Local Moran Index: an application in epidemiological coefficients of the COVID-19 pandemic in Brazil

Índice de Moran Local: uma aplicação em coeficientes epidemiológicos da pandemia de COVID-19 no Brasil

Índice de Moran local: una aplicación en coeficientes epidemiológicos de la pandemia COVID-19 en Brasil

Received: 03/02/2021 | Reviewed: 03/07/2021 | Accept: 03/08/2021 | Published: 03/16/2021

Henrique José de Paula Alves
ORCID: https://orcid.org/0000-0002-0124-3093
Instituto de Pesquisa Econômica Aplicada, Brazil
E-mail: jphenhrique@gmail.com

Felipe Augusto Fernandes
ORCID: https://orcid.org/0000-0003-2552-3433
Universidade Federal de Lavras, Brazil
E-mail: fernandesfelpest@gmail.com

Kelly Pereira de Lima
ORCID: https://orcid.org/0000-0002-2581-8525
Universidade Federal de Lavras, Brazil
E-mail: kelly.lima.88@gmail.com

Ben Dévide de Oliveira Batista
ORCID: https://orcid.org/0000-0001-7019-8794
Universidade Federal de São João del-Rei, Brazil
E-mail: ben.devide@ufsj.edu.br

Tales Jesus Fernandes
ORCID: https://orcid.org/0000-0002-1457-9653
Universidade Federal de Lavras, Brazil
E-mail: tales.jfernandes@ufla.br

Abstract
The COVID-19 pandemic spread quickly around the world in a frightening way. In Brazil, the third country in the world with the highest number of people infected and killed by the disease, it is important that the government health authorities identify the federation units that stand out in cases and deaths due to this disease for targeting resources. The Local Moran Index is a statistical tool that estimates those units of the federation that stands out the most with some statistical significance. We used the epidemiological coefficients of incidence, prevalence, and lethality to describe Brazil’s pandemic better today. We use R software to obtain maps and results.

Keywords: Brazil; COVID-19; Epidemiological coefficients; Local Moran Index.

Resumo
A pandemia da COVID-19 se espalhou rapidamente pelo mundo de uma forma assustadora. No Brasil, terceiro país do mundo com maior número de infectados e mortos pela doença, é importante que as autoridades sanitárias governamentais identifiquem as unidades da federação que se destacam nos casos e óbitos por essa doença para o direcionamento dos recursos. O Índice de Moran Local é uma ferramenta estatística que estima as unidades da federação que mais se destacam com alguma significância estatística. Usamos os coeficientes epidemiológicos de incidência, prevalência e letalidade para descrever melhor a pandemia no Brasil hoje. Usamos o software R para obter os mapas e resultados.

Palavras-chave: Brasil; COVID-19; Coeficientes epidemiológicos; Índice de Moran local.

Resumen
La pandemia de COVID-19 se ha extendido rápidamente por todo el mundo de una manera aterradora. En Brasil, el tercer país del mundo con mayor número de infectados y muertos por la enfermedad, es importante que las autoridades gubernamentales de salud identifiquen las unidades de la federación que se destacan en los casos y muertes por esta enfermedad para la focalización de recursos. El Índice de Moran Local es una herramienta estadística que estima las unidades de la federación que más se destacan con alguna significación estadística. Usamos los coeficientes...
epidemiológicos de incidencia, prevalencia y letalidad para describir mejor la pandemia en Brasil hoy. Usamos el software R para obtener los mapas y resultados.

Palabras clave: Brasil; COVID-19; Coeficientes epidemiológicos; Índice de Moran local.

1. Introduction

It is of great interest to the syndromic surveillance agencies to detect, monitor, and prevent emergencies such as epidemics, pandemics, and occurrences of crimes. Such actions are necessary for the protection measures for the population to be taken. However, there is a need to direct resources from these protection actions to areas or regions where the phenomenon in question is different from other regions due to the scarcity of resources. Such regions are called clusters, in which this term is related to a set of similar regions concerning some characteristics.

Among the different existing methods that seek to identify spatial association patterns, such as the local moving average, the Geary (Griffith et al., 2003) index, the Global and Local Moran Index stand out. It is essential to check some factors, such as the distribution of data, outliers presence, and the absence of stationarity (Pinto et al., 2014).

