

Fundamentos de Geoestatística

Dr. Rachid Muleia, PhD in Statistics

Universidade Eduardo Mondlane
Faculdade de Ciências
Departamento de Matemática e Informática

Fevereiro de 2023

Programa Temático

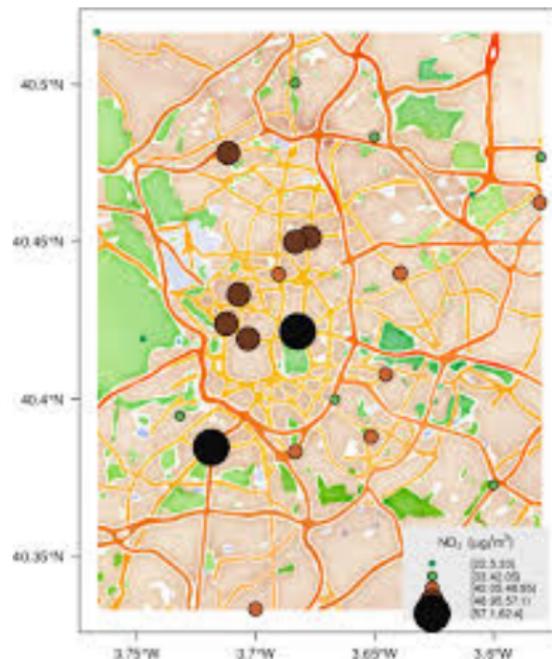
- 1 Introdução à Geoestatística
- 2 Abordagem probabilística e condições de estacionaridade
- 3 Análise exploratória de dados espaciais
- 4 Análise variográfica e estudo da anisotropia
- 5 Métodos de interpolação espacial

Avaliações

- A presença do estudante nas minhas aulas é de carácter “obrigatório”. O estudante não deve ter mais que 25% de ausências.
- Ao longo do semestre faremos dois testes. A resolução das fichas práticas também é uma avaliação.
- Para além dos testes, iremos ter 01 trabalho prático em grupo

$$\text{Nota de Frequência} = 0.3 \times \text{TE I} + 0.5 \times \text{TE II} + 0.2 \times \text{TP}$$

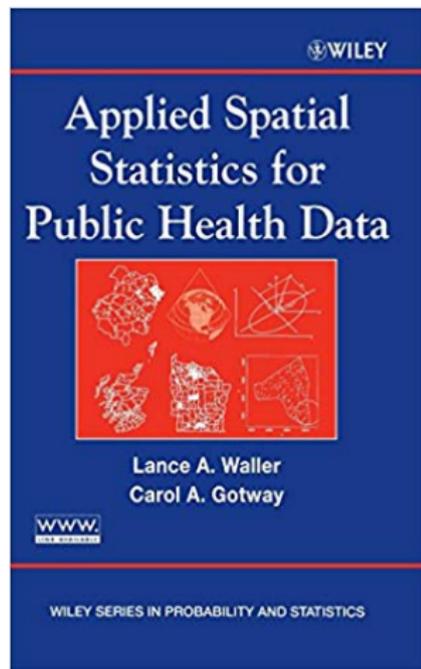
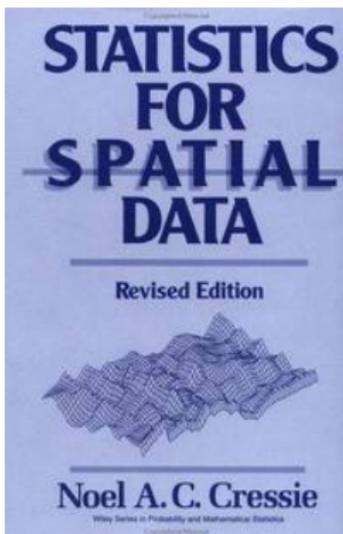
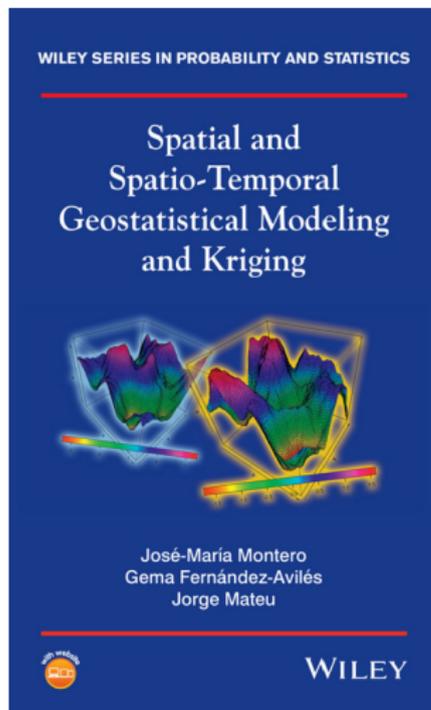
Ferramentas de análise



Porquê R??

- Software estatístico gratuito. É também uma linguagem de programação, e isto torna o programa flexível;
- Muito popular na comunidade académica. Actualmente conta com **2 milhões de usuários**
- Encontra-se disponível na internet para download em <https://cran.r-project.org/bin/windows/base/>
- Actualmente é acompanhado do Rstudio que, também, é grátis e pode ser baixado a partir da internet em <https://www.rstudio.com/products/rstudio/download/>

Referências recomendadas



Introdução à Geoestatística

- Geo-estatística lida com a caracterização de dados espaciais e espacio-temporal;
- É uma ciência que surge da necessidade de modelação de recursos geológicos- caracterização espacial da concentração de metais em jazigos;
- Primeiros desenvolvimentos foram dados na África do Sul por D.G. Krige, um engenheiro de minas, e H.S. Sichel, um estatístico, por volta de 1950;
- Posteriormente, Georges Matheron, um engenheiro francês, expandiu o trabalho iniciado por D.G. Krige, dando origem a modelos de krigagem;

Porquê geo-estatística

- Maior parte dos fenómenos/grandezas variam ao longo do espaço, e a geo-estatística ajuda a perceber e quantificar a variabilidade espacial de um dado fenómeno;
- Estatística clássica assume independência das observações, enquanto que dados geo-referenciados apresentam dependência espacial;
- Observações mais próximas apresentam maior dependência espacial, e as mais afastadas apresentam menor dependência espacial;
- Ignorar a dependência espacial pode condicionar a validade das análises estatística.

Processo espacial-Definições básicas

Variável regionalizada:

→ Toda variável distribuída no espaço é tida como “regionalizada” ou “espacial”. Por exemplo:

1. Concentração de fósforo nas machambas do vale do infulene.
2. Níveis de precipitação na cidade de Maputo.

→ Uma variável regionalizada pode ser vista como uma função $f(s)$ que assume um dado valor em cada ponto s , num espaço propriamente definido.

Função aleatória

→ Seja $Z(s_i)$ o valor observado de uma variável de interesse em uma posição s_i . Este valor pode ser tido como uma realização particular de uma variável aleatória $Z(s_i)$ em um ponto s_i .

→ O conjunto de variáveis aleatórias $\{Z(s) : s \in R\}$, onde R é uma região de interesse (ex: campo de plantação), é tido como uma **função aleatória**. Por exemplo: $\{Z(s) : s \in R\}$, onde:

- $Z(s)$ = valor observado no ponto $s = (x_1, x_2)$ ou $s = (x_1, x_2, x_3)$
- R = Conjunto de todos pontos sobre a área de interesse

Tipos de dados

Seja $s \in R$ uma localização genérica num espaço euclidiano de dimensão d e seja $\{Z(s) : s \in R\}$ função aleatória espacial, em que Z denota o atributo de interesse.

Dados geo-estatísticos (dados de superfície)

Dados geo-estatísticos surgem quando o domínio em estudo é um conjunto fixo e contínuo:

- $Z(s)$ pode ser observado em qualquer ponto do domínio D ;
- O domínio D é não-estocástico;

Dados de área/regionais (lattice data)

Os dados de área surgem, quando o domínio/região em estudo é de natureza discreta, e $Z(\mathbf{s})$ pode ser observado em locais fixos que possam ser devidamente enumerados. Os locais podem ser: Províncias, distritos, regiões. Para este tipo de dados, os dados, normalmente, aparecem de forma agregada.

Dados de ponto padrão

Diferentemente dos dados geoestatísticos e regionais, nos dados de ponto padrão a região de interesse não é fixa, mas sim aleatória. Este tipo de dado surge quando o interesse reside em estudar/analisar os locais onde os eventos de interesse ocorrem.

Estacionariedade

- Em geo-estatística o processo inferencial depende da estacionariedade da variável regionalizada.
- Pode-se pensar da estacionariedade como sinónimo de homogeneidade da variável regionalizada na região em estudo.
- Assume-se a estacionariedade, pois a variável regionalizada só pode assumir uma única realização (isto contradiz o conceito de uma variável aleatória).
- Comportamento regular dos momentos de uma função aleatória sobre uma região ou intervalo de tempo.

Estacionariedade estrita:

→ Uma função aleatória é estacionária de forma estrita, se a família de v.a's $Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_k)$ e $Z(\mathbf{s}_1 + \mathbf{h}), Z(\mathbf{s}_2 + \mathbf{h}), \dots, Z(\mathbf{s}_k + \mathbf{h})$ para $\forall k$ e \mathbf{h} a distribuição conjunta de probabilidades é a mesma.

→ A distribuição de conjunta de probabilidades de $\{Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_k)\}$ não depende de qualquer que seja a translação de \mathbf{h} .

→ A hipótese de estacionariedade estrita é bastante rigorosa, podendo se relaxar usando a hipótese de **estacionariedade de segunda ordem**.

Exemplo de estacionariedade estrita

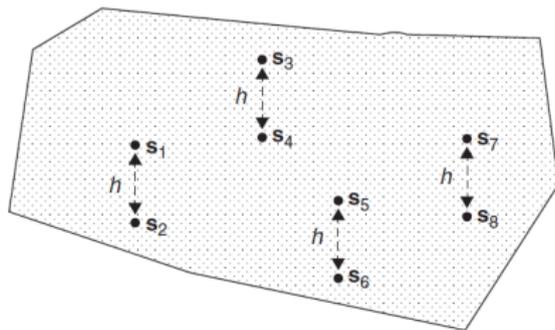


Figure 1: Quatro pares de pontos separados por uma distância h

No caso de estacionariedade estrita $\{Z(s_1), Z(s_2)\}$, $\{Z(s_3), Z(s_4)\}$ e $\{Z(s_5), Z(s_6)\}$ têm a mesma distribuição bivariada de probabilidades, pois a distância entre os pares $\{s_1, s_2\}$, $\{s_3, s_4\}$ e $\{s_5, s_6\}$ é a mesma.

Estacionariedade de segunda ordem:

→ A hipótese de estacionariedade de segunda ordem exige apenas que os dois primeiros momentos da função aleatória estejam definidos.

- $E(Z(\mathbf{s})) < \infty$ e $E(Z(\mathbf{s})) = \mu(\mathbf{s}) = \mu$, e não depende da posição \mathbf{s} :
- $C(Z(\mathbf{s}), Z(\mathbf{s} + \mathbf{h})) < \infty$,
- $C(Z(\mathbf{s}), Z(\mathbf{s} + \mathbf{h})) = C(\mathbf{h}), \forall \mathbf{s} \in R$ and \mathbf{h}

→ A estacionariedade de segunda ordem pode ser interpretada como se a variável regionalizada assumisse valores que flutuam em volta de um valor constante (média), e a variação dessas flutuações fosse a mesma em todo o domínio.

