

COLHEITA DE DADOS E INSIGHTS -  
DADOS VALIOSOS E MADUROS

# *A ESTATÍSTICA DE COORDENADAS E GEOLOCALIZAÇÕES*



6

## LISTA DE CÓDIGOS-FONTE

Código-fonte 1 - Exemplo da estatística espacial descritiva na Linguagem R .....	14
Código-fonte 2 - Exemplo de visualização de dados em R .....	17

EXEMPLO

## SUMÁRIO

1 A ESTATÍSTICA DE COORDENADAS E GEOLOCALIZAÇÕES .....	4
1.1 Conceitos Básicos da Análise Exploratória de Dados Espaciais.....	7
1.1.1 Dados Espaciais.....	7
1.1.2 Autocorrelação Espacial.....	9
1.1.3 Heterogeneidade Espacial .....	9
1.1.4 Visualização Espacial.....	10
2 ESTATÍSTICA ESPACIAL DESCRITIVA .....	11
2.1 Medidas de Centralidade.....	11
2.1.1 Medidas de Dispersão.....	11
2.1.2 Índices de Autocorrelação Espacial .....	12
2.2 Exemplo Aplicado à Linguagem R .....	12
2.3 Visualização de Dados Espaciais.....	14
2.3.1 Mapas.....	14
2.3.2 Gráficos de Dispersão Espacial .....	15
2.3.3 Diagramas de Moran.....	15
2.4 Exemplo Aplicado à Linguagem R .....	16
REFERÊNCIAS.....	18

## 1 A ESTATÍSTICA DE COORDENADAS E GEOLOCALIZAÇÕES

A Análise Exploratória de Dados Espaciais (AEDE) é um ramo da Ciência de Dados que explora e analisa os dados que possuem uma referência espacial, como coordenadas geográficas ou informações de localização (Haining, 2003). A AEDE permite identificar padrões, correlações e *insights* que podem ser perdidos em análises tradicionais que não consideram a dimensão espacial dos dados. É uma abordagem na estatística utilizada para sumarizar as características principais de um conjunto de dados, frequentemente com métodos visuais. Quando aplicada ao contexto espacial, essa abordagem se expande para investigar as relações geográficas e espaciais entre os dados.

A análise de dados espaciais representa um campo de crescente importância para profissionais da Ciência de Dados, Engenharia de Software e Inteligência Artificial (IA), marcando uma evolução nas habilidades demandadas nesses campos.

A análise exploratória de dados espaciais é importante para a IA por várias razões, entre elas:

- **Identificação de padrões espaciais:** AEDE permite a identificação de padrões espaciais nos dados, como agrupamentos ou dispersões, que podem ser determinantes para entender fenômenos geográficos e tomar decisões informadas.
- **Melhoria de modelos preditivos:** ao incorporar informações espaciais, os modelos de IA podem se tornar mais precisos e robustos. Por exemplo, em previsão de crimes ou análise de mercado imobiliário, a localização geográfica é uma variável fundamental.
- **Deteção de anomalias:** AEDE auxilia a detecção de anomalias espaciais que podem indicar problemas ou oportunidades não evidentes. Isso é útil em áreas como monitoramento ambiental, saúde pública e logística.
- **Integração de dados heterogêneos:** AEDE facilita a integração de diferentes tipos de dados (socioeconômicos, ambientais, demográficos) que têm um componente espacial, proporcionando uma visão mais completa e multifacetada dos problemas.