These methods are instrumental in analyzing area data (Monteiro et al., 2004) and were developed to detect regions where the distribution of the values present some specific pattern, being this pattern random or non-random forming, in the non-random case, groupings associated with their spatial location. The objective is to identify how much the value of an attribute in a given area is similar to the values of that same attribute in neighboring regions, according to some pre-established neighborhood criteria. It is then possible to state that the use of these methodologies in the analysis of health indicators contributes to improving the quality of the implementation of public health policies since it is possible to identify the spatial distribution of cases, allowing them to be used as tools to assist in the planning and monitoring these events (Pinto et al., 2014).

Hendricks and Mark-Carew (2017) used the local Moran index to identify clusters in the USA of Lyme disease. Koh et al. (2018) used this same technique to identify clusters with high obesity rates in the USA. Gehlen et al. (2019) applied the local and global Moran indices to identify Brazilian cities with increased rates of tuberculosis. Lew and Rigdon (2019) also applied the local Moran index to identify clusters of American individuals who had some mental illness, using Bayesian hierarchical models. Li et al. (2020) used global and local Moran indices in COVID-19 data in China to identify clusters of pneumonia cases.

Cordes and Castro (2020) used this same index, together with the circular scan method as in Kulldorff (1997), seeking to identify areas with low access to tests and a high load of COVID-19 cases in New York (USA). Huang et al. (2020) also used the Local Moran Index to show that COVID-19 infection is spatially dependent and has spread mainly from Hubei Province, Central China, to neighboring areas. The logistic model was used according to the trend of the available data. Nilima et al. (2020) used the local Moran index to investigate possible clusters of psycho-social factors caused by the lockdown implemented in India due to the COVID-19 pandemic.

Other works related to COVID-19 using the global and local Moran indexes can be seen in Kim and Castro (2020); Arashi et al. (2020); Lieberman-Cribbin et al. (2020); Kang et al. (2020); Yao et al. (2020).

In December 2019, the Chinese government announced that in Wuhan (Hubei, China) a new coronavirus was emerging, called SARS-CoV-2, whose disease caused in humans is called COVID-19, being declared by the World Health Organization (WHO) as a global public health emergency (Nassiri, 2020; Wang et al., 2020). This virus has RNA single-stranded with high potential for mutations, which favors infectivity and virulence (Amaral et al., 2020; Khailany et al., 2020; Nassiri, 2020; Velavan & Meyer, 2020). The main mechanism of entry of this virus in the cells of the hosts is through the
epithelial cells of the upper respiratory tract, and the type of infection caused can lead to signs and symptoms such as fever, dry cough, headache, fatigue, dyspnoea, shortness of breath, chills, and arthralgia, occurring on average between 5 and 6 days after incubation (Salathé et al., 2020; Khailany et al., 2020; Letko et al., 2020). The damage caused to public health is evident because vaccines and specific medications for the treatment of COVID-19 (Khailany et al., 2020; Nassiri, 2020; Velavan & Meyer, 2020; Salathé et al., 2020) still do not exist. Besides, many critically ill patients need hospital assistance and mechanical respirators, and SARS-CoV-2 has a high potential for transmission, which leads to the need for social distance and these characteristics have made COVID-19 was quickly considered a pandemic by WHO (Salathé et al., 2020).

In Brazil, this pandemic spread rapidly, and this occurred due to the challenges regarding the conditions of social vulnerability, housing, poor sanitation, and overcrowding in the home (Alves et al., 2020; Fernandes et al., 2020; Werneck & Carvalho, 2020). These factors make it difficult for millions of Brazilians to follow the WHO recommendations for partial and even total social isolation in some situations and hygiene with alcohol gel and respiratory masks, and these measures were effective in other countries (Werneck & Carvalho, 2020).

Given the above, we use the local Moran Index to identify Brazil’s areas where the incidence, prevalence, and lethality related to COVID-19 stand out from the other regions. We highlight that Brazil is currently the third country globally, with the most cases and deaths due to COVID-19, these indices were described according to Pereira et al. (2018) methodology. The main objective is to estimate and quantify the magnitude of the autocorrelation between the areas (Birch et al., 2009). We also highlight the use of an alternative form to the Moran scatter plot, known as Moran map and LISA map (Birch et al., 2009), where the term LISA refers to the “local indicator of spatial autocorrelation”.

