PALESTRA

Estado da Arte nas Avaliações por Regressões Espaciais

Rubens Alves Dantas

- **Professor de Engenharia de Avaliações e Avaliação Econômica de Empreendimentos, na Universidade Federal de Pernambuco – UFPE, desde 1981;**
- **Diretor da Dantas Engenharia de Avaliações.**

rubens@dantas.eng.br – www.dantas.eng.br

Promoção: IBAPE NACIONAL Instituto Brasileiro de Avaliações e Perícias de Engenharia

Realização: IBAPE BAHIA Instituto Brasileiro de Avaliações e Perícias de Engenharia da Bahia

1

Inferência Estatística

- Clássica
  - Regressão Tradicional
- Espacial
  - Regressão Espacial
  - Geoestatística

2

Efeitos Espaciais

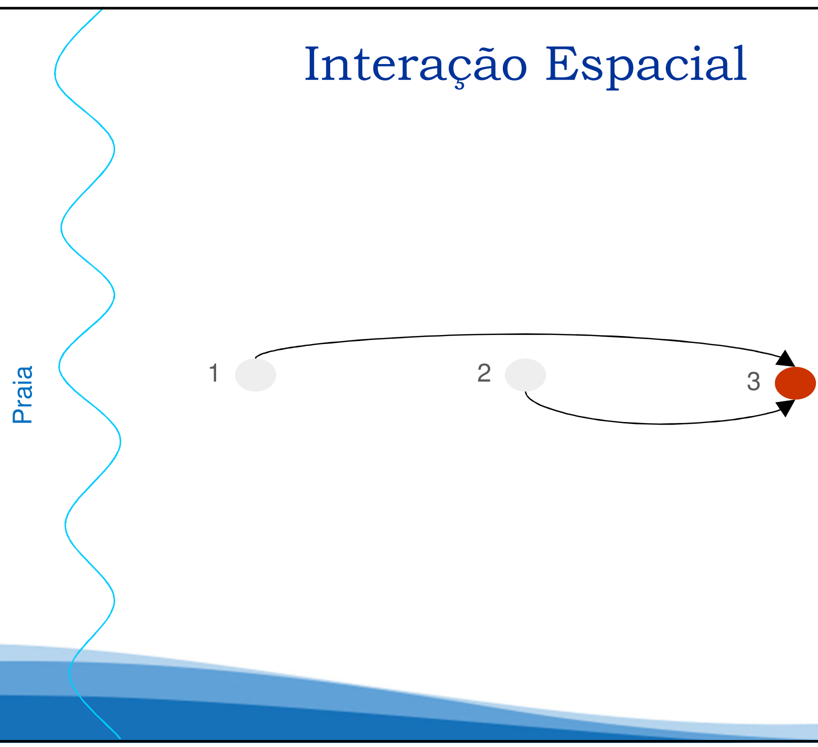
- Dependência Espacial:

"As coisas mais próximas são mais parecidas"
Waldo TOBLER

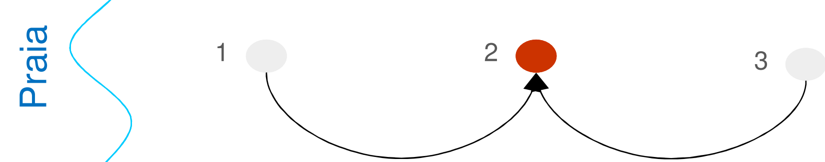
- Heterogeneidade Espacial:

Instabilidade dos parâmetros do Modelo no espaço

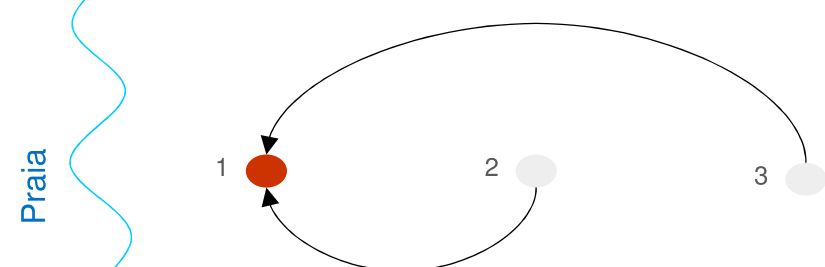
Interação Espacial



Interação Espacial



Interação Espacial



Matríz de Distancia

1 2 3

1hm 1hm

$$W^{d}=\begin{array}{c|ccc|} & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{array} \Rightarrow W^{id}=\begin{array}{c|ccc|} & 1 & 2 & 3 \\ 1 & 0 & 1 & 0.5 \\ 2 & 1 & 0 & 1 \\ 3 & 0.5 & 1 & 0 \end{array} \Rightarrow W^{s}=\begin{array}{c|ccc|} & 1 & 2 & 3 \\ 1 & 0 & 0.67 & 0.33 \\ 2 & 0.5 & 0 & 0.5 \\ 3 & 0.33 & 0.67 & 0 \end{array}$$

7

Variável Dependente Espacialmente Defasada

P1 = 100 P2 = 80 P3 = 70

1hm 1hm

$$W^{s} \times P=\begin{array}{c|ccc||c|} & 1 & 2 & 3 & \\ 1 & 0 & 0{,}67 & 0{,}33 & 100 \\ 2 & 0{,}5 & 0 & 0{,}5 & 80 \\ 3 & 0{,}33 & 0{,}67 & 0 & 70 \end{array} = \begin{vmatrix} 76{,}70 \\ 85 \\ 86{,}6 \end{vmatrix}$$

- Modelo Clássico: $P = b_o + b_1 * D$

- Modelo Defasagem Espacial: $P = b_o + b_1 * D + b_3 * W * P$

- Modelo Erro Espacial: $P = b_o + b_1 * D + b_3 * W * e$

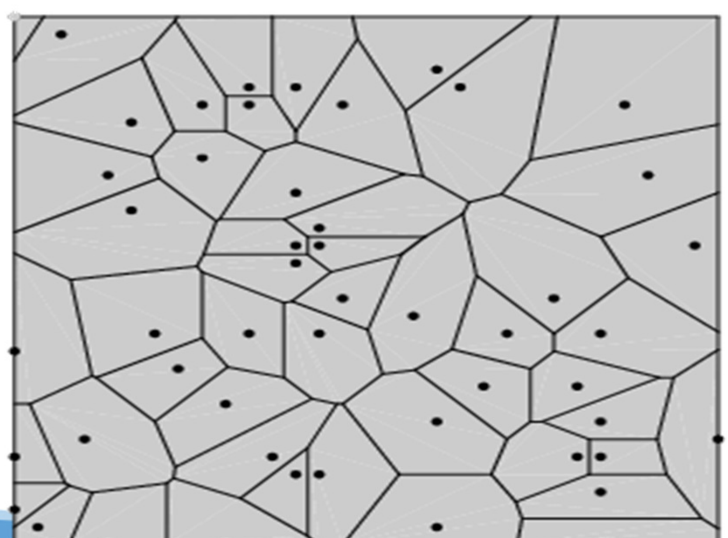
8

Dependência Espacial

- Distância (pontos)

- Contigüidade (polígonos)

Polígono de Voronoi



Matriz de Contiguidade W

Considera-se cada imóvel como uma unidade de vizinhança

- **w_{ij} = 1 se i e j são vizinhos;**
- **w_{ij} = 0 caso contrario**

A vizinhança pode ser do tipo:

- **Rook (Torre) : uma face comum**
- **Bishop (Bispo) : uma quina comum**
- **Queen (Rainha) : uma face ou quina comum**

