

NOTA TÉCNICA 01/2020

# Correlação espacial da covid-19 em municípios do estado do Paraná

Cassio Rolim (pesquisador convidado, Nesde/UFPR)

9 de maio de 2020

Av. Prefeito Lothário Meissner, 632, térreo  
Jardim Botânico, Curitiba – PR  
CEP: 80210-170  
Telefone: +55-41-3360-4442  
E-mail: [nesde@ufpr.br](mailto:nesde@ufpr.br)  
Website: <http://www.nesde.ufpr.br/>

## Ficha técnica

Núcleo de Estudos em Economia Social e Demografia Econômica, Universidade Federal do Paraná  
Secretaria do Estado da Saúde, Governo do Paraná

## Colaboradores

### Núcleo de Estudos em Economia Social e Demografia Econômica

Profa. Dra. Raquel Guimarães  
Prof. Dr. Junior Ruiz Garcia  
Profa. Dra. Angela Welters  
Profa. Dra. Denise Maria Maia  
Luis Alceu Paganotto  
Eron José Maranhão  
Rossana Ribeiro Ciminelli  
Cassio Rolim

### Secretaria do Estado de Saúde do Paraná, SESA/PR

Carlos Alberto Gebrim Preto (Secretário do Estado de Saúde do Paraná)  
Nestor Werner Junior (Diretor Geral da Secretaria do Estado de Saúde do Paraná)  
Maria Goretti David Lopes (Diretora de Atenção e Vigilância em Saúde)  
Dra. Acácia Maria Lourenço Francisco Nasr (Coordenadoria de Vigilância Epidemiológica)

Esse trabalho deve ser citado como:

ROLIM, Cassio. **Correlação espacial da covid-19 em municípios do estado do Paraná**. Nota Técnica Nesde/UFPR 01/2020. Curitiba: UFPR e SESA, 2020. doi 10.17605/OSF.IO/YTU3C



Este trabalho está licenciado sob uma Licença Internacional 4.0 Creative Commons - Atribuição Não Comercial.

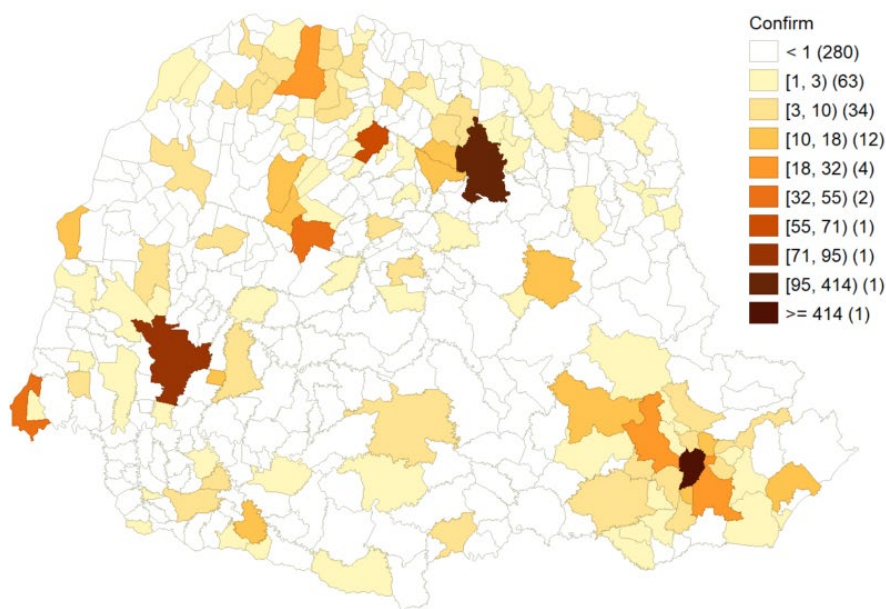
Para qualquer uso comercial, por favor contate-nos:  
[nesde@ufpr.br](mailto:nesde@ufpr.br)

As Notas Técnicas do Núcleo de Estudos em Economia Social e Demografia Econômica (Nesde/UFPR) recebem uma revisão limitada. As opiniões ou opiniões expressas neste documento não representam necessariamente as da Universidade Federal do Paraná ou de outras instituições que apoiam seu trabalho.