2. The Local Moran Index

The Local Moran Index is given by the following expression (Pinto et al., 2014)

\[ I_i = \frac{x_i - \bar{x}}{\sum_j w_{ij}(x_j - \bar{x})^2} \sum_{j \neq i} w_{ij}, \quad i, j = 1, 2, \ldots, m \]

where \( \bar{x} \) is the average sample rate of the \( m \) areas that make up the region, \( x_i, x_j \) are the sample rates of the \( i, j \) indexed areas, respectively, \( w_{ij} \in W \) are the weights of the \( i, j \) indexed areas and come from the normalized weight matrix \( W \), which is defined by

\[
W = \begin{cases} 
0, & \text{se } i = j \text{ or } i \text{ is not a neighbor of } j \\
1/\tau, & \text{otherwise.}
\end{cases}
\]

The term \( \tau \) represents the number of neighbors that the \( i \)-th area has. The local Moran Index, as in the expression (1), generally assumes values ranging from −1 to 1, that is, \(-1 \leq I_i \leq 1\). It can happen to exceed these limits, but this fact rarely happens. If the value of \( I_i \) is close to or equal to 1, it indicates that both the rates of the corresponding municipality and that of its neighborhood are above or below the global average, suggesting possible clusters, which are denoted by high-high or low-low. If the value attributed by \( I_i \) is close to or equal to −1, it indicates that the rate of the corresponding municipality is above the global average while that of its neighbors is below or vice versa, suggesting extreme values,
which are denoted by high-low or low-high. If $I_{k}$ assumes values close to or equal to 0 indicates no spatial association. Another important consideration for the calculation of the Local Moran Index is the neighborhood criterion and the order of contiguity to be considered. We chose the “queen” neighborhood criterion, which considers neighbors in all directions and also second-order contiguity. These better represent the reality of the studied phenomenon. Figure 1 illustrates this scenario.

**Figure 1** - Neighborhood criteria of the queen type considering second-order contiguity. The area of interest is in pink, its first-order neighbors are in green, and its second-order neighbors are yellow.

For example, be a municipality of interest (pink). Thus, with the “queen” neighborhood criterion, those municipalities with a common border in all directions are considered first-order neighbors (green). However, those municipalities with no common border with the municipality of interest but have a common border with their first order neighbors in any direction are considered second-order neighbors (yellow). The neighborhood order can be extended to larger values.

### 3. Diagnostic Epidemiological Measures

#### 3.1 The incidence coefficient

The incidence coefficient (IC) expresses the risk of new cases of disease occurring in a population over a period of time. Its expression is given by

$$IC = \frac{|NC|}{|POP|} \times 100,000,$$

where $|NC|$ expresses the new cases and $|POP|$ the population at risk.

#### 3.2 The prevalence coefficient

The prevalence coefficient (PC) expresses the risk of accumulated cases of a disease occurring in a population over a period of time. Its expression is given by

$$PC = \frac{|CC|}{|POP|} \times 100,000,$$

where $|CC|$ expresses the accumulated cases and $|POP|$ the population at risk.
3.3 The lethality coefficient

The lethality coefficient (LC) provides the risk of death from a disease in a population over a period of time. Its expression is given by

\[ PC = \frac{|CD|}{|CAS|} \times 100,000. \]

where |CD| expresses accumulated deaths and |CAS| the number of accumulated cases in this population.

4. The Dataset

The software R Core Team (2020) has a package called covid19br, whose function downloadCovid19 makes it possible to obtain updated data from COVID-19 in Brazil and the world through the link https://covid.saude.gov.br/. The Brazilian COVID-19 data is available at country, region, state, and city-levels, and we chose to work with Brazilian federation units (FU’s) with the “state” option. We decided to evaluate these coefficients only recently and chose new and accumulated cases and deaths in 2020-10-17, considering week 42 of COVID-19 infection, and we calculated these coefficients according to the section 3.

