



**UNIVERSIDADE ESTADUAL DA PARAÍBA  
CAMPUS I - CAMPINA GRANDE  
CENTRO DE CIÊNCIAS E TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA  
CURSO DE BACHARELADO EM ESTATÍSTICA**

**WESLEY ROSENO SARAIVA**

**ANÁLISE DA DISTRIBUIÇÃO ESPACIAL POR MEIO DE PROCESSOS PONTUAIS  
DAS ESCOLAS DO DISTRITO FEDERAL.**

**CAMPINA GRANDE - PB**

**2023**

WESLEY ROSENO SARAIVA

**ANÁLISE DA DISTRIBUIÇÃO ESPACIAL POR MEIO DE PROCESSOS PONTUAIS  
DAS ESCOLAS DO DISTRITO FEDERAL.**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Estatística.

**Orientador:** Prof. Dr. Ricardo Alves de Olinda

**CAMPINA GRANDE - PB  
2023**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

S243a Saraiva, Wesley Roseno.  
Análise da distribuição espacial por meio de processos pontuais das escolas do Distrito Federal. [manuscrito] / Wesley Roseno Saraiva. - 2023.  
26 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2024.

"Orientação : Prof. Dr. Ricardo Alves de Olinda, Departamento de Estatística - CCT. "

1. Estatística espacial. 2. Análise espacial. 3. Processos pontuais. I. Título

21. ed. CDD 519.5

WESLEY ROSENO SARAIVA

ANÁLISE DA DISTRIBUIÇÃO ESPACIAL POR MEIO DE PROCESSOS PONTUAIS DAS  
ESCOLAS DO DISTRITO FEDERAL.

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Estatística.

Trabalho aprovado em 18 de julho de 2023.

**BANCA EXAMINADORA**



---

Prof. Dr. Ricardo Alves de Olinda (Orientador)  
Universidade Estadual da Paraíba (UEPB)



---

Prof. Dr. Edwirde Luiz Silva Camêlo  
Universidade Estadual da Paraíba (UEPB)



---

Prof. Dr. Mácio Augusto de Albuquerque  
Universidade Estadual da Paraíba (UEPB)

Dedico a todos que contribuíram com a minha formação acadêmica, com as condições sociais necessárias para o acesso ao conhecimento. Dedico também aos cientistas que por meio de pesquisas avançam o conhecimento científico, como aos professores no ensino superior, médio e fundamental, além dos divulgadores científicos dos mais diversos meios de comunicação.

## **AGRADECIMENTOS**

Agradeço aos discentes e docentes do curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, que compartilham os seus conhecimentos, mesmo não havendo obrigação.

“As fábulas devem ser ensinadas como fábulas, mitos como mitos e milagres, como fantasias poéticas. Ensinar superstições como se fossem verdadeiras é terrível. A mente da criança aceita e acredita nelas, e somente com muita dor, e talvez tragédia, ele pode se livrar delas ao longo dos anos.”

(Hipátia de Alexandria)

## RESUMO

Este trabalho aborda um desafio fundamental na educação: a compreensão eficaz da distribuição espacial das escolas e seu impacto na alocação de recursos e no planejamento de políticas públicas. Tradicionalmente, as análises educacionais não consideram o componente espacial da distribuição das escolas, devido esses dados terem visibilidade só mais recentemente com a popularização da internet e o uso de dispositivos com GPS, como os *smartphones*, limitando assim a profundidade da nossa compreensão. Devido então a ausência destes dados de geolocalização, não era possível uma análise utilizando a abordagem dos processos pontuais, até o atual momento. Através desta análise, objetiva-se compreender se a distribuição espacial das escolas segue algum padrão específico ou se é aleatória, assim fornecendo, *insights* valiosos para a tomada de decisões políticas e estratégicas na área da educação. Os principais métodos usados para a análise incluem o teste de Kolmogorov-Smirnov, a Função K de Ripley, e a estimativa da taxa de eventos por meio de um Estimador de Kernel.

**Palavras-chave:** Estatística Espacial; Análise Espacial; Processos Pontuais.

## **ABSTRACT**

This work addresses a fundamental challenge in education: effectively understanding the spatial distribution of schools and its impact on resource allocation and public policy planning. Traditionally, educational analyzes do not consider the spatial component of school distribution, as this data only became visible more recently with the popularization of the internet and the use of GPS devices, such as smartphones, thus limiting the depth of our understanding. Due to the absence of this geolocation data, an analysis using the point process approach was not possible until now. Through this analysis, the aim is to understand whether the spatial distribution of schools follows a specific pattern or whether it is random, thus providing valuable insights for making political and strategic decisions in the area of education. The main methods used for the analysis include the Kolmogorov-Smirnov test, Ripley's K Function, and event rate estimation using a Kernel Estimator.

**Keywords:** Spatial Statistics; Spatial Analysis; Point Processes.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Valor observado na função K de Ripley. . . . .	15
Figura 2 – Brasil. (A) Em verde: Região Centro-Oeste do Brasil. (B) Em vermelho: Distrito Federal. . . . .	16
Figura 3 – Regiões administrativas do Distrito Federal. . . . .	17
Figura 4 – Escolas do Distrito Federal, representadas por pontos azuis. . . . .	17
Figura 5 – Visão com perspectiva tridimensional da coordenada 'X'. . . . .	18
Figura 6 – Teste de Kolmogorov-Smirnov para a coordenada X. . . . .	19
Figura 7 – Visão com perspectiva tridimensional da coordenada 'Y'. . . . .	20
Figura 8 – Teste de Kolmogorov-Smirnov para a coordenada Y. . . . .	20
Figura 9 – Kernel Gaussiano com o método de Silverman (SILVERMAN, 1986). . . . .	21
Figura 10 – Função K de Ripley com dados de geolocalização das escolas no Distrito Federal. . . . .	22
Figura 11 – Intensidade de agrupamento das escolas no Distrito Federal. . . . .	23

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>10</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> . . . . .	<b>11</b>
<b>2.1</b>	<b>Completa Aleatoriedade Espacial</b> . . . . .	<b>11</b>
<b>2.2</b>	<b>Teste de Kolmogorov-Smirnov</b> . . . . .	<b>12</b>
<b>2.3</b>	<b>Estimador de Kernel</b> . . . . .	<b>13</b>
<b>2.4</b>	<b>Função K de Ripley</b> . . . . .	<b>14</b>
<b>3</b>	<b>MATERIAL</b> . . . . .	<b>16</b>
<b>3.1</b>	<b>Base de dados</b> . . . . .	<b>16</b>
<b>3.1.1</b>	<i>Regiões Administrativas</i> . . . . .	<i>16</i>
<b>3.1.2</b>	<i>Escolas</i> . . . . .	<i>17</i>
<b>4</b>	<b>APLICAÇÃO</b> . . . . .	<b>18</b>
<b>4.1</b>	<b>Teste de Kolmogorov-Smirnov</b> . . . . .	<b>18</b>
<b>4.2</b>	<b>Estimador de Kernel</b> . . . . .	<b>21</b>
<b>4.3</b>	<b>Função K de Ripley</b> . . . . .	<b>21</b>
<b>5</b>	<b>CONCLUSÃO</b> . . . . .	<b>24</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>25</b>

## 1 INTRODUÇÃO

Nos últimos anos, o avanço da coleta de dados de geolocalização proporcionou uma quantidade crescente de informações espacialmente distribuídas, o que ampliou o alcance de diversas áreas de estudo, incluindo a educação. No entanto, apesar do aumento no volume dos dados, a exploração efetiva apresenta desafios significativos. Neste contexto, um problema chave é a falta de métodos eficazes para a análise de dados geoespaciais no setor educacional, particularmente em termos da distribuição espacial das escolas e como isso afeta a alocação de recursos e o planejamento de políticas públicas.