- **Auxilia na tomada de decisão:** compreender a distribuição espacial dos dados auxilia na tomada de decisões estratégicas, como a localização de novos empreendimentos, planejamento urbano e gestão de recursos naturais.
- **Personalização e recomendação:** em aplicações de IA, como sistemas de recomendação e personalização, a AEDE pode melhorar as recomendações ao considerar a proximidade geográfica e os comportamentos locais dos usuários.
- **Resolução de problemas complexos:** problemas complexos como redes de transporte, planejamento urbano e gestão de crises se beneficiam de uma análise espacial detalhada para identificar as melhores soluções possíveis. O advento dos smartphones, a popularização dos Sistemas de Informação Geográfica (SIG), e o surgimento de plataformas de dados abertos ampliaram o acesso a dados geoespaciais, transformando a maneira como as organizações tomam decisões estratégicas. Para os profissionais de IA isso significou:
  - **Acesso amplo a dados geoespaciais:** a disponibilidade massiva de dados geoespaciais através de smartphones e SIG ampliou o volume e a variedade de dados disponíveis para análise. Isso permite que os profissionais de IA desenvolvam modelos mais precisos e informados, utilizando dados geográficos em grande escala.
  - **Desenvolvimento de novos algoritmos:** o aumento dos dados geoespaciais disponíveis resultou em uma maior demanda por algoritmos especializados em análise espacial, como redes neurais convolucionais para processamento de imagens geográficas e algoritmos de aprendizado de máquina para análise de padrões espaciais.
  - **Melhoria na tomada de decisão:** a integração de dados geoespaciais em modelos de IA permite uma tomada de decisão mais precisa e contextualizada. Isso é crucial para aplicações como planejamento urbano, gestão de recursos naturais e otimização de rotas em logística.
  - **Avanços em sistemas de recomendação e personalização:** os dados geoespaciais são fundamentais para melhorar sistemas de

recomendação e personalização, considerando a localização do usuário e seus comportamentos locais. Isso é amplamente aplicado em serviços de delivery, marketing direcionado e turismo.

- **Inovação em aplicações de mobilidade e transporte:** a IA combinada com os dados geoespaciais impulsiona diversas inovações em transporte autônomo, sistemas de navegação inteligente e soluções de mobilidade urbana. Isso inclui algoritmos para previsão de tráfego, otimização de rotas e análise de padrões de movimento.
- **Integração com tecnologias emergentes:** o uso de dados geoespaciais em conjunto com tecnologias emergentes como Internet das Coisas (IoT, do inglês Internet of Things), Big Data e computação em nuvem está promovendo novas oportunidades para desenvolvimento de soluções avançadas, como monitoramento ambiental em tempo real e análise preditiva de desastres naturais.
- **Aprimoramento de sistemas de monitoramento e resposta a emergências:** os dados geoespaciais são essenciais para melhorar os sistemas de monitoramento e resposta rápida a emergências, como desastres naturais e crises sanitárias. Isso inclui o uso de IA para modelagem preditiva e gestão de recursos em tempo real.
- **Desafios e oportunidades éticas:** com o aumento no uso de dados geoespaciais surgem também desafios éticos relacionados à privacidade, segurança e uso responsável dos dados. Assim, profissionais de IA precisam desenvolver soluções que equilibram inovação e conformidade com regulamentos de proteção de dados.

A importância da análise de dados espaciais é evidenciada pela evolução dos sistemas baseados em localização. Aplicações como mapas digitais, aplicativos de navegação, monitoramento ambiental e análise de tráfego dependem de análises espaciais para fornecer serviços precisos e personalizados. Essas aplicações, além de serem tecnicamente desafiadoras, exigem também uma compreensão profunda de como os dados espaciais funcionam e interagem com os sistemas de software.

Além disso, a crescente tendência de cidades inteligentes (*smart cities*) e de IoT reforça ainda mais a necessidade de competências em análise de dados espaciais

para profissionais de IA. Dispositivos conectados geram continuamente dados geolocalizados que, se corretamente analisados, podem melhorar a qualidade de vida, otimizar operações urbanas e reduzir impactos ambientais.

Portanto, a análise de dados espaciais para profissionais de IA não é apenas uma habilidade adicional, mas sim uma área de conhecimento crítica que amplia o escopo de aplicação do software para resolver problemas reais e complexos. À medida que o mundo se torna cada vez mais conectado e geolocalizado, a habilidade de integrar análise espacial no desenvolvimento de software não apenas aumentará a relevância dos profissionais nesse campo, mas também capacitará organizações a aproveitar plenamente o potencial dos dados espaciais para inovação e impacto social.

### 1.1 Conceitos Básicos da Análise Exploratória de Dados Espaciais

Para uma compreensão de dados espaciais, vamos inicialmente discutir alguns conceitos para facilitar a compreensão de dados e análise na perspectiva espacial.

