

Estatística Espacial Geoestatística II

Raquel Menezes

Departamento de Matemática
Universidade do Minho

Outubro de 2023



Universidade do Minho
Escola de Ciências

Krigagem - predição espacial linear ótima

Dependendo dos pressupostos sobre $\mu(\cdot)$, tem-se:

- **Krigagem simples:** $\mu(\mathbf{x})$ é conhecido pouco realista
- **Krigagem ordinária:** $\mu(\mathbf{x}) = \mu$ é desconhecido
- **Krigagem com tendência externa**¹: $\mu(\mathbf{x}) = \sum_{j=0}^p f_j(\mathbf{x})\beta_j$ o mais realista

onde $\beta = (\beta_0, \dots, \beta_p)^t \in \mathbb{R}^{p+1}$ desconhecido e $\{f_j(\cdot) : j = 0, \dots, p\}$ são funções explicativas conhecidas (considere-se $f_0(\cdot) \equiv 1$)

Notas:

- Pesos $\lambda = (\lambda_0, \dots, \lambda_n)^t$ são obtidos como soluções das **equações de kriging**.
- Para além da predição $\hat{Y}(\mathbf{x}_0)$, também se obtém uma estimativa para $E[(\hat{Y}(\mathbf{x}_0) - Y(\mathbf{x}_0))^2]$, referida como **variância de kriging** $\sigma_K^2(\mathbf{x}_0)$.

¹Referido como “krigagem universal” se as covariáveis se restringem às próprias coordenadas.

Predição espacial via *kriging*

Na predição espacial, uma questão básica é, **dado um conjunto de n observações de $Y(\cdot)$ nos pontos \mathbf{x}_i , $i = 1, \dots, n$, qual é o valor obtido por $Y(\cdot)$ em \mathbf{x}_0** , onde os dados não estão disponíveis?
 \mathbf{x}_0 vai ser o ponto novo onde queremos prever

- Tendo em mente que algumas observações nas proximidades de \mathbf{x}_0 estão mais relacionadas do que outras ao valor verdadeiro no ponto \mathbf{x}_0 , pode-se adotar uma **média ponderada**:

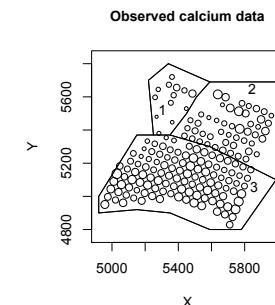
$$\hat{Y}(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i) + \lambda_0$$

onde λ_i são **funções de ponderação**, que determinam a **importância dos locais vizinhos** de \mathbf{x}_0

- Normalmente, a estimação **de um variograma válido desempenha aqui um papel decisivo**, já que é habitualmente usado para encontrar a solução ótima de λ_i . O método é chamado de **krigagem ordinária** e pode ser aplicado a um processo intrinsecamente estacionário.

Exemplo 3: Conteúdo de cálcio em amostras de solo

- Amostras de solo retiradas da camada de 0-20 cm em 178 locais dentro de uma determinada área de estudo dividida em três subáreas.
- **A primeira região** é normalmente inundada durante a estação das chuvas e não é usada como área experimental. Os níveis de cálcio representariam o conteúdo natural da região. **A segunda região** já recebeu fertilizantes há algum tempo e costuma ser ocupada por arrozais. **A terceira região** recebeu fertilizantes recentemente e é frequentemente utilizada como área experimental.



