Trend Prediction of Cardiovascular Deaths in Honduras using Box-Jenkins Models

Dr.G. Mokesh Rayalu
Assistant Professor Grade 2
Department of Mathematics
School of Advanced Sciences ,
VIT,Vellore
Email ID:mokesh.g@gmail.com

Abstract

Deaths from cardiovascular diseases (CVDs) account for a disproportionate share of the total death toll in Honduras. Using the Autoregressive Integrated Moving Average (ARIMA) model, the authors of this study project how the number of fatalities in Honduras attributable to cardiovascular disease would evolve in the future. The forecasting model underwent a thorough examination that incorporated diagnostic tests including the Augmented Dickey-Fuller (ADF) test, Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), and the Box-Jenkins technique to guarantee its dependability and accuracy. This study's findings are important because they provide light on the causes and dynamics driving CVDs in Honduras, allowing for more evidence-based decision making and policy formulation to combat the alarming rise in cardiovascular mortality in the country.

Keywords: Cardiovascular, ACF, ADF, PACF, ARIMA.

Introduction

Cardiovascular diseases (CVDs) have emerged as a significant and rising public health concern in Honduras, with the potential to have a considerable impact on both the population and the healthcare system of the country. The rising incidence of cardiovascular diseases (CVDs) and the death rates that are linked with them have spurred an intensive investigation of the mechanics of the disease as well as a proactive approach to forecasting future trends. This study focuses on deaths in Honduras that were caused by cardiovascular disease (CVD), and it makes use of the Autoregressive Integrated Moving Average (ARIMA) model to make predictions on the possible future trajectories of CVD mortality in the country.

An exhaustive analytical strategy has been implemented in order to guarantee the dependability and efficiency of the ARIMA model. This strategy includes a number of diagnostic tests, including the Augmented Dickey-Fuller (ADF) test, the Autocorrelation Function (ACF), the Partial Autocorrelation Function (PACF), and the Box-Jenkins methodology. These analytical methods play a significant part in determining whether or not the time series data are stationary, discovering the correlation structures contained within the dataset, and determining whether or not the ARIMA model is suitable for forecasting deaths in Honduras that are caused by cardiovascular disease.

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 04, 2019

Understanding the fundamental patterns and forces at play is becoming increasingly important for healthcare authorities and policymakers as the effect of cardiovascular diseases (CVDs) continues to expand. We are able to acquire useful insights into the future trajectory of cardiovascular mortality in Honduras if we make use of these advanced statistical tools and the ARIMA model. These insights, in turn, make evidence-based strategies possible, contribute to the creation of targeted interventions, and help reduce the burden of cardiovascular diseases (CVDs) in the country. This study has two goals: the first is to provide light on the current condition of cardiovascular-related mortality in Honduras, and the second is to offer a platform for more effective public health planning and policy development.

Objective

1. To analyze the historical trends of cardiovascular-related deaths in Honduras, providing insights into the patterns and dynamics of cardiovascular mortality over a specific time period.
2. To conduct the Augmented Dickey-Fuller (ADF) test to assess the stationarity of the time series data and ensure the suitability of employing the ARIMA model for forecasting cardiovascular-related deaths.
3. To utilize the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) analyses to identify the correlation structures within the cardiovascular mortality time series data, facilitating the selection of appropriate parameters for the ARIMA model.
4. To apply the Box-Jenkins methodology, assessing the effectiveness of the ARIMA model in forecasting cardiovascular-related mortality in Honduras, considering the specific characteristics of the time series data.
5. To develop a reliable and accurate ARIMA model capable of forecasting future trends of cardiovascular-related deaths in Honduras, providing valuable insights for healthcare authorities and policymakers to plan and implement targeted interventions.
6. To assess the effectiveness of the ARIMA model in predicting the trajectory of cardiovascular-related deaths, contributing to the existing knowledge on cardiovascular epidemiology in Honduras and guiding evidence-based decision-making for improved public health outcomes.