- Se a covariância existe, e é finita, então a variância está definida, e por sua vez é constante $V(Z(\mathbf{s})) = C(\mathbf{0})$,
- No caso de estacionariedade de segunda ordem, a covariância e o semivariograma são equivalentes:

$$\begin{aligned}
 \gamma(\mathbf{h}) &= \frac{1}{2} \text{Var}\{Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})\} \\
 &= \frac{1}{2} \left\{ \text{Var}[Z(\mathbf{s} + \mathbf{h})] + \text{Var}[Z(\mathbf{s})] - 2C[Z(\mathbf{s} + \mathbf{h}), Z(\mathbf{s})] \right\} \\
 &= \frac{1}{2}C(\mathbf{0}) + \frac{1}{2}C(\mathbf{0}) - \frac{2}{2}C(\mathbf{h}) \\
 &= C(\mathbf{0}) - C(\mathbf{h})
 \end{aligned}$$

Estacionariedade de intrínseca:

→ Uma função aleatória é intrinsecamente estacionária, se para qualquer translação \mathbf{h} , as primeiras diferenças $Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})$ são estacionárias de segunda ordem.

- $E(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) = 0$
- $V(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) = 2\gamma(\mathbf{s})$

→ A grandeza $2\gamma(\mathbf{s})$ é conhecida como variograma, e é um parâmetro de extrema importância em geo-estatística.

Variograma/Semivariograma

→ O objectivo primordial da geo-estatística é a previsão (Krigagem/interpolação espacial) do fenómeno em estudo em lugares não amostrados;

→ A interpolação espacial depende da análise variográfica;

→ O semivariograma é dados por :

$$\gamma(\mathbf{h}) = \frac{1}{2}V(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)), \forall \mathbf{s}_i, \mathbf{s}_j \in R$$

→ O variograma é usado para descrever a dependência espacial da variável regionalizada.

→ Se o processo for estacionário de segunda ordem e intrinsecamente estacionário, então:

$$\gamma(\mathbf{h}) = \frac{1}{2}V(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) = \frac{1}{2}E((Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}))^2),$$

mostra como a dissimilaridade entre $Z(\mathbf{s} + \mathbf{h})$ e $Z(\mathbf{s})$ aumenta com a distância \mathbf{h} .

→ O semivariograma que depende apenas da distância de separação das observações, e não da direção, é tido como **isotrópico**. Se este, por sua vez, for intrinsecamente estacionário, o processo é tido como **homogêneo**.

Semivariograma experimental

- Na prática usa-se o semivariograma experimental para estudar a dependência espacial dos dados;
- Usa os dados observados da variável regionalizada para estimar variabilidade espacial do fenómeno em estudo;
- O estimador do semivariograma é dado por:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2\#N(\mathbf{h})} \sum_{N(\mathbf{h})} (Z(\mathbf{s}_i + \mathbf{h}) - Z(\mathbf{s}_i))^2,$$

e é designado por **estimador clássico**. $\#N(\mathbf{h})$ representa o número de pares de observações que são separados por uma distância \mathbf{h} ;

→ A representação gráfica de $\hat{\gamma}(\mathbf{h})$ versus $|\mathbf{h}|$ é designada por **semivariograma experimental**;

→ O estimador do semivariograma $\hat{\gamma}(\mathbf{h})$ é um estimador não enviesado:

$$E(\hat{\gamma}(\mathbf{h})) = E\left(\frac{1}{2\#N(\mathbf{h})} \sum_{N(\mathbf{h})} (Z(\mathbf{s}_i + \mathbf{h}) - Z(\mathbf{s}_i))^2\right) = \gamma(\mathbf{h})$$

Geralmente, na prática, o semivariograma é calculado para distâncias inferiores a metade do diâmetro do domínio de estudo. Isto porque, o número de pares diminui com a distância, e para distâncias maiores, o número de pares não é suficiente para produzir estimativas credíveis.

```
path='C:/Users/Rachid/Dropbox/File requests/EXAME_ISCISA_2021/Fotos/ISCISA/Geostatistics Class Material/Lecture  
hiv.df_prop=read.csv(paste(path, 'hiv_prev.csv', sep='/'), header=TRUE); head(hiv.df_prop)
```

```
##      lat      long      prev  
## 1 -15.27855 32.12809 0.000000  
## 2 -25.95078 32.33560 25.000000  
## 3 -25.83233 32.34203 13.888889  
## 4 -25.34178 32.35649 14.893617  
## 5 -25.94245 32.44658  8.620690  
## 6 -25.82849 32.45118  6.976744
```

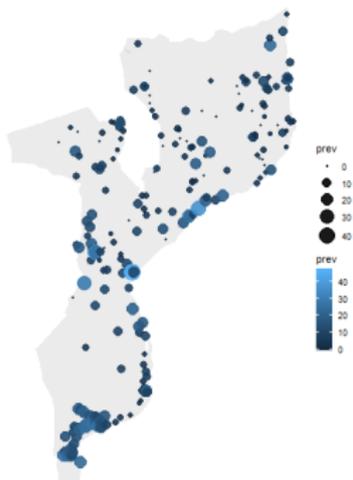


Figure 2: Prevalência de HIV/SIDA por área de enumeração (AE)

```
##install.packages('geoR') # primeiro instalar o pacote geoR
suppressMessages({library('geoR')}) # chamar o pacote geoR
hiv_geo=as.geodata(hiv.df_prop, coords.col = c(2,1),data.col = 3)
summary(hiv_geo)
```

```
## Number of data points: 270
```

```
##
```

```
## Coordinates summary
```

```
##           long           lat
```

```
## min 32.12809 -26.16241
```

```
## max 40.69896 -10.73105
```

```
##
```

```
## Distance summary
```

```
##           min           max
```

```
## 6.596969e-04 1.726818e+01
```

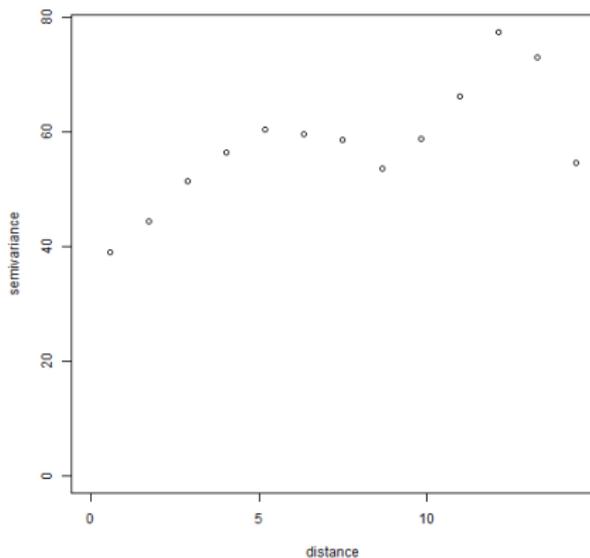
```
##
```

```
## Data summary
```

```
##           Min.    1st Qu.    Median      Mean    3rd Qu.      Max.
```

```
## 0.000000  2.797619  7.643622  8.864440 13.043478 46.511628
```

```
variograma=variog(hiv_geo, max.dist=15)  
plot(variograma)
```



Estimação robusta do semivariograma

O estimador clássico do semivariograma apresenta algumas desvantagens:

- Extremamente sensível a valores atípicos;
- $(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}))^2 \sim 2\gamma(\mathbf{h})\chi_1^2$, conseqüentemente, $2\gamma(\mathbf{h})$ tem uma distribuição bastante assimétrica;

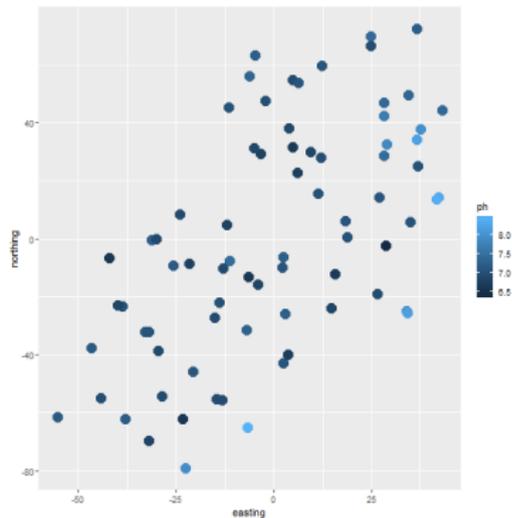
Como forma de remediar, Cressie and Hawkins (1980) propuseram :

$$2\bar{\gamma}(\mathbf{h}) = \frac{1}{0.457 + 0.494/N(\mathbf{h})} \left\{ \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [|Z(x_i + \mathbf{h}) - Z(x_i)|]^{1/2} \right\}^4$$

Exemplo-DATA BREAK: Dados de pH da Smoky Mountain

- O pH nas águas do riacho pode afectar os organismos do riacho
- Mudança do pH pode ser um indicador de poluição
- $0 < \text{pH} < 14$, $\text{pH} < 7$ -ácido, $\text{pH} > 7$ -alcalino

	northing	easting	ph	elevation
1	-79.19	-22.53	7.91	0.244
2	-61.66	-55.39	7.14	0.375
3	-69.63	-31.93	6.81	0.567
4	-54.82	-44.38	6.97	0.512
5	-62.14	-37.94	7.21	0.408
6	-55.29	-14.60	6.94	0.512
7	-62.27	-23.30	6.59	0.506
8	-55.78	-13.17	6.94	0.488
9	-54.48	-28.70	6.94	0.488
10	-65.00	-6.62	8.43	0.232



```

path = "/Users/rachid/Downloads/Requests/EXAMN-IBOIBA_2021/10005_IBOIBA/geoestatisticas/Geoestatistica/Geoestatistica"
ph_df=read.table(paste(path, 'ph_data.txt', sep='/'), sep=';', header=TRUE)
ph_geo=as.geodata(ph_df, coords.col=c(2,1), data.col=3)
variogram_ph=variog(ph_geo, max.dis=120)

## variog: computing omnidirectional variogram
variogram_ph

## $u
## [1] 4.614145 13.842436 23.070726 32.299017 41.527308 50.755598
## [7] 59.983889 69.212179 78.440470 87.668760 96.897051 106.125341
## [13] 115.353632
##
## $v
## [1] 0.0579369 0.1201206 0.1594079 0.1737550 0.1626211 0.2152035 0.2005385
## [8] 0.2732983 0.1865881 0.2541853 0.2295052 0.1987260 0.2042285
##
## $n
## [1] 42 189 258 339 311 288 286 258 193 174 134 98 79
##
## $sd
## [1] 0.1185887 0.2396956 0.3161247 0.2687749 0.3072735 0.3543915 0.3181760
## [8] 0.3716392 0.2824465 0.3649581 0.3585341 0.2872941 0.3110783
##
## $bins.lim
## [1] 1.000000e-12 9.228291e+00 1.845658e+01 2.768487e+01 3.691316e+01
## [6] 4.614145e+01 5.536974e+01 6.459803e+01 7.382632e+01 8.305462e+01
## [11] 9.228291e+01 1.015112e+02 1.107395e+02 1.199678e+02
##
## $ind.bin
## [1] TRUE TRUE
##
## $var.mark
## [1] 0.1930417
##
## $beta.ols
## [1] 7.1396

