Caatinga, 4(1/2):42-48, 1984

UMA METODOLOGIA APLICÁVEL AO ZONEAMENTO AGROCLIMATOLÓGICO[1]

MÁRIO BEZERRA FERNANDES

Professor Titular, Escola Superior de Agricultura de Mossoró

SINOPSE - É proposto um novo procedimento estatístico para classificar n entidades (locais), quando estes são possíveis de se apresentarem sob forma de equação de regressão do tipo

$$y_j = b_0 + \sum_{i=1}^{k} b_i x_i \qquad \text{ou}$$

$$y_j = b_0 + \sum_{i=1}^{k} b_i x^i, \qquad j = \{1, 2 \ldots n\}.$$

Dois locais (equações) são considerados similares se $y_j \leq y_j'$, com base no valor mínimo de t obtido a partir do vetor W de distâncias entre essas regressões. Se $y_j > y_j'$, H_0 é rejeitada e os 2 locais são de grupos distintos. Uma aplicação foi feita com dados pluviométricos mensais médios das cidades de Fortaleza-CE e Mossoró-RN.

INTRODUÇÃO

No momento, há uma tendência, especialmente no campo da Biologia sistemática para o emprego de métodos estatísticos destinados a caracterizar grupos ou entidades dentro de um universo de dados predominantemente dispersos.

Quando se examina os dados das normas pluviométricas mensais de várias localidades de uma grande região como o Nordeste do Brasil, verifica-se, através de plotação em escala apropriada, analogias entre certo número de locais.

A constatação de semelhança sugere a necessidade de grupar ou zonear a região relativamente a essa variável climática. Para se efetuar um zoneamento, impõe-se uma sofisticada análise estatística através da taxonomia numérica e análise de conglomeração (*cluster analysis*). O primeiro trabalho, utilizando a análise de conglomeração é devido a ABOU-EL-FITTOUH *et al.* (1969), destinado a classificar os locais da região cotonicultora dos EE.UU. Neste trabalho, foi adotado, como medida de similaridade entre 2 locais, o coeficiente de distância, calculado a partir de uma matriz de estimativa de interação cultivar x local. A partir de então, outras surgiram: MUNGOMERY *et al.* (1974) empregaram análise de conglomeração e ordenação no estudo de adaptação da cultura da soja; CAMPBELL & LAFEVER (1977)

[1]Recebido para publicação em 03.08.1981.

realizaram um estudo de similaridade de todo o leste dos EE.UU. no que concerne à produção de trigo, para tanto usando o método de conglomeração para agrupamento dos locais e os resultados são apresentados através de dendrogramas. Na bibliografia brasileira são encontrados MOTA *et al.* (1974) com uma proposta de zoneamento agroclimatológico com base no estudo dos fatores climáticos através da análise de variância e CORDEIRO & SILVA (1980) que realizaram um estudo sobre o zoneamento para a cultura do milho, na região central do Brasil, utilizando a metodologia de análise de conglomeração através da técnica de ligação média não ponderada. Este trabalho constitui a mais importante contribuição, no gênero, em língua portuguesa, por ser o primeiro a empregar os métodos de análise numérica para solucionar problemas de similaridade e classificação de entidades. Pela citação, pode-se concluir que a importância dos estudos realizados é função da metodologia empregada. Por outro lado, é sabido que um refinamento na técnica estatística fornece mais informações acerca do problema, objeto de estudo, e, em contra-partida, uma técnica assim traz dificuldades enormes inerentes ao seu uso (emprego de computador), à análise e a interpretação dos resultados. O presente trabalho propõe uma nova metodologia para classificação de entidades ou medir similaridades existentes entre elas através da análise multivariada, a partir de equação de regressão múltipla e polinomial. Trata-se de uma extensão do método apresentado por HEWETT & LABABIDI (1980), aos polinômios em X de qualquer grau.

O MÉTODO

Vamos considerar preliminarmente que os dados disponíveis possam se ajustar a duas equações de regressão:

$$\begin{cases} y_1 = b_1{}^0 + \sum_{i=1}^{k} b_{1i}\, x_{1i} \\ \text{e} \\ y_2 = b_2{}^0 + \sum_{i=1}^{k} b_{2i}\, x_{2i} \end{cases} \quad \text{ou}$$

$$\begin{cases} y_1 = b_1{}^0 + \sum_{i=1}^{k} b_{1i}\, x_1^i \\ \text{e} \\ y_2 = b_2{}^0 + \sum_{i=1}^{k} b_{2i}\, x_2^i \end{cases}$$

Nestas condições, devemos formular as seguintes hipóteses:

H_0: $y_1 \leq y_2$, para no mínimo um valor de $X \in R$.

