

Análise Espacial da Mortalidade Infantil em Porto Alegre-RS

Processos pontuais espaciais

Silvia Shimakura

Departamento de Estatística
Laboratório de Estatística e Geoinformação

07 de novembro, 2025



CONTEÚDO

- ▶ Problema
- ▶ Questões científicas
- ▶ Análise
 - ▶ de contagens
 - ▶ de pontos
- ▶ Aplicação
- ▶ Conclusões

PROBLEMA

► O quê?

- Óbitos neo-natais - NEO: 183 casos
- Óbitos pós-neonatais - PÓS: 116 casos

► Quando?

- Período: ano de 1998

► Onde?

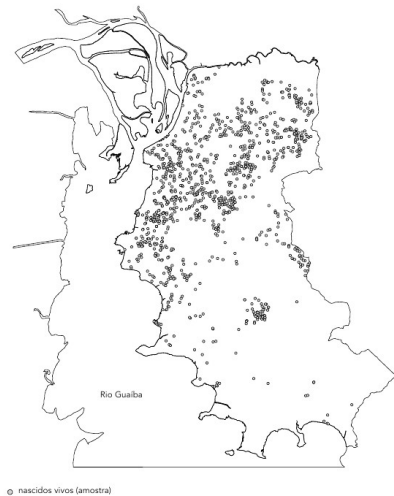
- Município de Porto Alegre - RS

► Covariáveis:

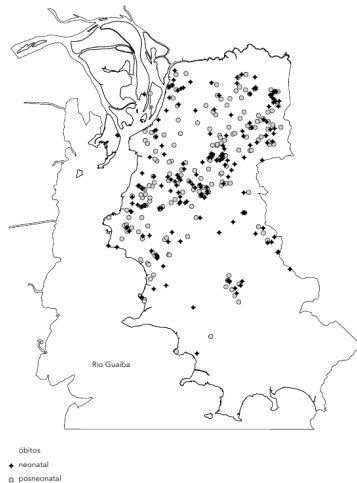
- bebê: Sexo, Peso ao nascer
- maternas: Idade, grau de instrução
- gestação: Idade gestacional, Tipo de gravidez, Tipo de parto
- coordenadas geográficas da residência

DISTRIBUIÇÃO ESPACIAL DOS DADOS

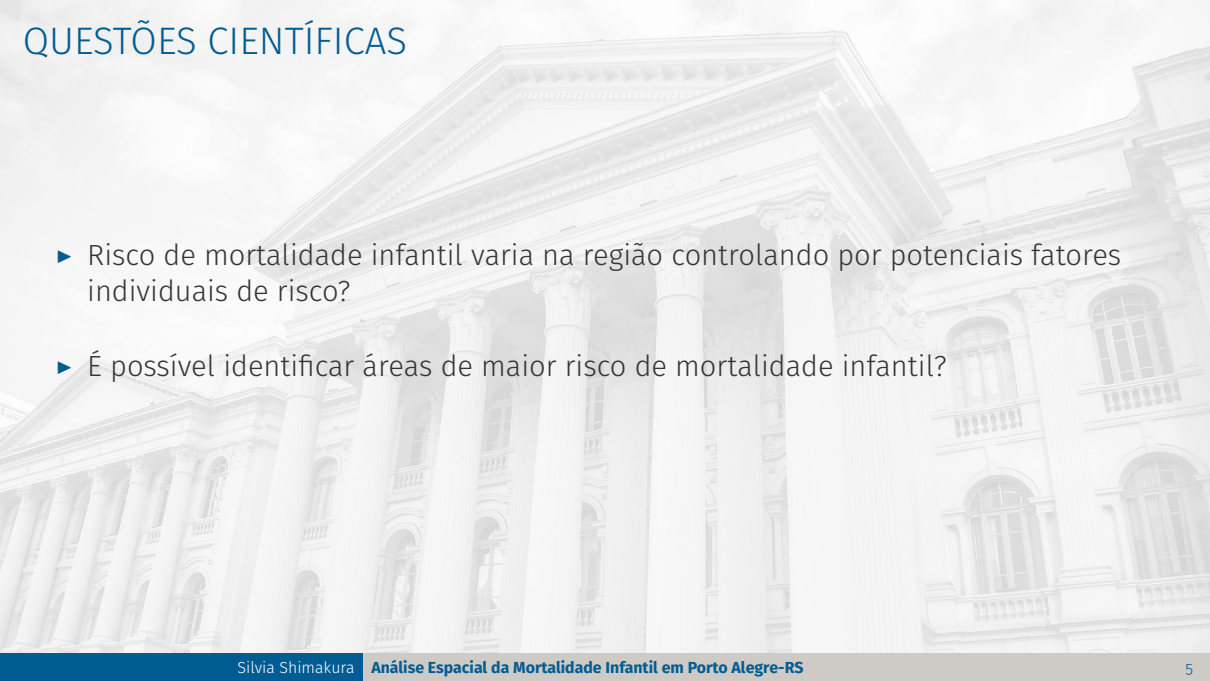
Distribuição da amostra de nascidos vivos. Porto Alegre, 1998.



