

ALFREDO JOSÉ BARRETO LUIZ - EMBRAPA/CNPMA

e.mail: alfredo@cnpma.embrapa.br

RESUMO

Em diversas áreas do conhecimento humano há interesse no estudo do padrão de distribuição espacial de pontos, sendo estes pontos nada mais que a localização dos objetos de estudo. Estes objetos podem ser dos mais variados tipos, tais como árvores dos Cerrados (MEIRELLES & LUIZ, 1995), unicórnios (HURLBERT, 1990), aranhas (HOFFMASTER, 1985), esconderijos de alimento de corvos (JAMES & VERBEEK, 1985), micro-artrópodes do solo (PANDE, 1984), urbanismo (SILK, 1979 e ROGERS, 1969), entre outros. Existem vários métodos para determinação da distribuição espacial de pontos, entretanto a descrição e a discussão desses métodos não fazem parte do escopo deste trabalho; para maiores detalhes existem boas revisões em MARTINS (1991) e LUIZ (1995). O objetivo deste trabalho é apresentar um programa SAS que, dadas as coordenadas cartesianas da posição de objetos, numa seção quadrada qualquer do espaço bidimensional, calcula a distância de cada objeto ao seu vizinho mais próximo e, utilizando esta distância, aplica o Teste de Clark e Evans (CLARK & EVANS, 1954) e o Teste de Aderência de Kolmogorov (CONOVER, 1971), permitindo determinar o padrão de distribuição espacial destes objetos. O programa foi utilizado para analisar dados simulados para casos típicos dos três padrões de distribuição espacial: agrupado, aleatório e regular. Os resultados e o programa são apresentados.

1. INTRODUÇÃO

O padrão com que se distribuem espacialmente os elementos de uma população é uma de suas características fundamentais. Muitas vezes o reconhecimento deste padrão permite fazer inferências sobre as suas causas, levando a uma melhor compreensão dos fenômenos envolvidos na sua formação. Entretanto, a determinação precisa do padrão de distribuição espacial de uma população ainda é uma matéria controvertida em diversas áreas, não havendo um teste de aceitação universal. O assunto torna-se ainda mais delicado pelo fato que, na maioria das vezes, não se pode estudar toda a população, mas apenas uma amostra dela.

Neste ponto cabe definir os três tipos de padrão de distribuição espacial: agrupado, aleatório e regular, que são aqui estudados. O padrão aleatório é gerado, sobre um plano, por um mecanismo que satisfaça as duas seguintes condições:

1. *Condição de Igual Probabilidade* - qualquer objeto (ponto) tem a mesma probabilidade de ocorrer em qualquer posição do plano. Para qualquer subregião do plano, portanto, a probabilidade de conter um objeto é igual a de qualquer outra subregião de mesma área.

2. *Condição de Independência* - a posição de um objeto no plano é independente da posição de qualquer outro objeto.

Os padrões agrupado e regular violam estas duas condições. Em ambos os casos há dependência entre a posição de um objeto e outro, sendo que no primeiro caso um objeto tem maior probabilidade de ocorrer em regiões próximas a outro, enquanto que no segundo um objeto só tem probabilidade não nula de ocorrer a uma distância fixa de outro.

Os pesquisadores tem abordado o problema de distinguir padrões de diversas maneiras, entre elas uma se destaca por sua facilidade de utilização prática e de interpretação; ela utiliza a distância entre um indivíduo e o seu vizinho mais próximo no espaço para determinar se o conjunto dos elementos de uma amostra aderem ou não a uma distribuição pré-determinada. Um dos métodos mais citados na literatura recente é o Teste de Clark e Evans, que utiliza a razão entre a média das distâncias ao vizinho mais próximo de uma amostra e o valor esperado para esta média, no caso de os elementos da população estarem distribuídos inteiramente ao acaso no espaço considerado (distribuição aleatória). Outra maneira de se avaliar o grau de aderência de dados amostrais à uma dada função de distribuição acumulada (f.d.a.) é através do Teste de Aderência de Kolmogorov. Neste trabalho é apresentado um programa SAS que permite a aplicação desses testes a conjuntos de dados das coordenadas cartesianas da posição de objetos, numa seção quadrada qualquer do espaço bidimensional, auxiliando na determinação do padrão de distribuição destes objetos em R^2 .

O programa desenvolvido foi aplicado a conjuntos de dados simulados, representando os três padrões de distribuição espacial típicos. A descrição dos testes, os dados simulados, o programa SAS, os resultados dos testes e sua interpretação são apresentados a seguir.