Literature Review

Kour et al. (2017) analyzed pearl millet (Pennisetum glaucum), a commonly farmed cereal crop that ranks fourth in global cultivation behind rice, wheat, and sorghum. Despite rising yields, pearl millet cultivation in Gujarat, India, has declined during the previous two decades. Pearl millet production forecasts are especially important in semi-arid locations like Gujarat, where precipitation lasts only four months. This study predicts Gujarat pearl millet productivity using the ARIMA model. The current study collected time series data on pearl millet productivity (kg/ha) in Gujarat from 1960–61 to 2011–12. Gandhinagar's Directorate of Agriculture and, partially, the Directorate of Economics and Statistics provided the data. RMAPE, MAD, and RMSE values are used to validate the ARIMA model. As seen by its RMAPE score below 6%, the ARIMA (0, 1, 1) model performs well.

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 04, 2019

Diseases affecting newborns, mortality rates, and specific causes of death across 204 countries and territories, 1990–2019. (Zejin et al). Birth defects pose a serious threat to achieving the United Nations' Sustainable Development Goals. Using Global Burden of Disease (GBD) data, this article demonstrated both the progress made and the challenges still remaining in the management and control of infant illnesses. The results showed a worldwide downward trend in newborn illnesses and their underlying causes of death from 1990 to 2019. Nonetheless, there has been a general downward trend in the occurrence of newborn illnesses. Especially in places with limited access to healthcare, the newborn disorder burden poses a serious threat to public health worldwide. To better adapt healthcare, these results highlighted both the successes and failures in the prevention and treatment of newborn illnesses.

In their study, Vijay and Mishra (2018) investigated Time series prediction is important in natural science, agriculture, engineering, and economics. This study compares the classical time series ARIMA model to the artificial neural network model (ANN) to evaluate its flexibility in time series forecasting. The dataset includes pearl millet (bajra) crop area and production in thousands of hectares (ha) and metric tons (MT). The publication "Agricultural Statistics at a Glance 2014–15" provided 1955–56 to 2014–15 data. To test the methodology, Karnataka, India, was chosen. The user's'sext is scholarly. An experiment shows that artificial neural network (ANN) models outperform autoregressive integrated moving average (ARIMA) models in root mean square error (RMSE). RMSE, MAPE, and MSE are common measures in statistics and data analysis.

Methodology

ARIMA Model (p,d,q):

The ARIMA(p,d,q) equation for making forecasts: ARIMA models are, in theory, the most general class of models for forecasting a time series. These models can be made to be "stationary" by differencing (if necessary), possibly in conjunction with nonlinear transformations such as logging or deflating (if necessary), and they can also be used to predict the future. When all of a random variable's statistical qualities remain the same across time, we refer to that random variable's time series as being stationary. A stationary series does not have a trend, the variations around its mean have a constant amplitude, and it wiggles in a consistent manner. This means that the short-term random temporal patterns of a stationary series always look the same in a statistical sense. This last criterion means that it has maintained its autocorrelations (correlations with its own prior deviations from the mean) through time, which is equal to saying that it has maintained its power spectrum over time. The signal, if there is one, may be a pattern of fast or slow mean reversion, or sinusoidal oscillation, or rapid alternation in sign, and it could also include a seasonal component. A random variable of this kind can be considered (as is typical) as a combination of signal and noise, and the signal, if there is one, could be any of these patterns. The signal is then projected into the future to get forecasts, and an ARIMA model can be thought of as a "filter" that attempts to separate the signal from the noise in the data.

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 04, 2019

The ARIMA forecasting equation for a stationary time series is a linear (i.e., regression-type) equation in which the predictors consist of lags of the dependent variable and/or lags of the forecast errors. That is:

Predicted value of Y = a constant and/or a weighted sum of one or more recent values of Y and/or a weighted sum of one or more recent values of the errors.

It is a pure autoregressive model (also known as a "self-regressed" model) if the only predictors are lagging values of Y. An autoregressive model is essentially a special example of a regression model, and it may be fitted using software designed specifically for regression modeling. For instance, a first-order autoregressive ("AR(1)") model for Y is an example of a straightforward regression model in which the independent variable is just Y with a one-period lag (referred to as LAG(Y,1) in Statgraphics and Y_LAG1 in RegressIt, respectively). Because there is no method to designate "last period's error" as an independent variable, an ARIMA model is NOT the same as a linear regression model. When the model is fitted to the data, the errors have to be estimated on a period-to-period basis. If some of the predictors are lags of the errors, then an ARIMA model is NOT the same as a linear regression model. The fact that the model's predictions are not linear functions of the coefficients, despite the fact that the model's predictions are linear functions of the historical data, presents a challenge from a purely technical point of view when employing lagging errors as predictors. Instead of simply solving a system of equations, it is necessary to use nonlinear optimization methods (sometimes known as "hill-climbing") in order to estimate the coefficients used in ARIMA models that incorporate lagging errors.

