ANÁLISE DE DISTRIBUIÇÕES ESPACIAIS EM ARQUEOLOGIA – UMA INTRODUÇÃO

JOSÉ ALEXANDRE FELIZOLA DINIZ[1]

ABSTRACT

The archaeological research has presented meaningful results with it's chronological approach to data, but has also almost forgotten the spatial structure of information. This paper tries to arise some elemental statistical possibilities to handle with spatial distributions and patterns. Median, average and standard deviation are set in a spatial context and transformed into spatial median, gravity center and standard distance to describe a hypothetic point distribution. The nearest neighbor and "**R**" statistics are used to define the point distribution pattern.

Palavras chave: análise espacial, centro de gravidade, estatística dos vizinhos mais próximos.

1 Professor do Núcleo de Pós-Graduaçãoem Geografia da UFS
Pesquisador 1-A do CNPq, 1991-1999
Diretor do Museu de Arqueologia de Xingó.

INTRODUÇÃO

Todas as ciências que lidam com distribuições espaciais enfrentam dois problemas: primeiro, o da descrição mais objetiva dessas distribuições e, segundo, o da identificação de padrões espaciais que permitam associação com hipóteses e teorias sobre o tema. A Arqueologia, ao tratar da posição espacial de sítios arqueológicos ou da distribuição de vestígios intra-sítios, enfrenta essas questões, embora, como coloca Hodder e Orton (op. cit., p. 11), os pré-historiadores tenham sempre se preocupado mais com a seqüência cronológica do que com a dimensão geográfica das culturas. Segundo esses autores, um tratamento mais acurado das distribuições espaciais em Arqueologia é importante por três razões (op.cit., p. 12):

> "Primeira, porque a investigação precedente nesse campo foi limitada em seus objetivos e métodos, com freqüência acríticos e de pouca utilidade para uma interpretação detalhada. Segunda, porque as valorações subjetivas podem ser perigosas; e, terceira, porque são necessários certos métodos para manejar a enorme quantidade de informações sobre distribuições que já começa a ser importante".

Para se iniciar um estudo mais objetivo de distribuições, pode-se começar por uma análise de pontos, que podem indicar sítios arqueológicos ou mesmo vestígios líticos, cerâmicos, restos humanos ou de animais no interior de determinado sítio. E, nesse primeiro nível de estudo, há três medidas espaciais que podem ser obtidas, exatamente no sentido da busca de maior objetividade recomendada por Hodder e Orton . Primeira, o centro da distribuição, quer o centro mediano, quer o centro de gravidade, este correspondendo à média; segunda, a distância padrão, correspondendo ao desvio padrão; terceira, o valor de r, que indica o padrão da distribuição, situado entre o maior agrupamento e a maior regularidade.

Nesses estudos, a Arqueologia deve se valer, não só dos seus trabalhos como das numerosas contribuições de outras ciências tradicionalmente mais voltadas a análises espaciais, como a Geologia, a Ecologia e a Geografia.

MEDIDAS DE TENDÊNCIA CENTRAL, DE VARIAÇÃO E SEUS EQUIVALENTES ESPACIAIS

Mediana, Média e Desvio Padrão

Em qualquer distribuição, a mediana é o ponto que a equilibra em termos do número de observações. Tendo-se, por exemplo, uma distribuição formada pelos números 3, 3, 5, 6, 8, 9, 15, 28 e 50, cujo número de observações, **n,** é igual a 9, a mediana corresponderá à quinta observação, ou seja, 8, que se coloca no centro, repartindo quatro observações para a esquerda e quatro para a direita. No caso de o **n** ser par, a mediana será a média aritmética das duas posições centrais.

A média aritmética já significa algo bem diferente. Seu valor é decorrente, não propriamente do número de observações, mas do valor delas, sendo intensamente atraída pelas observações de valores mais extremos. A média aritmética da distribuição **x** é calculada pela soma dos valores $\mathbf{x_i}$ observados, dividida por **n**, de forma:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

No caso do exemplo anterior, a média aritmética será:

$$\bar{x} = \frac{127}{9} = 14,1$$

Note-se que o valor da média, bem mais elevado do que o da mediana, foi atraído pelas observações mais à direita. Assim, a depender da distribuição e dos objetivos da descrição, a posição central pode ser melhor descrita pela média ou pela mediana.