## Correlação espacial da covid-19 em municípios do estado do Paraná

A figura 1 mostra o número acumulado de casos confirmados de infecção por Coronavirus-19 no estado do Paraná até o dia 24 de abril de 2020. São números absolutos e mostram claramente que o maior número de casos ocorre nas maiores cidades, Curitiba, Londrina, Maringá, Cascavel. Mostra também que até 24 de abril em 280 municípios não tinham sido registrados casos de infecção. Curitiba destaca-se com valores acima de 414 casos.

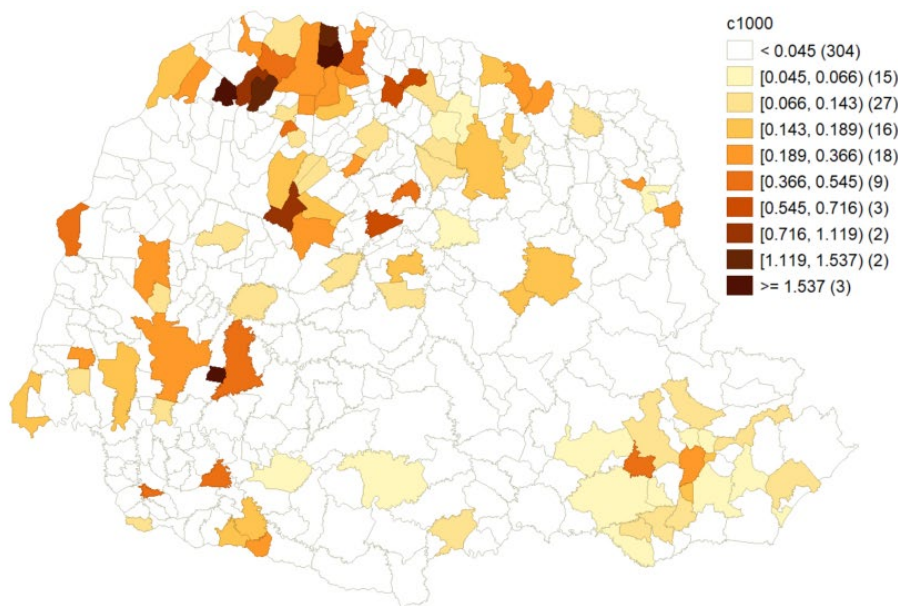
FIGURA 1: Casos confirmados da covid-19 no Estado do Paraná. Valores Acumulados até 24/04/2020



Fonte: SESA/UFPR. Elaboração própria

No entanto, tão importante como conhecer a distribuição de números absolutos é importante conhecer a distribuição relativa, a distribuição de casos ponderada pelo número de habitantes. Ou seja, mostrar a distribuição do Coeficiente de Incidência da infecção<sup>i</sup>. Assim sendo, torna-se possível comparar a incidência em Curitiba (0,214 infectados por 1.000 habitantes<sup>ii</sup>) com a da pequena Santa Mônica (1,758 infectados por 1.000 habitantes) sem o viés do peso populacional da metrópole. A figura 2 mostra que em termos de coeficiente de incidência o maior problema parece estar em outras áreas do estado. Isso traz implicações para a política de combate ao vírus: priorizar os números absolutos ou os relativos; ou ainda, uma combinação dos dois.