O Brasil, com suas dimensões continentais, apresenta uma distribuição escolar heterogênea, reforçando a necessidade de ferramentas apropriadas para analisar essa distribuição de forma aprofundada e compreensível. Tradicionalmente, as análises na educação não consideram o atributo da distribuição espacial das escolas, limitando assim nossa compreensão de como a geolocalização pode afetar a alocação de recursos e o planejamento educacional.

A análise espacial, em particular o estudo de processos pontuais, surge como uma ferramenta promissora para preencher essa lacuna. Os processos pontuais representam uma classe de modelos estocásticos que descrevem a distribuição espacial de eventos que ocorrem no espaço. Sendo que para este estudo foi escolhido o Distrito Federal, por haver dados públicos, de órgãos confiáveis e de fácil acesso. Nesse caso, cada escola no Distrito Federal é representada como um ponto em um espaço geográfico.

Vários fatores podem influenciar a distribuição espacial das escolas, como decisões de planejamento urbano, demanda da população e políticas educacionais. Assim, ao adotar a abordagem de processos pontuais, adquire-se informações valiosas sobre a estrutura de distribuição das escolas e fornecer *insights* significativos para tomadas de decisão políticas e estratégicas na área da educação.

A abordagem de processos pontuais é utilizada com sucesso em diversas áreas de estudo, onde os fenômenos de interesse são representados como pontos em um espaço. Por exemplo, as técnicas de processos pontuais são aplicadas para investigar a distribuição de unidades prisionais no Brasil (SOUZA, 2018), e na análise espacial de espécies arbóreas (OLIVEIRA, 2022).

O objetivo deste estudo é, então, analisar a distribuição espacial das escolas, aprofundando a compreensão de como elas estão espacialmente distribuídas. Será aplicada a análise de processos pontuais para estudar a localização das escolas no Distrito Federal do Brasil. Investigado se a distribuição espacial das escolas segue algum padrão específico ou se é aleatória. Com a compreensão desse padrão, se terá *insights* significativos para políticas públicas de educação, planejamento urbano e alocação de recursos.

## 2 FUNDAMENTAÇÃO TEÓRICA

Em contextos educacionais, a análise espacial é uma ferramenta valiosa para entender a distribuição e localização das escolas, em contraste a muitas abordagens tradicionais para análise de dados, como a análise de variância e modelos lineares, que não consideram o espaço ou a localização das observações (OYANA, 2020). Apesar de serem métodos consolidados e amplamente utilizados em várias áreas, tais abordagens podem não ser adequadas para estudar fenômenos que possuem uma forte componente espacial, como a localização de escolas.

A ideia fundamental por trás da estatística espacial é definida na primeira lei da geografia: “Todas as coisas estão relacionadas com todas as outras, mas coisas próximas estão mais relacionadas do que coisas distantes” (Tobler W., 1970). Isto implica que a localização de uma escola pode ser influenciada pela localização de outras escolas próximas. Portanto, é importante incorporar a dimensão espacial na análise de dados para obter uma compreensão mais completa da distribuição das escolas.

A análise de processos pontuais é uma abordagem que considera explicitamente a localização espacial das observações. Nesta abordagem, cada escola é representada como um ponto em um espaço geográfico, e o objetivo é analisar a distribuição desses pontos. A análise de processos pontuais é utilizada para investigar a correlação entre pontos. Esta abordagem tem origem na investigação de John Snow no século XIX, onde ele intuitivamente incorporou a categoria espaço nas suas análises para entender a propagação de uma epidemia de cólera em Londres (OLINDA, 2008).

A análise espacial de dados está inserida em um contexto mais amplo, que considera a localização espacial do fenômeno em estudo. Este contexto permite estudar, explorar e modelar fenômenos geográficos (DRUCK et al., 2004). A análise espacial é composta por um conjunto de procedimentos encadeados cuja finalidade é a escolha de um modelo inferencial que considera explicitamente os relacionamentos espaciais presentes no fenômeno. Assim, a análise de processos pontuais se encaixa dentro deste contexto mais amplo de análise espacial de dados.

Neste trabalho, a análise de processos pontuais será aplicada para estudar a localização das escolas no Distrito Federal do Brasil. O objetivo é investigar se a distribuição espacial das escolas segue algum padrão específico ou se é aleatória. Com a compreensão desses padrões, é possível fornecer *insights* úteis para políticas públicas de educação, planejamento urbano e alocação de recursos.

### 2.1 Completa Aleatoriedade Espacial

A Completa Aleatoriedade Espacial é uma propriedade fundamental na análise de processos pontuais. Ela representa a hipótese de que cada ponto é independente dos outros e está uniformemente distribuído no espaço de estudo. Matematicamente, um processo pontual  $\Phi$  no espaço Euclidiano  $R^d$  é dito ter Completa Aleatoriedade Espacial se a localização de cada ponto é independente e uniformemente distribuída (DIGGLE, 2013).

Em termos mais simples, um padrão de pontos exibe Completa Aleatoriedade Espacial se não houver nenhuma interação entre os pontos, isto é, a presença de um ponto em uma localização não afeta a probabilidade de existência de pontos em outras localidades. Na análise de processos pontuais, é frequentemente usado como um modelo nulo, fornecendo uma base de comparação para determinar se um padrão espacial observado exibe algum tipo de estrutura ou agrupamento.

Este conceito é frequentemente associado ao Processo de Poisson, um modelo estatístico que descreve eventos ocorrendo de forma independente a uma taxa constante. No contexto de processos pontuais, um Processo de Poisson espacial é um processo pontual com Completa Aleatoriedade Espacial.

No contexto deste trabalho, a hipótese testada é a de que a distribuição das escolas no Distrito Federal é completamente aleatória, ou seja, a localização de uma escola não influencia a localização de outras escolas. Isso se traduziria em um cenário onde as escolas estão uniformemente distribuídas em todo o Distrito Federal, sem padrões de agrupamento ou segregação.