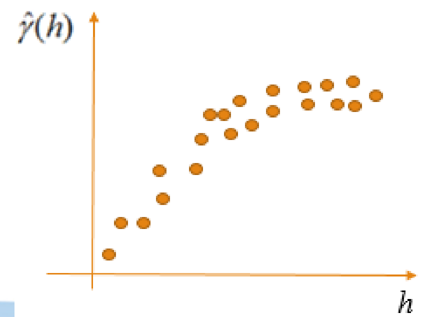
Matriz de Contiguidade tipo Rook para as Regiões do Brasil

$$W=\begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

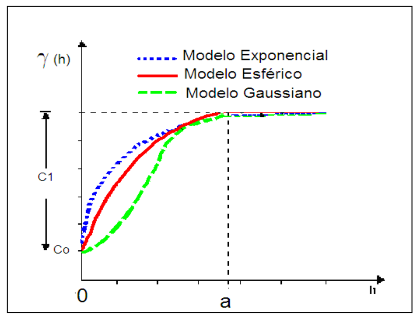
Matriz de Covariância Espacial - *Endógena*

$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [z(s_i) - z(s_{i+h})]^2$$

$$C_{ij}(h) = \begin{cases} 0 & for\ \ i = j\ and\ h = 0 \\ C_0 + C_1 - \gamma(h) & for\ \ i \neq j\ and\ 0 < h < a \\ 0 & for\ \ i \neq j\ and\ h > a \end{cases}$$



Variograma Empírico



Variograma Teórico

METODOLOGIA DE REGRESÃO ESPACIAL

1º passo: Diagnóstico da Dependencia Espacial

- Teste LM (erro)

- Teste LM (defasagem)

2º passo: Incorporação dos Efeitos Espaciais

- Modelo de Erro Espacial

- Modelo de Defasagem Espacial

14

Modelagem através da Regressão Espacial

Modelo de Erro Espacial

$Y = X\beta + u,$

onde $u = \lambda Wu + \varepsilon$ y $y \approx N(0, \sigma^2 I)$

Sendo:

W - Matriz de Pesos Espaciais

λ - coeficiente de autocorrelação espacial do erro u

Modelo de Defasagem Espacial

$Y = \rho WY + X\beta + \varepsilon,$

onde $y \approx N(0, \sigma^2 I)$

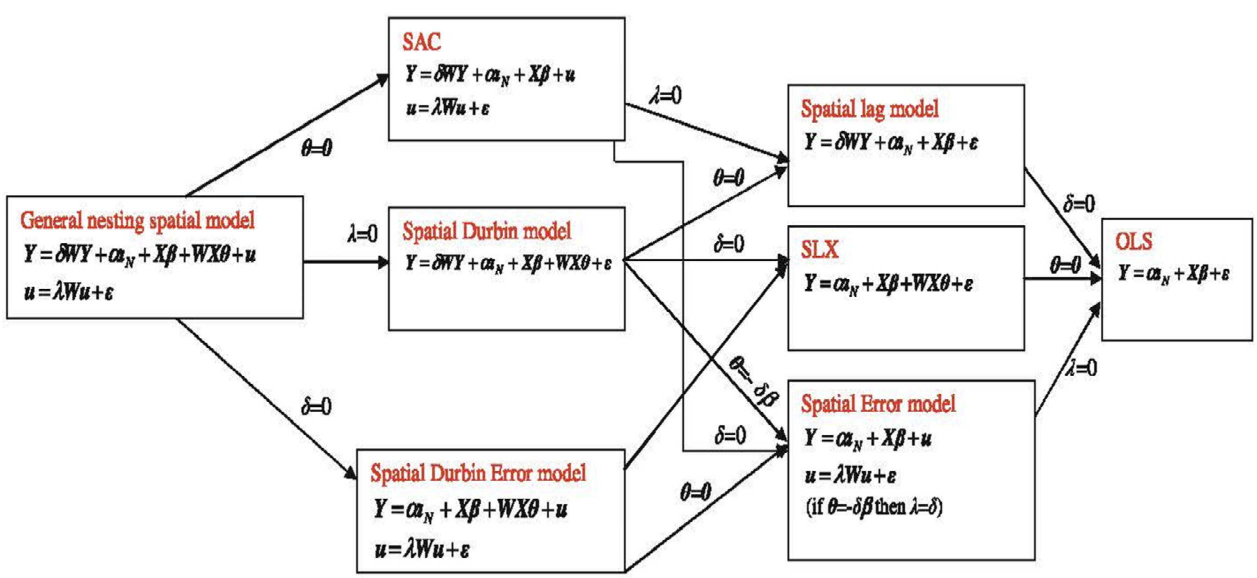
Sendo:

WY - variável dependente espacialmente defasada

ρ - coeficiente de autocorrelação espacial de WY

15

TIPOS DE MODELOS ESPACIAIS POSSÍVEIS



SAC = spatial autoregressive combined model, SLX = spatial lag of X model

Estimação dos Parâmetros: **Método da Máxima Verossimilhança**

- Funções de Log Verossimilhança: **Defasagem Espacial**

$$\ln L = -N/2 \ln 2\pi - N/2 \ln \sigma^2 + \ln\left|I - \rho W\right| - 1/2\sigma^2 \{[(I-\rho W)y - X\beta]'[(I-\rho W)y - X\beta]\}$$

- Funções de Log Verossimilhança: **Erro Espacial**

$$lnL\ (y/\lambda, \beta, \sigma^2) = -(N/2) ln 2\pi - (N/2) ln\ \sigma^2 + ln\ |I - \lambda w|\ (1/2\ \sigma^2)\ \{[(I - \lambda W)(y - X\beta)]'[(I - \lambda W)(y - X\beta)]$$

17

Diagnósticos pelos Testes dos Multiplicadores de Lagrange - LM

$$\text{LM (erro)} = \frac{[e'We/(e'/N)]^2}{[tr(W^2 + W'W)]} \overset{a}{\approx} \chi^2_{(1)}$$

$$\text{LM (defasagem)} = \{[e'Wy/(e'e/N)]^2\}/[(WX\beta)'(I - X(X'X)^{-1}X')(WX\beta)/\sigma^2] + tr(W^2 + W'W) \overset{a}{\approx} \chi^2_{(1)}$$

18

ESCOLHA DE MODELOS

I - Critério de informação de Akaike

AIC = - 2L + 2k

II - Critério de Schwartz

SC = - 2L + k * ln (n)

19

tyou.com | Cursos a Distância | GeoDa on Github

https://geodacenter.github.io

GeoDa

An Introduction to Spatial Data Analysis

Download | View on GitHub | Data | Documentation | Support | 中文

Introducing GeoDa 1.12

GeoDa is a free and open source software tool that serves as an introduction to spatial data analysis. It is designed to facilitate new insights from data analysis by exploring and modeling spatial patterns.

GeoDa was developed by Dr. Luc Anselin and his team. The program provides a user-friendly and graphical interface to methods of exploratory spatial data analysis (ESDA), such as spatial autocorrelation statistics for aggregate data (several thousand records), and basic spatial regression analysis for point and polygon data (tens of thousands of records). To work with big data in GeoDa it should first be aggregated to areal units.

Since its initial release in February 2003, GeoDa's user numbers have increased exponentially to over