#### 1.1.1 Dados Espaciais

Dados espaciais são fundamentais na análise geográfica e ambiental, possuindo a característica única de uma referência espacial (Haining, 2003). Essas referências podem ser coordenadas geográficas, sistemas de referência de grade, ou simplesmente informações de localização que vinculam cada ponto de dados a um local específico. Este atributo espacial permite que analistas e engenheiros de software explorem a dimensão geográfica dos dados, abrindo um amplo espectro de análises e aplicações, desde o mapeamento de fenômenos naturais até a otimização de rotas em sistemas logísticos.

Sem os dados espaciais, a inteligência artificial pode enfrentar diversos problemas, incluindo:

- **Perda de contexto geográfico:** a ausência de dados espaciais pode levar à falta de compreensão sobre o contexto geográfico, que é crucial para muitas aplicações. Por exemplo, modelos de previsão do tempo,

monitoramento ambiental e planejamento urbano dependem fortemente da localização para fornecer resultados precisos.

- **Modelos menos precisos:** muitos fenômenos têm um componente espacial importante e ignorar essa dimensão pode resultar em modelos menos precisos e úteis, uma vez que a localização pode ser um fator determinante para o comportamento dos dados.
- **Incapacidade de detectar padrões espaciais:** sem dados espaciais, a IA não pode detectar padrões de agrupamento ou dispersão que são essenciais para identificar tendências e anomalias. Isso é particularmente relevante em áreas como epidemiologia, criminologia e análise de mercado.
- **Tomada de decisão ineficiente:** decisões estratégicas em áreas como logística, marketing e gestão de recursos naturais podem ser ineficientes sem a consideração da dimensão espacial. Por exemplo, a otimização de rotas para entrega de mercadorias ou a seleção de locais para novos pontos de venda dependem fortemente da análise espacial.
- **Soluções incompletas ou inadequadas:** problemas complexos como gestão de desastres, planejamento de infraestrutura e desenvolvimento sustentável exigem uma compreensão profunda do espaço geográfico. Sem os dados espaciais, as soluções propostas podem ser incompletas ou inadequadas.
- **Perda de oportunidades de personalização:** aplicações que dependem de personalização, como serviços de recomendação e marketing direcionado, podem perder oportunidades significativas de oferecer conteúdo mais relevantes e específicos ao usuário caso não levem em consideração a localização geográfica.
- **Dificuldades em monitoramento e resposta a emergências:** a falta de dados espaciais pode dificultar o monitoramento eficaz e a resposta rápida a emergências, como desastres naturais e crises de saúde pública. A localização precisa dos eventos é crucial para a mobilização de recursos e intervenções.



- **Limitações na análise de Big Data:** muitos conjuntos de dados grandes incluem um componente espacial. Ignorar essa dimensão pode limitar a análise e a interpretação dos dados, impedindo *insights* valiosos que poderiam ser obtidos a partir da análise espacial.
- **Problemas em mobilidade e transporte:** sem os dados espaciais, soluções de mobilidade e transporte, como a otimização de rotas, a previsão de tráfego e o transporte autônomo, enfrentam desafios significativos em termos de precisão e eficiência.

### 1.1.2 Autocorrelação Espacial

A autocorrelação espacial descreve a tendência de valores similares a se agruparem no espaço, indicando uma dependência espacial entre os valores de uma variável. Em termos simples, significa que, em muitos casos, os dados espaciais próximos tendem a ser mais semelhantes do que dados que estão distantes um do outro. A compreensão da autocorrelação é crucial para a análise espacial, pois ajuda a identificar padrões de agrupamento ou dispersão que são fundamentais para a interpretação de fenômenos geográficos e a tomada de decisões baseada em localização.

### 1.1.3 Heterogeneidade Espacial

A heterogeneidade espacial refere-se à variação espacial dos valores de uma variável, enfatizando que os fenômenos geográficos, muitas vezes, não se distribuem uniformemente pelo espaço. Essa variação pode ser influenciada por uma multiplicidade de fatores ambientais, sociais ou econômicos, resultando em padrões complexos que são essenciais para entender a dinâmica espacial. A análise da heterogeneidade espacial permite aos pesquisadores e profissionais reconhecer áreas de mudança ou pontos de interesse específicos para estudos mais aprofundados ou intervenções direcionadas.

