10.18605/2175-7275/cereus.v13n4p248-261

ARTIGO ORIGINAL

<< Recebido em: 16/11/2021 Aceito em: 25/11/2021. >>

REVISTA CEREUS ISSN: 2175-7275

# Analysis of homicides and basic food basket cost time series from Ilhéus and Itabuna cities, Brasil

## *Análise de séries temporais de homicídios e custo da cesta básica das cidades de Ilhéus e Itabuna*

Mateus Santos Marinho[1], Dany Sanchez Dominguez[2], Murillo Almeida dos Santos Torres[3], Natália Luiza Freire Botelho[4]

[1] Mestrando em Modelagem Computacional, Universidade Estadual de Santa Cruz.

E-mail:

marinhoms.uesc@gmail.com

[2] Doutor em Modelagem Computacional, Universidade Estadual de Santa Cruz.

[3] Graduando em Ciência da Computação, Universidade Estadual de Santa Cruz.

[4] Mestranda em Modelagem Computacional, Universidade Estadual de Santa Cruz.

## RESUMO

A violência figura entre os maiores problemas enfrentados pela gestão pública brasileira e o aumento do número de homicídios em cidades do interior é um fator alarmante para a atuação das instituições de segurança pública. Entender suas causas é um grande desafio para o desenvolvimento social e analisar os dados disponíveis dos indicadores de homicídios pode constituir um método eficiente que resulte em conclusões apropriadas. Esse trabalho tem o objetivo de aplicar técnicas computacionais para analisar as séries de dados históricos de homicídios e também de custo da cesta básica das cidades de Ilhéus e Itabuna, Bahia, Brasil, para posterior aplicação em modelos de predição com aprendizado profundo. Os dados foram coletados das bases públicas do Sistema de Informações sobre Mortalidade (SIM) e do Projeto de Acompanhamento de Custo da Cesta Básica, da Universidade Estadual de Santa Cruz. Os dados foram tratados e transformados em séries temporais univariadas, que passaram por análise exploratória, em busca de extrair estatísticas e características como tendência e sazonalidade e também testá-las quanto a estacionariedade. Também verificou-se a existência de correlação entre as séries. Os resultados apontaram a não estacionariedade das séries, com verificação de tendências estocásticas e determinística, e além disso, os valores de correlação entre as séries de homicídios e de custo da cesta básica mostraram-se irrelevantes.

**Palavras-chave**: Segurança pública. Cesta básica. Séries temporais.

## ABSTRACT

Violence is among the biggest problems faced by Brazilian public administration and the increase in the number of homicides in interior cities is an alarming factor for the performance of public security institutions. Understanding its causes is a major challenge for social development and analyzing the available data on homicide indicators can be an efficient method that results in appropriate conclusions. This work aims to apply computational techniques to analyze historical data series on homicides and also on the cost of the basic food basket in the cities of Ilhéus and Itabuna, Bahia, Brazil, for later application in prediction models with deep learning. Data were collected from the public databases of the Mortality Information System (SIM) and the Basic Basket Cost Monitoring Project, at the State University of Santa Cruz. Data were processed and transformed into univariate time series, which underwent exploratory analysis, in order to extract statistics and characteristics such as trends and seasonality and also test them for stationarity. It was also verified the existence of correlation between the series. The results showed the non-stationarity of the series, with verification of stochastic and deterministic trends, and in addition, the correlation values between the series of homicides and the cost of the basic food basket proved to be irrelevant.

**Keywords**: Public security. Basic food basket. Time series.

## 1. INTRODUCTION

Violence is among the biggest problems faced by Brazilian public management, especially in areas and regions with low socioeconomic indicators. The Global Peace Index (INSTITUTE FOR ECONOMICS & PEACE, 2019) showed that Brazil is at a great disadvantage compared to some social indicators and highlighted that, in addition to the worsening of violent crimes, the country has a considerably high homicide rate. State of Bahia, according to (CERQUEIRA et al., 2019), ranked 5th in the highest homicide rates in the states of the Northeast region, a region in which the homicide rate presents the highest proportion in coastal cities. The prioritization of the use of police coping force at the expense of intelligence and investigation in the state helps fueling the cycle of violence.

Another factor, highlighted by (REICHENHEIM et al*.,* 2011), is the phenomenon of internalization in the concentration of homicides, which was previously more present in large metropolises. Contributing to this process are the difficulties of public institutions to contain the advance, as organized crime seeks new spheres of action, and the degradation of urban socioeconomic indicators that drive problems such as poverty, social inequality and lack of access to goods and services, which are no longer exclusive problems in metropolitan regions.

Understanding the causes of violence has become a major challenge for managers and researchers when the objective is to propose adequate solutions to mitigate this factor that hinders the development of any society. Taking advantage of modern computer modeling techniques applied to data, it becomes feasible to represent socioeconomic and public safety data in the form of time series and analyze them in order to prepare them to generate predictive models for violence, as well as find the causal relationship between these data.

The cities of Ilhéus and Itabuna, located in the southern region of the state of Bahia, like other sub-regions in the Northeast of Brazil, have significant rates of urban violence and low socioeconomic indicators. The objective of this work is to analyze the time series of homicides and cost of the basic food basket in the microregion Ilhéus-Itabuna and extract their characteristics and behaviors that are relevant for further application in prediction models with deep learning, in addition to identifying whether there is causality between the observed variables. These time series are important resources that can help

the region's governments to identify and meet economic and social needs that mitigate violence in the short and long term.