Nem sempre se tem todos os pontos observados numa distribuição. Às vezes, os dados estão agrupados e tem-se apenas as classes e a

Tabela 1

Classes de notas	**(n)** Nº de alunos por classe	**(pm)** Ponto médio da classe	**n.p**
0 a 2,0	2	1	2
de 2,1 a 4,0	6	3	18
de 4,1 a 6,0	9	5	45
de 6,1 a 8,0	8	7	56
de 8,1 a 10,0	5	9	45
	30	-	$\sum$ 166

freqüência em cada uma delas, como no exemplo seguinte, onde são observadas as notas dos 30 alunos de uma turma, já agrupadas:

Pode ser assumido que a distribuição no interior de cada classe é regular, sendo bem apresentada pelo ponto médio. Assim, obtém-se o somatório das notas pelo produto do **n** de cada classe pelo ponto médio, como se observa na tabela 1. Dividindo-se 166 por 30 obtém-se 5,5, que seria a média da turma. É esse procedimento de cálculo da média para dados agrupados que será utilizado para obtenção do centro de gravidade.

Duas distribuições podem ter médias iguais e serem completamente diferentes. Uma outra turma, em que cada um dos trinta alunos tivesse obtido a nota 5,5, teria também esse valor como média, mas o significado seria bem diferente, pois não teria alunos péssimos ou excelentes, como a anteriormente descrita, mas apenas estudantes medíocres. É preciso, então, que a descrição seja completada por alguma medida de distribuição em torno da média, ou seja, de sua variabilidade, destacando-se dentre elas o desvio padrão. Recorrendo-se ao primeiro exem-

Tabela 2

x_i	(xi - $\bar{x}$)	(xi - $\bar{x}$)^2
3	11,1	123,2
3	11,1	123,2
5	9,1	82,8
6	8,1	65,6
8	6,1	37,2
9	5,1	26,0
15	0,9	0,81
28	13,9	193,2
50	35,9	1.288,8
n = 9	9	9
$\sum$ = 127	-	1.940,8

plo dado de distribuição, cuja média foi 14,1, chega-se ao exposto na primeira coluna da tabela 2.

Como o desvio padrão é calculado por

$$s = \sqrt{\frac{\sum\left(x_i - \bar{x}\right)^2}{n-1}}$$

de forma que é necessário o cálculo da diferença entre cada observação e a média, exatamente para se verificar o nível geral de afastamento ao ponto central (coluna 2 da Tabela 2), diferença sem sinal, já o resultado será elevado ao quadrado na coluna 3. Substituindo-se a fórmula pelos valores encontrados, tem-se:

$$s = \sqrt{\frac{1.940,8}{8}} = \sqrt{242,6} = 15,6$$

que corresponde a um padrão de distribuição das observações em torno da média, aumentando na medida do afastamento das observações. Nas distribuições normais, o intervalo entre a média e um desvio padrão positivo e negativo tende a abranger 66,6% das observações.[2]

MEDIANA ESPACIAL E CENTRO DE GRAVIDADE

Nessa introdução à análise espacial em Arqueologia, trabalha-se com o exemplo hipotético da área x, que conta com 18 observações, al-

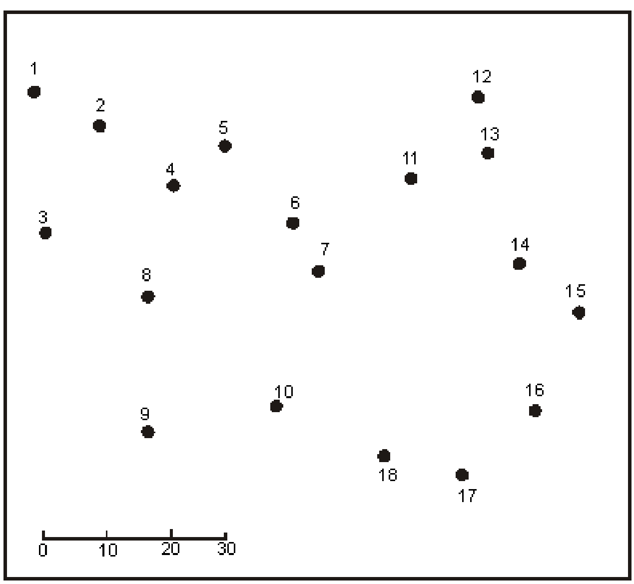

Fig.1 - Sítios Arqueológicos de povoamento da área X

deias pré-históricas que se constituem em sítios arqueológicos atuais (Fig. 1).