```

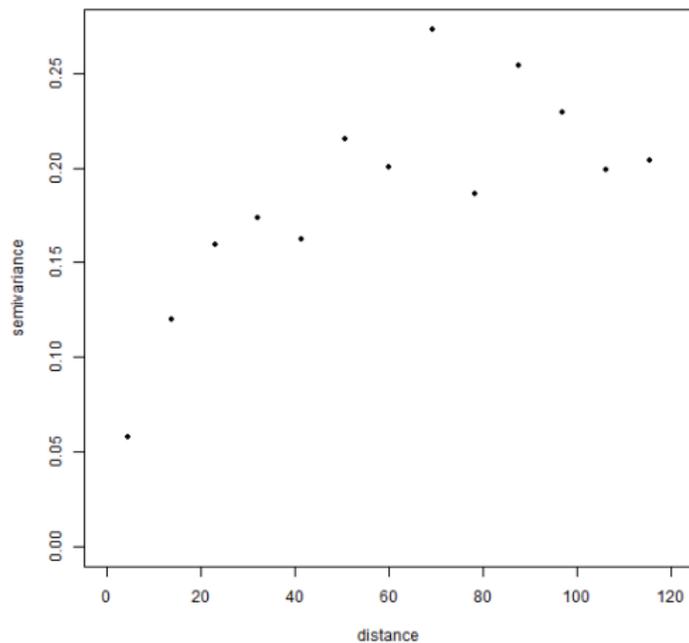


Figure 4: Semivariograma empirico para os valores do pH

Características do semivariograma

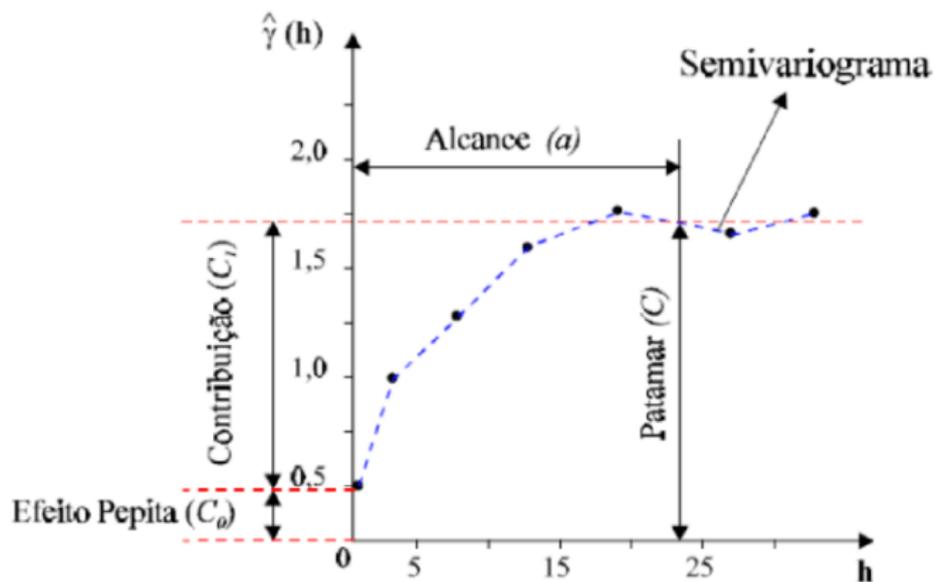


Figure 5: Parâmetros do semivariograma.

Efeito pepita- Valor do semivariograma para $h = 0$. Representa micro variações devido a erros de medição.

Amplitude- Distância em que o variograma atinge o patamar, ou por outra, distância a partir da qual os dados não estão correlacionados.

Patamar/soleira-Valor do semivariograma para uma distância igual a amplitude.

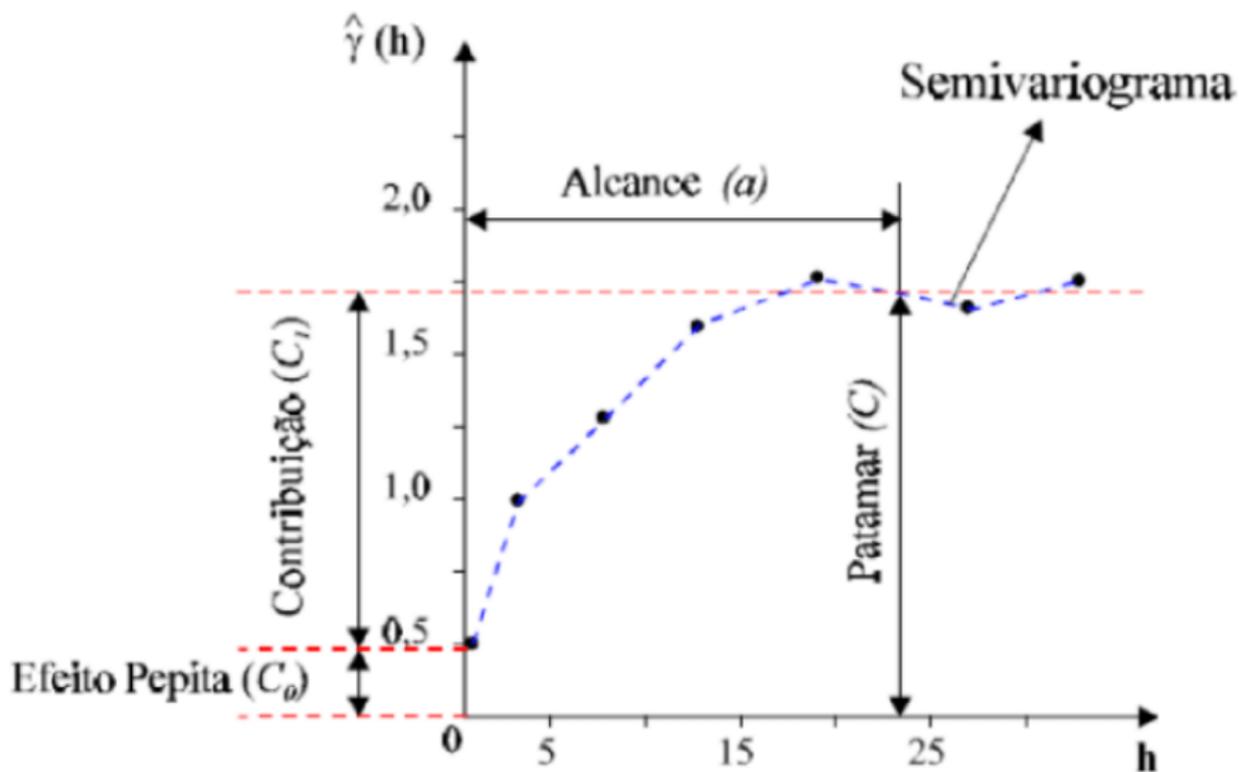


Figure 6: Características do semivariograma.

Ajustamento do modelo de semivariogram

→ O processo de interpolação espacial por meio de Krigagem depende da escolha de um modelo teórico de semivariograma.

→ O modelo teórico de semivariograma deve cumprir algumas condições

- $\gamma(\mathbf{h}) = \gamma(-\mathbf{h})$
- $\gamma(\mathbf{0}) = 0$, visto que, $V(Z(\mathbf{s}) - Z(\mathbf{s})) = 0$
- $\gamma(\cdot)$ deve ser negativa definida, isto é,

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j) \leq 0$$

- O modelo deve ser isotrópico, isto é, o semivariograma depende apenas da distância de separação entre as observações.

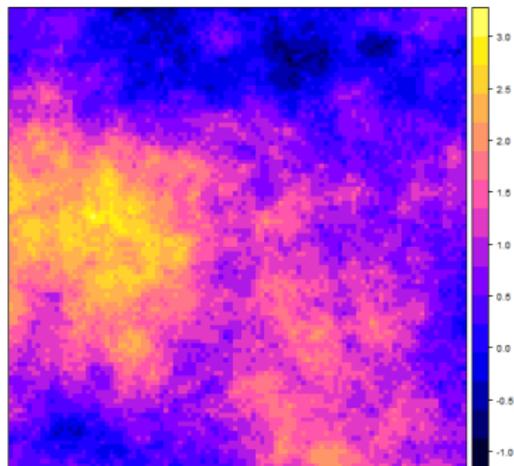
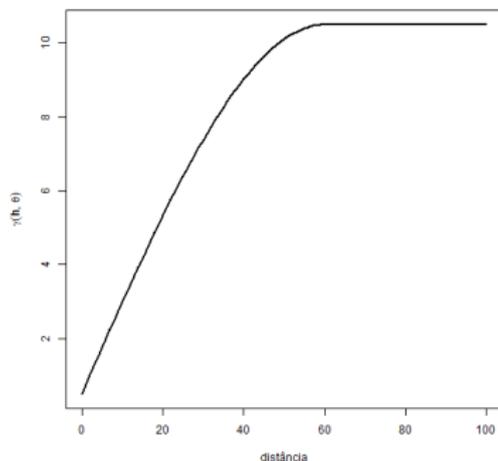
→ Para além das propriedades acima mencionadas, é preciso encontrar um modelo que tenha uma curvatura similar a do variograma empírico.

→ O semivariograma experimental não pode ser usado, pois, não cumpre com as condições acima descritas de um modelo de semivariograma teórico.

→ Existem vários modelos de semivariogramas. Para esta disciplina, vamos-nos limitar nos modelos paramétricos (modelos com uma expressão matemática analiticamente definida).

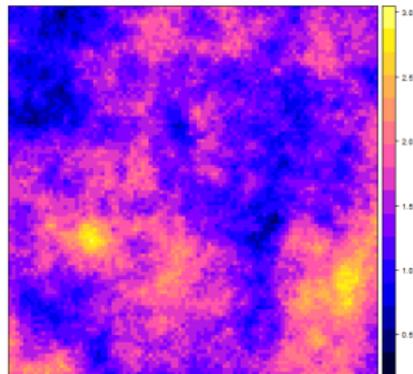
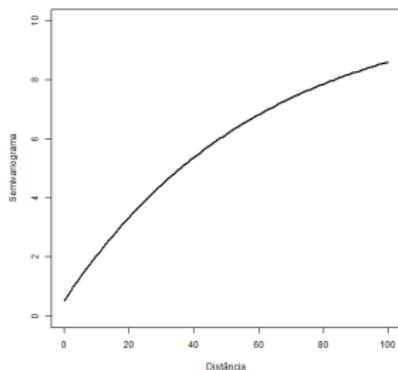
Modelo esférico

$$\gamma(\mathbf{h}, \boldsymbol{\theta}) = \begin{cases} 0, & h = 0 \\ c_0 + c_s [1.5(h/a_s) - 0.5(h/a_s)^3], & 0 < h < a_s \\ c_0 + c_s, & h \geq a_s \end{cases}$$



Modelo exponencial

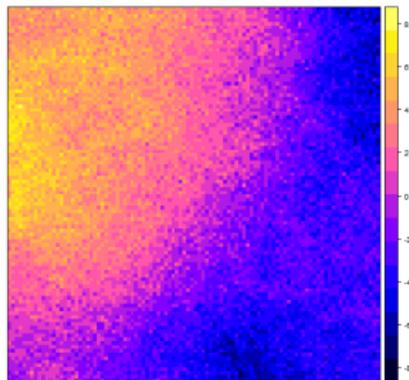
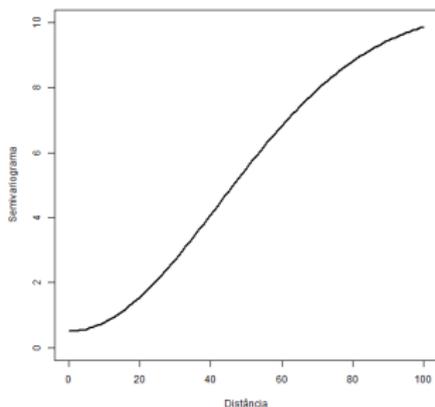
$$\gamma(\mathbf{h}, \boldsymbol{\theta}) = \begin{cases} 0, & h = 0 \\ c_0 + c_e [1 - \exp(-h/a_e)], & h > a_e \end{cases}$$



- A soleira é alcançada assintoticamente quando $|\mathbf{h}| \rightarrow \infty$
- Amplitude prática/efectiva $a' = 3a$ corresponde a distância na qual o valor do semivariogram é 95% da soleira

Modelo Gaussiano

$$\gamma(\mathbf{h}, \boldsymbol{\theta}) = \begin{cases} 0, & h = 0 \\ c_0 + c_g[1 - \exp((-h/a_g)^2)], & h > a_g \end{cases}$$



- A soleira também é alcançada assintoticamente
- Calcule a amplitude efectiva para o modelo Gaussiano (Exercício)

→ **Mínimos quadrados ordinários**- A estimação é feita minimizando a seguinte expressão:

$$\min \sum_h (\gamma(\mathbf{h}) - \hat{\gamma}(\mathbf{h}))^2$$

→ **Mínimos quadrados ponderados**-Os quadrados dos resíduos são atribuídos um ponderador.