Brazilian FUs are represented on the maps by the classification code according to Instituto Brasileiro de Geografia e Estatística (IBGE): 11 – Rondônia (RO), 12 – Acre (AC), 13 – Amazonas (AM), 14 – Roraima (RR), 15 – Pará (PA), 16 – Amapá (AP), 17 – Tocantins (TO), 21 – Maranhão (MA), 22 – Piauí (PI), 23 – Ceará (CE), 24 – Rio Grande do Norte (RN), 25 – Paraíba (PB), 26 – Pernambuco (PE), 27 – Alagoas (AL), 28 – Sergipe (SE), 29 – Bahia (BA), 31 – Minas Gerais (MG), 32 – Espírito Santo (ES), 33 – Rio de Janeiro (RJ), 35 – São Paulo (SP), 41 – Paraná (PR), 42 – Santa Catarina (SC), 43 – Rio Grande do Sul (RS), 50 – Mato Grosso do Sul (MS), 51 – Mato Grosso (MT), 52 – Goiás (GO), 53 – Distrito Federal (DF). We present our results about the significance of these coefficients via the local Moran index through the Moran map and LISA map, as in section 1.

5. Result

It is of great importance for the bodies responsible for public health in some locality to know the rate of manifestation of a specific disease (incidence) to assess situations of cause and effect, the prevalence of the number of cases of that same disease to estimate the burden that same cause for the population or relate the number of deaths from a given reason and the number of people who were affected by such disease (lethality) by estimating the severity of the disease.

Figure 2 illustrates, according to the local Moran index, the Brazilian FU’s where the incidence coefficient stands out from the others and the location of a possible outlier. Concerning its neighbors that also have a high incidence (red) (Figure 2 (a)), we note that the state of Rio Grande do Sul (43) and also the state of Amapá (16) form the areas that stand out (high-high) concerning the high incidence in Brazil (Figure 2 (b)). This means that COVID-19 today in Brazil manifests more in the Rio Grande do Sul (43) and Amapá (16). The state of Maranhão (21), on the other hand, is an outlier (Figure 2 (b)), since this location has a high incidence concerning its neighbors that have a low incidence (high-low) (Figure 2 (a)). This means that there is a significant manifestation of 20% of the disease concerning its neighbors in this state. In the other FU’s, the incidence was not significant (20%) concerning its neighbors.
Figure 2 - Moran amp and LISA map of the significant clusters according to the incidence coefficient in Brazilian FU's per 100,000 in habitants ($\alpha = 20\%$). The numbers on the maps refer to the code of each FU according to IBGE classification.

![Moran map and LISA map](image)

Fonte: Authors.

Figure 3 illustrates, according to the local Moran index, the Brazilian FU’s where the prevalence coefficient stands out from the others and the location of a possible outlier. Regarding its neighbors that also have a high incidence (red) (Figure 3 (a)), we note that the state of Santa Catarina (42) stands out (high-high) concerning the high prevalence in Brazil (Figure 3 (b)), meaning that this state has the greatest burden related to the duration of the COVID-19 pandemic in Brazil, with a significance of 20%. A curious fact is that this state was significant despite having a low incidence neighbor (Paraná (41)). We also noticed that Paraná (41) was not significant on the LISA map (Figure 3 (b)), even though it has high prevalence areas (42, 35). The Distrito Federal (53) is an outlier (Figure 3 (b)) since this location has a high prevalence concerning its neighbors that have low prevalence (high-low) (Figure 3(b)). Therefore, concerning its closest neighbors with a low prevalence, the Distrito Federal (53) presents a more significant pandemic burden. On the other, FU’s prevalence was not significant in relation to its neighbors.
Figure 3 - Moran amp and LISA map of the significant clusters according to the prevalence coefficient in Brazilian FU’s per 100,000 in habitants ($\alpha = 20\%$). The numbers on the maps refer to the code of each FU according to IBGE classification.

![Moran map](image1)

![LISA map](image2)

Fonte: Authors.

Figure 4 illustrates, according to the local Moran index, the Brazilian FU’s where the lethality coefficient stands out from the others and the location of a possible outlier. Concerning its neighbors that also present high incidence (high-high) (Figure 4 (a)), we note that the states of São Paulo (35), Espírito Santo (32), Bahia (29) and also the Distrito Federal (53) are the areas that stand out (high-high) concerning high lethality in Brazil (Figure 4 (b)). In these regions, COVID-19 manifests itself more severely than in other areas. On the other, FU’s lethality was not significant in relation to its neighbors.