#### 1.1.4 Visualização Espacial

Visualização espacial é uma ferramenta para análise de dados geoespaciais, fornecendo meios para representar dados espaciais através de mapas, gráficos e outras formas visuais. A capacidade de visualizar dados espaciais não apenas facilita a compreensão dos padrões e relações geográficas, mas também apoia a comunicação eficaz de *insights* complexos para públicos variados. Técnicas avançadas de visualização, como mapas interativos, visualizações 3D e mapas de calor, são essenciais para explorar a riqueza dos dados espaciais, permitindo aos analistas revelar tendências ocultas, identificar outliers ou confirmar hipóteses sobre a distribuição espacial dos fenômenos estudados.

A compreensão desses conceitos básicos é essencial para aprofundar no campo da análise exploratória de dados espaciais. Além de fornecerem a base sobre a qual técnicas mais avançadas são construídas, permitindo aos profissionais de engenharia de software e analistas de dados explorar o potencial completo dos dados espaciais para resolver problemas complexos e contribuir para a tomada de decisões baseada em evidências geográficas.

## 2 ESTATÍSTICA ESPACIAL DESCRITIVA

A estatística espacial descritiva constitui um conjunto fundamental de ferramentas e métodos destinados à análise e interpretação de dados geoespaciais (Anselin, 1999). Este segmento aborda as medidas de centralidade e dispersão adaptadas ao contexto espacial, assim como os índices de autocorrelação espacial, proporcionando uma visão abrangente sobre como quantificar padrões e variações em dados que possuem uma dimensão geográfica.

### 2.1 Medidas de Centralidade

- **Média espacial:** diferentemente da média aritmética simples, a média espacial leva em consideração a localização geográfica dos dados, ponderando-os pela sua posição no espaço. Além disso, ajuda a identificar o centro de gravidade de um conjunto de pontos, oferecendo uma visão centralizada de onde os valores tendem a se aglomerar geograficamente.
- **Mediana espacial:** a mediana espacial é o ponto que minimiza a distância total até todos os outros pontos no espaço. Esta medida é particularmente útil para identificar um centro ou localização representativa que é menos sensível a valores extremos ou outliers espaciais.
- **Moda espacial:** embora mais complexa de calcular em contextos espaciais, a moda espacial refere-se ao valor (ou valores) que ocorre com mais frequência dentro de uma região geográfica específica. Identificar a moda pode revelar os valores mais comuns em áreas específicas, destacando padrões de concentração.

#### 2.1.1 Medidas de Dispersão

- **Desvio Padrão Espacial e Variância Espacial:** essas medidas quantificam a variação ou dispersão dos valores em relação à média espacial. Indicam quão espalhados estão os dados geograficamente, ajudando a entender a distribuição espacial e a identificar áreas de alta variabilidade.

- **Coeficiente de Variação Espacial:** essa medida normaliza a variância espacial pela média espacial, proporcionando uma métrica de dispersão relativa que é independente da magnitude dos valores dos dados. É útil para comparar a variabilidade entre conjuntos de dados com diferentes escalas de valores.

### 2.1.2 Índices de Autocorrelação Espacial

- **Índice de Moran:** um dos índices mais comuns, o Índice de Moran, mede a autocorrelação espacial global de uma variável, indicando se há um padrão de agrupamento ou dispersão aleatória no espaço.
- **Índice de Geary:** semelhante ao Índice de Moran, o Índice de Geary foca nas diferenças entre valores adjacentes, oferecendo uma perspectiva sobre a heterogeneidade ou homogeneidade espacial.
- **Índice de LISA (Local Indicators of Spatial Association):** diferentemente dos índices globais, o LISA identifica a autocorrelação espacial em nível local, permitindo a detecção de clusters ou outliers espaciais.

Estas ferramentas da estatística espacial descritiva são essenciais para explorar e entender a natureza dos dados geoespaciais. Ao aplicar essas medidas e índices, os analistas podem desvendar padrões espaciais complexos, oferecendo *insights* valiosos para a tomada de decisões em diversos campos.

## 2.2 Exemplo Aplicado à Linguagem R

Para elaborar um exemplo completo que utiliza conceitos da estatística espacial descritiva na Linguagem R, vamos utilizar um conjunto de dados fictício que representa localizações geográficas (coordenadas) e um valor associado a cada localização, como poderia ser, por exemplo, o preço médio de imóveis por área. Utilizaremos pacotes como `sp`, `rgeos`, e `spdep` para calcular medidas de centralidade, dispersão e índices de autocorrelação espacial.