Começando a análise, traça-se um sistema de coordenadas, podendo-se utilizar, para isso, o quadro de delimitação da área como referência. A determinação da mediana espacial é simples, já que essa estaria exatamente no meio dos 18 sítios, separando-se duas classes de nove sítios a partir da ordenada e duas classes de igual freqüência a partir da abscissa. Como se observa na figura 2, a mediana espacial se localiza nas proximidades do sítio número sete (Cole & King, pp. 213/217).

No cálculo do centro de gravidade, o mesmo sistema de coordenadas será utilizado para o estabelecimento de classes, tanto na ordenada quanto na abscissa. Na abscissa, o eixo de x, as classes são chamadas de colunas, como se observa na figura 3. É conveniente, visando facilitar o cálculo, que as classes tenham intervalos pares, para que o ponto médio seja um número inteiro. Começa-se o cálculo pela contagem do número de pontos em cada classe. No caso em estudo, por exemplo, na classe de 0 a 2, com ponto médio 1, localizam-se três sítios: o 1, o 2 e o 3. Todas as classes apresentam as freqüências apresentadas na tabela 3, notando-se que o sítio nº 13 colocou-se exatamente sob a reta

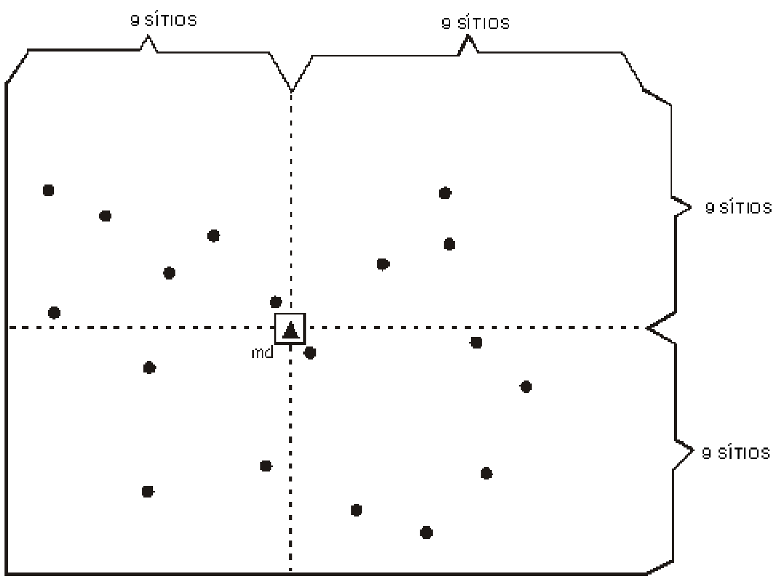

Fig.2 - Mediana da distribuição dos sítios

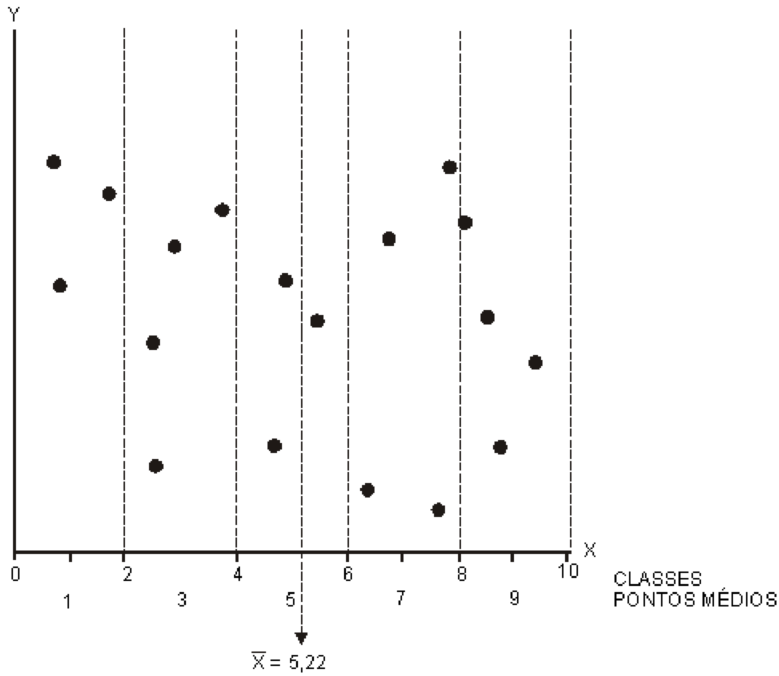

Fig. 3 - Pontos por classes de colunas

Tabela 3

Freqüência de Sítios nas Colunas

Classes	Ponto Médio (pm)	Freqüência (f)	f.pm
0 a 2	1	3	3
2 a 4	3	4	12