H_a: $y_1 > y_2$, para todos os valores de $X \in R$.

R representa uma região finita no espaço K-dimensional, tal que $X_{i*} \leq X_i \leq X_i^*$, onde X_{i*} e X_i^* representam os valores máximo e mínimo, respectivamente, de X_i observados na amostra.

Note-se que haverá similaridade entre y_1 e y_2 se H_0 for aceita, caso contrário elas serão significativamente diferentes com base no teste t ao nível de 5% de probabilidade. Na obtenção do teste, são indispensáveis as seguintes etapas:

a) Obtêm-se as estimativas dos vetores de parâmetros

$$\hat{\beta}_1' = \left[\hat{b}_{01},\ \hat{b}_{11} \ldots \hat{b}_{k1}\right] \text{ e}$$

$$\hat{\beta}_2' = \left[b_{02},\ b_{12} \ldots b_{k2}\right].$$

b) Estima-se a variância residual:

$$S^2 = \left[(y_1'y_1 - \hat{\beta}_1'x_1'y_1) + (y_2'y_2 - \hat{\beta}_2'x_2'y_2)\right] / \left[N_1 + N_2 - 2(k+1)\right]$$

onde, y é o vetor de observação e y', o vetor transposto;
x, a matriz das variáveis independentes, sendo x' sua transposta;
N_1, o nº de observação de y_1;
N_2, o nº de observação de y_2; e
k, o nº de parâmetros.

c) Calcula-se o vetor $\hat{w}$ de distância entre dois hiperplanos com 2^k pontos sobre cada um deles. Os pontos são fórmados pelas seguintes combinações:

$P_1\ (x_1^*,\ x_2^*,\ x_3^*,\ x_k^*)$,

$P_2\ (x_{1*},\ x_2^*,\ x_3^* \ldots x_k^*)$, ...

... $P_{2k}\ (x_{1*},\ x_{2*},\ x_{3*} \ldots x_{k*})$.

Então, $\hat{w} = D \cdot \hat{\beta}$, onde

$$D = \begin{bmatrix} 1 & x_1^* & x_2^* & x_3^* \cdots & x_k^* & 1 & -x_1^* & -x_2^* & -x_3^* \cdots & -x_k^* \\ 1 & x_{1*} & x_2^* & x_3^* \cdots & x_k^* & -1 & -x_{1*} & -x_2^* & -x_3^* \cdots & -x_k^* \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1*} & x_{2*} & x_{3*} \cdots & x_{k*} & -1 & -x_{1*} & -x_{2*} & -x_{3*} \cdots & -x_{k*} \end{bmatrix}_{2^k \times (2k+2)}$$

Os valores X_1, X_2, X_3 ... X_k podem significar também X, X^2, X^3 ... ... X^k, respectivamente, conforme veremos adiante.

d) Calcula-se as matrizes V_1 e V_2 de variância e covariância de $\hat{\beta}_1$ e β_2, respectivamente, que tem dimensões $(k+1) \cdot (k+1)$.

e) Obtem-se em seguida matriz E de variância e covariância de $\hat{w}$, de dimensões $(2^k \cdot 2^k)$, expressa por:

$$E = D \cdot \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} \cdot D'$$

A partir dos elementos e_{ii}, calcula-se os t_i valores correspondentes às w_i, distâncias $i = 1$ ... ... 2^k, através da fórmula:

$$t_i = \frac{w_i}{\sqrt{e_{ii} \cdot s^2}}$$

A hipótese H_0 será rejeitada se o valor mínimo $(t_1,\ t_2 \ldots t_{k^2}) > t_{1-\alpha}\ (g)$, onde $t_{1-\alpha}\ (g)$ representa o percente de ordem $1 - \alpha$ da distribuição de t com g graus de liberdade. Em outras palavras, a rejeição de H_0 indica que as duas entidades y_1 e y_2 não podem constar num mesmo grupo, só ocorrendo quando essa hipótese for aceita, e isto caracterizará a similaridade. É claro que num estudo de zoneamento existirão y_1, $y_2 \ldots y_n$ correspondentes aos n locais, havendo portanto $\frac{n(n-1)}{2}$ comparações que obrigatoriamente serão submetidas a idêntico tratamento. Em consequência, poderá ser distinguido um nº de grupos compreendido entre 1 e n. É natural que dentro de cada grupo haverão possibilidades de numa segunda etapa identificar sub-grupos, bastando usar adequadamente a referida técnica.

UMA APLICAÇÃO EM DADOS PLUVIOMÉTRICOS

Com o intuito de emprestar maior simplicidade num exemplo prático, foram selecionadas as normais pluviométricas mensais de Fortaleza-CE e Mossoró-RN, conforme quadro 1. A falta de normalidade nos dados originais obrigou a que se aplicasse uma transformação de modo a atender a essa exigência, mostrando-se plenamente satisfatória, ao nível de 5%, a logarítimica L(y), através do teste de Wilk.