2. MATERIAL E MÉTODO

2.1. Teste de Clark e Evans

O teste estatístico proposto por Clark e Evans é dado por

$$C = \frac{\bar{X} - \hat{\bar{X}}}{\hat{S}_{\bar{X}}},$$

onde $\bar{X}$ é a média observada, $\hat{\bar{X}}$ é a média esperada para a distância ao vizinho mais próximo, quando os objetos estão distribuídos de forma aleatória no espaço, e $\hat{S}_{\bar{X}}$ é o desvio padrão da média esperado. Esta estatística tem distribuição normal com média zero e variância um. A hipótese nula do teste é a de que os objetos estão distribuídos de forma aleatória no espaço, sendo rejeitada para valores de C nas caudas da distribuição. Se a hipótese nula é rejeitada, utiliza-se um índice R , onde

$$R = \frac{\bar{X}}{\hat{\bar{X}}}, \text{ como indicativo de agregação } (R < 1) \text{ ou de regularidade } (R > 1).$$

Os valores esperados $\hat{\bar{x}}$ e $\hat{\bar{S}}$, propostos por Clark e Evans não são exatos para regiões limitadas, devido ao efeito de borda, o que gera um viés do teste em favor da regularidade (SINCLAIR, 1985). Uma estimativa mais precisa destes valores foi estabelecida por DONNELLY (1978), mas ainda neste caso, a determinação não era exata. Neste trabalho são utilizados os valores exatos obtidos através do desenvolvimento teórico apresentado por LUIZ (1995), mantendo-se a estrutura do teste.

2.2. Teste de aderência de Kolmogorov

No artigo onde apresentaram pela primeira vez o seu teste, Clark e Evans ressaltaram o seguinte:

"... Os testes de significância apresentados aqui são baseados na diferença entre $\bar{x}$ e $\hat{\bar{x}}$. É teoricamente possível, no entanto, existirem populações distribuídas não aleatoriamente, nas quais $\bar{x}$ e $\hat{\bar{x}}$ sejam iguais. A distribuição de tais populações não pode ter a sua não aleatoriedade detectada por estes testes. Neste caso, o afastamento da aleatoriedade pode ser determinado pela comparação das distribuições de frequência das distâncias ao vizinho mais próximo, observada e esperada, através de um teste χ^2 . A maioria das investigações de populações naturais, entretanto, não requererão testes tão sensíveis." (CLARK & EVANS, 1954).

Desde então, a experiência tem mostrado que a última afirmação não era totalmente correta. Por este motivo, neste trabalho também será aplicado um teste para comparar as f.d.a. observada e esperada. O teste escolhido foi o de Kolmogorov, por ser exato para distribuições contínuas e conservador em caso contrário (CONOVER, 1971).

Para a aplicação deste teste, precisamos da f.d.a. observada $F_{(x)}$, e da f.d.a. esperada $F_{(x)}^*$, que são utilizadas no cálculo da estatística teste T_1 , que é dada por

$$T_1 = \sup_x \left| F_{(x)} - F_{(x)}^* \right|.$$

A hipótese nula, que supõe completa aleatoriedade na distribuição dos objetos, é rejeitada a um nível de significância α se a estatística teste T_1 excede o valor de tabela para $p=1-\alpha$. Em CONOVER (1971) são fornecidos os valores da estatística teste de Kolmogorov para níveis de significância de 1, 2, 5, 10 e 20%.

2.3. Programa SAS para simulação de dados

Este programa SAS tem por finalidade simular as coordenadas cartesianas da posição ocupada por objetos distribuídos de forma aleatória, regular ou agrupada, em uma seção quadrada do plano, bastando para isso que se forneça a medida do lado da seção quadrada (L), o número de objetos (NOBJ) e o número de grupos (NGRU) a serem simulados. Para simular um conjunto de objetos distribuídos aleatoriamente, o número de grupos deve ser igual a zero (NGRU=0); para objetos com distribuição regular, o número de grupos deve ser igual ao número de objetos (NGRU=NOBJ); e para distribuição agrupada, o número de grupos deve ser aquele desejado, devendo ficar entre zero e o número de objetos ($0 < \text{NGRU} < \text{NOBJ}$). O programa é apresentado a seguir:

OPTIONS LS=80 PS=64;

TITLE'COORDENADAS CARTESIANAS SIMULADAS';