- Pode se usar número de pontos que distam h , isto é, $N(\mathbf{h})$

$$\min \sum_h N(\mathbf{h})(\gamma(\mathbf{h}) - \hat{\gamma}(\mathbf{h}))^2$$

- MQP de Cressie tem como ponderador $\frac{N(\mathbf{h})}{\gamma(\mathbf{h})^2}$, logo, a estimação é feita usando:

$$\min \sum_h N(\mathbf{h}) \left(\frac{\hat{\gamma}(\mathbf{h})}{\gamma(\mathbf{h})} - 1 \right)^2$$

Exemplo

```
ph_geo=as.geodata(ph_df,coords.col=c(2,1),data.col=3)
variogram_ph=variog(ph_geo,uvec = seq(0,120,l=10),messages = FALSE)
fit1=variofit(variogram_ph,cov.model="sph",ini.cov.pars=c(0.23,60),
             fix.nugget=FALSE,nugget=0,weights='equal',messages=FALSE)
fit2=variofit(variogram_ph,cov.model="exponential",ini.cov.pars=c(0.23,60),
             fix.nugget=FALSE,nugget=0,weights='equal',messages=FALSE)

fit1
```

```
## variofit: model parameters estimated by OLS (ordinary least squares):
## covariance model is: spherical
## parameter estimates:
##   tausq sigmasq   phi
##  0.0434  0.1786 60.3042
## Practical Range with cor=0.05 for asymptotic range: 60.30417
##
## variofit: minimised sum of squares = 0.0021

fit2
```

```
## variofit: model parameters estimated by OLS (ordinary least squares):
## covariance model is: exponential
## parameter estimates:
##   tausq sigmasq   phi
##  0.0675  0.2077 59.9998
## Practical Range with cor=0.05 for asymptotic range: 179.7434
##
## variofit: minimised sum of squares = 0.0056
```

```
ph_geo=as.geodata(ph_df,coords.col=c(2,1),data.col=3)
variogram_ph=variog(ph_geo,uvec = seq(0,120,l=10),messages = FALSE)
fit3=variofit(variogram_ph,cov.model="sph",ini.cov.pars=c(0.23,100),
             fix.nugget=FALSE,nugget=0,weights='npairs',messages=FALSE)
fit4=variofit(variogram_ph,cov.model="exponential",ini.cov.pars=c(0.23,60),
             fix.nugget=FALSE,nugget=0,weights='npairs',messages=FALSE)
fit3
```

```
## variofit: model parameters estimated by WLS (weighted least squares):
## covariance model is: spherical
## parameter estimates:
##   tausq sigmasq   phi
## 0.0615 0.1622 65.9554
## Practical Range with cor=0.05 for asymptotic range: 65.95537
##
## variofit: minimised weighted sum of squares = 0.4801
```

```
fit4
```

```
## variofit: model parameters estimated by WLS (weighted least squares):
## covariance model is: exponential
## parameter estimates:
##   tausq sigmasq   phi
## 0.0279 0.2049 25.1745
## Practical Range with cor=0.05 for asymptotic range: 75.4161
##
## variofit: minimised weighted sum of squares = 0.607
```

```
# modelo sem efeito pepita
ph_geo=as.geodata(ph_df,coords.col=c(1,2),data.col=3)
variogram_ph=variog(ph_geo,uvec = seq(0,120,l=10),messages = FALSE)
fit31=variofit(variogram_ph,cov.model="sph",ini.cov.pars=c(0.23,100),
              fix.nugget=TRUE,nugget=0,weights='npairs',messages=FALSE)
fit41=variofit(variogram_ph,cov.model="exponential",ini.cov.pars=c(0.23,60),
              fix.nugget=TRUE,nugget=0,weights='npairs',messages=FALSE)

fit31
```

```
## variofit: model parameters estimated by WLS (weighted least squares):
## covariance model is: spherical
## fixed value for tausq = 0
## parameter estimates:
## sigmasq      phi
## 0.2215 54.5730
## Practical Range with cor=0.05 for asymptotic range: 54.57299
##
## variofit: minimised weighted sum of squares = 0.9122
```

```
fit41
```

```
## variofit: model parameters estimated by WLS (weighted least squares):
## covariance model is: exponential
## fixed value for tausq = 0
## parameter estimates:
## sigmasq      phi
## 0.2295 21.7987
## Practical Range with cor=0.05 for asymptotic range: 65.30313
##
## variofit: minimised weighted sum of squares = 0.6449
```

```
ph_geo=as.geodata(ph_df,coords.col=c(2,1),data.col=3)
variogram_ph=variog(ph_geo,uvec = seq(0,120,l=10),messages = FALSE)
fit5=variofit(variogram_ph,cov.model="sph",ini.cov.pars=c(0.23,60),
             fix.nugget=FALSE,nugget=0.01,weights='cressie',messages=FALSE)
fit6=variofit(variogram_ph,cov.model="exponential",ini.cov.pars=c(0.23,60),
             fix.nugget=FALSE,nugget=0.01,weights='cressie',messages=FALSE)
fit5
```

```
## variofit: model parameters estimated by WLS (weighted least squares):
## covariance model is: spherical
## parameter estimates:
##   tausq sigmasq   phi
##  0.052  0.170  60.000
## Practical Range with cor=0.05 for asymptotic range: 59.99996
##
## variofit: minimised weighted sum of squares = 14.7584
```

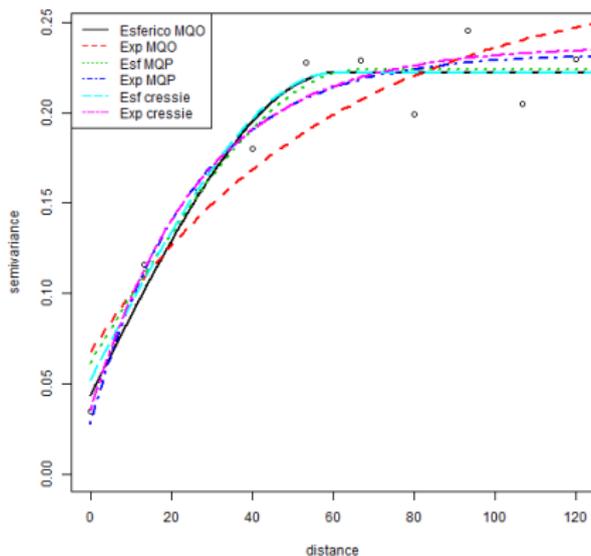
```
fit6
```

```
## variofit: model parameters estimated by WLS (weighted least squares):
## covariance model is: exponential
## parameter estimates:
##   tausq sigmasq   phi
##  0.0352  0.2015  27.0918
## Practical Range with cor=0.05 for asymptotic range: 81.15985
##
## variofit: minimised weighted sum of squares = 13.1579
```

```

var_fit=list(fit1,fit2,fit3,fit4,fit5,fit6)
plot(variogram_ph)
for(i in 1:6){
  lines(var_fit[[i]],lty=i,col=i, lwd=2)
}
legend('topleft',legend = c('Esferico MQO', 'Exp MQO', 'Esf MQP',
                             'Exp MQP', 'Esf cressie', 'Exp cressie'), lty=1:6,col=1:6)

```

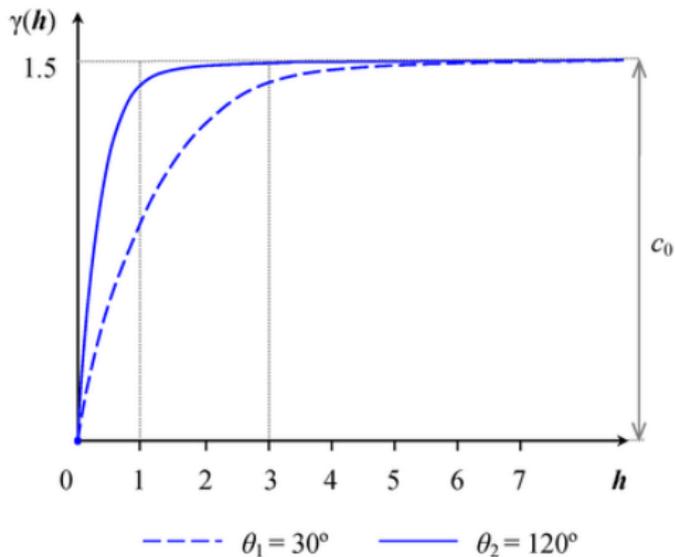


- As estimativas do semivariograma, isto é $\hat{\gamma}(\mathbf{h})$, estão correlacionadas e têm variâncias diferentes.
- MQO pressupõe que as observações sejam independentes com variância constante.
- Como alternativa, pode-se usar o **método dos mínimos quadrados ponderados**.
- Uma outra alternativa é o método dos mínimos quadrados generalizados (MQG).

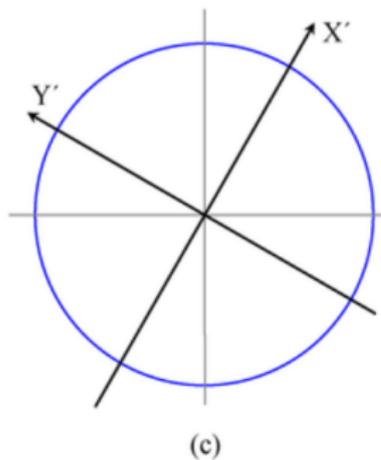
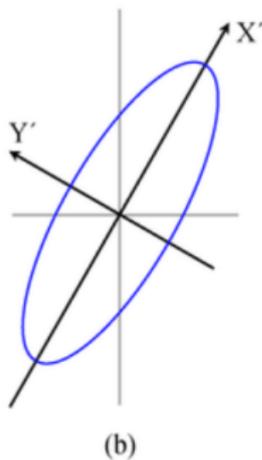
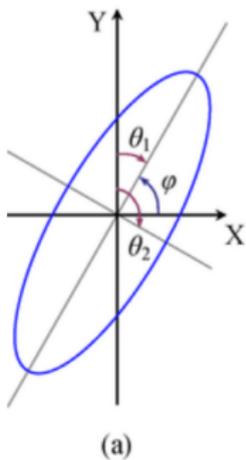
Anisotropia

- Os modelos de semivariograma vistos até aqui são para fenómenos isotrópicos, isto é, que dependem apenas da separação (distância) entre as observações.
- Na prática, maior parte dos processos geoespaciais é de natureza anisotrópica, o que quer dizer que para além da distância, dependente também da direcção.
- A anisotropia pode ser identificada com base no semivariograma experimental, calculado para várias direcções.