Approaches for predicting crimes through complex systems using machine learning, also evaluating the factors that most contribute to violence, were developed by (ALVES; RIBEIRO; RODRIGUES, 2018). On the other hand, the prediction of criminal occurrences also using deep learning with neural networks with crime data and special geolocation was addressed in (YU et al*.,* 2011), (KANG; KANG, 2017), (STEC; KLABJAN, 2018), (SAFAT; ASGHAR; GILLANI, 2021).

## 2. MATERIALS AND METHODS

The methodology steps of this work consisted on: data collection in public databases; data extraction, cleaning and tabulation; representation in the form of univariate time series and exploratory data analysis.

### 2.1. Survey and Data Collection

The time series of homicides in the cities of Ilhéus and Itabuna were extracted from the Tabnet system, from DATASUS (Ministry of Health), through which it is possible to access data from the Mortality Information System (SIM). Monthly data on mortality from homicides in the two cities for the period 1996 to 2018 were extracted. The search criteria used were for data on deaths by place of occurrence characterized as homicides according to the CID-10 definition (10th Revision of the International Code of Diseases ) for deaths due to aggressions and legal interventions (X85-Y09 and Y35-Y36), according to the criteria defined for homicides in (CERQUEIRA et al ., 2019).

The cost of the basic food basket time series for the cities of Ilhéus and Itabuna were extracted in the domain of the Project Monitoring the Cost of the Basic Food Basket (ACCB) of UESC. The series consist of the monthly price of the basic basket, which consists of a sum of the average prices of several products that make up the basket and their respective quantities. The range of monthly data is from 2005 to 2020.

Each of these series was initially tabulated in comma-separated data files (.csv) and later loaded into data structures called dataframes, created by the Pandas library of the Python language.

## 2.2. Exploratory Data Analysis

According to (MALIK et al., 2016), data professionals and scientists use up to 80% of the time to prepare data for prediction algorithms. This is because the problem of identifying and removing discontinuity in data is one of the most faced in preparing data for long-term forecasting. And, after all, the accuracy of the prediction results depends on the quality of the data that will be processed.

The exploration consisted of an analysis of the time series in order to investigate series statistics, possible discontinuities, identify components such as trend and seasonality and classify the series in terms of stationarity. For this, graphical visualization resources were used (temporal graphs, dispersion, histograms, moving averages, annual series) and also the application of the stationarity hypothesis tests ADF (Augmented Dickey-Fuller), KPSS (Kwiatkowski-Phillips-Schmidt-Shin) ) and Phillips-Perron.

The ADF and Phillips-Perron tests test the null hypothesis that there is a unit root in the data generation process, thus assuming that the series is based on a stochastic trend random walk model according to Eq. (1), in which ρ = 1 (which characterizes the unit root) and $Y_t$ is a realization of the series at a given time t, $Y_{t-1}$ is the last realization, $u_t$ is a variable that characterizes the model error.

$$Y_t = \rho Y_{t-1} + u_t \quad (1)$$

Consequently, due to the presence of the trend, if the results of the ADF and Phillips-Perron tests do not reject the presence of the unit root, the time series is not stationary. On the other hand, the null hypothesis of the KPSS test states that the time series is stationary. The tests were applied both in their standard format, adding a constant to the process, as well as adding a trend, and the values of the statistical results against the critical values were verified at three confidence levels (90%, 95% and 99%) and the result of the p-value returned.

As unit root tests establish the presence of a stochastic trend and it is known that some time series can be generated by a deterministic trend process, the studies by (STADNYTSKA, 2010) and (IJOMAH; ENEGESELE, 2017) with the ADF test were used to define the characteristic trend type in case of non-stationarity.

### 2.3. Correlation Between Time Series

The dependence between the observed points of two variables can have an impact on the search for causality between two time series. Thus, the correlation between the series was evaluated based on visual analysis with superimposed graphs and through the Pearson correlation coefficient metric.

Since the correlation between two series may not occur at the same point in time, as one phenomenon often precedes another, Pearson's correlation coefficients were also calculated for the series related to their monthly variations (differentiated) in temporal lags of 1, 2, 4, 6 and 12 months. The correlation values then inform whether there is a relationship between the series over the course of their observed time, and whether there is a relationship between the previous event of one variable and the future of another.

## 3. RESULTS AND DISCUSSIONS

### 3.1. Exploratory Data Analysis

Figures 1 and 2 show the temporal graph of the monthly homicide series in Ilhéus and Itabuna, respectively, in the period 1996 to 2018. It is possible to see, in Figure 1, that the homicide series in Ilhéus presents the occurrence of two maximum points in 2011 (27 occurrences in March and 26 in December). The series of homicides in Itabuna has a maximum point in the year 2016 (30 homicides in the month of April). The statistics for the two series are shown in Table 1.

In the period analyzed, as shown in Table 1, the city of Itabuna had the highest average number of homicides/month, as well as standard deviation and maximum number of homicides in one month, in relation to the city of Ilhéus. From the graphs in Figures 1 and 2, it is not possible to extract apparent seasonality, but both cities seem to present different trend behaviors at different time intervals. In Figure 1, it is possible to assume four different behaviors of the trend over time: a period of constancy, with no apparent trend, from 1996 to 2003, ascendancy from 2004 to 2011, descent from 2012 to 2015 and, again, a stability from 2016 to 2018. In Figure 2, the presumptive behaviors are: stability

from 1996 to 2000, ascendancy from 2001 to 2009, and, again, a constancy from 2010 to 2018.