## A estatística de coordenadas e geolocalizações

```
# Carregar bibliotecas necessárias
library(sp)
library(spdep)

# Carregar conjunto de dados meuse
data(meuse)

# Converter o conjunto de dados meuse para um objeto
SpatialPointsDataFrame
coordinates(meuse) <- c("x", "y")

# Calcular os vizinhos mais próximos usando a distância de
Euclides
knn <- knearneigh(coordinates(meuse), k = 6)

# Converter os vizinhos mais próximos em uma matriz de pesos
espaciais
listw <- knn2nb(knn)

# Tornar a matriz de pesos espaciais em uma matriz binária
de pesos espaciais
listw <- nb2listw(listw, style = "B")

# Medidas de Centralidade
# Calcular a média espacial
media_espacial <- mean(meuse$zinc, na.rm = TRUE)
print(media_espacial)

# Calcular a mediana espacial
mediana_espacial <- median(meuse$zinc)
print(mediana_espacial)

# Medidas de Dispersão
# Calcular o desvio padrão espacial e a variância espacial
desvio_padrao_espacial <- sd(meuse$zinc)
print(desvio_padrao_espacial)
variacao_espacial <- var(meuse$zinc)
print(variacao_espacial)

# Calcular o coeficiente de variação espacial
coeficiente_variacao_espacial <- desvio_padrao_espacial /
media_espacial
print(coeficiente_variacao_espacial)

# Índices de Autocorrelação Espacial
# Calcular o Índice de Moran
indice_moran <- moran.test(meuse$zinc, listw)
print(indice_moran)

# Calcular o Índice de Geary
indice_geary <- geary.test(meuse$zinc, listw)
```

```
print(indice_geary)

# Calcular o Índice de LISA
indice_lisa <- localmoran(meuse$zinc, listw)
print(indice_lisa)
```

Código-fonte 1 - Exemplo da estatística espacial descritiva na Linguagem R  
Fonte: Elaborado pelo autor (2025)

O código cria uma estrutura básica para trabalhar com dados espaciais em R, incluindo cálculos de medidas de centralidade e dispersão, além da preparação para análise de autocorrelação espacial. Deve-se notar que a análise de autocorrelação espacial, como o Índice de Moran, exige dados e contextos mais complexos para serem significativos, especialmente em termos de definição de vizinhança e estrutura espacial.

### 2.3 Visualização de Dados Espaciais

A visualização de dados espaciais é um pilar fundamental na análise exploratória de dados espaciais, fornecendo *insights* visuais que apoiam a interpretação e a comunicação de padrões, tendências e relações presentes nos dados geoespaciais (Anselin, 1999). Esta seção explora três técnicas chave de visualização: mapas, gráficos de dispersão espacial e diagramas de Moran, cada uma oferecendo uma perspectiva única para analisar e apresentar dados espaciais.

#### 2.3.1 Mapas

Os mapas são a forma mais direta e intuitiva de representar dados espaciais, permitindo a visualização da distribuição geográfica de variáveis, a identificação de padrões espaciais e a compreensão do contexto geográfico das análises. Mapas podem variar desde simples representações pontuais até mapas temáticos complexos que incorporam várias camadas de dados, como densidade populacional, uso do solo ou índices socioeconômicos. São ferramentas essenciais para qualquer análise espacial, facilitando a identificação de agrupamentos, anomalias e tendências espaciais.

### 2.3.2 Gráficos de Dispersão Espacial

Gráficos de dispersão espacial estendem o conceito de gráficos de dispersão tradicionais ao incorporar a dimensão espacial, mostrando a relação entre duas variáveis enquanto consideram a localização espacial dos pontos de dados. Esses gráficos podem revelar padrões de correlação espacial ou ajudar a identificar áreas onde as relações entre variáveis são mais fortes ou mais fracas. Ao visualizar como as variáveis interagem em diferentes locais, os gráficos de dispersão espacial são uma ferramenta para a análise exploratória, especialmente em estágios iniciais de pesquisa quando hipóteses estão sendo formuladas e testadas.

