

Ordinary least squares regression | Geographically Weighted Regression

O OLS é um modelo global de regressão, enquanto o GWR é um modelo local (permite que as relações que estão a ser modeladas variem no espaço).

Exemplo: modelar as chamadas de emergência

0. Desactivar os processos em fundo (Geoprocessing>Geoprocessing Options).



1. Testar a possibilidade de as chamadas de emergência aumentarem nas áreas com mais população

1.1. Abrir a ferramenta OLS;

1.2. Definir o id;

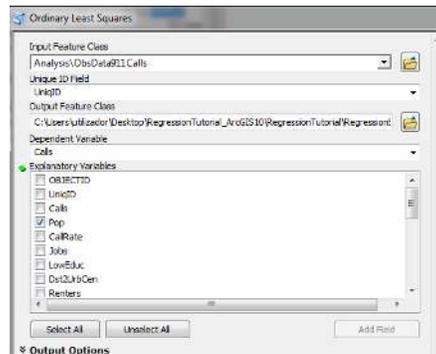
1.3. Variável dependente (neste caso chamadas de emergência);

1.4. Identificar as variáveis independentes (neste caso apenas a pop);

1.5. O OLS mostra o desempenho do modelo, com a diferença entre os valores observados e estimados;

1.6. Quando o modelo está adequado os desvios entre o observado e o esperado apresentam-se distribuídos aleatoriamente. Neste caso há um cluster de áreas em que o observado é superior ao previsto;

1.7. O outro output mostra o relatório com as estatísticas do modelo. Com o R2 de 0.39, vê-mos que este modelo explica 39%



2. Testar a possibilidade de as chamadas emergência serem explicadas pela pop, jobs, baixa educação, distância ao centro

2.1. Repetir o processo, acrescentando estas 4 variáveis como independentes;

2.2. O novo modelo explica 0.83;

2.3. A diferença entre observados e esperados aparece distribuída de forma mais homogênea, sem cluster;

2.4. Para confirmar pode-se correr o spatial autocorrelation Moran I (Input Feature Class: OLS do modelo com as 4 variáveis; Input Field: StdResid; Generate Report: checked ON; Conceptualization of Spatial Relationships: Inverse Distance; Distance Method: Euclidean Distance; Standardization: ROW)

2.5. O valor do I Moran é baixo e não é estatisticamente significativo, portanto os resíduos do OLS estão distribuídos de forma aleatória;

3. Validar o modelo

3.1. Os coeficientes de cada variável no modelo devem indicar a relação esperada (Summary of OLS results>coefficient) ;

- 3.2. As variáveis explicativas não devem ser redundantes (Summary of OLS results>VIF: se alguma variável é superior a 7,5 é pq é redundante; deve-se começar a remover as que tiverem VIF superiores);

The screenshot shows the following OLS results:

Variable	Coefficient	StdError	t-Statistic	Probability	Robust_SE	Robust_t	Robust_Pe	VIF [1]
Intercept	86.082979	0.875151	98.363521	0.000000*	0.813152	105.863324	0.000000*	-----
NVENTACCID	-110.520016	12.213013	-9.049366	0.000000*	14.544464	-7.598769	0.000000*	2.351229
NBUICIDE	-138.221155	18.180324	-7.602788	0.000000*	29.800993	-4.638139	0.000011*	1.556498
NLUNGCANC	-47.045741	12.076316	-3.895703	0.000172*	13.536130	-3.475568	0.000732*	1.051207
NDIABETES	-33.429850	13.805975	-2.421405	0.017044*	14.732174	-2.269173	0.025148*	1.400358
NBELOWPOV	-14.408804	3.633873	-3.965137	0.000134*	4.125643	-3.492499	0.000692*	3.232363

OLS Diagnostics:

Number of Observations:	119	Number of Variables:	6
Degrees of Freedom:	113	Akaike's Information Criterion (AIC) [2]:	524.97620
Multiple R-Squared [2]:	0.870551	Adjusted R-Squared [2]:	0.864823
Joint F-Statistic [3]:	151.985705	Prob(>F), (5,112) degrees of freedom:	0.000000*
Joint Wald Statistic [4]:	496.057428	Prob(>chi-squared), (5) degrees of freedom:	0.000000*
Koenker (BP) Statistic [5]:	21.590491	Prob(>chi-squared), (5) degrees of freedom:	0.000026*
Jarque-Bera Statistic [6]:	4.207198	Prob(>chi-squared), (2) degrees of freedom:	0.122017

Numbered callouts in the image point to:

- 1: Coefficients have the expected sign. (points to the negative coefficients)
- 2: No redundancy among explanatory variables. (points to the VIF column)
- 3: Coefficients are statistically significant. (points to the Probability column)
- 4: Residuals are normally distributed. (points to the Koenker (BP) Statistic)
- 5: Strong Adjusted R-Square value. (points to the Adjusted R-Squared value)
- 6: Residuals are not spatially autocorrelated. (points to the Jarque-Bera Statistic)

- 3.3. Verificar se todas as variáveis explicativas são significativas

(Summary of OLS results>Probability; Summary of OLS results>Robust: se têm "*" é pq são significativas). Se alguma variável não é significativa em principio pode ser retirada;

- 3.4. Garantir que o Jarque-Bera teste **não** é significativo. Se for positivo o modelo encontra-se enviesado;
- 3.5. Avaliar o desempenho do modelo (Summary of OLS results>Akaike's information criteria). Este critério é utilizado para comparar modelos, o que tiver menores valores é o mais adequado;
- 3.6. Os resíduos devem estar distribuídos espacialmente de forma aleatória (através do I. Morin).

4. GWR

- 4.1. Qd o teste Koenker é estatisticamente significativo as variáveis independentes (uma ou todas) são não-estacionárias, e o GWR pode melhorar o desempenho do modelo
- 4.2. Correr o GWR utilizando as mesmas variáveis independentes e dependente; Kernal type: ADAPTIVE; o Bandwidth method: AICs.
- 4.3. A capacidade explicativa do modelo aumentou para 86% e o Akaike criteria diminuiu;
- 4.4. Verificar se os std resíduos apresentam uma distribuição espacial aleatória;
- 4.5. Gera diversas colunas com os coeficientes para cada variável explicativa. Resultados mais elevados nestes coeficientes indicam áreas onde a capacidade explicativa da variável em causa é superior ou inferior.

5. GWR predictors

- 5.1. Não analisei esta parte