4 a 6	5	3	15
6 a 8	7	4	28
8 a 10	9	4	36
S	-	18	94

divisória das duas últimas classe. Nesse caso, optou-se por colocá-lo na penúltima. Aliás, esse pode ser um procedimento geral adotado, ou seja, o de se optar pelo posicionamento de pontos intermediários na classe de menor ponto médio.

$$\bar{x} = \frac{94}{18} = 5,22$$

Esse valor equilibra a distribuição dos pontos no eixo de x e para lá deve ser transportado (vide figura 3).

O mesmo procedimento deve ser adotado para o eixo das ordenadas, y, como se vê na figura 4. A freqüência dos sítios nas barras e o

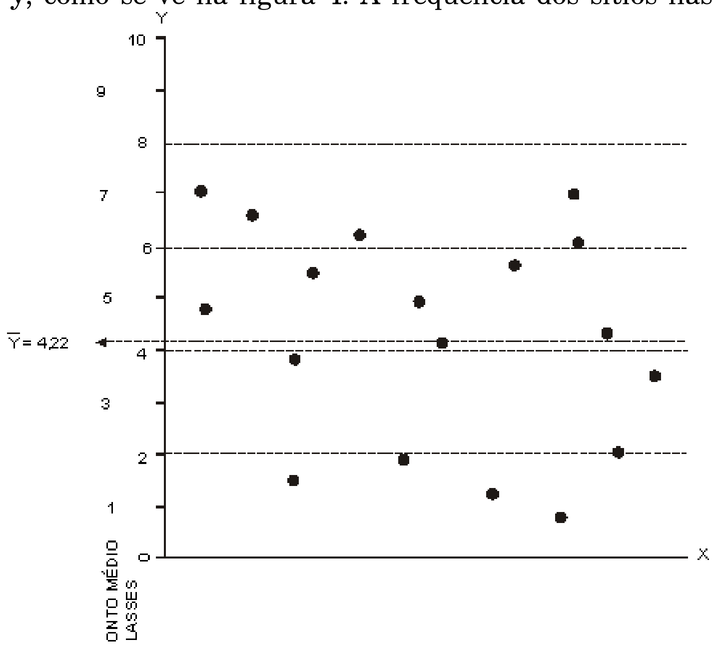

Fig. 4 - Pontos por classes de barras

Tabela 4

Freqüência de Sítios nas barras

Classes	Ponto Médio (pm)	Freqüência (f)	f.p
0 a 2	1	5	5
2 a 4	3	2	6
4 a 6	5	6	30
6 a 8	7	5	35
8 a 10	9	0	0
Σ	-	18	76

cálculo da média estão a seguir, notando-se que os sítios 10 e 16 foram colocados na primeira barra, adotando-se o procedimento anteriormente estabelecido.

$$\bar{y} = \frac{76}{18} = 4{,}22$$

Esse é o valor que equilibra a distribuição dos pontos no eixo de y. O centro de gravidade da distribuição dos pontos é definido, então, pelo cruzamento das linhas com valores de x = 5,22 e y = 4,22, ficando bem próximo ao sítio nº 7, como se observa na figura 5. É interessante observar que o ponto mediano e o centro de gravidade praticamente coincidiram no espaço, mostrando que a distribuição espacial desses sítios tende a apresentar uma certa regularidade.