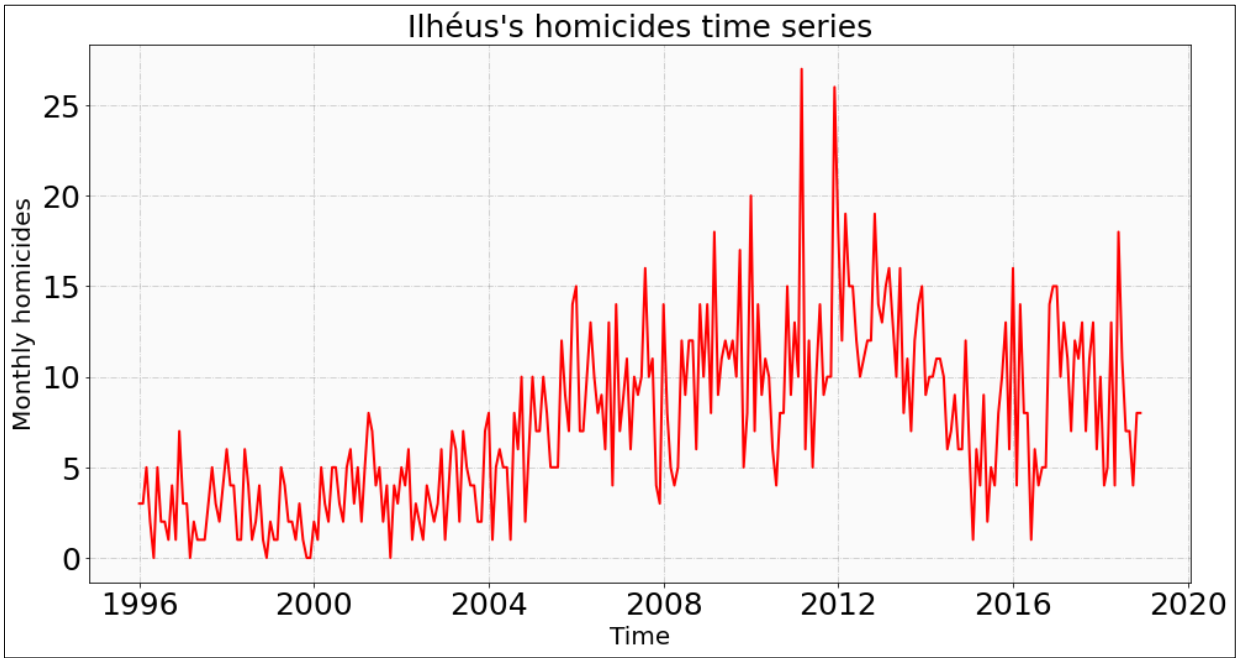


**Figure 1.** Time graph of the homicide series in Ilhéus

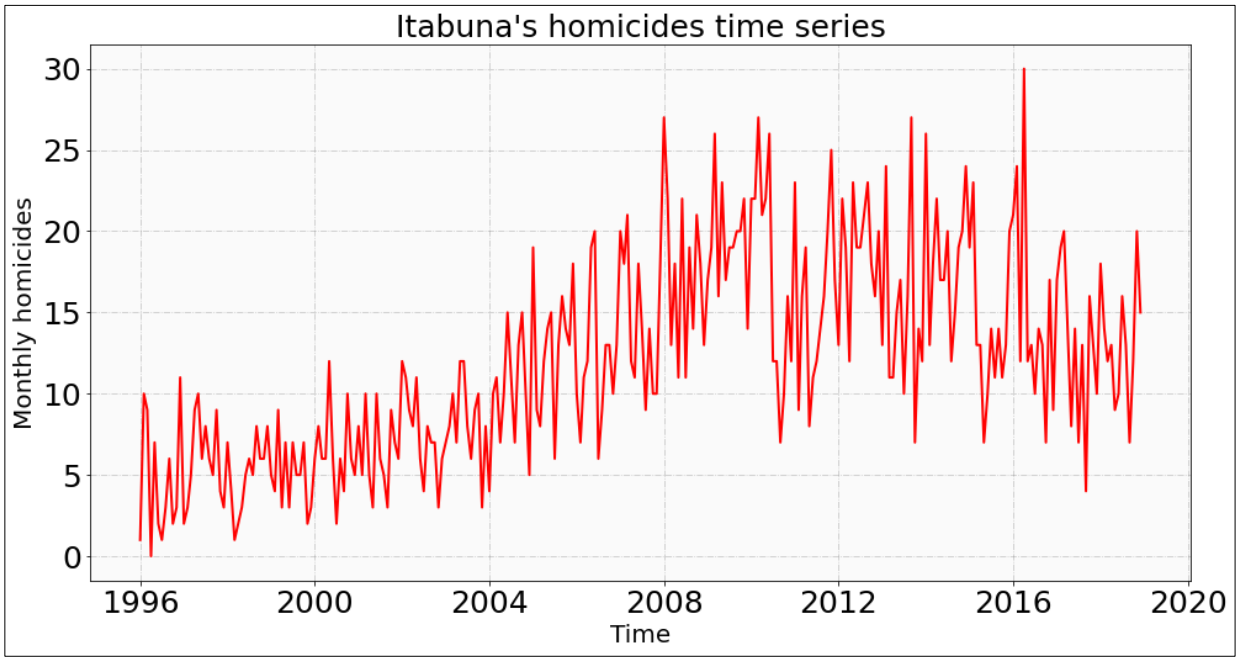


**Figure 2.** Time graph of the homicide series in Itabuna

**Table 1.** Homicide time series statistics.

| **Ilhéus's homicides time series** | | **Itabuna's homicides time series** | |
|---|---|---|---|
| **Statistics** | **Result** | **Statistics** | **Result** |
| Observations | 276 | Observations | 276 |
| Average (homicides/month) | 7.36 | Average (homicides/month) | 12.01 |
| Standard deviation | 4.82 | Standard deviation | 6.28 |
| Minimum | 0 | Minimum | 0 |
| 1st quartile | 4.00 | 1st quartile | 7.00 |
| Median | 6.50 | Median | 11.00 |
| 3rd quartile | 10.25 | 3rd quartile | 17.00 |
| Maximum | 27 | Maximum | 30 |

The presence of a trend is already an indicator of non-stationarity in the series. Long-term changes in data behavior can mean the existence of structural breaks, another source of non-stationarity in a series' data. The term structural break is a concept studied and analyzed by econometrics and means that there is one or more changes at the level of the series, in its dispersion and/or inclination (SHIKIDA; PAIVA; ARAÚJO, 2016).

Long-term changes in behavior are clearer when smoothing by moving averages with sliding windows of 2, 4, 6 and 12 months is applied, as in Figure 3, for the Ilhéus series. The histogram with asymmetric distribution, as shown in Figure 4 for the data from Itabuna, corroborates the hypothesis of non-stationarity. The variation of annual averages and variances highlights the heteroskedasticity of the system, as in Figure 5 for the city of Ilhéus, and consequently the lack of stability in the behavior of these metrics. In addition, the autocorrelation function (ACF) and partial autocorrelation (PACF) correlograms, such as those in Figure 6 of the Itabuna data, demonstrate strong evidence of an autoregressive process in the data.