Se a hipótese de Completa Aleatoriedade Espacial for rejeitada, isso significará que existe algum tipo de padrão na distribuição das escolas, o que poderia ser o resultado de vários fatores, como políticas de planejamento urbano, demanda da população ou políticas educacionais.

Um exemplo de uma situação no contexto deste trabalho seria se, ao traçar um mapa do Distrito Federal, fosse identificado escolas uniformemente distribuídas em toda a região, sem grandes aglomerações ou vazios de escolas em áreas específicas. Isso significaria que a probabilidade de encontrar uma escola em qualquer local específico do Distrito Federal é a mesma, independentemente da localização das outras escolas.

## 2.2 Teste de Kolmogorov-Smirnov

O teste de Kolmogorov-Smirnov (K-S) é uma metodologia não-paramétrica utilizada para comparar uma amostra com uma distribuição teórica (teste de uma amostra) ou duas amostras entre si (teste de duas amostras). Em particular, para processos pontuais, o teste K-S pode ser empregado para testar a hipótese de que os pontos seguem uma distribuição específica ao longo do espaço (STEPHENS, 1992).

A estatística do teste K-S é definida como o supremo das diferenças absolutas entre a Função de Distribuição Acumulada Empírica (FDAC) e a Função de Distribuição Acumulada Teórica (FDAT), isto é, a máxima diferença entre as duas funções de distribuição. Para uma amostra de tamanho  $n$ , se  $F_n(x)$  é a FDAC e  $F(x)$  é a FDAT, a estatística do teste K-S é dada por:

$$D_n = \sup_x |F_n(x) - F(x)| \quad (2.1)$$

onde  $\sup_x$  denota o supremo (ou máximo) da diferença absoluta entre  $F_n(x)$  e  $F(x)$ .

No contexto desta tese, as hipóteses para o teste K-S são definidas da seguinte forma:

- $H_0$ : A distribuição das escolas no Distrito Federal é completamente aleatória. Essa é a hipótese nula, que supõe que a localização de uma escola não influencia a localização de outras escolas.
- $H_1$ : A distribuição das escolas no Distrito Federal não é completamente aleatória. Essa é a hipótese alternativa, que supõe que exista algum tipo de padrão na distribuição das escolas.

Se o valor de  $D_n$  for significativamente grande, a hipótese nula  $H_0$  é rejeitada, indicando que a amostra não segue a distribuição teórica, portanto assumisse a hipótese alternativa  $H_1$ , de que existe algum tipo de padrão na distribuição das escolas. Vale ressaltar que o teste K-S requer um número relativamente grande de pontos de dados para rejeitar adequadamente a hipótese nula.

### 2.3 Estimador de Kernel

Um componente, chave na análise de processos pontuais, é a estimativa da taxa de eventos subjacentes. Uma abordagem comum para fazer isso é através do uso de um Estimador de Kernel.

Um Estimador de Kernel é um método não-paramétrico usado para estimar a função de densidade de probabilidade de uma variável aleatória. Em termos gerais, um Estimador de Kernel suaviza a contagem de eventos em uma região local, fornecendo uma estimativa da densidade de eventos naquela região (CHEN, 2017).

A fórmula geral para um Estimador de Kernel é:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-X_i}{h} \right)^2} \quad (2.2)$$

Em primeiro lugar,  $x$  refere-se ao ponto específico no qual a densidade é estimada. Em outras palavras, este é o local onde se deseja entender a densidade de probabilidade subjacente à nossa amostra de dados.

Em seguida, temos  $n$ , que denota o número total de pontos de dados em nossa amostra. Este valor é essencial para normalizar a estimativa, garantindo que ela seja um reflexo preciso da densidade de probabilidade, independentemente do tamanho da amostra.

Os  $X_i$  representam os pontos de dados em si. Estes são os valores individuais que estão sendo utilizados para calcular a estimativa de densidade. Cada um desses pontos contribui para a estimativa global de densidade.

Já a escolha da função Kernel e da largura de banda são dois aspectos importantes na aplicação de um Estimador de Kernel. Existem várias funções Kernel que podem ser usadas, incluindo o Kernel Gaussiano, o Kernel Epanechnikov, entre outros. A escolha da largura de banda pode ser feita por vários métodos, porém é comumente utilizado o método de Silverman (SILVERMAN, 1986). A função de Kernel gaussiano tem uma forma de sino suave, com maior peso nos pontos próximos ao centro. É amplamente utilizada devido às suas propriedades suavizantes e simétricas.

Na análise de processos pontuais, o Estimador de Kernel pode ser usado para estimar a taxa de eventos subjacente em um processo pontual. Esta taxa de eventos pode então ser usada para análises posteriores, como a modelagem de dependências temporais ou espaciais nos dados.