DATA UM;

L= ;

NOBJ= ;

NGRU= ;

A=L*L;

LB=0.1*L;

LI=0.8*L;

IF NGRU=0 THEN DO

I=1 TO NOBJ;

X=L*RANUNI(-33117);

Y=L*RANUNI(-57735);

OUTPUT;

END;

IF NGRU=NOBJ THEN DO;

D0=SQRT(NOBJ);

D1=INT(D0);

IF D1=D0 THEN D2=D1;

ELSE D2=D1+1;

NOBJI=D1*D2;

LGX=L/D1;

VARX=(0.1*LGX)**2;

LGY=L/D2;

VARY=(0.1*LGY)**2;

DO J=1 TO D1;

DO K=1 TO D2;

X=((J*LGX)-(LGX/2))+SQRT(VARX)*RANNOR(-11973);

Y=((K*LGY)-(LGY/2))+SQRT(VARY)*RANNOR(-73391);

OUTPUT;

END;

END;

END;

IF 0<NGRU<NOBJ THEN DO;

NN=INT(NOBJ/NGRU);

VAR=0.05*A/NOBJ;

DO J=1 TO NGRU;

CX=LB+(LI*RANUNI(-93579));

CY=LB+(LI*RANUNI(-13591));

DO K=1 TO NN;

X=CX+SQRT(VAR)*RANNOR(-13975);

$Y = CY + \text{SQRT}(\text{VAR}) * \text{RANNOR}(-77559);$

OUTPUT;

END;

END;

END;

PROC PRINT DATA=UM NOOBS;

FORMAT X Y 5.2;

VAR X Y;

RUN;

Esse programa foi utilizado para simular três conjuntos de 25 objetos, sendo um de cada tipo de distribuição espacial. Os dados gerados são apresentados abaixo:

Objeto	Padrão aleatório		Padrão regular		Padrão agrupado	
	X	Y	X	Y	X	Y
1	7,16	59,61	8,41	12,47	35,79	59,40
2	8,54	71,25	8,69	86,88	37,15	59,19
3	10,24	39,46	9,08	47,34	37,29	58,98
4	19,27	78,71	11,52	65,71	37,42	65,76
5	21,59	3,36	11,61	34,12	37,66	60,87
6	28,82	85,72	27,15	50,71	38,32	63,29
7	32,32	59,34	28,11	13,01	38,68	61,20
8	35,83	55,71	29,69	74,08	38,74	61,19
9	38,36	2,13	29,84	92,60	39,43	50,85
10	46,15	11,85	30,96	30,81	39,87	48,67
11	50,49	90,57	47,37	29,78	40,47	64,02
12	55,07	11,59	49,75	49,40	40,75	63,01
13	60,20	74,83	51,49	88,26	41,26	60,25
14	65,12	7,78	52,14	70,80	41,73	57,10
15	69,00	52,42	52,48	9,60	42,13	58,20
16	69,51	37,18	68,05	68,47	43,31	62,12
17	71,99	3,41	69,90	8,69	43,36	55,21
18	72,49	47,17	71,75	48,67	43,99	59,64
19	78,06	62,16	71,99	87,18	44,54	60,17
20	80,82	28,54	72,59	31,30	44,94	60,38
21	88,77	85,69	86,50	86,74	45,44	56,75
22	95,57	57,61	88,79	27,46	46,39	61,73
23	96,53	84,07	90,35	50,90	48,47	60,71
24	99,13	43,37	91,39	10,07	50,13	60,77
25	99,42	55,46	91,45	69,04	51,46	64,08