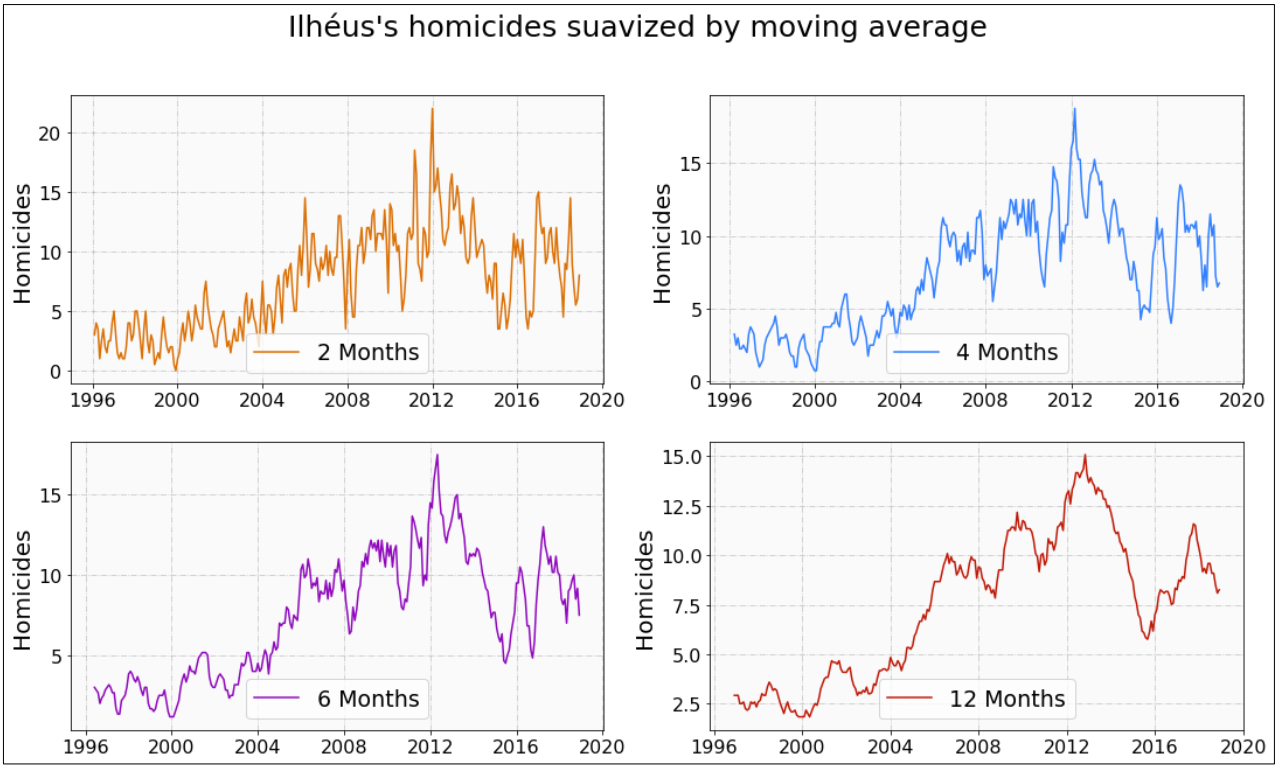


**Figure 3.** Smoothing by moving averages of the Ilhéus homicide series

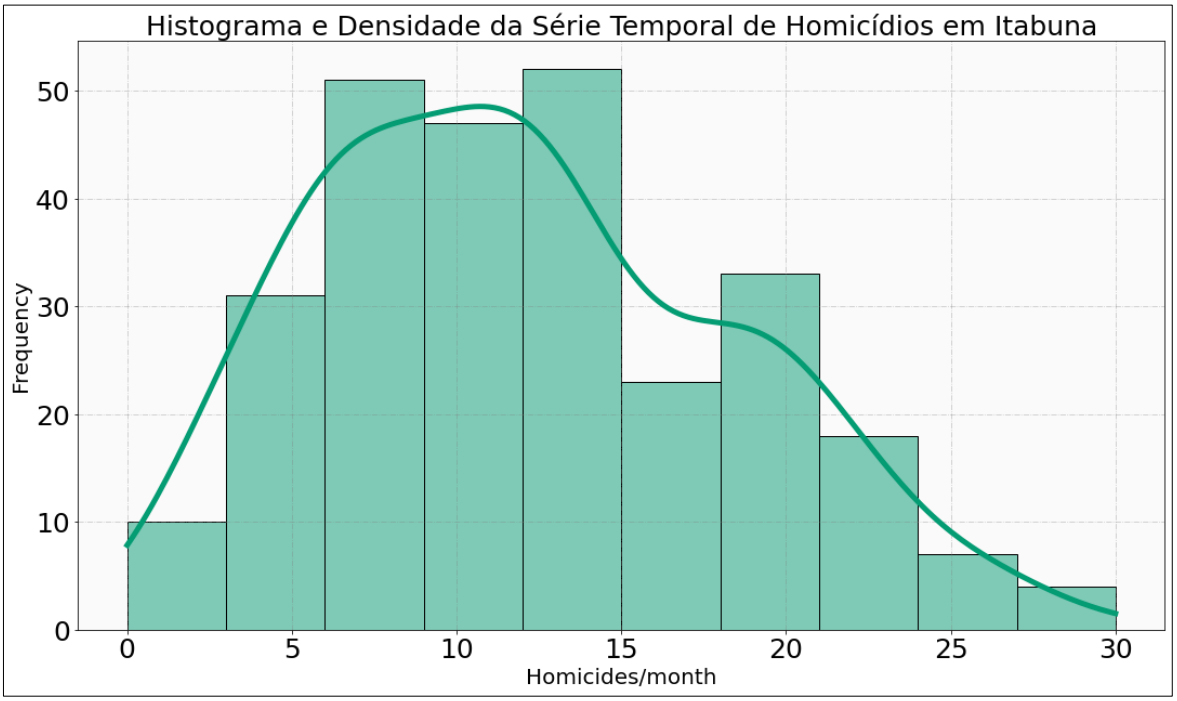


**Figure 4.** Histogram of data from the series of homicides in Itabuna

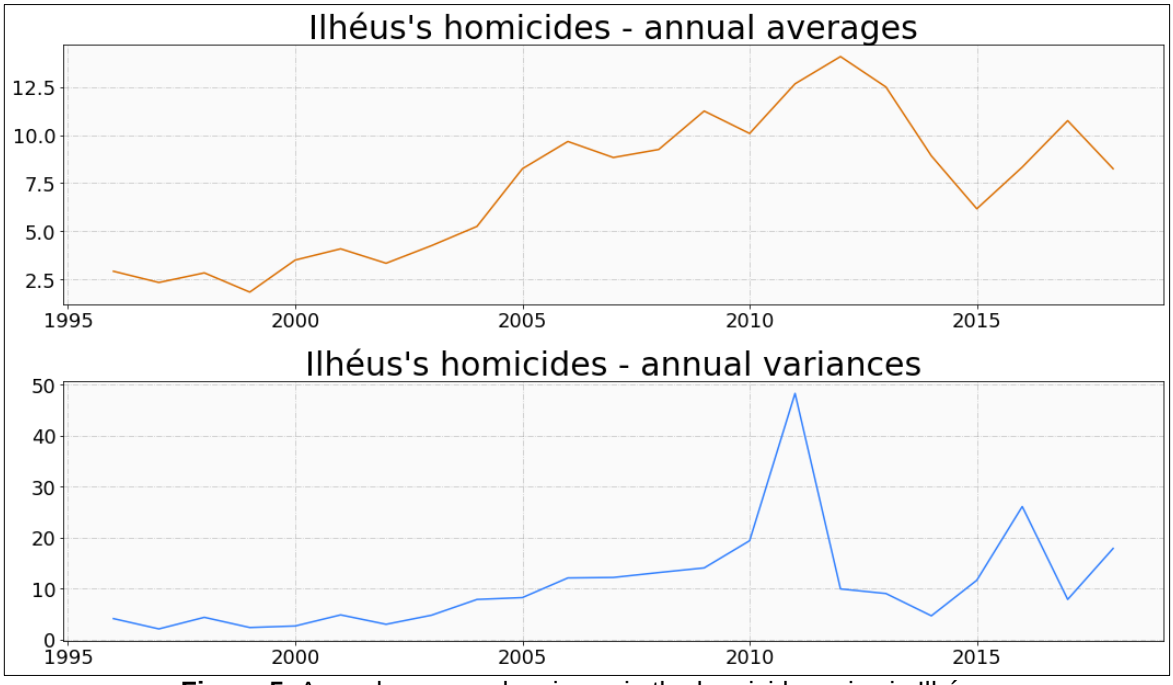


**Figure 5.** Annual mean and variance in the homicide series in Ilhéus

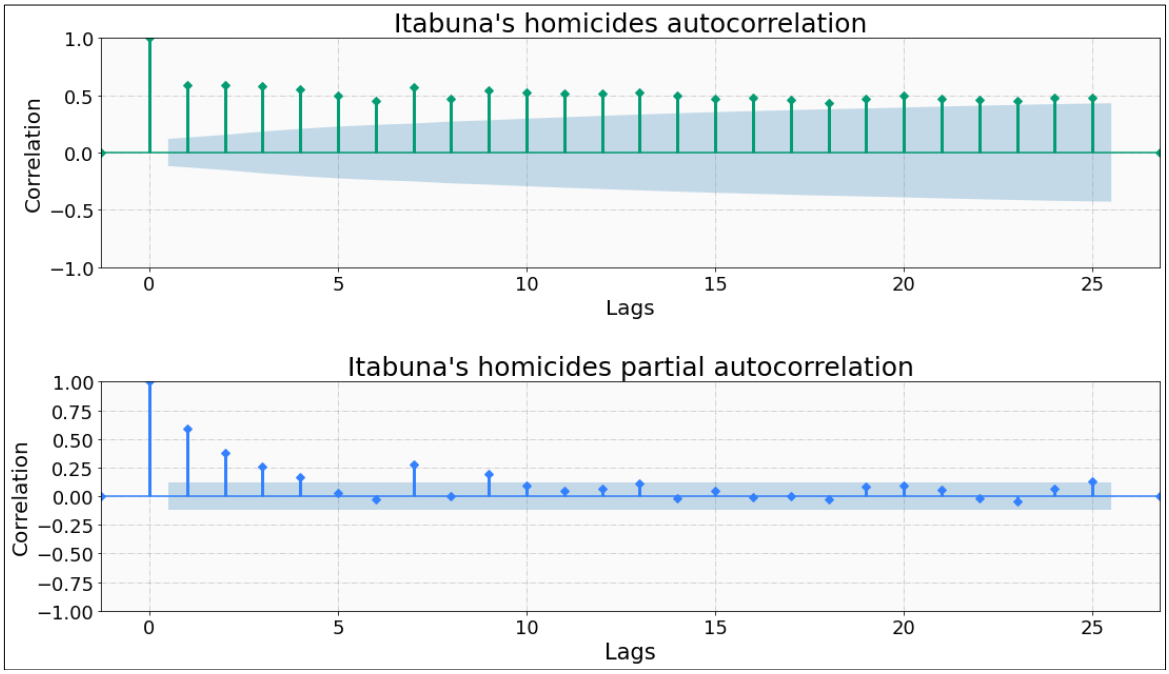


**Figure 6.** Correlograms of the Itabuna homicide series

Given the evidence observed, the ADF, KPSS and Phillips-Perron tests were performed in the hope of results that pointed to the non-stationarity of both series. The result of applying the tests for the homicide time series in Ilhéus is shown in Table 2, for the application scenarios with constant (C) and with constant and trend (C+T). From the

results for the p-value in Table 2, the ADF test confirms the null hypothesis of the existence of a unit root and the KPSS test rejects the null hypothesis of stationarity. On the other hand, the Phillips-Perron test rejected the existence of a unit root in the