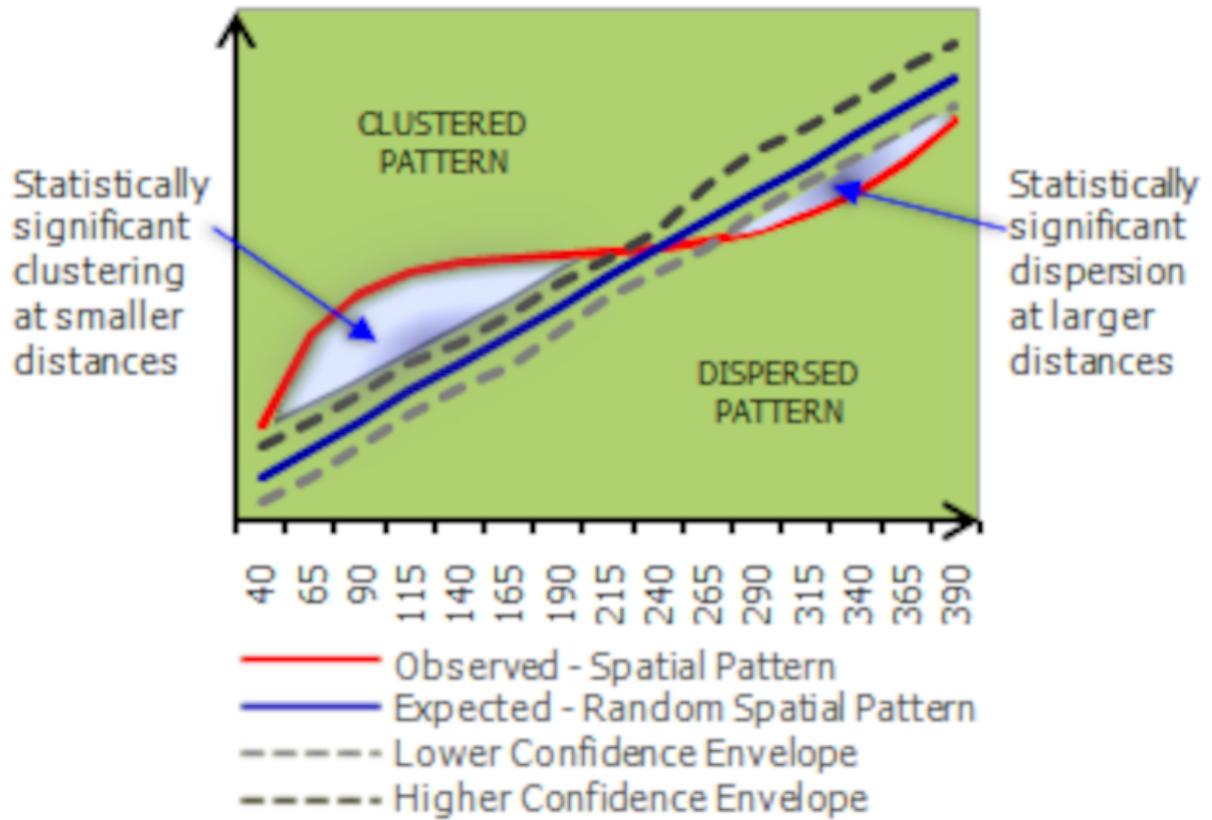
## 2.4 Função K de Ripley

A função K de Ripley é uma ferramenta estatística proposta por Brian D. Ripley em 1977 (RIPLEY, 1977), utilizada para analisar padrões espaciais de dados de pontos. Ela é uma maneira eficaz de sumarizar a dependência espacial, ou seja, o agrupamento ou dispersão de características, em uma variedade de distâncias. Em muitos estudos de análise de padrões de características, a seleção de uma escala de análise apropriada é necessária. Ao explorar padrões espaciais em múltiplas distâncias e escalas espaciais, os padrões mudam, muitas vezes refletindo a dominância de processos espaciais específicos. A função K de Ripley ilustra como o agrupamento espacial ou dispersão de centros de características muda quando o tamanho do bairro muda.

Ao utilizar a função K de Ripley, especifica-se o número de distâncias a serem avaliadas e, opcionalmente, uma distância inicial e/ou incremento de distância. Com essas informações, a ferramenta calcula o número médio de características vizinhas associadas a cada característica; características vizinhas são aquelas mais próximas do que a distância que está sendo avaliada. À medida que a distância de avaliação aumenta, cada característica geralmente terá mais vizinhos. Se o número médio de vizinhos para uma distância de avaliação específica for maior do que a concentração média de características em toda a área de estudo, a distribuição é considerada agrupada nessa distância.

Quando o valor K observado é maior que o valor K esperado para uma distância específica, a distribuição é mais agrupada do que uma distribuição aleatória naquela distância. Quando o valor K observado é menor que o valor K esperado, a distribuição é mais dispersa do que uma distribuição aleatória naquela distância. Quando o valor K observado é maior do que o valor do envelope de confiança superior, o agrupamento espacial para aquela distância é estatisticamente significativo. Quando o valor K observado é menor do que o valor do envelope de confiança inferior, a dispersão espacial para aquela distância é estatisticamente significativa (ArcGIS Pro, 2023).

Figura 1 – Valor observado na função K de Ripley.



Fonte: ArcGIS Pro (<https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-multi-distance-spatial-cluster-analysis-ripl.htm>, acessado em junho de 2023).

### 3 MATERIAL

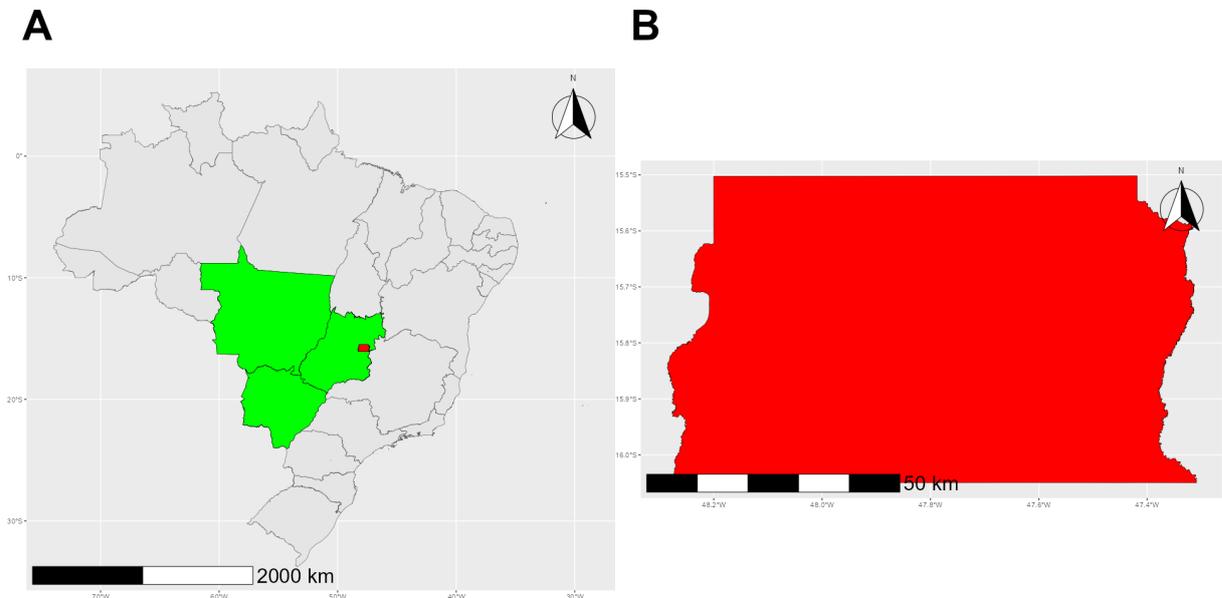
#### 3.1 Base de dados

A primeira atividade do trabalho foi a pesquisa por informações de geolocalização das escolas que possuem registro na secretaria de educação do Distrito Federal, e também a informações de geolocalização do próprio Distrito Federal. Os dados consistem de dois arquivos disponibilizados pelo Sistema Distrital de Informações Ambientais – SISDIA (Secretaria de Estado de Saúde do Distrito Federal, 2023). Sendo o primeiro constituído pela geolocalização do Distrito Federal e das suas regiões administrativas<sup>1</sup>. E o segundo *Shapefile* constituído pela geolocalização das escolas no Distrito Federal.<sup>2</sup>

##### 3.1.1 Regiões Administrativas

As regiões administrativas do Distrito Federal são divisões administrativas do Distrito Federal. O Distrito Federal é uma das 27 unidades federativas do Brasil. Localizado na Região Centro-Oeste (Figura 2), é a menor unidade federal brasileira e a única que não possui municípios, sendo dividida em 35 regiões administrativas (Figura 3). A capital federal do Brasil, Brasília, que também é a sede do governo do Distrito Federal, está localizada em seu território, sendo identificada também pelo nome de “Plano Piloto”.

Figura 2 – Brasil. (A) Em verde: Região Centro-Oeste do Brasil. (B) Em vermelho: Distrito Federal.