Anisotropia geométrica- Ocorre quando a amplitude varia consoante a direcção mas a soleira matem-se constante.



Correcção da anisotropia



Converter a anisotropia geométrica em um processo isotrópico

Para o caso, em que, os eixos da elipse não coincidem com os eixos do sistema cartesiano ortogonal, antes de calcular as distâncias transformadas, é preciso aplicar a rotação dos eixos, de tal maneira que estes coincidam.

$$\begin{pmatrix} \Delta x' \\ \Delta y' \end{pmatrix} = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi/R & \cos \phi/R \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \quad (1)$$

onde $R = a_{min}/a_{max}$, $\Delta x = x_2 - x_1$ e $\Delta y = y_2 - y_1$ e $\phi = 90^\circ - \theta_1$.
As novas distâncias serão calculadas usando a fórmula:

$$\|\mathbf{h}'\| = \sqrt{(\Delta x')^2 + (\Delta y')^2}$$

Os valores do semivariograma teórico são calculados com base nas distâncias isotropadas.

→ A fórmula apresentada no slide anterior é generalista. Todavia, para o caso onde as eixos da elipse coincidem com os eixos do sistema cartesiano ortogonal, as distâncias isotropadas podem ser calculados com base na seguinte fórmula:

$$\|\mathbf{h}'\| = \sqrt{(\Delta x/a_x)^2 + (\Delta y/a_y)^2}$$

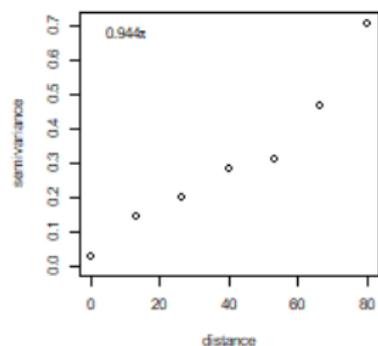
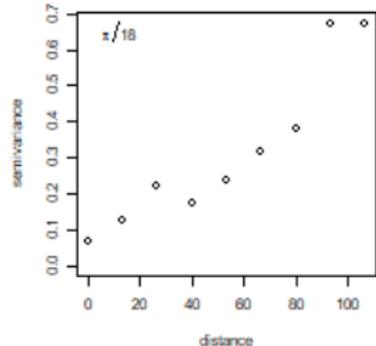
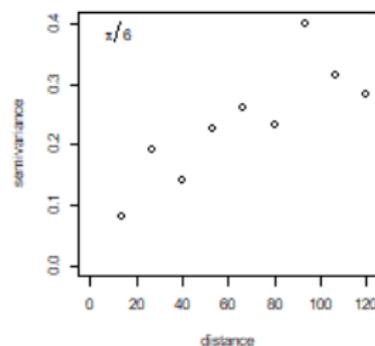
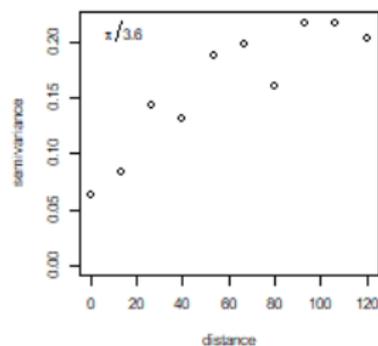
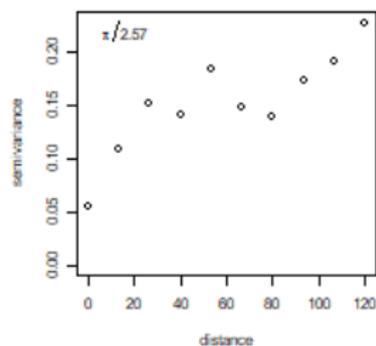
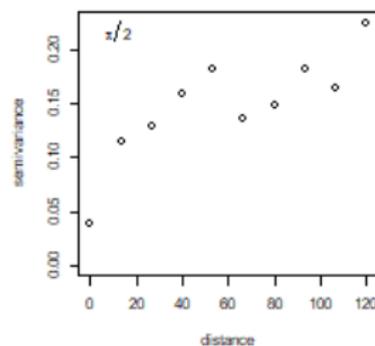
→ A amplitude do semivariograma nas várias direcções terá uma amplitude igual a 1.

→ Se optarmos por transformar a anisotropia num semivariograma de referência (por exemplo, o de maior amplitude em vez do semivariograma com amplitude igual a 1), então a distância \mathbf{h} “isotropizada” fica igual a :

$$\|\mathbf{h}'\| = \sqrt{\Delta x (a_x/a_x)^2 + \Delta y (a_x/a_y)^2}$$

onde a_x é a amplitude se semivariograma de referência e $r_x = a_x/a_x$, $r_y = a_x/a_y$, são os factores de anisotropia nos 2 eixos principais.

Exemplo- pH do Smoky Mountain



```
var1=variog(ph_geo, uvec = seq(0,120,l=10),messages = FALSE, direction=pi/2)
var2=variog(ph_geo, uvec = seq(0,120,l=10),messages = FALSE, direction=pi/2.57)
var3=variog(ph_geo, uvec = seq(0,120,l=10),messages = FALSE, direction=pi/3.6)
var4=variog(ph_geo, uvec = seq(0,120,l=10),messages = FALSE, direction=pi/6)
var5=variog(ph_geo, uvec = seq(0,120,l=10),messages = FALSE, direction=pi/18)
var6=variog(ph_geo, uvec = seq(0,120,l=10),messages = FALSE, direction=0.944*pi)
```

```
fit1=variofit(var1, cov.model="exponential", ini.cov.pars=c(0.23,60),
  fix.nugget=FALSE, nugget=0.02, weights='npairs', messages=FALSE)
fit2=variofit(var2, cov.model="exponential", ini.cov.pars=c(0.23,60),
  fix.nugget=FALSE, nugget=0.02, weights='npairs', messages=FALSE)
fit3=variofit(var3, cov.model="exponential", ini.cov.pars=c(0.23,60),
  fix.nugget=FALSE, nugget=0.02, weights='npairs', messages=FALSE)
fit4=variofit(var4, cov.model="exponential", ini.cov.pars=c(0.23,60),
  fix.nugget=FALSE, nugget=0.02, weights='npairs', messages=FALSE)
fit5=variofit(var5, cov.model="exponential", ini.cov.pars=c(0.23,60),
  fix.nugget=FALSE, nugget=0.02, weights='npairs', messages=FALSE)
fit6=variofit(var6, cov.model="exponential", ini.cov.pars=c(0.23,60),
  fix.nugget=FALSE, nugget=0.02, weights='npairs', messages=FALSE)
fit1; fit2
```

```
## variofit: model parameters estimated by WLS (weighted least squares):
## covariance model is: exponential
## parameter estimates:
##      tausq      sigmasq      phi
##      0.0436      725.8673 139439.7379
## Practical Range with cor=0.05 for asymptotic range: 417724.1
##
## variofit: minimised weighted sum of squares = 0.9011
```

```
## variofit: model parameters estimated by WLS (weighted least squares):
## covariance model is: exponential
## parameter estimates:
##      tausq      sigmasq      phi
##      0.0475      122.5758 31611.0551
## Practical Range with cor=0.05 for asymptotic range: 94698.26
##
## variofit: minimised weighted sum of squares = 1.0231
```

```
## variofit: model parameters estimated by WLS (weighted least squares):  
## covariance model is: exponential  
## parameter estimates:  
##   tausq sigmasq   phi  
## 0.0300 0.2817 58.4009  
## Practical Range with cor=0.05 for asymptotic range: 174.9536  
##  
## variofit: minimised weighted sum of squares = 1.1912  
  
## variofit: model parameters estimated by WLS (weighted least squares):  
## covariance model is: exponential  
## parameter estimates:  
##   tausq sigmasq   phi  
## 0.0667 0.1424 60.0115  
## Practical Range with cor=0.05 for asymptotic range: 179.7783  
##  
## variofit: minimised weighted sum of squares = 0.2707  
  
## variofit: model parameters estimated by WLS (weighted least squares):  
## covariance model is: exponential  
## parameter estimates:  
##   tausq sigmasq   phi  
## 0.0854 0.0980 60.0260  
## Practical Range with cor=0.05 for asymptotic range: 179.8217  
##  
## variofit: minimised weighted sum of squares = 0.6764  
  
## variofit: model parameters estimated by WLS (weighted least squares):  
## covariance model is: exponential  
## parameter estimates:  
##   tausq   sigmasq     phi  
## 0.0921 16.7275 10045.3723  
## Practical Range with cor=0.05 for asymptotic range: 30093.25  
##  
## variofit: minimised weighted sum of squares = 0.75
```

- Os semivariogramas experimentais apresentados na página 51 sugerem que o processo é anisotrópico;
- O modelo exponencial apresenta melhor ajuste na direcção $\pi/2$;
- Os semivariogramas experimentais nas direcções $\pi/6$, $\pi/18$ e 0.944π , apresentam uma **tendência**, isto é, a medida que a distância aumenta o semivariograma aumenta;
- Quando o processo apresenta uma tendência, antes de calcular o semivariograma é preciso remover a tendência;
- Sendo o processo anisotrópico, o ideal é ajustar um semivariograma anisotrópico;

Remoção da tendência (DRIFT)

→ A presença notável da tendência faz com que o pressuposto da estacionariedade seja violado;

→ A tendência torna as estimativas do semivariograma **viciadas**. Consequentemente, pode afectar negativamente o exercício de interpolação espacial.

Como identificar??

→ A tendência pode ser identificada a partir do semivariograma experimental, através de um comportamento crescente acima da soleira.