O CÁLCULO DA DISTÂNCIA PADRÃO

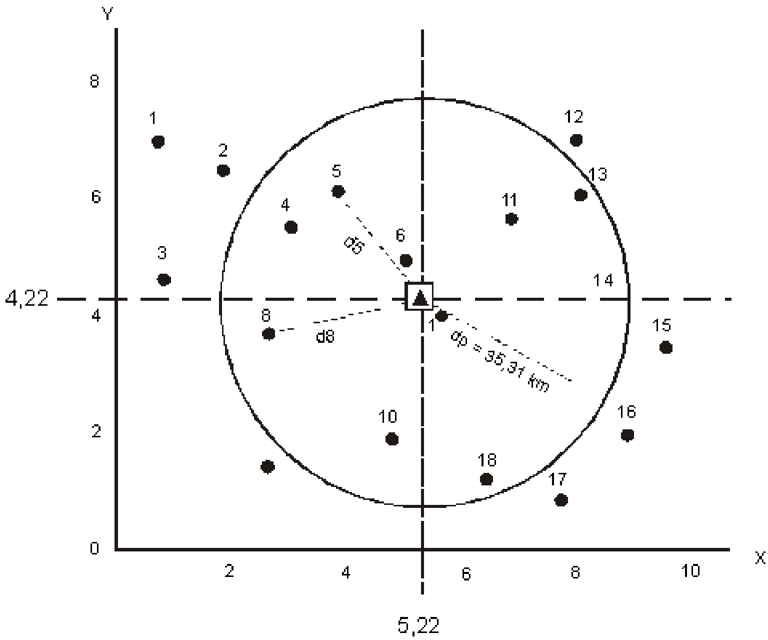

Fig.5 - Centro de gravidade e distância padrão

Para completar a descrição da distribuição dos sítios arqueológicos na área x, resta o cálculo da distância padrão, a ser expressa pelo raio de um círculo que, ao indicar a variação em torno da média, delimitará um espaço no qual tenderiam a estar concentrados, aproximadamente, 66,6% dos pontos da distribuição.

Lembrando da fórmula do desvio padrão, pode ser elaborada a tabela 5, onde os valores $x - \overline{x}$ (o valor de cada ponto da variável subtraído da média) vai corresponder, no caso de distribuições espaciais, **à distância** entre cada ponto x e o centro de gravidade, já que diferença, afastamento e distância são equivalentes. Na figura 5 estão indicadas, como exemplo, as distâncias $d_5 = x_5$ e $d_8 = x_8$.

Substituindo-se os valores na fórmula do desvio padrão, temos:

$$dp = \sqrt{\frac{21.197}{18-1}} - \sqrt{\frac{21.197}{17}} = \sqrt{1.246{,}88} = 35{,}31\text{km}$$

Tabela 5

x_i	(xi - $\bar{x}$) (km)	(xi - $\bar{x}$)^2	x_i	(xi - $\bar{x}$) (km)	(xi - $\bar{x}$)^2
1	53	2.809	10	24	576
2	41	1.681	11	22	484
3	44	1.936	12	39	1.521
4	26	676	13	34	1.156
5	24	576	14	33	1.089
6	7	49	15	44	1.936
7	1	1	16	42	1.764
8	27	729	17	41	1.681
9	38	1.444	18	33	1.089
-	-	-	-		Σ 21.197

Note-se que o círculo definido por esse raio engloba **55,6%** dos sítios arqueológicos da área x.