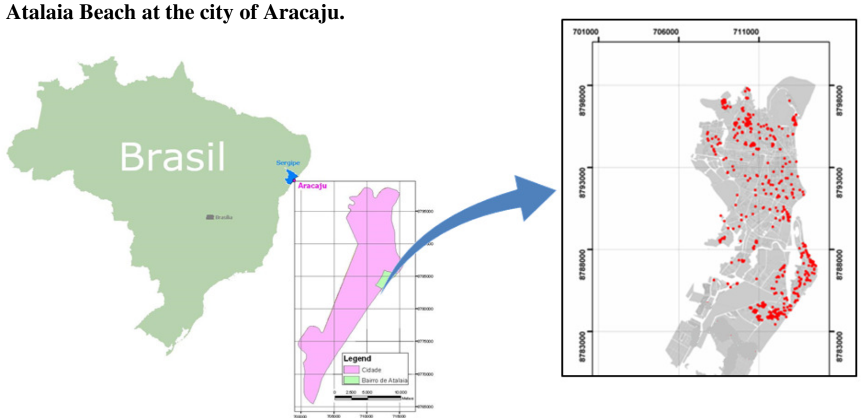
```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set                  :   Columbus Crime
Dependent Variable        :        hoval  Number of Observations:   49
Mean dependent var        :      38.4362  Number of Variables   :    3
S.D. dependent var        :      18.2767  Degrees of Freedom    :   46

R-squared                 :     0.349514  F-statistic           :     12.3582
Adjusted R-squared        :     0.321232  Prob(F-statistic)     : 5.0637e-005
Sum squared residual:            10647  Log likelihood        :    -201.368
Sigma-square              :      231.457  Akaike info criterion :     408.735
S.E. of regression        :      15.2137  Schwarz criterion     :     414.411
Sigma-square ML           :      217.286
S.E of regression ML:            14.7406

-----------------------------------------------------------------------------
       Variable       Coefficient      Std.Error    t-Statistic    Probability
-----------------------------------------------------------------------------
          CONSTANT       46.4282        13.1918       3.51948        0.00099
               inc      0.628984        0.53591       1.17367        0.24657
             crime     -0.484889       0.182673      -2.65441        0.01087
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   12.537555
TEST ON NORMALITY OF ERRORS
TEST                   DF             VALUE                PROB
Jarque-Bera             2              39.7062             0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                   DF             VALUE                PROB
Breusch-Pagan test      2               5.7668             0.05594
Koenker-Bassett test    2               2.2700             0.32142

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : WColumbusRook
   (row-standardized weights)
TEST                              MI/DF         VALUE            PROB
Moran's I (error)                 0.2071         2.5040          0.01228
Lagrange Multiplier (lag)           1            1.8195          0.17737
Robust LM (lag)                     1            0.6978          0.40351
Lagrange Multiplier (error)         1            3.9929          0.04569
Robust LM (error)                   1            2.8713          0.09017
```

```
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set                  : Columbus Crime
Spatial Weight            : WColumbusRook
Dependent Variable        :       hoval  Number of Observations:   49
Mean dependent var        :     38.4362  Number of Variables   :    4
S.D. dependent var        :     18.2767  Degrees of Freedom    :   45
Lag coeff.    (Rho)       :     0.23063

R-squared                 :    0.382337  Log likelihood        :    -200.439
Sq. Correlation           : -            Akaike info criterion :     408.879
Sigma-square              :     206.322  Schwarz criterion     :     416.446
S.E of regression         :     14.3639

-----------------------------------------------------------------------------
       Variable       Coefficient      Std.Error       z-value     Probability
-----------------------------------------------------------------------------
           W_hoval       0.23063       0.154956        1.48836       0.13666
          CONSTANT       37.2681        13.9456        2.67239       0.00753
               inc      0.563156       0.513164        1.09742       0.27246
             crime      -0.45102       0.177697       -2.53815       0.01114
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                      DF       VALUE          PROB
Breusch-Pagan test                        2         5.9099        0.05208

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : WColumbusRook
TEST                                      DF       VALUE          PROB
Likelihood Ratio Test                     1         1.8567        0.17300
```

```
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set              : Columbus Crime
Spatial Weight        : WColumbusRook
Dependent Variable    :        hoval  Number of Observations:   49
Mean dependent var    :    38.436224  Number of Variables   :    3
S.D. dependent var    :    18.276669  Degrees of Freedom    :   46
Lag coeff.  (Lambda)  :     0.390498

R-squared             :     0.428509  R-squared (BUSE)      : -
Sq. Correlation       : -             Log likelihood        : -199.232493
Sigma-square          :      190.899  Akaike info criterion :     404.465
S.E of regression     :      13.8166  Schwarz criterion     :      410.14

-----------------------------------------------------------------------------
       Variable        Coefficient      Std.Error        z-value     Probability
-----------------------------------------------------------------------------
          CONSTANT       48.0083         12.3126         3.89912      0.00010
               inc      0.711532        0.497923           1.429      0.15300
             crime     -0.559458        0.178773        -3.12943      0.00175
            LAMBDA      0.390498        0.157587         2.47798      0.01321
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                        DF      VALUE        PROB
Breusch-Pagan test                          2         5.6443     0.05948

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : WColumbusRook
TEST                                        DF      VALUE        PROB
Likelihood Ratio Test                       1         4.2705     0.03878
```