- A forma funcional da tendência, geralmente, é desconhecida;
- Se a tendência for linear, ajusta-se um modelo linear aos dados, e extraem-se os resíduos. Ex:

$$Z(x, y) = \beta_0 + \beta_1 x + \beta_2 y$$

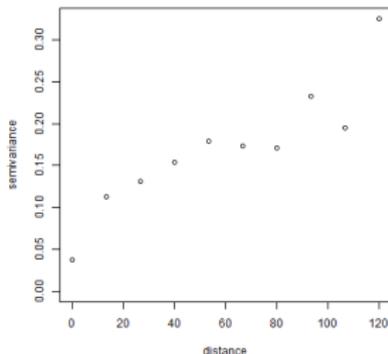
- A tendência pode, também, ser identificada a partir do diagrama de dispersão;

Exemplo- remoção da tendência- Smoky Mountain

```
trend=lm(ph~easting+northing, data=ph_df)
summary(trend)

##
## Call:
## lm(formula = ph ~ easting + northing, data = ph_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.03489 -0.23082 -0.05789  0.20716  1.16752
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.138924   0.045898 155.539 < 2e-16 ***
## easting      0.009396   0.002217   4.239 6.57e-05 ***
## northing    -0.002858   0.001462  -1.954  0.0546 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3975 on 72 degrees of freedom
## Multiple R-squared:  0.2037, Adjusted R-squared:  0.1816
## F-statistic: 9.209 on 2 and 72 DF,  p-value: 0.0002746

res=trend$residuals
res_df=cbind(ph_df,res)
res_geo=as.geodata(res_df,coords.col=c(2,1),data.col=5)
```



```
res_variog=variog(res_geo, uvec = seq(0,120,l=10), messages = FALSE, direction=0.944*pi)
res_fit=variofit(res_variog, cov.model="exponential", ini.cov.pars=c(0.23,60),
  fix.nugget=FALSE, nugget=0.02, weights='npairs', messages=FALSE)
res_fit
```

```
## variofit: model parameters estimated by WLS (weighted least squares):
## covariance model is: exponential
## parameter estimates:
##   tausq  sigmasq   phi
##  0.0906  0.4164 263.4559
## Practical Range with cor=0.05 for asymptotic range: 789.2432
##
## variofit: minimised weighted sum of squares = 0.2248
```

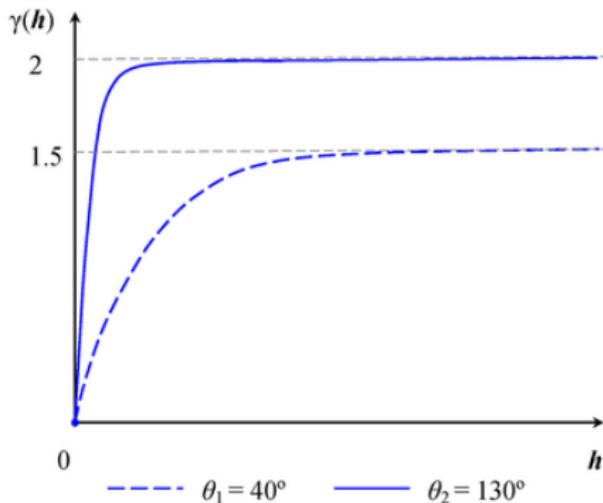
```
var_trend=variog(ph_geo,uvec = seq(0,120,l=10),messages = FALSE, direction=0.944*pi,trend = "1st")
fit_trend=variofit(var_trend,cov.model="exponential",ini.cov.pars=c(0.23,60),
                  fix.nugget=FALSE,nugget=0.02,weights='npairs',messages=FALSE)
fit_trend
```

```
## variofit: model parameters estimated by WLS (weighted least squares):
## covariance model is: exponential
## parameter estimates:
##   tausq  sigmasq   phi
##  0.0906  0.4164 263.4559
## Practical Range with cor=0.05 for asymptotic range: 789.2432
##
## variofit: minimised weighted sum of squares = 0.2248
```

```
var_trend.quadra=variog(ph_geo,uvec = seq(0,120,l=10),messages = FALSE, direction=0.944*pi,trend = "2nd")
fit_trend.quadra=variofit(var_trend.quadra,cov.model="exponential",ini.cov.pars=c(0.23,60),
                          fix.nugget=FALSE,nugget=0.02,weights='npairs',messages=FALSE)
fit_trend.quadra
```

```
## variofit: model parameters estimated by WLS (weighted least squares):
## covariance model is: exponential
## parameter estimates:
##   tausq  sigmasq   phi
##  0.0958  0.0560 60.0282
## Practical Range with cor=0.05 for asymptotic range: 179.8283
##
## variofit: minimised weighted sum of squares = 0.4085
```

Anisotropia zonal- ocorre quando, para várias direcções a soleira é diferente. Geralmente, este tipo de anisotropia é difícil de encontrar na prática. O que tem acontecido, é encontrar uma combinação da anisotropia geométrica e zonal.



Interpolação espacial

- Estimar valores em locais não amostrados usando observações vizinhas;
- Assume-se que o atributo de interesse é contínuo em todo domínio;
- Assume-se que o atributo é espacialmente dependente, com maior similaridade para dados mais próximos, e maior dissimilaridade para dados mais afastados;

Pode-se destacar dois tipos de interpolação: **determinística** e **estatística**.

Interpolação inversa da distância ponderada (IVD)

→ A IVD é simplesmente a média ponderada das observações vizinhas;

→ A interpolação de um ponto não amostrado pode ser calculada usando:

$$Z(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i)$$

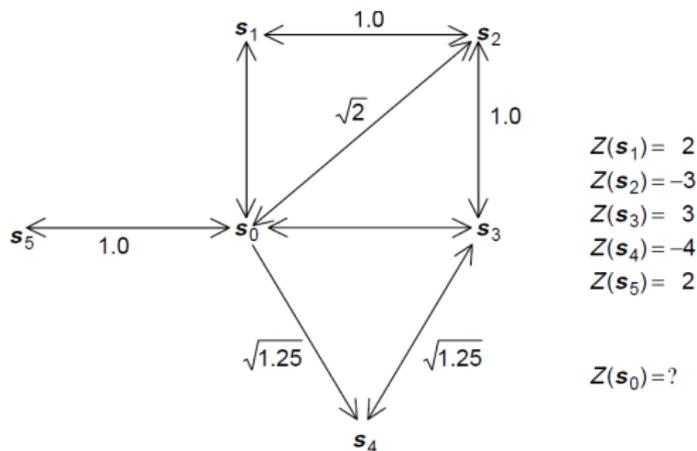
λ_i representa o peso atribuído observação amostrada. O λ_i depende da distância entre o valor amostrado e o ponto a ser estimado, e é dado por:

$$\lambda_i = \frac{d_{0,i}^{-p}}{\sum_i^n d_{0,i}^{-p}}$$

$d_{0,i}$ representa a distância entre o ponto \mathbf{s}_0 e \mathbf{s}_i

- A soma dos pesos deve ser igual a 1, isto é, $\sum_i^n \lambda_i$;
- Geralmente, para p usa-se valores entre 1 e 3;
- $p = 2$ dá origem à **inverso do quadrado da distância**;
- Assumir $p = 0$ é mesmo que usar a média aritmética para estimar o valor de $Z(s_0)$;
- A medida que o p aumenta, os pesos vão diminuindo em função da distância;

Exemplo-IVD



- Uma estimativa razoável para s_0 é a média aritmética;
- Usando IVD as os valores em s_2 e s_4 terão menor; contribuição.
- Considerando $p = 2$, tem-se que $Z(s_0) = 0.53$;

Exemplo-Ilustração em R

```
library(sp)
library(gstat)
# valores amostrados
X = c(61,63,64,68,71,73,75)
Y = c(139,140,129,128,140,141,128)
Z = c(477,696,227,646,606,791,783)
# locais não amostrados. Qual é o valor de Z1
X1 = 65; Y1 = 137

obser_dt = data.frame(X,Y,Z)
coordinates(obser_dt)= ~ X + Y
nao_obser_dt = data.frame(X1,Y1)
coordinates(nao_obser_dt)=~ X1 + Y1
idwmodel = idw(Z ~1, obser_dt,nao_obser_dt,
               maxdist = Inf, idp = 2)
```

```
## [inverse distance weighted interpolation]
```

```
predZ= idwmodel@data$var1.pred
predZ
```

```
## [1] 597.6204
```

Este exemplo foi extraído em <https://rpubs.com/hungle510/202761>

Exemplo: IDW para interpolação de HIV

```
path='C:/Users/lucp8943/Dropbox/Geostatistics Class Material/Lecture notes'
hiv.df_prop=read.csv(paste(path, 'hiv_prev.csv', sep='/'), header=TRUE); head(hiv.df_prop)
shp1<-readOGR("MZGE52FL.shp")
shp<-readOGR("MOZ-level_1.shp")
# create outer boundary
shp2 <- gUnaryUnion(shp)
plot(shp)
plot(shp2, add=T, border = 'red')

# create coordinates of outer boundary
extractCoords <- function(sp.df)
{
  results <- list()
  for(i in 1:length(sp.df@polygons[[1]]@Polygons))
  {
    results[[i]] <- sp.df@polygons[[1]]@Polygons[[i]]@coords
  }
  results <- Reduce(rbind, results)
  results
}
coord<-extractCoords(shp2)

# make grid within area
plot.x = seq(30.21786,40.84447,0.1)
plot.y = seq(-26.86867,-10.47188,0.1)
x.pred = expand.grid(plot.x, plot.y)[,1]
y.pred = expand.grid(plot.x, plot.y)[,2]
pred.loc = data.frame(cbind(x.pred,y.pred))
pred.loc2<-SpatialPoints(pred.loc)
point.in.polygon = over(pred.loc2,shp2)
plot(shp2)
points(pred.loc2[!is.na(point.in.polygon),],cex=0.2)
coordinates(hiv.df_prop)= ~ long +lat
coordinates(pred.loc)=-x.pred+y.pred
idwmodel = idw(prev ~1, hiv.df_prop,pred.loc,
```

Exemplo- IVD para interpolação do HIV (Cont.)

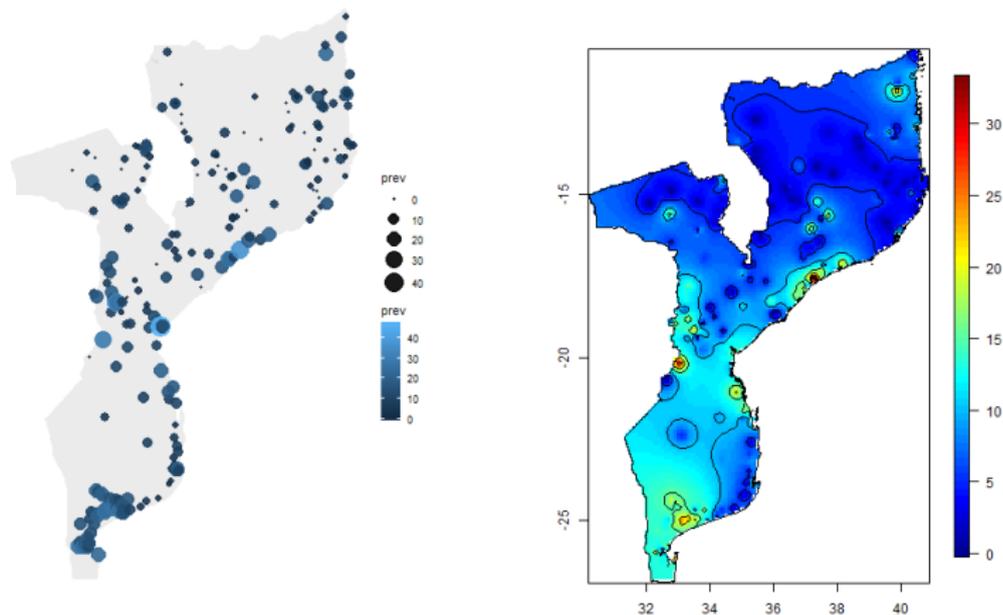


Figure 7: Do lado esquerdo temos valores observados e do lado direito temos valores previsto para todo o domínio

Interpolação por Krigagem

- **Krigagem**- Técnica geoestatística para interpolação espacial;
- Igual ao IVD, a Krigagem usa valores vizinhos para prever locais não amostrados ou estimar a média sobre um determinado bloco;

$$Z(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i)$$

- A Krigagem é conhecida como um interpolador exacto;

Salientar que, a qualidade da interpolação depende de vários factores, nomeadamente:

- i) tamanho da amostra e a qualidade dos dados;
- ii) localização das observações (uniforme ou agrupados);
- iii) distância entre os pontos observados e o ponto a ser previsto;
- iv) continuidade espacial da variável aleatória;

Os métodos de Krigagem levam uma vantagem sobre os outros métodos de interpolação, pois para além de considerar a característica geométrica do processo, consideram também a estrutura espacial do fenómeno.