Distribuição da localização dos óbitos, neonatais e pós-neonatais. Porto Alegre, 1998.



QUESTÕES CIENTÍFICAS

A faded background image of a grand classical building with a prominent portico supported by tall columns. The building has multiple stories with arched windows and decorative moldings.

- ▶ Risco de mortalidade infantil varia na região controlando por potenciais fatores individuais de risco?
- ▶ É possível identificar áreas de maior risco de mortalidade infantil?

ANÁLISE DE CONTAGENS

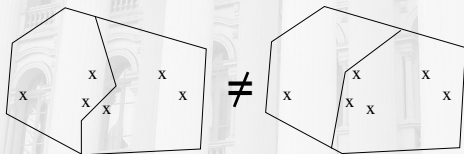
- ▶ Casos agregados em contagens $(Y_1, Y_2, \dots, Y_k)$ dentro de k estratos (ex: bairros)
- ▶ Suposição:
 - ▶ $Y_i \sim \text{Poisson}(\lambda_i)$
$$\lambda_i = E_i \times \theta_i$$
 - E_i : número esperado de casos (offset baseado na população em risco)
 - θ_i : risco relativo no estrato i .
 - ▶ $\log \lambda_i = \log E_i + \log \theta_i = \log E_i + \beta x_i + u_i$,
 - ▶ $(u_1, u_2, \dots, u_k)$ efeitos aleatórios espacialmente estruturados (ex: Markov Random Field)

Desvantagens

- ▶ Perda de informação ao agregar os dados em contagens.
- ▶ Métodos não fazem distinção entre os padrões:

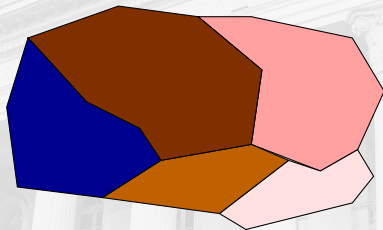


- ▶ Modelos carecem de interpretação espacial apropriada → áreas definidas arbitrariamente



Desvantagens

- Superfície estimada é discreta.



- Potencial falácia ecológica: associação entre variáveis em nível individual não pode ser inferida a partir de associações observadas em nível agregado.

ANÁLISE DE PONTOS: PROCESSOS PONTUAIS ESPACIAIS

- ▶ Definição da medida de risco.
- ▶ Comparação da distribuição espacial:

casos x controles

- ▶ **Suposições:**

- ▶ Localizações dos casos $\sim \text{Poisson}(\lambda_1(s))$
- ▶ Localizações dos controles $\sim \text{Poisson}(\lambda_2(s))$

- ▶ **Objetivo:** Estimar $\rho(s)$

$$\rho(s) = \log \left\{ \frac{\lambda_1(s)}{\lambda_2(s)} \right\}$$

- ▶ s : coordenada espacial da observação.

Regressão Binária (Kelsall & Diggle, 1995)

- Atribui-se ao indivíduo i :

$$y_i = \begin{cases} 1 & \text{se } i \text{ for um caso,} \\ 0 & \text{c.c.} \end{cases}$$

- Assume-se que $(Y_i | s_i = s) \sim \text{Bernoulli}(p(s))$

$$p(s) = \frac{\lambda_1(s)}{\lambda_1(s) + \lambda_2(s)}$$

↓

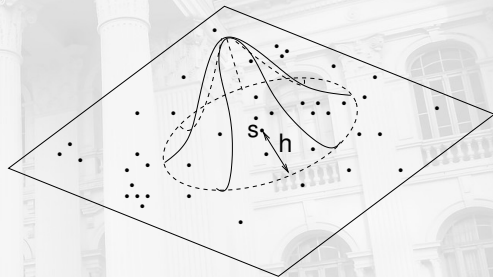
$$\text{logit}\{p(s)\} = \log\left(\frac{p(s)}{1-p(s)}\right) = \log\left(\frac{\lambda_1(s)}{\lambda_2(s)}\right) = \rho(s)$$

► Estimador Kernel

Média ponderada de y_i com pesos definidos em termos das distâncias de s para s_i .

Kernel Gaussiano:

$$\hat{p}(s) = \frac{\sum_{i=1}^n K(s_i - s; h) y_i}{\sum_{i=1}^n K(s_i - s; h)}$$



► Estimador Kernel

$$\hat{p}(s) = \frac{\sum_{i=1}^n K(s_i - s; h) y_i}{\sum_{i=1}^n K(s_i - s; h)}$$



covariáveis???