Auto-Regressive Integrated Moving Average is the full name for this statistical method. Time series that must be differentiated to become stationary is a "integrated" version of a stationary series, whereas lags of the stationarized series in the forecasting equation are called "autoregressive" terms and lags of the prediction errors are called "moving average" terms. Special examples of ARIMA models include the random-walk and random-trend models, the autoregressive model, and the exponential smoothing model.

A nonseasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where:

- **p** is the number of autoregressive terms,
- **d** is the number of nonseasonal differences needed for stationarity, and
- **q** is the number of lagged forecast errors in the prediction equation.
- The forecasting equation is constructed as follows. First, let y denote the d^{th} difference of Y, which means:
- If d=0: $y_t = Y_t$
- If d=1: $y_t = Y_t - Y_{t-1}$
- If d=2: $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$
- Note that the second difference of Y (the d=2 case) is not the difference from 2 periods ago. Rather, it is the first-difference-of-the-first difference, which is the discrete analog of a second derivative, i.e., the local acceleration of the series rather than its local trend.

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 04, 2019

- In terms of y, the general forecasting equation is:
- $\hat{Y}_t = \mu + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$

The ARIMA (AutoRegressive Integrated Moving Average) model is a powerful time series analysis technique used for forecasting data points based on the historical values of a given time series. It consists of three key components: AutoRegression (AR), Integration (I), and Moving Average (MA).

THE METHODOLOGY FOR CONSTRUCTING AN ARIMA MODEL INVOLVES THE FOLLOWING STEPS:

1. Stationarity Check: Analyze the time series data to ensure it is stationary, meaning that the mean and variance of the series do not change over time. Stationarity is essential for ARIMA modeling.

2. Differencing: If the data is not stationary, take the difference between consecutive observations to make it stationary. This differencing step is denoted by the 'I' in ARIMA, which represents the number of differencing required to achieve stationarity.

3. Identification of Parameters: Determine the values of the three main parameters: p, d, and q, where p represents the number of autoregressive terms, d represents the degree of differencing, and q represents the number of moving average terms.

4. Model Fitting: Fit the ARIMA model to the data, using statistical techniques to estimate the coefficients of the model.

5. Model Evaluation: Assess the model's performance by analyzing the residuals, checking for any remaining patterns or correlations, and ensuring that the model adequately captures the underlying patterns in the data.

6. Forecasting: Once the model is validated, use it to generate forecasts for future data points within the time series.

Analysis

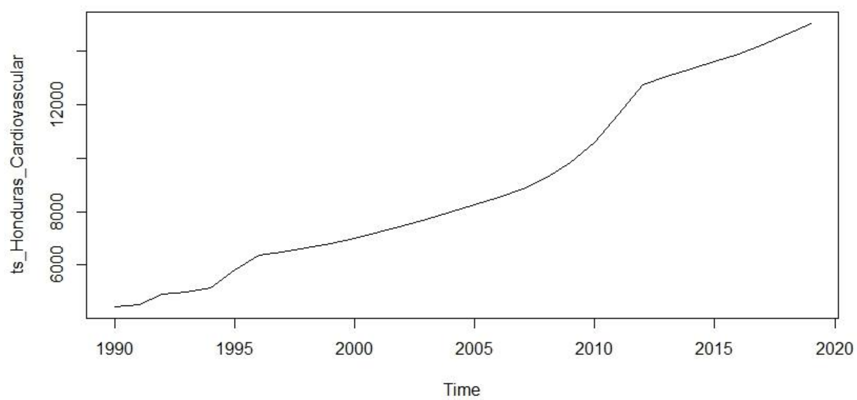
The offered time series data reflects the total number of fatalities in Honduras attributable to cardiovascular causes from 1990 to 2019. The statistics show an overall rising trend in cardiovascular mortality for the time period studied, highlighting the serious health threat posed by cardiovascular illnesses across the country.