Noção de vizinhança

Krigagem usa valores circunvizinhos para estimação/previsão de uma local não amostrado.

1. Todos valores amostrados são incluídos na interpolação;
 - ▶ O impacto das observações distantes é insignificante;
 - ▶ Inclusão de todas observações exige que o processo seja estacionário em todo o domínio;
2. Apenas os pontos mais próximos do local a ser interpolado é que são considerados;
 - ▶ Apenas exige-se que o processo seja estacionário de segunda ordem ou quase intrínsecamente estacionário;

Não existe uma regra clara para definir a dimensão da vizinhança, contudo Webster & Oliver (2001) sugerem algumas directrizes:

1. Se os dados forem densos e o semivariograma tiver efeito pepita menor, então o raio da vizinhança pode ser igual a amplitude ou amplitude prática;
2. Se o efeito pepita for maior, observações distantes do ponto a ser previsto irão ter um impacto significativo na interpolação e por essa razão devem ser incluídos na vizinhança;
3. Por outra, pode-se definir a vizinhança em termos de número mínimo e máximo de observações próximos do ponto a ser interpolado. Geralmente recomenda-se um mínimo de $n \approx 7$ e um máximo $n \approx 20$.

Krigagem ordinária

Considere-se que uma função aleatória seja estacionária de segunda ordem, então

$$E[Z(\mathbf{s})] = \mu$$

com covariância definida por:

$$E[Z(\mathbf{s})Z(\mathbf{s} + \mathbf{h})] - \mu^2$$

e variograma dado por

$$E\left([Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})]^2\right)$$

A interpolação por Krigagem é dada por

$$Z^*(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i)$$

A interpolação depende dos valores dos λ_i

1. Estimador não enviesado

$$E[Z^*(\mathbf{s}_0)] = \mu \sum_{i=1}^n \lambda_i = E[Z(\mathbf{s})]$$

2. Variância do erro de estimação (erro quadrático médio) deve ser mínima. O que significa minimizar $E[(Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0))^2]$ sujeito a restrição $\sum_{i=1}^n \lambda_i = 1$

Usando o método dos multiplicadores de Lagrange os pesos (λ_i) podem ser calculados usando a seguinte expressão:

$$\boldsymbol{\lambda} = \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}_0,$$

onde

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ m \end{bmatrix}, \boldsymbol{\Gamma} = \begin{bmatrix} \gamma(\mathbf{s}_1 - \mathbf{s}_1) & \cdots & \gamma(\mathbf{s}_1 - \mathbf{s}_n) & 1 \\ \gamma(\mathbf{s}_2 - \mathbf{s}_1) & \cdots & \gamma(\mathbf{s}_2 - \mathbf{s}_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(\mathbf{s}_n - \mathbf{s}_1) & \cdots & \gamma(\mathbf{s}_n - \mathbf{s}_n) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}, \boldsymbol{\gamma}_0 = \begin{bmatrix} \gamma(\mathbf{s}_0 - \mathbf{s}_1) \\ \gamma(\mathbf{s}_0 - \mathbf{s}_2) \\ \vdots \\ \gamma(\mathbf{s}_0 - \mathbf{s}_n) \\ 1 \end{bmatrix}$$

A variância de estimativa resultante da Krigagem é dada por:

$$\sigma^2(\mathbf{s}_0) = \boldsymbol{\lambda}^T \boldsymbol{\gamma}_0 = \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) + m$$

Assumindo-se que o erro de previsão $Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)$ segue distribuição normal, então o intervalo de previsão a 95% será dado por:

$$(Z^*(\mathbf{s}_0 \pm 1.96\sigma(\mathbf{s}_0)))$$

Exemplo - Krigagem ordinária (pontual)

→ Considere exercício no slide 67.

→ Considere, igualmente, um semivariograma esférico com $c_0 = 0$, $c_s = 1$ e $a = 1.5$

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ m \end{bmatrix} = \begin{bmatrix} 0 & 0.852 & 0.995 & 1.00 & 0.995 & 1 \\ 0.852 & 0 & 0.852 & 1.00 & 1.00 & 1 \\ 0.995 & 0.852 & 0 & 0.911 & 1.00 & 1 \\ 1.00 & 1.00 & 0.911 & 0 & 1.00 & 1 \\ 0.995 & 1.00 & 1.00 & 1.00 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0.852 \\ 0.995 \\ 0.852 \\ 0.911 \\ 0.852 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.256809 \\ 0.049996 \\ 0.241484 \\ 0.186251 \\ 0.265459 \\ 0.118743 \end{bmatrix}$$

$$Z^*(s_0) = 0.257 \times 62 + 0.050 \times (-3) + 0.241 \times 3 + 0.186 \times (-4) + 0.265 \times 2 = 0.88$$

Exemplo: Dados do pH do Smoky Mountain

```
ph_geo=as.geodata(ph_df,coords.col=c(2,1),data.col=3)
variogram_ph=variog(ph_geo,uvec = seq(0,120,l=10),messages = FALSE)
fit3=variofit(variogram_ph,cov.model="sph",ini.cov.pars=c(0.23,100),
             fix.nugget=FALSE,nugget=0,weights='npairs',messages=FALSE)
x.range=as.integer(range(ph_df[,1]))
y.range=as.integer(range(ph_df[,2]))
grd=expand.grid(x=seq(from=x.range[1], to=x.range[2], length.out=50),
                y=seq(from=y.range[1], to=y.range[2], length.out =50))
kg_ph=krige.conv(ph_geo,locations = grd, krige = krige.control(obj.m = fit3))
```

```
## krige.conv: model with constant mean
## krige.conv: Kriging performed using global neighbourhood
```

```
kg_ph$predict
```

```
## [1] 7.308242 7.286815 7.257535 7.223387 7.192816 7.167306 7.146242 7.126020
## [9] 7.103787 7.080872 7.056975 7.031615 7.019327 7.007065 6.989325 6.968901
## [17] 6.950733 6.942323 6.948501 6.976496 7.018352 7.066827 7.179687 7.291974
## [25] 7.372481 7.422265 7.450354 7.467347 7.482102 7.498229 7.516474 7.536553
## [33] 7.557614 7.578573 7.597988 7.615639 7.628826 7.636905 7.640956 7.639331
## [41] 7.630393 7.619320 7.607137 7.594626 7.584877 7.576486 7.561943 7.537759
## [49] 7.504155 7.467748 7.309336 7.287117 7.257313 7.222630 7.191364 7.165117
## [57] 7.143185 7.121556 7.098915 7.075831 7.051864 7.028314 7.013050 6.999818
## [65] 6.983992 6.966790 6.952468 6.945976 6.951796 6.974345 7.008500 7.058825
## [73] 7.144774 7.236176 7.307538 7.355848 7.387020 7.409348 7.429696 7.451964
## [81] 7.476419 7.502606 7.529284 7.555393 7.580196 7.601843 7.618469 7.630029
## [89] 7.638311 7.639510 7.633150 7.623603 7.613173 7.602654 7.595528 7.589950
## [97] 7.577693 7.555989 7.523764 7.486218 7.309814 7.286994 7.257229 7.222357
## [105] 7.190180 7.163195 7.140387 7.118218 7.095798 7.073251 7.050310 7.028878
## [113] 7.011938 6.997234 6.982175 6.967052 6.955654 6.950624 6.958060 6.979060
## [121] 7.010043 7.056114 7.121813 7.192826 7.252086 7.295008 7.325643 7.350741
## [129] 7.375376 7.403145 7.433952 7.466363 7.499084 7.530616 7.560657 7.587046
## [137] 7.607370 7.622286 7.634286 7.638217 7.634277 7.626454 7.616518 7.607537
## [145] 7.602812 7.599602 7.589568 7.570810 7.540986 7.503959 7.463967 7.286722
## [153] 7.257610 7.222312 7.189241 7.161268 7.132327 7.115007 7.094105 7.076604
```

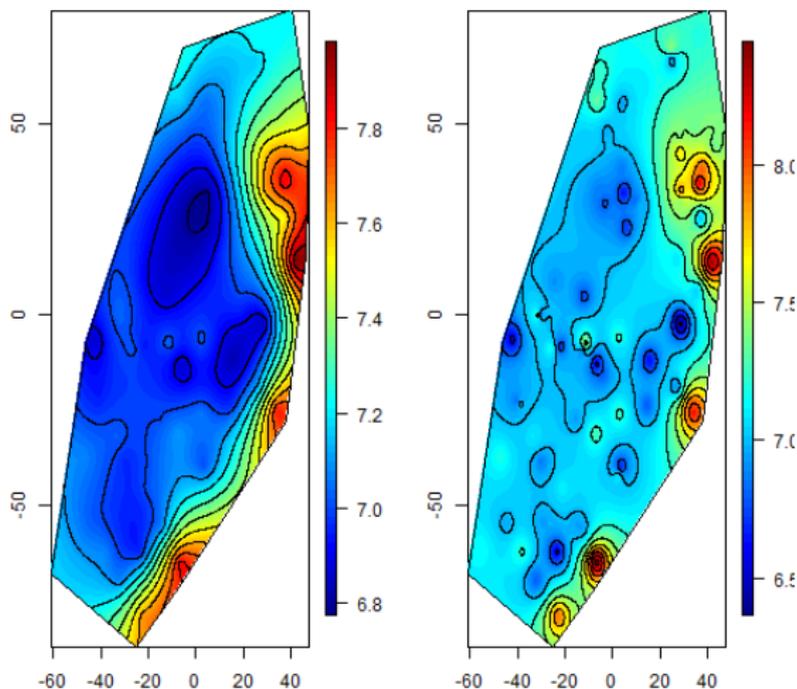


Figure 8: A figura mostra as interpolações, do lado esquerdo por Krigagem e do lado direito por Inverso da distância.