Empirical Evidence using Spatial Covaviance Matrix -WC

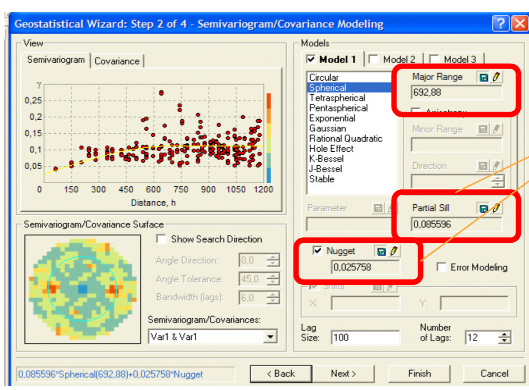
❖ Dataset used for illustration consists of 479 land lots transactions distributed along Atalaia Beach at the city of Aracaju.



Available Information

- ❖ Land Lot Area (AR): Log-transformed;
- ❖ Sector (SE): Proxy Variable (average income of the head of household) Log-transformed;
- ❖ Topography (TO): Dummy indicating (plan)
- ❖ Location (SI): Dummy variable (corner)
- ❖ Paving: Dummy variable (pave)
- ❖ Frontage (FR): Quantitative variable Land-use Coefficient (CA): Variable interval that identifies the relationship between the land area and the area that can be constructed - Log-transformed;
- ❖ Axis (EI): Ordinal variable 3 - principal axis, 2 - secondary axis and 1 for the other cases;
- ❖ Nature of the event.

Variogram



(0.025758 / 0.111354)% = 23%

- ❖ **23% of the total variability of the residuals is due to the random component;**
- ❖ **77% are explained by the presence of the spatial autocorrelation in the residuals.**

SPATIAL DEPENDENCE DIAGNOSE

	LMRobust Error		LMRobust lag	
WD	8.9330	0.0030	8.9110	0.0030
WG	6.0110	0.0140	109.7527	0.0000

	WD		WC	
	SAR	SEM	SAR	SEM
AIC	412,91	412,69	358,54	324,16
SC	452,62	458,58	414,42	370,05

Literatura Recente

- ALMEIDA, Eduardo. Econometria Espacial Aplicada. Campinas: Alínea, 2012.
- ANSELIN, Luc e REY, J. Sergio. Modern Spatial Econometrics in Practice. Chicago: GeoDa Press LLC, 2014.
- ARBIA, Giusepe. A Primer for Spatial Econometrics With Applications in R. London: Palgrave Macmillan, 2014.
- BIVAND, Edzer Pebesma, GOMEZ-RUBIO, Virgilio. Applied spatial data analysis with R, 2. ed. New York: Springer, 2013.
- REY, S.J. e ANSELIN, L. PySAL: A Python Library of Spatial Analytical Methods. [S.I.]: Review of Regional Studies, 2007.

Obrigado

rubens@dantas.eng.br

www.dantas.eng.br