2.4. Programa SAS para análise dos dados

Este programa SAS calcula as distâncias euclidianas de cada objeto ao seu vizinho mais próximo, o valor da estatística teste T_1 de Kolmogorov, as médias esperadas e observadas destas distâncias e os seus desvios padrões, e o valor da estatística C do teste de Clark e Evans, seu nível de significância (P) e o índice R, quando temos N pontos distribuídos em uma área quadrada de lado L, bastando para isso fornecer os valores de N, de L e as coordenadas X e Y da posição de cada objeto. A seguir é apresentado o programa:

```
OPTIONS LS=80 PS=64;

TITLE1 'DADOS DE _____';

DATA UM;

INPUT X Y;

CARDS;

...

;

PROC IML WORKSIZE=414;

    USE UM;

    READ ALL VAR{X} INTO X;

    READ ALL VAR{Y} INTO Y;

    N=NROW(X);

    X=REPEAT(X,1,N);

    Y=REPEAT(Y,1,N);

    M=(X-X`)*(X-X`)+(Y-Y`)*(Y-Y`);

    M=M+(I(N)*1000000);

    M=SQRT(M);

    DO I=1 TO N;

        MINIMO=MINIMO//MIN(M[I,]);

    END;

    NOME={X};

    CREATE DOIS FROM MINIMO[COLNAME=NOME];

    APPEND FROM MINIMO[COLNAME=NOME];

    CLOSE DOIS;

QUIT;

PROC SORT DATA=DOIS OUT=DOIS;

    BY X;
```

```

DATA DOIS;
SET DOIS;
I = _N;
L = ;
N = ;
PI = ARCCOS(-1);
IF X>L THEN
FXE=1-(((1 - ((PI*(X**2))/(L**2)))+(8*(X**3))/(3*(L**3))))
- ((X**4)/(2*(L**4)))*((N-1));
IF X>=L THEN
FXE=1-(((2/3)-(((2*(X**2))/(L**2))* ( ARCSIN (((2*(L**2))/(X**2))-1))))
-((4*SQRT((X**2)-(L**2))/(L)-(8*SQRT((X**2)-(L**2))*((3*(L**3))))
+ ((2*(X**2))/(L**2)))+(X**4)/(2*(L**4)))*((N-1));
DATA TRES;
N = ;
DO I=1 TO N;
FXO=I/N;
OUTPUT;
END;
DATA QUATRO;
MERGE DOIS TRES;
BY I;
T1=ABS(FXO-FXE);
PROC MEANS N MAX DATA=QUATRO NOPRINT;
VAR T1;
OUTPUT OUT=KOLM N=N MAX=T1;
PROC PRINT DATA=KOLM NOOBS;
FORMAT T1 5.3;
VAR N T1;
TITLE2' VALOR DA ESTATISTICA TESTE DE KOLMOGOROV (T1)';
RUN;
PROC MEANS MEAN STDERR DATA=QUATRO NOPRINT;
VAR X;
OUTPUT OUT=SEISA MEAN=XO STDERR=SXO;

```

```

DATA SEISA;
K=1;
SET SEISA;
DATA CINCO;
N =
;
L =
;
R = L*SQR(2);
PI = ARCCOS(-1);
D = R/5000;
DO I=0, TO R BY D;
IF I<L THEN
U = (((PI*(1**2))/(L**2)) - (((8*(1**3))/(3*(L**3)))+(1**4)/(2*(L**4)))));
IF I<L THEN
DU = (((2*PI*I)/(L**2)) - (((8*(1**2))/(L**3)))+(2*(1**3))/(L**4)));
IF I>=L THEN
U = (((2*(1**2))/(L**2))* (ARSIN(((2*(L**2))/(1**2))-1))) +
((4*SQR(1**2)-(L**2))/(L)+(8*SQR(1**2)-(L**2))**3)/(3*(L**3)))
- ((2*(1**2))/(L**2))-((1**4)/(2*(L**4)))+(1/3));
IF I>=L THEN
DU = (((4*I)/(L**2))* (ARSIN(((2*(L**2))/(1**2))-1))) +
((8*I*SQR(1**2)-(L**2))/(L**2)) - ((4*I)/(L**2)) - ((2*(1**3))/(L**4)))
EX = D*(N-1)*I*((1-U)**(N-2))*DU;
EX2 = D*(N-1)*I*((1-U)**(N-2))*DU;
OUTPUT;
END;
PROC MEANS SUM DATA=CINCO NOPRINT;
VAR EX EX2;
OUTPUT OUT=SEISB SUM=XE EX2;
DATA SEISB;
SET SEISB;
K=1;

```

DATA SEIS;

MERGE SEISA SEISB;

BY K;

N = ;

$SXE = \sqrt{(EX2 - (XE**2))/N}$;

$C = (XO - XE)/SXE$;

IF C < 0 THEN P=PROBNORM (C);

ELSE P=1-PROBNORM(C);

R=XO/XE;

PROC PRINT DATA=SEIS NOOBS;

FORMAT C P R 5.3;

VAR C P R;

TITLE2' VALOR DA ESTATISTICA TESTE DE CLARK E EVANS (C);'

TITLE3' SEU NIVEL DE SIGNIFICANCIA (P) E O INDICE (R)';

RUN;