Figure 4 - Moran amp and LISA map of the significant clusters according to the lethality coefficient in Brazilian FU’s per 100,000 in habitants ($\alpha = 20\%$). The numbers on the maps refer to the code of each FU according to IBGE classification.

![Moran map](image3)

![LISA map](image4)

Fonte: Authors.
We then present in Table 1 the descriptive measures of the clusters and outliers found by the Local Moran Index. We highlight the population of these areas and the new and accumulated cases and deaths, and the values of each of these locations’ coefficients.

Table 1 - Descriptive measures of clusters and outliers detected by the Local Moran Index for each of the coefficients per 100,000 in inhabitants.

| Coefficient | Identification | State | Population | New cases | Accumulated cases | Accumulated deaths |
|-------------|----------------|-------|------------|-----------|-------------------|-------------------|
| Incidence   | Cluster        | 43    | 11,377,239 | 2,129     | 222,692           | 5,342             |
|             |                | 16    | 845,731    | 78        | 50,15             | 731               |
|             | Outlier        | 21    | 7,075,181  | 327       | 180,887           | 3,923             |
| Prevalence  | Cluster        | 42    | 7,164,88   | 1,832     | 234,765           | 2,966             |
| Lethality   | Cluster        | 35    | 45,919,049 | 5,394     | 1,062,634         | 37,992            |
|             |                | 32    | 4,018,650  | 825       | 143,918           | 3,709             |
|             |                | 29    | 14,873,064 | 1,799     | 334,697           | 7,288             |
|             |                | 53    | 3,015,268  | 728       | 204,304           | 3,539             |

Fonte: Authors.

Regarding the incidence, we note that the Rio Grande do Sul (43) and also Amapá (16) are the states that present the greatest severity (incidence) concerning the pandemic. Maranhão (21) also fits this situation. Santa Catarina (42) is the region with the highest-burden concerning COVID-19 (prevalence). We note that São Paulo (35) is the most populous state, with the highest number of new and accumulated cases and deaths. This is why this state is the most problematic in Brazil, and for that reason, it presents high lethality, which represents the seriousness of COVID-19 in this location. We also note that the state of Bahia (29) also has serious problems with the pandemic.

6. Conclusion

The COVID-19 pandemic in Brazil is currently seriously manifesting itself. The local Moran Index presented the Brazilian FU’s that stands out concerning the epidemiological coefficients of incidence, prevalence, and lethality, with emphasis on the state of São Paulo (35), which presented a higher coefficient of lethality, indicating that in this location, the disease manifests itself more seriously. As future work, we intend to evaluate these same coefficients evolution over time using some space-time methodology.

Acknowledgments

To the Brazilian National Council for Scientific and Technological Development (CNPq) for the scholarship granted to Felipe Augusto Fernandes.

References

Alves, H. J. de P., Fernandes, F. A., Lima, K. P. de, Batista, B. D. de O. & Fernandes, T. J. (2020). A pandemia COVID-19 no Brasil: uma aplicação dos k-means método de agrupamento. Research, Society and Development, 9(10), e5829109059, 1-21.10.33448 / rsd-v9i10.9059.

Amaral, M., Conceição, K., Andrade, M., & Padovani, C. (2020). Generalized growth curve model for covid-19 in brazilian states. Revista Brasileira de Biometria, 38(2), 125-146.10.28951/tbb.v38i2.481

Arashi, M., Bekker, A., Salehi, M., Millard, S., Erasmus, B., Cronje, T., & Golpaygani, M., (2020). Spatial analysis and prediction of covid-19 spread in south africa after lockdown. arXiv preprint arXiv:2005.06596.
Werneck, G. L., & Carvalho, M. S. (2020). A pandemia de covid-19 no brasil: crônica de uma crise sanitária anunciada. Cadernos de Saúde Pública, 36(5):e00068820, 1-4. https://doi.org/10.1590/0102-311X00068820

Yao, Y., Pan, J., Wang, W., Liu, Z., Kan, H., Qiu, Y., Meng, X. & Wang, W. (2020). Association of particulate matter pollution and case fatality rate of covid-19 in 49 chinese cities. Science of the Total Environment, 741, 140396. https://doi.org/10.1016/j.scitotenv.2020.140396