FIGURA 2: Coeficiente de Incidência da covid-19 para cada 1000 habitantes, Valores acumulados até 24/04/2020

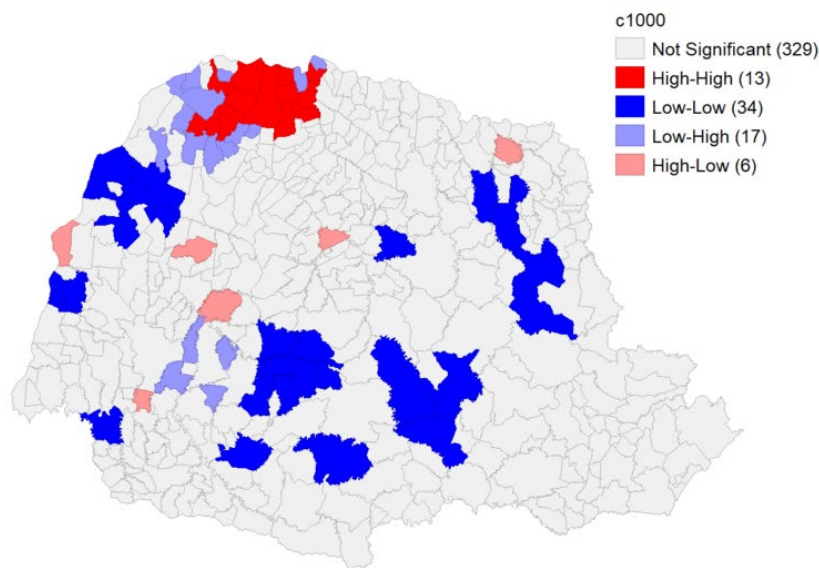


Fonte: SESA/UFPR. Elaboração própria

Considerar a evolução das infecções tendo em conta apenas os números totais para o estado do Paraná podem ser ilusórios, uma vez que o grande peso de Curitiba irá ofuscar tendências diferenciadas ao longo do território. Dessa forma, se a difusão for contida em Curitiba, os números agregados poderão apresentar um relativo alívio, não obstante a expansão acelerada em outras partes do estado. Daí a importância de analisá-los em termos relativos e tendo em conta os grupos de municípios.

Tendo em conta valores relativos ao tamanho da população, o coeficiente de incidência por 1.000 pessoas, é possível dar um passo além e utilizar instrumentos da econometria espacial que permite descobrir grupos de municípios com valores semelhantes. A visualização desses grupos em um mapa possibilita a percepção de áreas críticas e com potencial de expansão do vírus diferenciado no território paranaense. Isso poderá auxiliar as opções mais específicas da política de combate ao vírus.

FIGURA 3: Clusters de Municípios Paranaenses com Coeficiente de Incidência da covid-19 por 1.000 habitantes. Situação com os valores acumulados em 24 de Abril de 2020.



Fonte: SESA/UFPR. Elaboração própria

A Figura 3 mostra os grupos de municípios do Paraná segundo a variável Coeficiente de Incidência por 1000 habitantes da covid-19. Esses grupos foram obtidos através da aplicação de os indicadores de autocorrelação espacial local (LISA). Eles são uma decomposição ao nível local do conhecido I de Moran, indicador geral de autocorrelação espacial em um conjunto de unidades territoriais. Esses indicadores mostram como a variável coeficiente de incidência no município  $i$  varia com a variação dessa mesma variável coeficiente de incidência nos municípios vizinhos. Apenas aparecem os grupos em que a autocorrelação espacial é estatisticamente significativa (Vide a Nota Técnica). Note-se que os valores elevados não se referem a números absolutos, mas sim ao número de contaminados em relação à população do município.

As áreas em vermelho são os grupos de municípios em que o coeficiente de incidência é elevado (Alto-Alto), tanto no município  $i$  como nos seus vizinhos. As áreas em azul escuro representam os grupos de municípios em que esse fenômeno é reduzido (Baixo-Baixo). As áreas em cinza claro, não significativas, são aquelas em que o valor da variável não difere, estatisticamente, da média do conjunto de municípios do estado. As áreas em azul claro (Baixo-Alto) e as em rosa (Alto-Baixo), indicam pontos extremos (outliers) em relação aos seus vizinhos. Note-se que Curitiba não aparece nesses grupos, porque o seu coeficiente de incidência não é

muito elevado, o dos seus vizinhos é muito menor e, portanto, não apresentam significância estatística.