3. RESULTADOS E DISCUSSÃO

Os três conjuntos de dados simulados foram analisados pelo programa apresentado, produzindo os seguintes resultados:

Padrão	Estatística teste de Kolmogorov (T_1)	Nível de significância de T_1	Estatística teste de Clark e Evans (C)	Nível de significância de C	Índice R
Agrupado	0,911	<0,01	-8,000	<0,001	0,138
Aleatório	0,102	>0,20	-0,055	0,478	0,994
Regular	0,672	<0,01	5,701	<0,001	1,614

A interpretação dos resultados acima é a de que para os primeiro e terceiro padrões simulados, os dois testes utilizados rejeitaram a hipótese nula da distribuição aleatória dos objetos, e o índice R, com um valor menor do que um para o primeiro e maior que um para o terceiro, indica padrões agrupado e regular, respectivamente, acertando nos dois casos. Para o segundo conjunto de dados simulados, nenhum dos dois testes foi capaz de rejeitar a hipótese de aleatoriedade, coincidindo com um índice com valor muito próximo a um, o que também está correto.

Conclui-se que os testes e o programa apresentados são úteis na determinação do padrão de distribuição espacial de objetos em R^2 .

4.BIBLIOGRAFIA

- CLARK, J. P.; EVANS, F. C. Distance to nearest neighbor as a measure of spatial relationship in populations. Ecology, v.35, n.4, p.445-453, 1954.
- CLARK, J. P.; EVANS, F. C. On some aspects of spatial pattern in biological populations. Science, v.121, p.397-398, 1955.
- CONOVER, W. J. Practical nonparametric statistics. New York: John Wiley & Sons, 1971. 462p.
- DONNELLY, K. P. Simulations to determine the variance and edge effect of total nearest-neighbor distance. In: HODDER, I., ed. Simulation studies in archaeology. London: Cambridge University Press, 1978. p.91-95
- GOODALL, D. W. Quantitative aspects of plant distribution. Biological Reviews, v.27, p.195-245, 1952.
- GOODALL, D. W. Statistical plant ecology. Annual Review of Ecology and Systematics, v.1, p.99-124, 1970.
- HOFFMASTER, D. K. Community composition and nearest neighbor relationships in orb-weaving spiders: The product of aggression? Behav. Ecol. Sociobiol., v.16, p.349-353, 1985.
- HULBERT, S. H. Spatial distribution of montane unicorn. Oikos, v.58, n.3, p.257-271, 1990.
- JAMES, P. C.; VERBEEK, N. A. M. Clam storage in a North-western Crow (*Corvus caurinus*): dispersion and sequencing. Can. J. Zool., v.63, n.4, p.857-860, 1985.
- LUIZ, A. J. B. Determinação da distribuição espacial de pontos usando a distância ao vizinho mais próximo: aplicação em populações vegetais. Brasília: UnB, 1995. 84p. Tese de mestrado.
- MARTINS, F. R. Estrutura de uma floresta mesófila. Campinas: UNICAMP, 1991. 246p. Série teses.
- MEIRELLES, M. L.; LUIZ, A. J. B.. Padrões espaciais de árvores de um cerrado de Brasília, DF. Revista Brasileira de Botânica, v.18, n.2, p. 187-191. dez. 1995.
- PANDE, Y. D. Studies on the horizontal distribution of oribatei employing the technique of soil sectioning. J. Soil. Biol. Ecol., v.4, n.1, p.22-29, 1984.
- RIPLEY, B. D. Spatial statistics. New York: John Wiley & Sons, 1981. 252p.
- RIPLEY, B. D. Tests of 'randomness' for spatial point patterns. Journal of the Royal Statistical Society, Séries B, v.41, n.3, p.368-374, 1979.
- ROGERS, A. Quadrat analysis of urban dispersion: 1.Theoretical techniques. Environment and Planning, v.1, p.47-80, 1969.
- SAS INSTITUTE. SAS/IML user's guide, release 6.03 edition. Cary, NC, 1988. 357p.
- SAS INSTITUTE. SAS language guide for personal computers, release 6.03 edition. Cary, NC, 1988. 558p.
- SAS INSTITUTE. SAS procedures guide, release 6.03 edition. Cary, NC, 1988. 441p.
- SILK, J. Statistical concepts in geography. London: George Allen & Unwin, 1979. 276p.
- SINCLAIR, D. F. On tests of spatial randomness using mean nearest neighbor distance. Ecology, v.66, n.3, p.1084-1085, 1985.