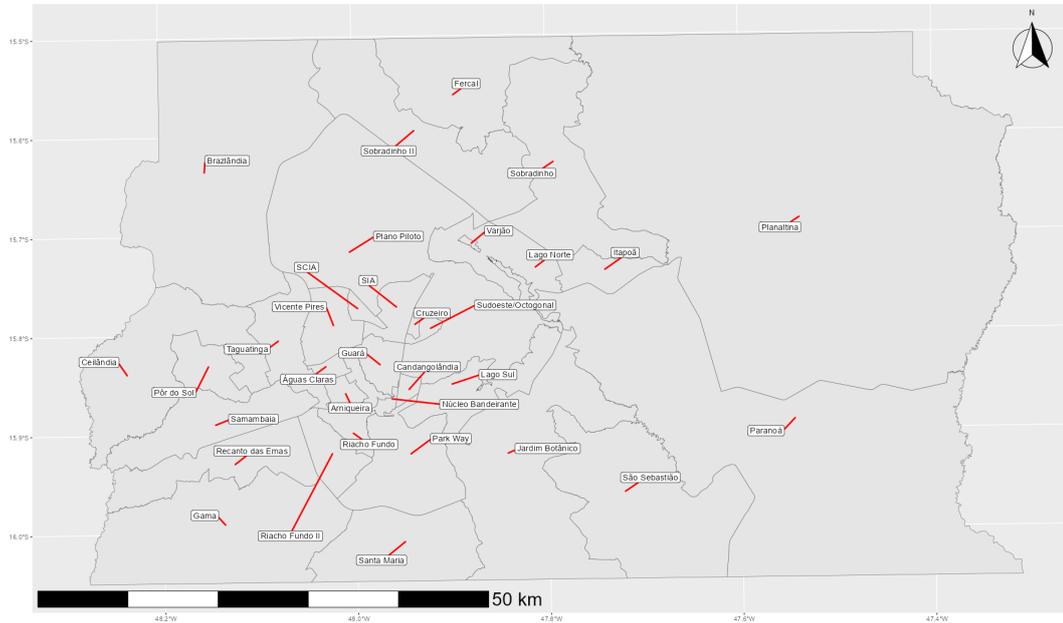
Fonte: Elaborado pelo autor, 2023.

<sup>1</sup> <<https://sisdia.df.gov.br/portal/home/item.html?id=ae46ce239eb14d66abd1d967e22bbdf4>>

<sup>2</sup> <<https://sisdia.df.gov.br/portal/home/item.html?id=0ad4d607b99b4bfdb144b4e6d255ef7e>>

O Distrito Federal possui 5.760 km<sup>2</sup>, sendo a menor unidade federativa do Brasil (IBGE, 2022), com mais de 3,8 milhões de habitantes (IBGE, 2022).

Figura 3 – Regiões administrativas do Distrito Federal.

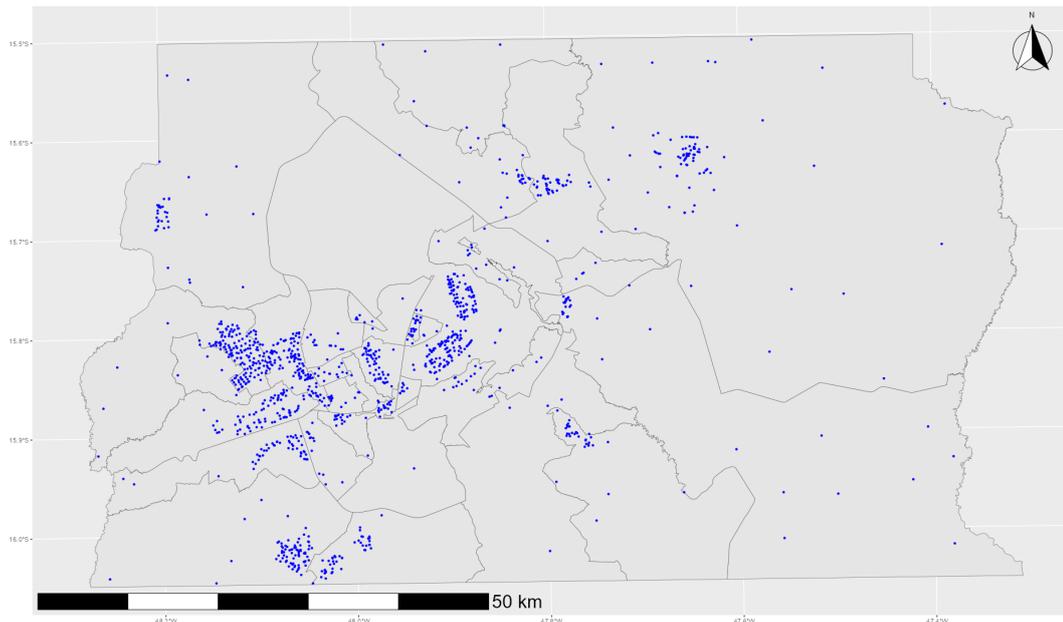


Fonte: Elaborado pelo autor, 2023.

### 3.1.2 Escolas

Os dados referentes às escolas são de 1196 escolas do Distrito Federal, com as informações da localização espacial de cada uma, conforme Figura 4.

Figura 4 – Escolas do Distrito Federal, representadas por pontos azuis.



Fonte: Elaborado pelo autor, 2023.

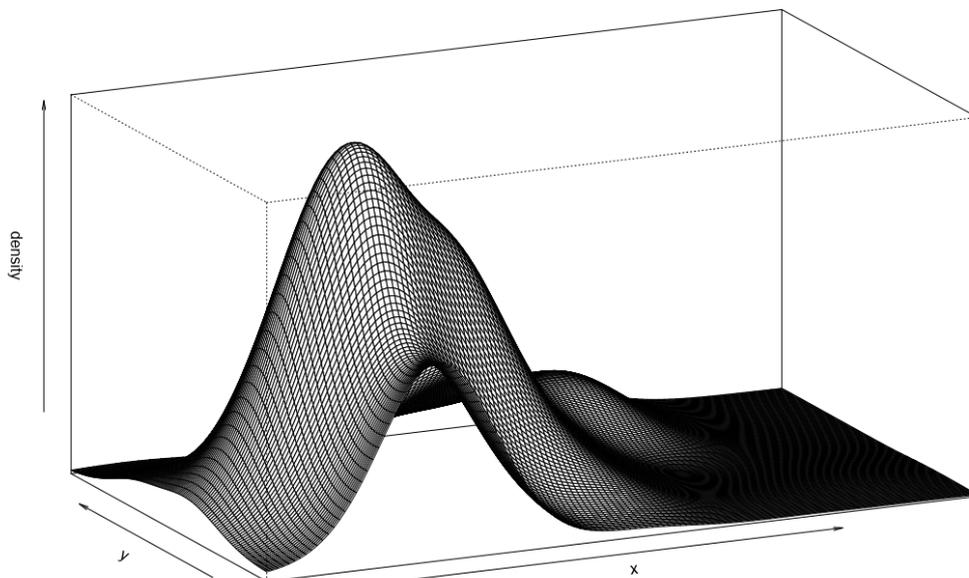
## 4 APLICAÇÃO

Após a obtenção dos dados no formato *Shapefile*, utilizou-se o software R versão 4.1.3, para análise dos dados. Foi verificado a presença de 1196 escolas distribuídas por todo o Distrito Federal, conforme apresentado na Figura 4. Observa-se que as escolas se distribuem em toda área de estudo com indícios de heterogeneidade espacial, com a região Centro-Oeste do Distrito Federal apresentado maior concentração de escolas, porém tal observação visual não é suficiente para definir com confiança a distribuição espacial que ocorre nas escolas, sendo necessária a utilização dos métodos de processos pontuais (OYANA, 2020).