### 2.3.3 Diagramas de Moran

Os diagramas de Moran são uma ferramenta específica para visualizar a autocorrelação espacial entre os valores de uma variável, oferecendo uma representação gráfica da medida de Moran's I. Geralmente plotam o valor de uma variável contra sua média espacial ponderada (lag espacial), com a linha de referência indicando a expectativa sob a hipótese de independência espacial. O lag espacial refere-se à medida que captura o efeito de influência ou correlação de uma variável em relação às suas localizações geográficas vizinhas, quantificando como os valores de uma variável em um ponto específico são relacionados aos valores dessa mesma variável nos pontos ao seu redor no espaço. Pontos que se desviam significativamente desta linha podem indicar a presença de autocorrelação espacial positiva ou negativa. Diagramas de Moran são particularmente úteis para identificar padrões de autocorrelação local, permitindo aos analistas explorar e investigar as dinâmicas espaciais subjacentes aos dados.

Cada uma dessas técnicas de visualização desempenha um papel único na análise de dados espaciais, permitindo aos pesquisadores e analistas explorar, interpretar e comunicar a complexidade e os padrões inerentes aos dados geoespaciais. Ao utilizar mapas, gráficos de dispersão espacial e diagramas de Moran, é possível obter uma compreensão profunda da estrutura espacial dos dados, facilitando a identificação de relações espaciais significativas, padrões de agrupamento e potenciais áreas de interesse para investigações futuras. Essas

ferramentas de visualização não apenas aprimoram a análise exploratória de dados espaciais, mas também promovem uma comunicação eficaz dos resultados, crucial para a tomada de decisões baseada em evidências em diversos domínios aplicados.

### 2.4 Exemplo Aplicado à Linguagem R

Para ilustrar a visualização de dados espaciais na Linguagem R:

```
library(geostan)
library(ggplot2)
library(gridExtra)
data("georgia")

# A função sp_diag() é usada para visualizar a distribuição
espacial de uma variável
# (neste caso, a porcentagem de faculdades) em um mapa
geográfico (representado pelo objeto georgia).
# Permite a análise de padrões espaciais e a identificação
de agrupamentos ou dispersões dessa variável
# na área geográfica estudada.
sp_diag(georgia$college, georgia, name = "College (%)")

# Cria uma matriz de pesos espaciais no estilo binário a
partir do objeto georgia.
# Essa matriz será usada para avaliar a autocorrelação
espacial da variável de porcentagem de faculdades.
W <- shape2mat(georgia, style = "W")

# Gera um gráfico de dispersão para o Diagrama de Moran,
utilizando a variável de porcentagem de faculdades
# e a matriz de pesos espaciais W para avaliar a
autocorrelação espacial.
moran_plot(georgia$college, W)

# Calcula o índice de autocorrelação espacial de Moran para
a variável de porcentagem de faculdades,
# utilizando a matriz de pesos espaciais W.
mc(georgia$college, W)

# Cria outra matriz de pesos espaciais no estilo binário a
partir do objeto georgia.
# Essa matriz será usada para avaliar a autocorrelação
espacial da variável de porcentagem de faculdades.
A <- shape2mat(georgia, "B")

# Gera outro gráfico de dispersão para o Diagrama de Moran,
utilizando a mesma variável e uma matriz de
```



## A estatística de coordenadas e geolocalizações

```
# pesos espaciais diferente (A) para avaliar a  
autocorrelação espacial.  
moran_plot(georgia$college, A)
```

Código-fonte 2 - Exemplo de visualização de dados em R

Fonte: Elaborado pelo autor (2025)

Este código proporciona um exemplo de como visualizar dados espaciais em R, desde a criação de um mapa para representar valores variáveis em localizações específicas até a análise de relações espaciais através de gráficos de dispersão e o diagrama de Moran para autocorrelação espacial. Essas técnicas são essenciais para explorar visualmente dados espaciais, identificar padrões e correlações espaciais e comunicar efetivamente os resultados de análises espaciais.

## REFERÊNCIAS

ANSELIN, L. **Spatial data analysis and GIS**. New York: CRC Press, 1999.

HAINING, R. **Spatial data analysis: Theory and practice**. England: Cambridge University Press, 2003.

EMAP