A BUSCA DE PADRÕES DE DISTRIBUIÇÃO

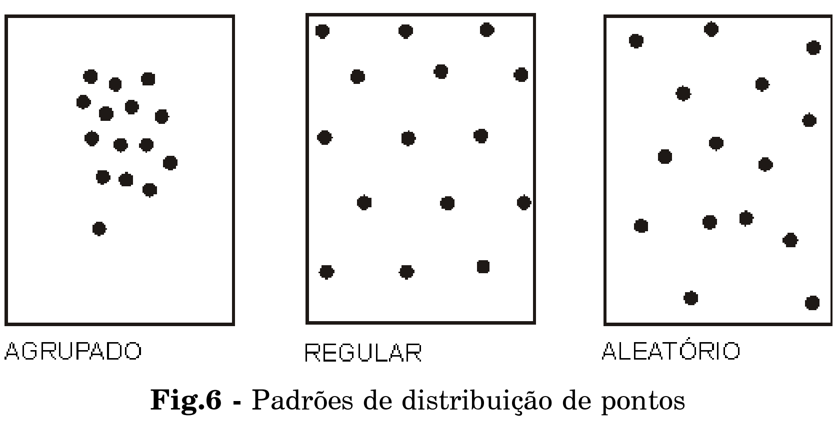

Fig.6 - Padrões de distribuição de pontos

Da simples descrição da distribuição pode-se passar para uma análise em que algumas hipóteses possam ser verificadas. É possível saber de forma precisa, fugindo-se de conclusões apenas baseadas na simples observação, se a distribuição tende ao agrupamento, à regularidade ou à aleatoriedade, como padrões básicos apresentados na figura 6.

Numa área definida qualquer, os sítios arqueológicos que indicam antigas aldeias podem estar agrupados, certamente em decorrência de algum fator que recomendou às antigas populações que se concentrassem em algum ponto da área (presença de água, solo mais fértil etc). Padrões mais regulares de distribuição de recursos poderiam sugerir a essas populações uma localização mais eqüidistante das aldeias, que chegaria ao máximo com assentamentos nos vértices de um hexágono, no típico modelo de Christaller. Mas é possível encontrar, também, distribuições aleatórias no mundo real. É possível encontrar-se uma situação em que cada assentamento tenha uma justificativa para sua localização individual mas não haver explicação para o conjunto das aldeias na área, ou seja, a relação entre elas não definir um padrão intencional.

Um procedimento utilizado para determinar o padrão de distribuição de pontos é o cálculo de R, que mede o desvio entre o afastamento real dos pontos e o que ocorreria se os mesmos estivessem posicionados de forma aleatória (King, p. 160). Através do cálculo de probabilidades, sabe-se que uma distribuição aleatória teria a distância entre os pontos, r(E), calculada por

$$\frac{1}{2} \cdot \frac{1}{\sqrt{p}},$$

onde p é a densidade dos pontos. Já o cálculo da distância entre os pontos na distribuição observada pressupõe a análise dos vizinhos mais próximos de cada ponto, indicada por uma média, r(A).

Observando-se a figura 7, vê-se que as setas indicam o vizinho mais próximo de cada aldeia da nossa área hipotética. A mais próxima da aldeia 2 é a aldeia 4, da 9 é a 10 e assim sucessivamente. As aldeias 6 e 7 são, reciprocamente, as mais próximas, enquanto a aldeia 3 é igualmente próxima das aldeias 2 e 8. Apenas a observação da citada figura permite um levantamento da hipótese de que os contatos na área x se faziam, prioritariamente, em quatro grupos de aldeias, relativamente isolados entre si. Note-se, inclusive, a posição privilegiada da aldeia 16,