### 4.1 Teste de Kolmogorov-Smirnov

O primeiro teste a ser feito foi o teste de Kolmogorov-Smirnov, com 5% de significância, para verificar se a distribuição dos dados refuta a hipótese de Completa Aleatoriedade Espacial. O teste foi aplicado em cada uma das coordenadas espaciais 'X' e 'Y', que se referem a longitude e latitude, respectivamente, na projeção UTM (ZANETTI, 2017).

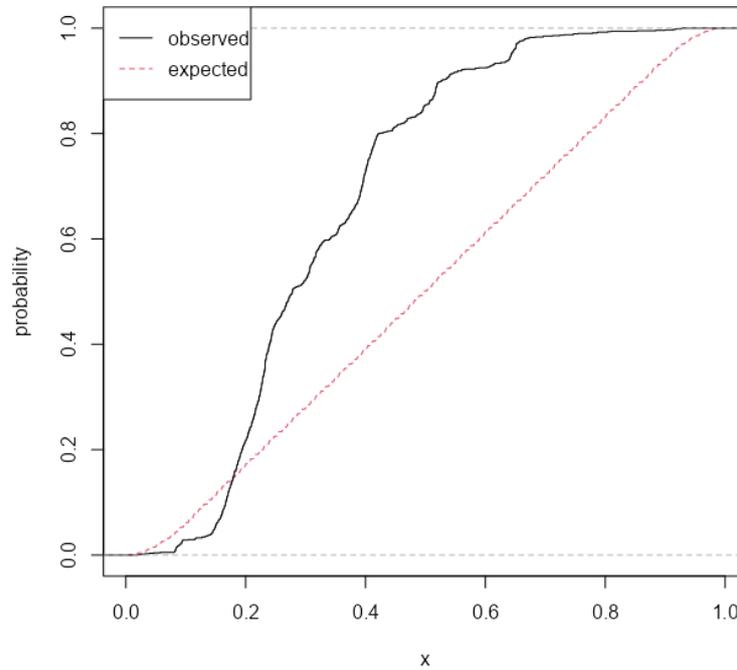
Figura 5 – Visão com perspectiva tridimensional da coordenada 'X'.



Fonte: Elaborado pelo autor, 2023.

É verificado na Figura 5 que há uma maior concentração de escolas na região da coordenada espacial 'X' com valores abaixo do valor médio possível para essa coordenada.

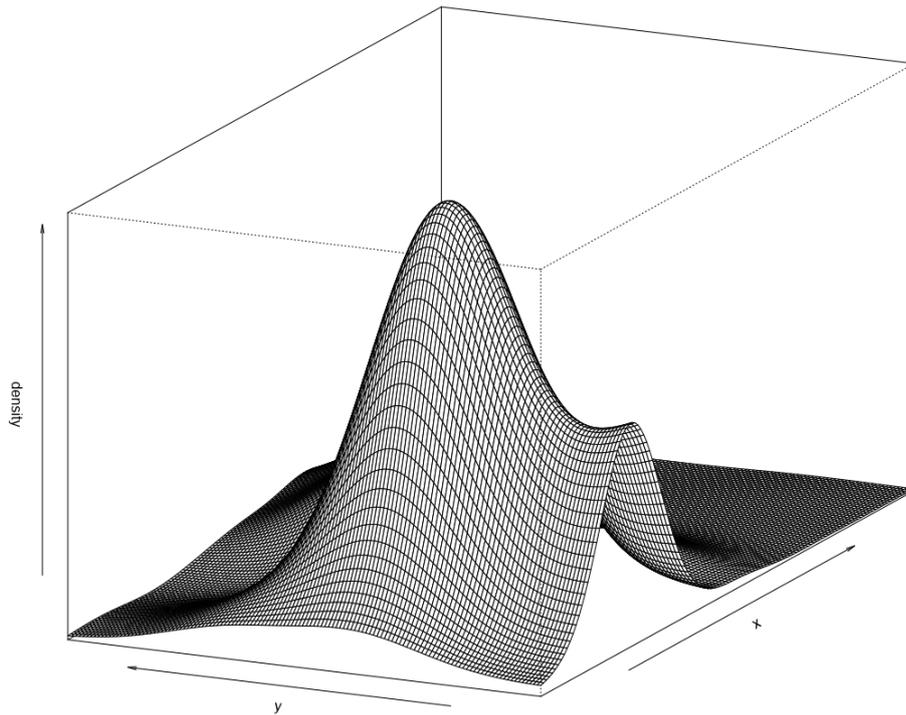
Figura 6 – Teste de Kolmogorov-Smirnov para a coordenada X.



Fonte: Elaborado pelo autor, 2023.

É verificado na Figura 6 que ao efetuar o teste de Kolmogorov-Smirnov para a coordenada espacial 'X', existe uma tendência de acúmulo da presença de escolas até o valor de 0,4 da coordenada espacial 'X' ao colocar a mesma em uma escala de 0 a 1. Onde para a coordenada 'X' o supremo das diferenças absolutas entre a Função de Distribuição Acumulada Empírica (FDAC) e a Função de Distribuição Acumulada Teórica (FDAT) atingiu o valor de 0.38526, e com um p-value menor que  $10^{-15}$ , apresentando-se a conclusão de rejeição da hipótese de Completa Aleatoriedade Espacial.

