derivation of nearest neighbor distances, *Journal of Regional Science*, nº 2, p. 81-87 (citado por Harvey, D. In: *Explanation in Geography*, Edward Arnold, Londres, 1969).
- Dacey, M. F. (1962), Analisis of central place and point patterns by a nearest neighbor method. In: Norborg, K. (Ed.), *IGU Symposium in urban Geography* (citado por Harvey, D. In: *Explanation in Geography*, Edward Arnold, Londres, 1969).
- Dice, L. R. (1952), Measure of the spacing between individuals within a population, *Contrib. Lab. Vert. Biol.*, Univ. de Michigan, vol. 55, p. 1-23 (citado por Clark e Evans, 1954).
- Gerardi, L. H. O. e Sanchez, M. C. (1975), Considerações sobre a aplicação da análise das superfícies de tendência em Geografia, *Boletim de Geografia Teorética*, vol. 5, nº 9/10, p. 63-82.
- Getis, A. (1964), Temporal analysis of land use patterns with the use of nearest neighbor and quadrat methods, *Annals of the Association of American Geographers*, vol. 54, p. 391-399.
- Hammond, R. e McCullagh, P. (1974), Tests for distributions in space. In: Hammond e McCullagh, *Quantitative Techniques in Geography*, Clarendon Press — Oxford, cap. 9, p. 237-249.
- Harvey, D. (1969), *Explanation in Geography*, Edward Arnold, Londres.
- King, L. J. (1968), A quantitative expression of the patterns of urban settlements in selected areas of the United States. In: Berry, B. J. L. e Marble, D. F., *Spatial Analysis*, Prentice-Hall Inc., Englewood Cliffs, Nova Jersey, p. 159-167.
- Skellam, J. G. (1952), Studies in statistical ecology — I. Spatial pattern, *Biometricka*, vol. 39, p. 346-362 (citado por Clark e Evans, 1954).

RESUMO

A partir da constatação da distribuição desigual dos eventos sobre a superfície da Terra e da necessidade e importância de expressar de forma objetiva diferentes graus de concentração ou dispersão, os autores retomam a técnica de análise de vizinhança proposta por Clark e Evans (1954). Baseada na medida da distância de um indivíduo a seu vizinho mais próximo, esta

técnica é aplicada a três distribuições de pontos no Estado de São Paulo, obtidos através de três critérios diferentes. Os autores discutem os resultados obtidos e verificam a relação dos índices calculados com o grau de ajustamento de Superfícies de Tendência.

ABSTRACT

Nearest neighbor analysis: a measure of spatial point patterns. On the basis of the necessity of a measure of different spatial point patterns, the authors apply the Clark & Evans (1954) technique of nearest neighbor analysis to three distributions in São Paulo State. The results are discussed and the authors try to analyse the relations of the nearest neighbor indices to the degree of fitness of trend surface.

BOLETIM DE GEOGRAFIA TEORÉTICA

Encontram-se em disponibilidade os seguintes exemplares do Boletim de Geografia Teorética:

vol. 3, n.º 5	, 1973	Cr\$ 10,00
3. n.º 6	, 1973	Cr\$ 10,00
vol. 4, n.º 7-8	, 1974	Cr\$ 20,00
vol. 5, n.º 9-10	, 1975	Cr\$ 20,00

Os pedidos, acompanhados de cheques nominiais, devem ser endereçados para:

Associação de Geografia Teorética
Caixa Postal, 178
13500 — Rio Claro (SP) — Brasil