No Noroeste do estado destaca-se um grande agrupamento com alto coeficiente de incidência. Quase todos os municípios pertencem à microrregião de Paranavaí. Provavelmente é a área que demanda maior urgência de intervenção das autoridades, particularmente porque ela parece estar se expandindo. O mesmo acontece com os outliers. São pontos com baixo coeficiente de incidência relativa, mas com vizinhos onde ele é alto ou então pontos com alto coeficiente de incidência, mas com vizinhos onde ele é baixo. Tanto em um caso com em outro, deixada à própria sorte, a contaminação da população não tardará a ocorrer. Além disso, são particularmente preocupantes aqueles situados nas fronteiras do Paraná, como Guaíra e os, ao Sul, na fronteira com Santa Catarina. Para melhor compreender esses pontos seria importante ampliar a análise englobando os municípios dos estados vizinhos do Paraná para descobrir eventuais trajetórias espaciais do vírus.

## Metodologia

A distribuição de informações em um mapa pode ser analisada visualmente ou analiticamente. A grande diferença entre a análise visual e a analítica é que os resultados desta última podem ter a sua significância estatística testada. Um dos objetivos dessas análises é caracterizar os fenômenos de dependência (autocorrelação) e heterogeneidade espacial. O objetivo é ver se os fenômenos ocorrem aleatoriamente no espaço ou se há uma correlação espacial na sua ocorrência. Um dos instrumentos mais utilizado para a autocorrelação espacial é a estatística I de Moran e a heterogeneidade espacial pode ser vista com o uso da LISA (*Local indicators of spatial association*) que é uma decomposição local do I de Moran.

O índice de Moran, por sua vez, mede como a variável  $z$  no município  $i$  varia com a variação dessa mesma variável  $z$  nos municípios vizinhos. Em outras palavras, é um produto cruzado entre uma variável e a sua defasagem espacial (*spatial lag*), expressas de maneira centrada (desvio em relação à média). No município  $i$ , a variável  $z_i = (x_i - x^*)$ , em que  $x^*$  é a média da variável  $x$ . Para calculá-lo é necessário previamente definir uma matriz de pesos espaciais. O tipo da matriz de pesos definida é essencial para os resultados (Anselin, 2005).

Uma vez definida a estrutura de vizinhanças com a matriz de pesos espaciais  $W$ , a variável de defasagem espacial é a soma ponderada dos valores observados nas localidades vizinhas. Neste caso está sendo usada, para os 399 municípios do Paraná, uma matriz de vizinhança tipo rainha<sup>iii</sup>. Em média cada município tem 5,6 vizinhos.

No caso de uma matriz de pesos espaciais normalizada, o intervalo de variação de  $I$  será:

$$-1 \text{ (dispersão perfeita)} \leq I \leq 1 \text{ (correlação perfeita)}.$$

O valor zero indica padrão espacial aleatório; valores positivos indicarão uma correlação positiva com os vizinhos e valores negativos, uma correlação negativa (inversa)<sup>[2]</sup>. Valores positivos indica que valores semelhantes aparecem próximos no espaço formando clusters; valores negativos indica que os valores entre os vizinhos são diferentes, isto é, os valores semelhantes estão dispersos no espaço; ausência de autocorrelação (valores nulos) significam aleatoriedade no padrão espacial de distribuição daquela variável.

A significância estatística pode ser obtida tanto por uma aproximação à curva normal, uma vez que  $I$  tem uma distribuição assintoticamente normal, como por um processo aleatório. Esta última, mais comum nos exercícios empíricos, baseia-se em permutações aleatórias

Por sua vez, os indicadores de autocorrelação espacial local irão permitir a decomposição dos indicadores globais (que consideram todas as observações). Anselin (1995) apresenta o LISA para o  $I$  de Moran como:

$$I_i = z_i \sum_j w_{ij} z_j$$

O resultado dos cálculos do indicador local de Moran (LISA) é apresentado em um mapa de clusters que, serão HH(valores altos no centro e na vizinhança); LL (valores baixos no centro e na vizinhança); HL (valor alto no centro e baixo na vizinhança); LH (valor baixo no centro e alto na vizinhança). Apenas os valores com significância estatística são considerados. O mapa também apresenta localidades consideradas não significantes que tendem a ser a maioria. A interpretação para essas localidades é que a variável em análise não é estatisticamente diferente da média nos outros municípios. (Almeida, 2012, p. 129)