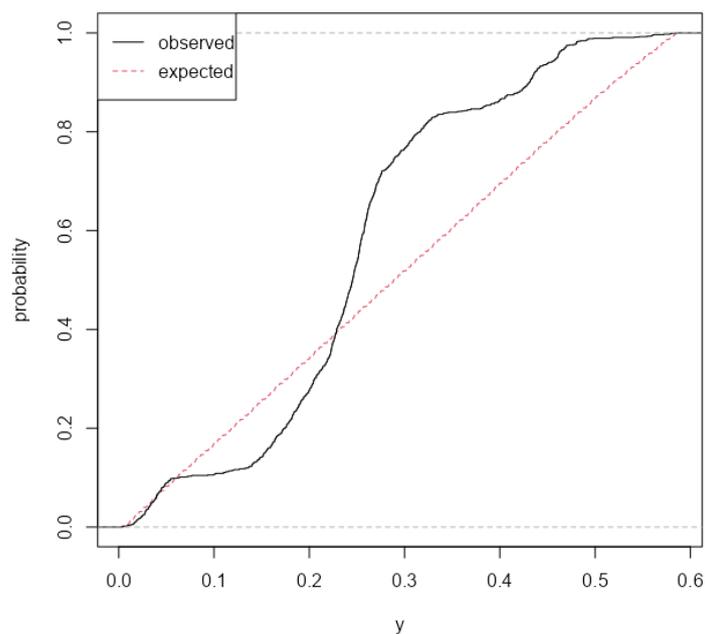
Figura 7 – Visão com perspectiva tridimensional da coordenada 'Y'.



Fonte: Elaborado pelo autor, 2023.

É verificado na Figura 7 que há uma maior concentração de escolas na região da coordenada espacial 'Y' com valores similares ou um pouco abaixo do valor médio possível para essa coordenada.

Figura 8 – Teste de Kolmogorov-Smirnov para a coordenada Y.



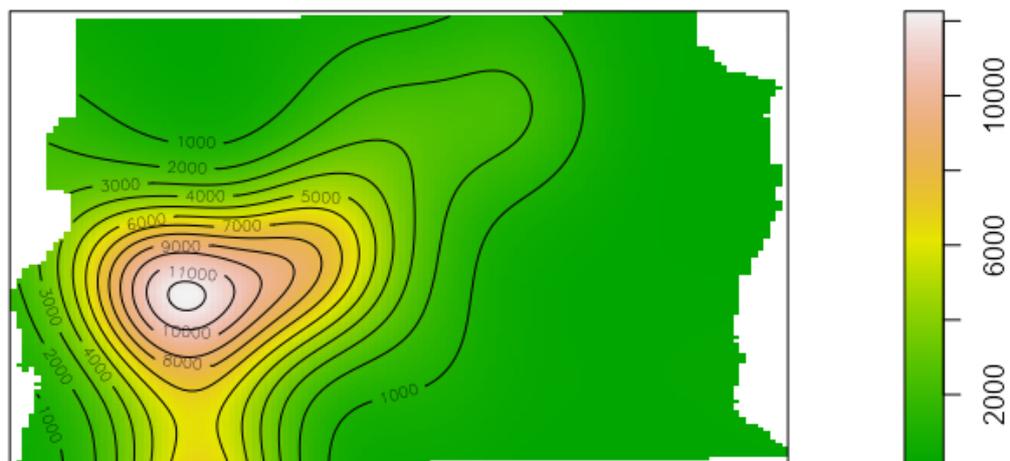
Fonte: Elaborado pelo autor, 2023.

É verificado na Figura 8 que ao efetuar o teste de Kolmogorov-Smirnov para a coordenada espacial 'Y', existe uma tendência de acúmulo da presença de escolas até o valor de 0,3 da coordenada espacial 'X' ao colocar a mesma em uma escala de 0 a 0,6. Onde para a coordenada 'Y' o supremo das diferenças absolutas entre a Função de Distribuição Acumulada Empírica (FDAC) e a Função de Distribuição Acumulada Teórica (FDAT) atingiu o valor de 0,26259, e com um p-value também menor que  $10^{-15}$ , apresentando-se também a conclusão de rejeição da hipótese de Completa Aleatoriedade Espacial.

## 4.2 Estimador de Kernel

Após rejeitada a hipótese de Completa Aleatoriedade Espacial, foi utilizado o estimador de Kernel descrito na seção 2.3.

Figura 9 – Kernel Gaussiano com o método de Silverman (SILVERMAN, 1986).



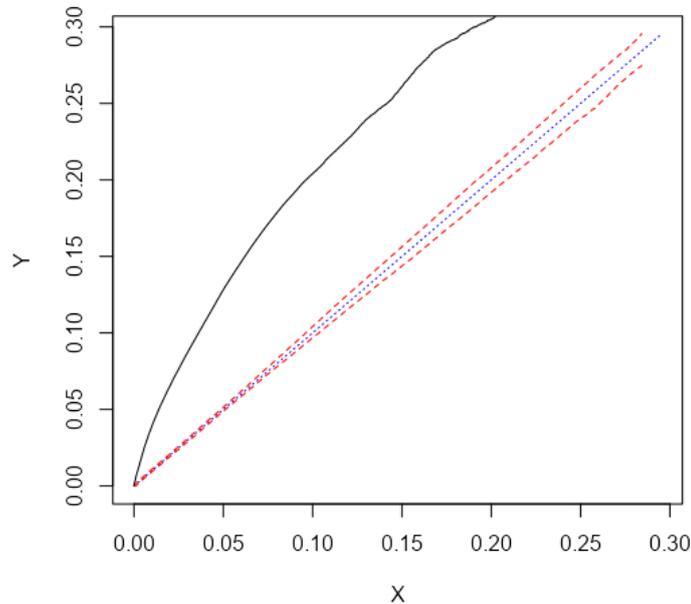
Fonte: Elaborado pelo autor, 2023.

Onde na Figura 4.2 há uma maior unidade de pontos por área, densidade, na região Centro-Oeste do próprio Distrito Federal. O que corrobora as observações ocorridas nos testes de Kolmogorov-Smirnov das Figuras 6 e 8, em relação às coordenadas espaciais 'X' e 'Y', respectivamente, com indicativo de concentração de escolas na mesma região apresentada no Estimador de Kernel.

## 4.3 Função K de Ripley

E finalizando a análise foi verificada a comparação da distribuição dos dados na Função K de Ripley, comparado a dados aleatórios, objetificando-se identificar o padrão espacial das escolas no Distrito Federal.

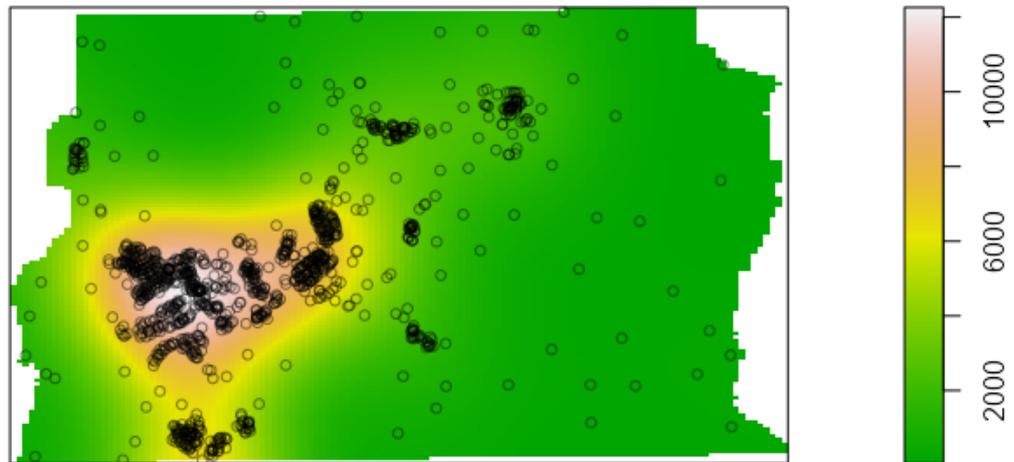
Figura 10 – Função K de Ripley com dados de geolocalização das escolas no Distrito Federal.