data. The unexpected result in relation to this test may indicate a bias due to the existence of structural break, since, as stated by (PERRON, 1989), a problem associated with the presence of structural break is the bias of unit root statistical test results. The results for the time series of homicides in Itabuna had a similar behavior.

**Table 2.** Stationarity tests applied to the homicides series in Ilhéus.

| Test | Statistical result | p-value | Critical values 10% | Critical values 5% | Critical values 1% |
|---|---|---|---|---|---|
| KPSS (C) | 1.895 | 0.0001 | 0.347 | 0.461 | 0.742 |
| ADF (C) | -1.734 | 0.4135 | -2.572 | -2.872 | -3.455 |
| Phillips-Perron (C) | -13.180 | 1.1995e-24 | -2.572 | -2.872 | -3.454 |
| KPSS (C+T) | 0.409 | 0.0002 | 0.119 | 0.147 | 0.217 |
| ADF (C+T) | -1.932 | 0.6379 | -3.136 | -3.427 | -3.993 |
| Phillips-Perron (C+T) | -16.110 | 1.1137e-22 | -3.136 | -3.426 | -3.992 |

According to (STADNYTSKA, 2010) and (IJOMAH; ENEGESELE, 2017), non-stationary processes that have a unit root present a stochastic tendency. As the ADF test suggested the existence of a unit root in the data generation processes of the two series of homicides, the original series of homicides was differentiated and it was verified that the series of monthly variations, based on the stationarity tests, they were stationary.

From the analysis of the behavior of the evolution of data month by month, the following seasonality was detected for the series of homicides in Ilhéus: the months of January and March registered, in most cases, an increase in the case of homicides, and February and April registered falls; there was also a recurrent decrease in cases in the June-July period and an increase in the December-January period. For the series of homicides in Itabuna, it was verified that there is a decrease in occurrences in the period June-July in most years and an increase in occurrences in the period December-January.

The temporal graphs of the basic food basket cost series for Ilhéus and Itabuna, respectively, in Figures 7a and 7b, show an upward behavior of the data in the long run.