Neste exemplo, devemos investigar qual o tipo de equação de regressão que melhor se ajusta aos dados transformados (y) ao longo dos meses (x), isto é, $y = f(x)$. Esta investigação é muito simples, no caso vertente, dado que, a variável independente, $x = \{1, 2, 3... 12\}$ permitirá a aplicação dos polinômios ortogonais. Analisando os locais separadamente, tem-se o que se apresenta nos quadros 2 e 3. Ressalte-se que a análise foi conduzida até o 3º grau, não havendo comprovação de existência desse componente, a equação a ser ajustada, para cada local será: $y = b_0 + b_1x + b_2x^2$.

A partir desse momento, seguem-se as etapas retro-estabelecidas:

a) Os vetores $\hat{\beta}_1$ e $\hat{\beta}_2$ correspondem às equações de Fortaleza e Mossoró, respectivamente:

$$\beta_1' = \left[0{,}4928;\ 1{,}4701;\ -\ 0{,}1099\right] \quad \text{e}$$

$$\beta_2' = \left[-\ 1{,}6746;\ 1{,}9836;\ -\ 0{,}1505\right].$$

Considerando a transformação realizada, a expressão das equações, para os dois locais estudados, será:

$$y_1 = e^{(0{,}4928\ +\ 1{,}4701x\ -\ 0{,}1099x^2)}$$

e

$$y_2 = e^{(-\ 1{,}6746\ +\ 1{,}9836x\ -\ 0{,}1505x^2)},$$

onde e representa a base neperiana.

b) A variância residual conjunta será:

$$S^2 = (1{,}4324 + 0{,}4922)/18 = 0{,}1069$$

c) O vetor $\hat{W}$ das distâncias, que é

QUADRO 1 - Normais pluviométricas mensais das cidades de Fortaleza-CE e Mossoró-RN (y e L(y)).

Meses / Locais	Out	Nov	Dez	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set
Fortaleza	9,0 2,1972	15,6 2,7472	32,2 3,4719	88,8 4,4864	183,6 5,2127	307,0 5,7268	333,7 5,8102	212,7 5,3599	105,6 4,6597	41,5 3,7257	19,0 2,9444	16,9 2,8273
Mossoró	1,4 0,3365	5,1 1,6292	13,4 2,5922	45,3 3,8133	98,3 4,5880	147,5 4,9938	159,1 5,0695	104,7 4,6511	38,6 3,6553	17,8 2,8792	9,1 2,2083	1,5 0,4505

Fonte: Balanço Hídrico do Brasil, DNM - MA. - 1972.

expresso por $\hat{W} = D\hat{\beta}_{12}$, onde a matriz D é:

$$D = \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 144 & -1 & -1 & -144 \\ 1 & 12 & 1 & -1 & -12 & -1 \\ 1 & 12 & 144 & -1 & -12 & -144 \end{bmatrix};$$

então,

$$\hat{W}' = \left[1{,}6945;\ 15{,}5003;\ -39540;\ 1{,}8518\right]$$

d) As matrizes V_1 e V_2 de variância de $\hat{\beta}_1$ e $\hat{\beta}_2$ são:

$$V_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/572 & 0 \\ 0 & 0 & 1/12012 \end{bmatrix} \cdot S^2 \quad \text{e}$$

$$V_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/572 & 0 \\ 0 & 0 & 1/12012 \end{bmatrix} \cdot S^2$$

e) A matriz de variância e covariân-

QUADRO 2 - Análise de regressão para os dados de Fortaleza.

C.V.	gl	S.Q.	Q.M.	F
Reg. 1º grau	1	0,2465	-	1,54 n.s.
Reg. 2º grau	1	16,1208	-	101,3***
Resíduo	6	1,4324	0,1591	
	11	17,7997		

*** significativo, P = 1%.

QUADRO 3 - Análise de regressão para os dados de Mossoró.