Durante o processo do estudo da continuidade espacial observou-se que, este processo apresenta um comportamento anisotrópico.

```
# Desenho da malha para efeitos de Krigagem
# esta malha contem os pontos que serão considerados
# como locais não amostrados , para os quais serão feitas
# as interpolações
plot.x = seq(-60.93033 , 47.48567,0.1)
plot.y = seq(-87.08972,79.58228,0.1)
x.pred = expand.grid(plot.x, plot.y)[,1]
y.pred = expand.grid(plot.x, plot.y)[,2]
pred.loc = data.frame(cbind(x.pred,y.pred))
pred.loc2<-SpatialPoints(pred.loc)
point.in.polygon = over(pred.loc2,sps)
plot(sps)
points(pred.loc2[!is.na(point.in.polygon),],cex=0.2)
plot.index = matrix(point.in.polygon, nrow=length(plot.x), ncol=length(plot.y))

# Krigagem com correcção da anisotropia
kg_ph=krige.conv(ph_geo,locations = pred.loc,
                krige = krige.control(cov.model ='exponential',nugget =0,
                cov.pars =c(0.2725,36.25),aniso.pars = c(7*pi/18,36.25/16.93) ))

pred.kg=kg_ph$predict

# visualização das intgerpolações por Krigagem
plot.mean.kg= matrix(pred.kg, nrow=length(plot.x), ncol=length(plot.y))
plot.mean.kg[is.na(plot.index)] = NA
image.plot(x=plot.x,y=plot.y,z=plot.mean.kg,xlab="",ylab="")
contour(x=plot.x,y=plot.y,z=plot.mean.kg,xlab="",ylab="",add=TRUE, drawlabels=FALSE)
plot(sps,add=T)
```

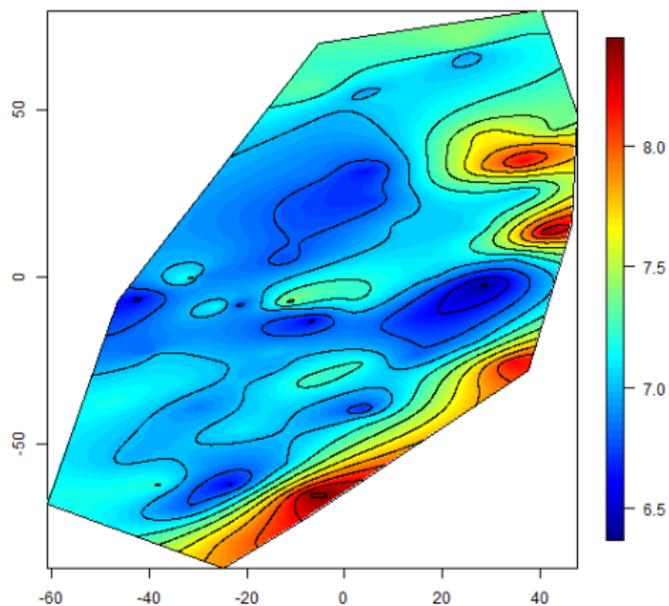


Figure 9: Distribuição espacial do pH. Valores interpolados com correcção da anisotropia geométrica

Krigagem Simples

- Este método não difere tanto da Krigagem Ordinária;
- Assume que a média do processo é conhecida;
- O processo de interpolação leva em consideração o conhecimento da média;

$$Z^*(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) + \left(1 - \sum_{i=1}^n \lambda_i\right) \mu$$

Visto que a média é conhecida o cálculo dos pesos será feito apenas com base na função de covariância.

$$\boldsymbol{\lambda} = \mathbf{C}^{-1} \mathbf{c}_0,$$

onde

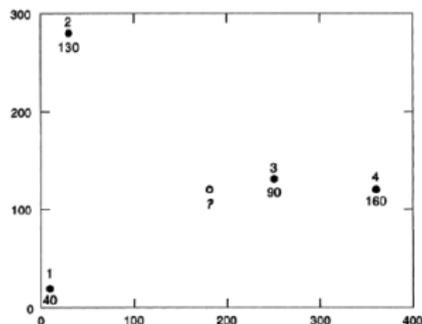
$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix}, \mathbf{C} = \begin{bmatrix} C(\mathbf{s}_1 - \mathbf{s}_1) & \cdots & C(\mathbf{s}_1 - \mathbf{s}_n) \\ C(\mathbf{s}_2 - \mathbf{s}_1) & \cdots & C(\mathbf{s}_2 - \mathbf{s}_n) \\ \vdots & \ddots & \vdots \\ C(\mathbf{s}_n - \mathbf{s}_1) & \cdots & C(\mathbf{s}_n - \mathbf{s}_n) \end{bmatrix}, \mathbf{c}_0 = \begin{bmatrix} C(\mathbf{s}_0 - \mathbf{s}_1) \\ C(\mathbf{s}_0 - \mathbf{s}_2) \\ \vdots \\ C(\mathbf{s}_0 - \mathbf{s}_n) \end{bmatrix}$$

A variância da estimativa para krigagem simples é dada por:

$$\sigma_{KS}^2 = C(0) - \boldsymbol{\lambda}^T \mathbf{c}_0$$

Exemplo Krigagem Simples

Considere a figura (lado esquerdo) e a tabela abaixo (lado direito)



índice	X	Y	Medições
1	10	20	40
2	30	280	130
3	250	130	90
4	360	120	160

Assuma que o atributo em estudo tem uma média de 110 e uma função de covariância $C(h) = 2000 \exp\left(-\frac{h}{250}\right)$. Se o ponto a ser estimado 'e $x_0 = (180, 120)$. Calcule os pesos e o valor do atributo para o ponto não amostrado.

```

x=c(10,30,250,360,180)
y=c(20,280,130,120,120)
XY=as.matrix(cbind(x,y))
dist=matrix(data=NA, nrow=5,ncol=5)

for(i in 1:5){
  for(j in 1:5){
    dist[i,j]=sqrt((XY[i,1]-XY[j,1])^2+(XY[i,2]-XY[j,2])^2)
  }
}

Cov=matrix(data=NA, nrow=5,ncol=5)
for(i in 1:5){
  for(j in 1:5){
    Cov[i,j]=2000*exp(-dist[i,j]/250)
  }
}

Cov1=Cov[1:4,1:4]
Cov0=Cov[1:4,5]
Cov0=t(Cov0)
w=solve(Cov1)%*%t(Cov0)
w

```

```

##           [,1]
## [1,]  0.184679065
## [2,]  0.128482048
## [3,]  0.645838236
## [4,] -0.001128155

```

Krigagem Universal

→ A Krigagem ordinária, assim como, a Krigagem simples assumem que a média do processo no campo aleatório é constante.

→ Na prática, campos ambientais e geológicos muitas vezes apresentam valores médios não constantes (a média do processo não é constante em todo espaço aleatório).

$$Z(\mathbf{x}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s})$$

onde, $\mu(\mathbf{s})$ é uma função que depende da localização (\mathbf{s}), e $\epsilon(\mathbf{s})$ é um processo estacionário de segunda ordem (com média zero).

A componente $\mu(s)$ caracteriza a tendência do processo (e designa-se por *drift*). Suponha que $\mu(s)$ pode ser representado como uma combinação linear de funções conhecidas $\{f_l(s), l = 1, \dots, k\}$, com coeficientes desconhecidos $\{a_l\}$

$$\mu(s) = \sum_{l=1}^k a_l f_l(s)$$

A média do processo bem como a covariância podem ser expressas da seguinte forma:

$$E[Z(s)] = \sum_{l=1}^k a_l f_l(s)$$

$$E\{[Z(s_1) - \mu(s_1)][Z(s_2) - \mu(s_2)]\} = E[\epsilon(s_1)\epsilon(s_2)] = C(s_1 - s_2)$$

Tal como a Krigagem Ordinária, a **Krigagem Universal** também é um interpolador que resulta da combinação linear das observações circundantes ao local a ser estimado,

$$Z^*(s_0) = \sum_{i=1}^n \lambda_i Z(s_i)$$

onde λ_i é escolhido de tal maneira que o estimador seja não enviesado e erro de estimação seja mínimo. O estimador será não enviesado, se e somente se, $E[Z^*(s_0)] = E[Z(s_0)]$, ou

$$\sum_{i=1}^n \lambda_i \mu(s_i) = \mu(s_0) \Rightarrow \sum_{l=0}^k a_l \sum_{i=1}^n \lambda_i f_l(s_i) = \sum_{l=0}^k a_l f_l(s_0)$$

O interpolador só será não-enviesado, se e somente se,
 $\sum_{i=1}^n \lambda_i f_l(s_i) = f_l(s_0)$.

A variância para este interpolador pode ser dada por:

$$\sigma_{KU}^2(x_0) = E\{[Z^*(s_0) - Z(s_0)]^2\} = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(s_i, s_j) - 2 \sum_{i=1}^n \lambda_i C(s_i, s_0) + C(0) \quad (2)$$

Esta expressão deve ser minimizada sob a condição de não-enviesamento $\sum_{i=1}^n \lambda_i f_l(s_i) = f_l(s_0)$. isto pode ser feito usando multiplicador de lagrange, que irá resultar no seguinte sistema de equações:

$$\begin{cases} \sum_{j=1}^n \lambda_j C(\mathbf{s}_i - \mathbf{s}_j) - \sum_{l=0}^k \alpha_l f_l(s_i) = C(\mathbf{s}_i - \mathbf{s}_0), & i = 1, 2, \dots, n \\ \sum_{i=1}^n \lambda_i f_l(s_i) = f_l(s_0), & l = 0, 1, 2, \dots, k \end{cases} \quad (3)$$

Krigagem Indicatriz

A krigagem indicatriz consiste na aplicação da krigagem ordinária numa variável transformada, isto é, uma variável resultante da aplicação da função não linear:

$$I(\mathbf{s}, z_T) = \begin{cases} 1, & \text{se } Z(\mathbf{s}) \leq z_T \\ 0, & \text{se } Z(\mathbf{s}) > z_T, \end{cases}$$

tal que

$$P(I(\mathbf{s}, z_T) = 1) = P(Z(\mathbf{s}) \leq z_T) = F_{Z(\mathbf{s})}(z_T)$$
$$P(I(\mathbf{s}, z_T) = 0) = P(Z(\mathbf{s}) > z_T) = 1 - F_{Z(\mathbf{s})}(z_T)$$

Krigagem Indicatriz

Se assumirmos que a variável regionalizada em estudo é estacionária de segunda ordem, então, tem-se:

$$E(I(\mathbf{s}, z_T)) = F_{Z(\mathbf{s})}(z_T) \quad (4)$$

$$V(I(\mathbf{s}, z_T)) = F_{Z(\mathbf{s})}(z_T)(1 - F_{Z(\mathbf{s})}(z_T)) \quad (5)$$

$$\begin{aligned} C_{z_T}(\mathbf{h}) &= E(I(\mathbf{s}, z_T)I(\mathbf{s} + \mathbf{h}, z_T)) - E(I(\mathbf{s}, z_T))E(I(\mathbf{s} + \mathbf{h}, z_T)) \\ &= P(I(\mathbf{s}, z_T)I(\mathbf{s} + \mathbf{h}, z_T) = 1) - P(I(\mathbf{s}, z_T) = 1)P(I(\mathbf{s} + \mathbf{h}, z_T) = 1) \\ &= F_{Z(\mathbf{s}), Z(\mathbf{s} + \mathbf{h})}(z_T) - (F_{Z(\mathbf{s})}(z_T))^2 \end{aligned} \quad (6)$$

Krigagem Indicatriz

A relação entre o variograma e covariograma fica definido da seguinte forma

$$\begin{aligned}\gamma(\mathbf{h}) &= C_{z_T}(\mathbf{0}) - C_{z_T}(\mathbf{h}) \\ &= V(I(\mathbf{s}_0, z_T)) - C_{z_T}(\mathbf{h}, z_T) \\ &= F_{Z(\mathbf{s})}(z_T) - F_{Z(\mathbf{s}), Z(\mathbf{s}+\mathbf{h})}(z_T)\end{aligned}\tag{7}$$

A interpolação por krigagem ordinário para uma variável regionalizada indicadora é dada por :

$$I(\mathbf{s}_0, z_T) = \sum_{i=1}^n \lambda_i I(\mathbf{s}_i, z_T)$$

Os λ_i são obtidos seguindo o mesmo procedimento da krigagem ordinária

A krigagem indicatriz requer variograma da variável indicadora para cada teor de corte

Krigagem indicatriz - variancia da estimativa

$$V(I^*(\mathbf{s}, z_T)) = I^*(\mathbf{s}, z_T)(1 - I^*(\mathbf{s}, z_T))$$