Smoothing by moving averages with sliding windows of the cost series of the Itabuna food basket in Figure 8, inclusive, indicates that this series may have a deterministic trend component, represented from a model containing a constant that varies as a function of time.

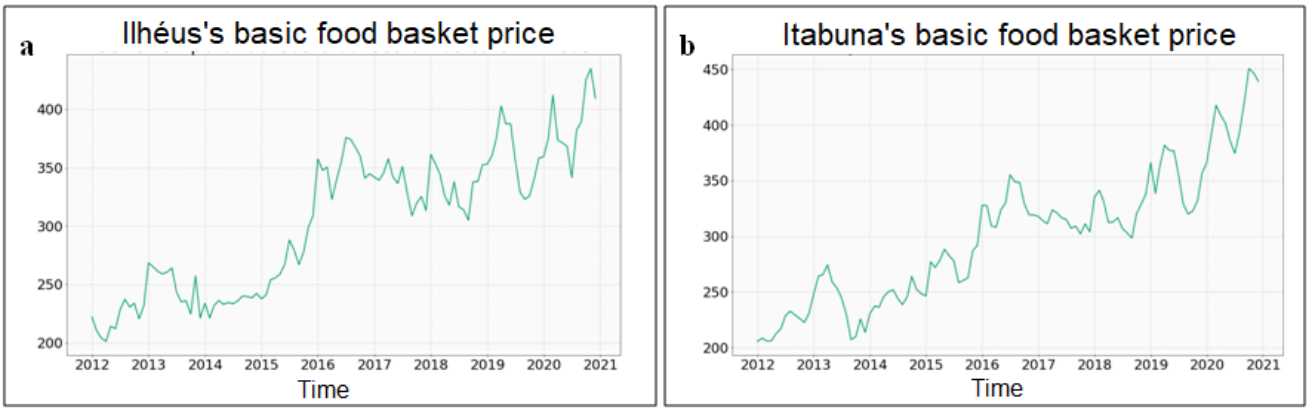


**Figure 7.** Time graph of the cost series of the basic food basket of Ilhéus (a) and Itabuna (b)

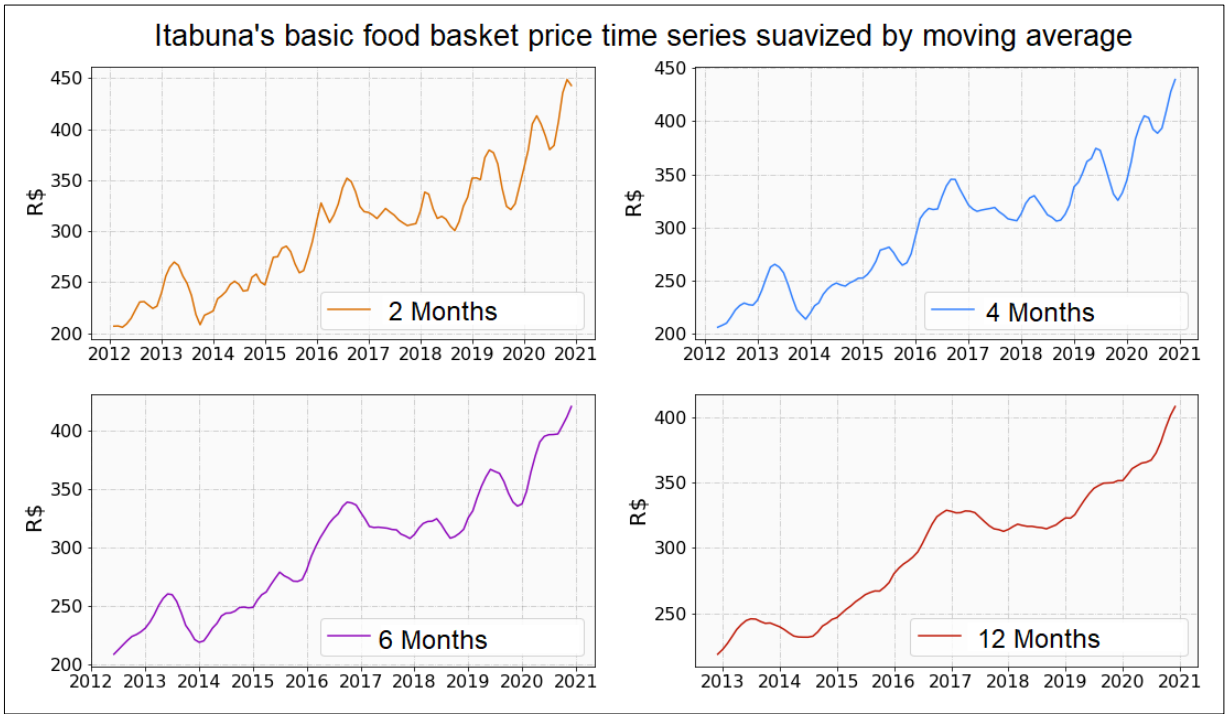


**Figure 8.** Smoothing by moving averages of the cost series of Itabuna's basic food basket

The application of the ADF and Phillips-Perron tests confirmed the null hypothesis that the time series of cost of the basic food basket in Ilhéus has a unit root and the KPSS test rejected the stationarity hypothesis for this same series. When applying the ADF test,

only the constant variable in the test regression model resulted in a p-value of 0.655. When adding the trend to the test model (ie, considering the variables constant and the trend) the p-value results in 0.079 (greater than 5%). In the same way as for the homicide series, the technique of differentiation in the series was applied and the series resulting from monthly variations was shown to be stationary. Thus, the unit root process detected by the ADF test characterizes the existence of a stochastic trend.