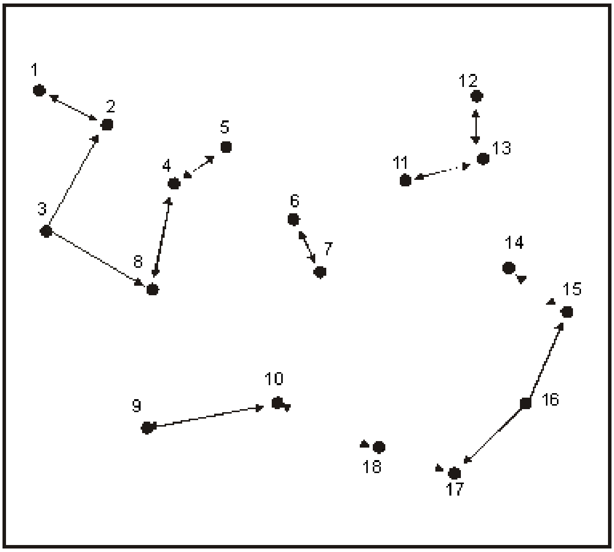

Fig.7 - Distância mais próxima entre aldeias

Tabela 6

Aldeia	Vizinho mais próximo	d (km)	Aldeia	Vizinho mais próximo	d (km)
1	2	12	10	18	17
2	1	12	11	11	11
3	2 e 8	20	12	13	7
4	5	10	13	12	7
5	4	10	14	15	10
6	7	7	15	14	10
7	6	7	16	15 e 17	15
8	4	16	17	18	12
9	10	20	18	17	12

através da qual poderiam ser efetuadas relações entre as aldeias 14 e 15, mais ao norte, e o grupo formado pelas aldeias 9, 10, 18 e 17. A distância, em quilômetros, entre cada aldeia e seu vizinho mais próximo, indicada por **d**, está expressa na tabela 4. No caso de haver mais de um vizinho mais próximo, opta-se por um deles, já que as distâncias são iguais.
A média das distâncias entre os vizinhos mais próximos é expressa por

$$r(A) = \frac{\sum r}{n} = \frac{215}{18} = 11{,}9\text{km}$$

ou seja, cada habitante da área precisaria deslocar-se, em média, 11,9km para alcançar a aldeia mais próxima à sua.

Pode-se, em seguida, comparar-se a distância real obtida com a hipotética, que ocorreria caso a distribuição fosse aleatória. Como a área **x** tem 9.000km^2 e a densidade **p** de pontos é de 0,002 pontos por km^2, a distância em quilômetro entre os sítios, caso a distribuição fosse aleatória seria:

$$r(E) = \frac{1}{2} \cdot \frac{1}{\sqrt{p}} = \frac{1}{2} \cdot \frac{1}{\sqrt{0{,}002}} = \frac{1}{0{,}09} = 11{,}1\text{km}$$

bem próxima da observada. Dividindo-se uma pela outra obtém-se o valor de R:

$$R = \frac{r(A)}{r(E)} = \frac{11{,}9}{11{,}1} = 1{,}07$$

notando-se que quanto mais próximo de 1, maior a tendência à aleatoriedade. De fato, sabe-se que o valor de R vai variar de zero, indicando o máximo agrupamento a 2,15, que representaria uma distribuição haxagonal de pontos. Pelo resultado obtido fica, então, estabelecido que a distribuição das aldeias na área **x** é aleatória.

CONSIDERAÇÕES FINAIS

As ferramentas apresentadas podem ser úteis para descrição e análise em Arqueologia. Podem servir para estabelecer pontos de partida em estudos espaciais podem ajudar no estabelecimento de hipóteses, mas devem ser empregadas num caráter exploratório, pois podem apresentar problemas de aplicação, sobretudo em decorrência da delimitação da área de estudo. No caso da Geografia, as áreas são sempre unidades administrativas de significado real, e as conclusões apenas a elas se referem. No caso específico dos estudos arqueológicos, torna-se mais difícil essa delimitação, podendo-se recorrer, para tal, a fronteiras de unidades ambientais. De qualquer modo, fica estabelecido que a validade dos resultados obtidos na análise é limitada à área especificada. Sua alteração, acrescentando ou reduzindo o número de observações, mudaria os valores de tendência central e de variabilidade, e o simples aumento ou redução da superfície estudada, por afetar a densidade de pontos, alteraria o cálculo de R(Hodder e Orton , p. 53).

REFERÊNCIAS

COLE, John P. e KING, Cuchlaine A M., **Quantitative Geography,** Wiley & Sons, Londres, 1968.

HODDER, Ian e ORTON, Clive, **Analisis espacial in Arqueologia.** Editorial Crítica, Barcelona, 1990.

KING, Leslie J., "A quantitative expression of the pattern of urban settlements in selected areas of the United States", in BERRY e MARBLE, **Spatial Analysis – A Reader in Statistical Geography.** Prentice Hall, Englewood Cliffs, 1968.