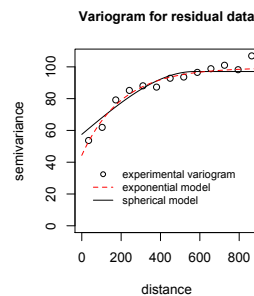
Exemplo 3: Conteúdo de cálcio

Area 1 é a área de referência
pq não aparece na fórmula

$$\text{Calcium}(x) = \alpha + \beta_1 \text{IF}(\text{Area2}(x)) + \beta_2 \text{IF}(\text{Area3}(x)) + \beta_3 \text{Latitude}(x) + S(x) + \epsilon(x)$$

	estimate	s.e.	p-value
α	163.69	37.07	<0.001
β_1	5.96	3.17	0.061
β_2	7.19	4.18	0.086
β_3	-0.023	0.007	<0.001

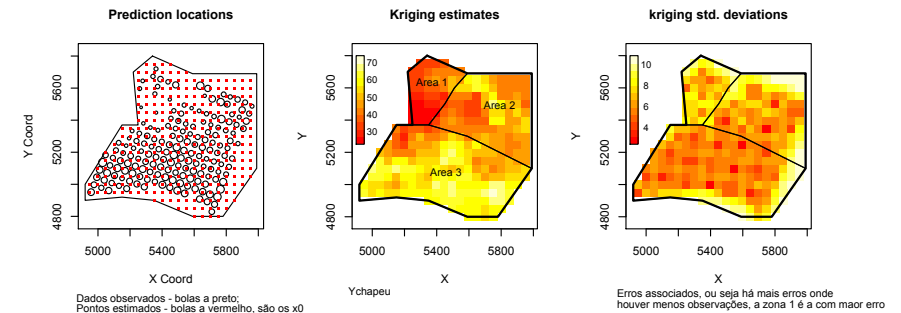
a latitude(beta3) é relevante, a área se considerarmos um alfa até 5% não é significativa



Nota: Para a modelação de dados, iremos considerar um modelo exponencial com $\hat{\tau}^2=44.32$, $\hat{\sigma}^2=55.2$ e $\hat{\phi}=202.5$.

5 / 8

Exemplo 3: Cálcio – kriging com tendência externa



- Considerando a superfície/tendência estimada anteriormente, prosseguimos com a previsão espacial sobre uma grelha (pontos vermelhos no painel esquerdo).
- As estimativas de krigagem, no painel central, confirmam que a Area 3 é a que apresenta os maiores valores de cálcio, enquanto a Area 1 é a que apresenta os menores valores.
- Desvios padrão de previsão (um indicador de erro), no painel direito, são maiores em locais com menos pontos amostrados.

6 / 8

Diagnóstico do modelo – validação-cruzada

- Suponha que um variograma $\gamma_{\hat{\theta}}$ foi ajustado aos dados $\{Y(\mathbf{x}_i) : i = 1, \dots, n\}$, e que conhecemos algum método de predição baseado em $\gamma_{\hat{\theta}}$ (por exemplo, um método de kriging).
- Uma forma de diagnosticar a qualidade do ajuste passa por fazer a **validação cruzada**:
 - Estima-se a observação $Y(\mathbf{x})$ em cada \mathbf{x}_i à custa dos dados vizinhos $Y_j = Y(\mathbf{x}_j)$, $j \neq i$, **excluindo-se a si próprio**.
 - Para cada \mathbf{x}_i , obtem-se uma predição $\hat{Y}_{-i} = \hat{Y}(\mathbf{x}_i)$ e a respectiva variância σ_{-i}^2 .
 - Calcula-se um erro predição $PE_i = Y_i - \hat{Y}_{-i}$.

7 / 8

Validação-cruzada (cont.)

A proximidade dos valores estimados \hat{Y}_{-i} e verdadeiros Y_i pode avaliada por:

- erro quadrático médio (EQM)**

$$EQM = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_{-i})^2, \quad (1)$$

- erro quadrático médio standardizado (EQMS)**

$$EQMS = \frac{1}{n} \sum_{i=1}^n \frac{(Y_i - \hat{Y}_{-i})^2}{\sigma_{-i}^2}. \quad (2)$$

Para avaliar a qualidade de ajuste do variograma $\gamma_{\hat{\theta}}$:

- a média em (1) deve ser aprox. 0 e a raiz da expressão dada em (2) deve ser aprox. 1.
- Alternativamente, pode-se analisar o histograma dos erros standardizados e confirmar que seguem (aprox.) uma normal standard.

8 / 8