A matriz de pesos espaciais utilizada para a construção da figura 3 é uma matriz de contiguidade tipo Queen. É o tipo mais encontrado na literatura. Nas figuras abaixo apresentamos os resultados para o mesmo Coeficiente de Incidência de covid-19 por 1.000 habitantes em 24 de abril de 2020, porém utilizando diferentes matrizes.

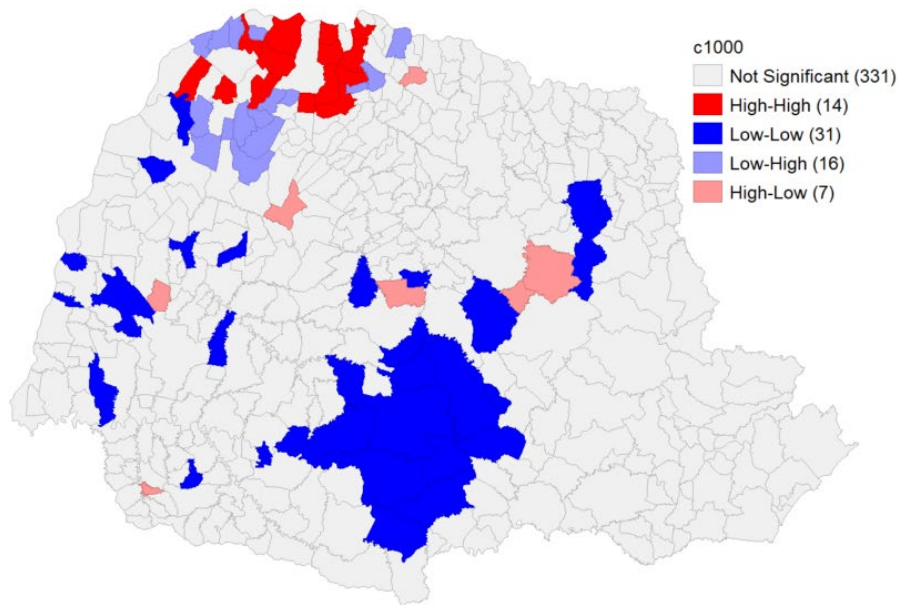
A figura A1 utiliza uma matriz Queen, porém com distância topológica 2. A diferença é que nesta a vizinhança, além dos vizinhos imediatos, também leva em conta os vizinhos um pouco mais distantes. Exemplificando, um vizinho imediato de Curitiba seria o município de Fazenda Rio Grande (distância topológica 1); para atingir o

vizinho seguinte, Mandirituba (distância topológica 2), é necessário atravessar o território de Fazenda Rio Grande. No caso da distância topológica 1, cada município apresentou em média 5,58 vizinhos (mínimo de 2 e máximo de 12). Para a distância topológica 2, cada município apresentou em média 12,54 vizinhos (mínimo de 3 e máximo de 25).

As figuras A2, A3 e A4, utilizaram matrizes de peso espacial baseadas na distância. Foram matrizes do tipo k vizinhos mais próximos (*k-nearest neighbours*). São matrizes em que é considerado um número fixo de vizinhos mais próximos, neste caso, 3, 5 e 10 vizinhos.