Fonte: Elaborado pelo autor, 2023.

Na figura 4.3 a função teórica é representada pela linha azul, que indica o padrão de dados aleatórios. O envelope de confiança, para o padrão dos dados serem aleatórios, é representado pelas linhas vermelhas, das quais a linha vermelha superior à linha azul (função teórica), representa o limite superior, e a linha vermelha inferior representa o limite inferior do envelope de confiança. Já a função empírica é representada pela linha preta, sendo os valores dos dados observados. Onde ver que a distribuição dos dados na Função K de Ripley obteve uma curva superior ao envelope, o que se configura como dados agrupados, para todos os valores de distâncias entre as escolas. Estas observações de aglomeração, confirmadas pela Função K de Ripley, estão expressas na plotagem das localizações das escolas sobre a função de densidade de probabilidade obtida com o Estimador de Kernel na Figura 4.3.

Figura 11 – Intensidade de agrupamento das escolas no Distrito Federal.



Fonte: Elaborado pelo autor, 2023.

A identificação de agrupamento das escolas no Distrito Federal se assemelha à conclusão para a identificação de agrupamento das unidades prisionais no Brasil (SOUZA, 2018), onde foi verificado a concentração de prisões nas regiões Sudoeste e Nordeste do Brasil.

## 5 CONCLUSÃO

Este estudo explorou a distribuição espacial das escolas no Distrito Federal por meio de técnicas de análise de processos pontuais, objetivando verificar a hipótese de uma distribuição aleatória. Os resultados rejeitam a hipótese de aleatoriedade, em favor da hipótese de agrupamento das escolas mais densamente na região Centro-Oeste do Distrito Federal. Com essa região, abrangendo várias regiões administrativas, possui uma maior densidade de escolas em comparação com as outras regiões do Distrito Federal.

A prevalência de escolas na região Centro-Oeste pode ser influenciada por diversos fatores, como densidade populacional mais alta e uma concentração de serviços e infraestrutura urbana. Esta concentração não necessariamente reflete desigualdades na alocação de escolas entre as regiões administrativas, mas destaca a heterogeneidade na distribuição espacial das unidades escolares pelo Distrito Federal.

## REFERÊNCIAS

- ArcGIS Pro. *How Multi-Distance Spatial Cluster Analysis (Ripley's K-function) works*. 2023. [Online; acessado em 30-Junho-2023]. Disponível em: <<https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-multi-distance-spatial-cluster-analysis-ripl.htm>>. Citado na página 14.
- CHEN, Y.-C. A tutorial on kernel density estimation and recent advances. *Biostatistics & Epidemiology*, Taylor & Francis, v. 1, n. 1, p. 161–187, 2017. Disponível em: <<https://doi.org/10.1080/24709360.2017.1396742>>. Citado na página 13.
- DIGGLE, P. J. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition*. 3. ed. Chapman and Hall/CRC, 2013. (Chapman & Hall/CRC Monographs on Statistics & Applied Probability). ISBN 978-1-4665-6024-6, 978-1-4665-6023-9. Disponível em: <<http://gen.lib.rus.ec/book/index.php?md5=1ede7fe066c86cbc9be30afb43502515>>. Citado na página 11.
- DRUCK, S. et al. (Ed.). *Análise Espacial de Dados Geográficos*. Brasília: EMBRAPA, 2004. Edição em papel: EMBRAPA, Brasília, 2004, disponível na Livraria Virtual da EMBRAPA. ISBN 85-7383-260-6. Citado na página 11.
- OLINDA, R. A. d. *Métodos para análise de independência entre marcas e pontos em processos pontuais marcados*. Dissertação (Mestrado) — Universidade Federal de Lavras, Lavras, Minas Gerais, Brasil, 2008. Citado na página 11.
- OLIVEIRA, W. A. d. *Métodos de processos pontuais para análise espacial de espécies arbóreas nativas de um fragmento florestal da região de Lavras-MG*. Dissertação (Mestrado) — Universidade Federal de Lavras, Lavras, Minas Gerais, Brasil, 2022. Citado na página 10.
- OYANA, T. J. *Spatial analysis with R: statistics, visualization, and computational methods*. Second edition. CRC Press, 2020. ISBN 9781003021643, 9781000173451, 1000173453, 9781000173468, 1000173461, 9781000173475, 100017347X, 1003021646. Disponível em: <<http://gen.lib.rus.ec/book/index.php?md5=7D1EF573F0218A877A867666C72AE75A>>. Citado 2 vezes nas páginas 11 e 18.
- RIPLEY, B. D. Modelling spatial patterns. *Journal of the Royal Statistical Society: Series B (Methodological)*, v. 39, n. 2, p. 172–192, 1977. Disponível em: <<https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1977.tb01615.x>>. Citado na página 14.
- Secretaria de Estado de Saúde do Distrito Federal. *SISDIA - Página Inicial*. 2023. [Online; acessado em 30-Junho-2023]. Disponível em: <<https://sisdia.df.gov.br/home/>>. Citado na página 16.
- SILVERMAN, B. *Density Estimation for Statistics and Data Analysis*. Taylor & Francis, 1986. (Chapman & Hall/CRC Monographs on Statistics & Applied Probability). ISBN 9780412246203. Disponível em: <<https://books.google.com.br/books?id=e-xsrjsL7WkC>>. Citado 3 vezes nas páginas 8, 13 e 21.
- SOUZA, R. d. O. *Processos Pontuais Espaciais para dados das unidades prisionais no Brasil*. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, Instituto de Matemática, Rio de Janeiro, Rio de Janeiro, Brasil, 2018. Citado 2 vezes nas páginas 10 e 23.

STEPHENS, M. A. Introduction to kolmogorov (1933) on the empirical determination of a distribution. In: \_\_\_\_\_. *Breakthroughs in Statistics: Methodology and Distribution*. New York, NY: Springer New York, 1992. p. 93–105. ISBN 978-1-4612-4380-9. Disponível em: <[https://doi.org/10.1007/978-1-4612-4380-9\\_9](https://doi.org/10.1007/978-1-4612-4380-9_9)>. Citado na página 12.

ZANETTI, M. A. Z. Sistema de coordenadas universal transverso de mercator-sistema utm. 2017. Citado na página 18.