► Modelo Aditivo Generalizado (Kelsall & Diggle, 1998)

$$\text{logit}\{p(s, \mathbf{x})\} = \mathbf{x}\boldsymbol{\beta} + g(s)$$

- $\boldsymbol{\beta}$ estimado por máxima verossimilhança para $g(\cdot)$ fixo.
- $g(\cdot)$ estimado por regressão kernel ponderada para $\boldsymbol{\beta}$ fixo com validação cruzada para seleção de h .

SIGNIFICÂNCIA

A superfície estimada varia significativamente na região?

- ▶ $H_0 : g(s) = c$ (risco constante no espaço)

Construção de regiões de tolerância e Teste Monte Carlo.

1. Gerar m conjuntos de dados sob H_0
2. Estimar $g_l, l = 1, \dots, m$, e calcular

$$t_l = \frac{\sum_{i=1}^n (\hat{g}_l(s_i) - \bar{g}_l)^2}{n}$$

3. Construir superfície p-valor. Para cada s

$$\text{p-valor}(s) = \text{Prop}(\hat{g}_l(s) < \hat{g}_{obs}(s))$$

4. Adicionar contornos 0.05 e 0.95 da superfície ao mapa de $\hat{g}_{obs}$
5. Denotar $k = \# (t_l > t_{obs}) \rightarrow p = \frac{(k+1)}{(m+1)}$ (p-valor do teste).

RESULTADOS

NEONATAL ($h_{vc} = 3km$)

fator	efeito	se	p-valor
Sexo (Masc)	-.367	.2713	.176
Peso	-.002	.0002	.000
Idade	-.013	.0197	.506
Instrução (10. grau inc.)	.072	.2753	.794
Idade gest (Prematuro)	1.17	.3737	.002
Gravidez (Gemelar)	-.201	.6558	.760
Parto (Vaginal)	-.532	.2838	.061

PÓS-NEONATAL ($h_{vc} = 1.9km$)

fator	efeito	se	p-valor
Sexo (Masc)	.300	.2303	.194
Peso	-.001	.0002	.000
Idade	-.017	.0171	.317
Instrução (10. grau inc.)	.803	.2422	.001
Idade gest (Prematuro)	.562	.3654	.125
Gravidez (Gemelar)	.973	.5451	.074
Parto (Vaginal)	.059	.2523	.816

MAPAS DE RISCO ESPACIAL

Mapas de risco para a mortalidade infantil, controlando para fatores individuais – largura de banda estimada por validação cruzada. Porto Alegre, 1998.

Figura 3a: Neonatal - cv ($p = 0,012$)

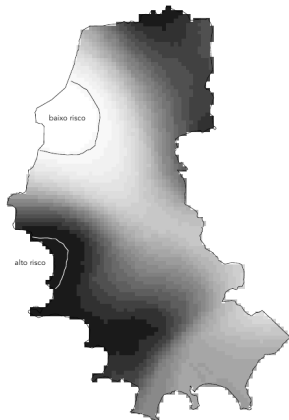


Figura 3b: Pós-neonatal - cv ($p = 0,084$)



CONCLUSÕES

Quanto a análise do efeito de covariáveis

- ▶ Risco de mortalidade infantil é elevado entre crianças:
 - ▶ com baixo peso no nascimento: $RR=1.002$
 - ▶ prematuras para **NEO**: $RR=3.2$
 - ▶ cujas mães tem baixo nível instrucional para **PÓS**: $RR=2.2$

Quanto a análise do efeito espacial

- ▶ **NEO**: significativo com risco elevado à Sudoeste.
- ▶ **PÓS**: marginalmente significativo com risco elevado à Noroeste, Leste e Oeste.

Quanto ao método

- ▶ Desvantagens:

- ▶ Requer controles.
- ▶ Requer métodos computacionalmente intensivos.

- ▶ Vantagens:

- ▶ Interpretação espacial natural do modelo.
- ▶ Permite fácil incorporação de covariáveis.
- ▶ Permite testar significância de variações espaciais do risco.
- ▶ Permite identificar regiões de riscos extremos.
- ▶ Implementado no R como pacote (não-oficial) spatgam.

Referências

- ▶ Diggle, P. *Statistical Analysis of Spatial Point Patterns*. London: Academic Press, 2nd Ed. 2004.
- ▶ Kelsall, J; Diggle, P. Spatial variation in risk of disease: A nonparametric binary regression approach. *Applied Statistics*, 47:559-573, 1998.
- ▶ KELSALL, J. E. Kernel Smoothing for Application in Environmental Epidemiology. Ph.D. Thesis, Lancaster: Department of Mathematics and Statistics, Lancaster University, 1992.
- ▶ KELSALL, J. E., DIGGLE, P. J. Kernel estimation of relative risk. *Bernoulli*, 1:3-16, 1995.
- ▶ KELSALL, J. E., DIGGLE, P. J. Non-parametric estimation of spatial variation in relative risk. *Statistics in Medicine*, 14:2335-2342, 1995.