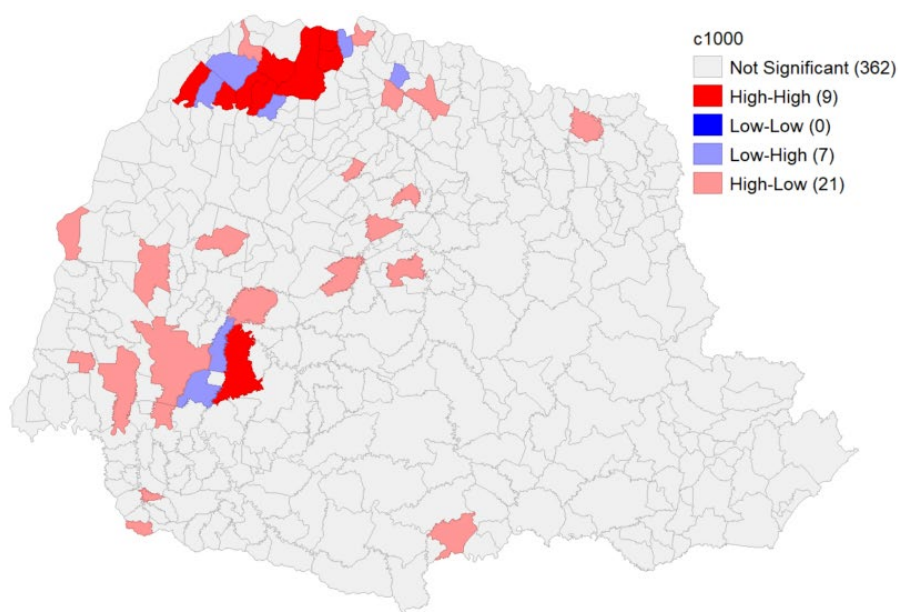
O que observa é que neste exercício, qualquer que seja o tipo de matriz de pesos espaciais considerada, o grupo de municípios com maior coeficiente de incidência da corona vírus-19 detectado no noroeste e no oeste do estado, tende a ser consistente. Não será surpresa se for neles que irá ocorrer a maior variação no coeficiente de incidência nos próximos dias.