C.V.	gl	S.Q.	Q.M.	F
Reg. 1º grau	1	0,1105	-	2,02 n.s.
Reg. 2º grau	1	30,2124	-	552,5***
Resíduo	6	0,4922	0,05468	
Total	11	30,8151		

*** significativo, P = 1%.

cia de $\hat{W}$, expressa por

$$E = D \cdot V_{12} \cdot D' \ ,$$

onde V_{12} é a matriz composta do seguinte modo:

$$V_{12} = \begin{bmatrix} V_1 & 0 \\ & \\ 0 & V_2 \end{bmatrix} \cdot S^2 =$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/572 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/12012 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/572 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/12012 \end{bmatrix} \cdot S^2$$

Como há interesse apenas nos elementos da diagonal de E, que correspondem às variâncias de W, omite-se o cálculo das covariâncias, daí resultando

$$\text{Diag. } E = \begin{bmatrix} 2{,}0036 & \cdots & \cdots & \cdots \\ \cdots & 5{,}4560 & \cdots & \cdots \\ \cdots & \cdots & 2{,}5036 & \cdots \\ \cdots & \cdots & \cdots & 5{,}9560 \end{bmatrix} \cdot S^2$$

A partir desses resultados, calcula-se os valores do teste t através de

$$t_i = \frac{w_i}{\sqrt{e_{ii} \cdot s^2}} \ ; \ t_1 = 3{,}66; \ t_2 = 20{,}29;$$

$t_3 = 7{,}64$ e $t_4 = 2{,}32$; o valor mínimo $(t_1, t_2, t_3, t_4) = 2{,}32$ e como $t_{95\%}(18) = 2{,}101$, podemos rejeitar H_0 com uma probabilidade de erro menor que 5%.

Conclui-se portanto que as duas entidades y_1 (Fortaleza) e y_2 (Mossoró) são significativamente dissimilares.

LITERATURA CITADA

ABOU-EL-FITTOUH, H.; RAWLINGS, J. C. & MILLER, P. A.; 1969. Classifications of environments to control genotype by environment interaction with application to cotton. *Crop Sci.*, 9: 135-40.

CAMPBELL, L. G. & LAFEVER, H. N.; 1977. Cultivar x environment interaction in soft red winter yeld tests. *Crop Sci.*, 17: 604-8.

CORDEIRO, C. M. T.; 1980. Estudo do zoneamento da região Centro-Sul do Brasil para a cultura do milho. *Pesq. Agropec. Bras.* Brasília, 15(2): 191-205.

DRAPER, N. R. & SMITH, H.; 1966. *Applied Regression Analysis*. New York, John Willey & Sons Inc. 407 p.

FISHER, R. A. & YATES. F.; 1971. *Tabelas Estatísticas*. São Paulo, USP. Ed. Polígono. 150 p.

HEWETT, J. E. & LABABIDI, Z.; 1980. Comparison of two populations with multivariate data. *Biometrika*, 36: 671-5.

MINISTÉRIO DA AGRICULTURA; 1972. *Balanço Hídrico do Brasil*, Rio de Janeiro.

MOTA, F. S.; BEIRSDORF, M. I. C.; MOTA, W. A. & WESTIPHALEN, S. L.; 1974. *Zoneamento agroclimático do Rio Grande do Sul e Santa Catarina*. Pelotas, Secretaria da Agricultura, 121 p. (IPEAS, circular nº 50).

MUNGOMERY, V. E.; SHORTER, R. & BYTH, D. E.; 1974. Genotype-environment interactions and environmental adaptation. I. Pattern analysis application to soybean populations. *Aust. J. Agric. Res.* 25: 59-72.

ABSTRACT

A new statistical method for classifying n entities (places), when they can be presented acording to a regression equation of the type

$$y_j = b_0 + \sum_{i=1}^{k} b_i x_i \quad \text{or}$$

$$y_j = b_0 + \sum_{i=1}^{k} b_i x^i , \quad \text{where} \quad j = \{1, 2 \ldots n\}.$$

Two places (equations) are considered to be similar if $y_j \leq y_j'$, on the basis of the minimum value of t by means of the distance vector w between those two regressions. If $y_j > y_j'$, H_0 is rejected and the two places are said to be distinct. The new method was tested using mean monthly pluviometrical data registered for the cities of Fortaleza and Mossoró, situated respectively in the States of Ceará and Rio Grande do Norte, Northeastern Brazil.

RÉSUMÉ

On a présenté une procedure méthodologique pour rechercher des ressemblances entre les régressions y_j et y_j', oú

$$y_j = b_0 + \sum_{i=1}^{k} b_i x_i \quad \text{ou}$$

$$y_j = b_0 + \sum_{i=1}^{k} b_i x^i , \quad \text{oú} \quad j = \{1, 2 \ldots n\}$$

qui doivent réprésenter des lieux ou des entités.

Deux endroits sont similaires quand $y_j \leq y_j'$; au contraire, on refuse l'hypothése H_0. La décision est prise en regardant la valeur minimum de t obtenie du vecteur de distances, nommé w. En fin, on fait une application avec les données pluviométriques des villes de Fortaleza et Mossoró, chez les États du Ceará et du Rio Grande do Norte, respectivement, dans la région Nord-Est du Brésil.