For the cost series of the Itabuna food basket, the result of the ADF test suggested the rejection of the null hypothesis with a p-value of 0.009 and a statistical result of -4.003, with critical values of -4.05 (1%), - 3.45 (5%) and -3.15 (10%). When applying the test again, considering only the constant variable, a result that suggested the non-rejection of the null hypothesis was obtained (p-value of 0.918 and statistical result of -0.349). This contradictory behavior is explained by (STADNYTSKA, 2010) and (IJOMAH; ENEGESELE, 2017) as characteristic of processes that tend to be deterministic in nature.

Once the hypothesis of the tendency of the series to be deterministic in nature was corroborated by the ADF tests, a trend removal model was developed in order to stationarize the series. The trend of the series was modeled using a simple polynomial regression with degree 3 polynomials. After modeling and removal of the trend, stationarity tests were applied on the residual and all results suggested the stationarity of the series.

Seasonality periods were not identified for the basic food basket cost series.

### 3.2. Correlation analysis between series

The results of the Pearson correlation coefficients calculated between the series of homicides and the cost of the basic food basket showed a very low correlation, with values of -0.044 for Ilhéus and -0.041 for Itabuna. The homicide series have a coefficient of about 0.20 between them, which can be considered a low correlation in general terms. The basic food basket cost series have a correlation coefficient of 0.99 between them.

Correlating the homicide series with the cost series of the basic food basket lagged in time, in lags of 1, 2, 4, 6, 10 and 12 months, no significant correlation coefficient value was found, and all values were below the absolute value of 0.06.

## 4. CONCLUSIONS

The time series of homicides and cost of the basic food basket analyzed here present some sources of non-stationarity verified in the long-term behavior. With the exception of the cost series of Itabuna's basic food basket, which has a deterministic trend component, the other three analyzed series present a stochastic trend component. Furthermore, only the homicide series showed signs of seasonality. At this moment, for the scenario visualized, the series of homicides and cost of the basic basket of both cities are not correlated.

The identification of these behaviors, as well as the discontinuities present in the series data, are important in the elaboration of prediction models. In this case, the application of time series analysis techniques of socioeconomic and public safety data from the microregion Ilhéus - Itabuna can result in knowledge of the behavior of the data over time and in a more careful perception of the causal relationship between the series, in addition to that, when combined with the generation of predictions, they can represent mechanisms that support the action of public security institutions in the fight against violence and inequality.

Finally, the absence of stationarity in the series sets a good scenario for the application of predictive models with deep learning. One of the advantages of using deep learning for time series is precisely that there is no requirement for series stationarity , as happens in most traditional models; another advantage is the development of specific neural network architectures for chronological data.

## REFERÊNCIAS

Alves, L. G. A; Ribeiro, H. V.; Rodrigues, F. A. **Crime prediction through urban metrics and statistical learning**. Physica A, v. 505, p. 435 – 443, 2018.

Cerqueira, D. *et al.* **Atlas da Violência: retrato dos municípios brasileiros**. In: Atlas da Violência, p. 52, 2019.

Chahim, A.; Cunha, M. A.; Knight, P. T.; Pinto, S. L. **e-gov.br: a Próxima Revolução Brasileira**. São Paulo: PrenticeHall, 2004.

Ijomah, M. A.; Enegesele, D. On the Use of Unit Root Test to Differentiate Between Deterministic and Stochastic Trend in Time Series Analysis**. American Scientific Research Journal for Engineering, Technology, and Sciences**, v. 27, p. 234-246, 2017.

Institute for Economics & Peace. **Global Peace Index: Measuring Peace in a Complex World**. Institute for Economics & Peace, p. 1–99, 2019.

Kang, H-W; Kang, H-B. Prediction of crime occurrence from multi-modal data using deep learning. **PLoS ONE**, v. 12, n. 4, 2017.

Malik, H., Davis, I. J., Godfrey, M. W.; Neuse, D.; Manskovskii, S. Connecting the dots: anomaly and discontinuity detection in large-scale systems. **J Ambient Intell Human Comput**, v. 7, p. 509–522, 2016.

Perron, P. The great crash, the oil-price shock, and the unit-root. **Econometrica**, v. 57, p. 1361–1401, 1989.

Reichenheim, M. E. *et al.* Violence and injuries in Brazil: The effect, progress made, and challenges ahead. **The Lancet**, v. 377, p. 1962–1975, 2011.

Safat, W.; Asghar, S.; Gillani, S. A. Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques. **IEEE Access**, v. 9, p. 70080-70094, 2021.

Shikida, C.; Paiva, G. L.; Araújo Jr., A. F. Análise de quebras estruturais na série do preço do boi gordo no estado de São Paulo. **Economia Aplicada**, v. 20, p. 265-286, 2016.

Stadnytska, T. **Deterministic or stochastic trend: Decision on the basis of the augmented dickey-fuller test**. Methodology, v. 6, p. 83–92, 2010.

Stec, A.; Klabjan, D. **Forecasting Crime with Deep Learning**, p. 1–20, 2018.

Yu, C. H. *et al.* **Crime forecasting using data mining techniques**. Proceedings - IEEE International Conference on Data Mining, ICDM, n. December, p. 779–786, 2011.