FIGURA 4: Matriz Queen com distância topológica 2



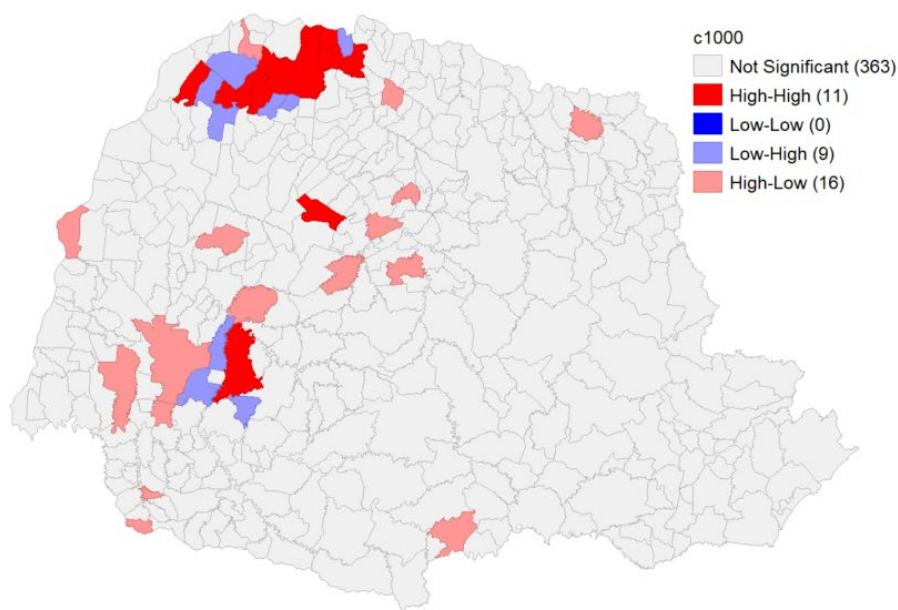
Fonte: SESA/UFPR. Elaboração própria

FIGURA 5: Matriz 3 vizinhos mais próximos



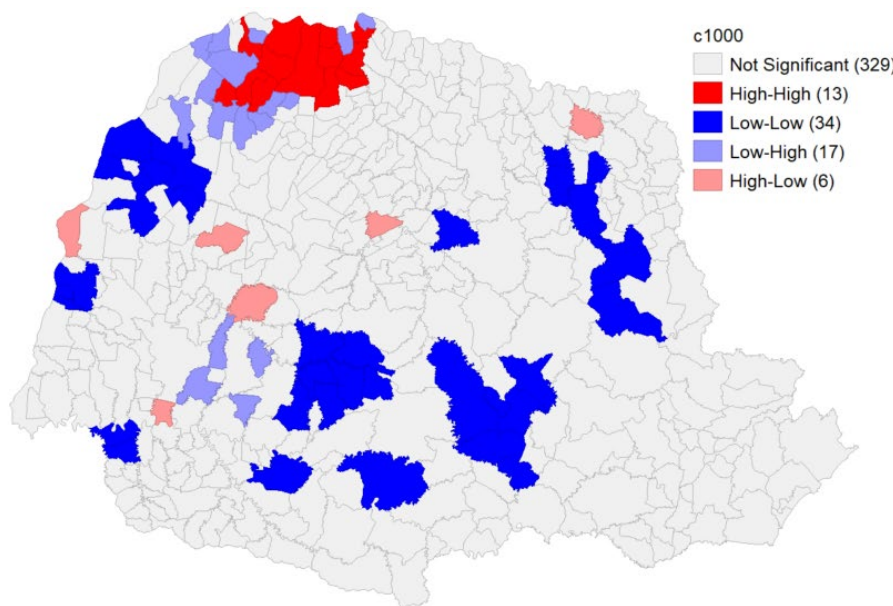
Fonte: SESA/UFPR. Elaboração própria

FIGURA 6: Matriz 5 vizinhos mais próximos



Fonte: SESA/UFPR. Elaboração própria

FIGURA 7: Matriz 10 vizinhos mais próximos



Fonte: SESA/UFPR. Elaboração própria

TABELA 1: Composição dos Grupos de Municípios segundo a Figura 3 (24/04/2020)

Alto-Alto (H-H)	Baixo-Baixo (L-L)	Baixo-Alto (L-H)	Alto-Baixo (H-L)
São João do Caiuá	Pinhão	Campo Bonito	São João do Ivaí
Santo Antônio do Caiuá	Chopinzinho	Catanduvas	Marumbi
Amaporã	Umuarama	Cidade Gaúcha	Guaíra
Planaltina do Paraná	Ibaiti	Inajá	Pinhal de São Bento
Guairaçá	Palmital	Loanda	Céu Azul
Paranacity		Santa Isabel do Ivaí	Goioerê
Alto Paraná		Tapira	União da Vitória
Cruzeiro do Sul			Iretama
Paranavaí			Barracão

## Referências

- Almeida, E. (2012). *Econometria Espacial Aplicada* (1st ed.).  
<https://www.amazon.com/Econometria-Espacial-Aplicada-Eduardo-Almeida/dp/8575166018>
- Anselin, L., Syabri, I., & Kho, Y. (2006). GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis*, 38(1), 5–22. <https://doi.org/10.1111/j.0016-7363.2005.00671.x>

Anselin, L. (2020). *GeoDa on Github: An Introduction to Spatial Data Analysis*.  
<https://geodacenter.github.io/>

---

<sup>i</sup> Segundo os manuais de epidemiologia, o termo técnico correto seria prevalência. A incidência refere-se a novos casos dentro de um intervalo de tempo. Dessa forma  $\text{Prevalência por 1000} = (\text{número de casos na população na data } i / \text{população na data } i) \times 1.000$ . (Gordis, 2015, p. 38–60). Entretanto manteremos incidência, tal como o faz a Secretaria de Saúde do Paraná.

<sup>ii</sup> Geralmente esses indicadores referem-se a 100 mil ou até um milhão de habitantes. Porém, como mais de 75% dos municípios do Paraná tem menos de 20 mil habitantes, preferiu-se apresentar os valores tendo como referência 1.000 habitantes.

<sup>iii</sup> Em analogia aos movimentos da rainha no jogo de xadrez. A matriz foi criada pelo programa GEODA.