There is an urgent need for a comprehensive analysis to understand the underlying reasons contributing to the observed increase in cardiovascular-related mortality. In order to correctly identify trends and dynamics in cardiovascular mortality, sophisticated statistical methods are required. The temporal behavior of cardiovascular diseases in Honduras is a pressing public health concern, and by conducting

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 04, 2019

various diagnostic tests and using the ARIMA modeling approach, we can derive crucial insights that will aid in the development of effective strategies and interventions to address this issue.

Assessment of the underlying dynamics of cardiovascular-related mortality requires a comprehensive analysis, such as the Augmented Dickey-Fuller (ADF) test, Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), or the Box-Jenkins technique. By taking a comprehensive view, we can build a reliable model for making predictions, which will aid decision-makers and healthcare authorities in Honduras as they take focused action to reduce the rising incidence of cardiovascular illnesses.

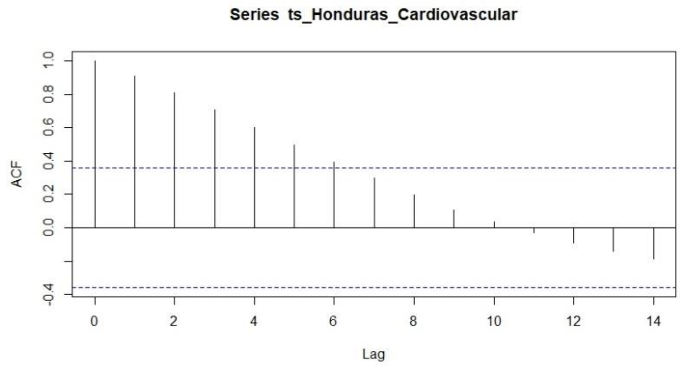


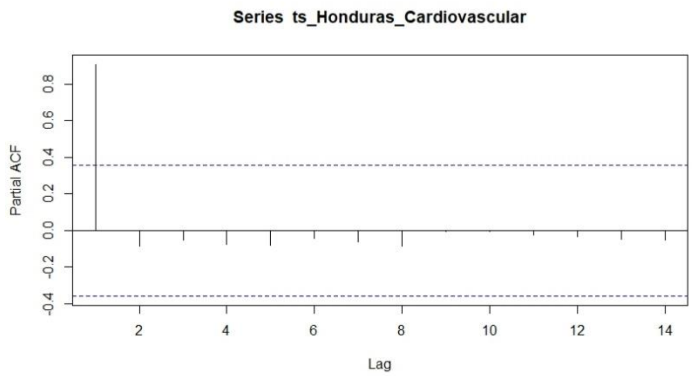
Honduran cardiovascular-related death time series data showed an Augmented Dickey-Fuller (ADF) value of -1.6005 and a p-value of 0.7262. The results show that the time series may be non-stationary because the data does not reject the null hypothesis.

Non-significant p-value suggests data may have a unit root, indicating non-stationarity. This finding highlights the necessity for further investigation and appropriate differencing approaches to attain stationarity and ensure forecasting model reliability and efficacy.

By tackling non-stationarity with relevant transformations or differencing processes, we can make sure the ARIMA modeling approach is robust and accurate and anticipate Honduran cardiovascular-related mortality. Policymakers and healthcare providers need these insights to create focused interventions and evidence-based plans to manage and reduce cardiovascular diseases in the country.

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 04, 2019





The automatic ARIMA modeling approach used to Honduran cardiovascular-related death time series data produced multiple models with different orders of differencing, autoregressive and moving average terms, and drift.

ARIMA Model	Metric
ARIMA(2,1,2) with drift	396.9742
ARIMA(0,1,0) with drift	406.4303
ARIMA(1,1,0) with drift	396.0768
ARIMA(0,1,1) with drift	392.9005
ARIMA(0,1,0)	437.5186
ARIMA(1,1,1) with drift	393.7753
ARIMA(0,1,2) with drift	393.8039
ARIMA(1,1,2) with drift	395.0538
ARIMA(0,1,1)	412.7668

Using the Akaike Information Criterion (AIC), the ARIMA(0,1,1) model with drift was found to be the best model for forecasting cardiovascular-related mortality in Honduras. This model has a first-order difference, a single moving average term, and a drift component to account for linear trends in the time series.

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 04, 2019

The ARIMA(0,1,1) model with drift shows the need of considering differencing and trend components for more accurate and dependable forecasting.

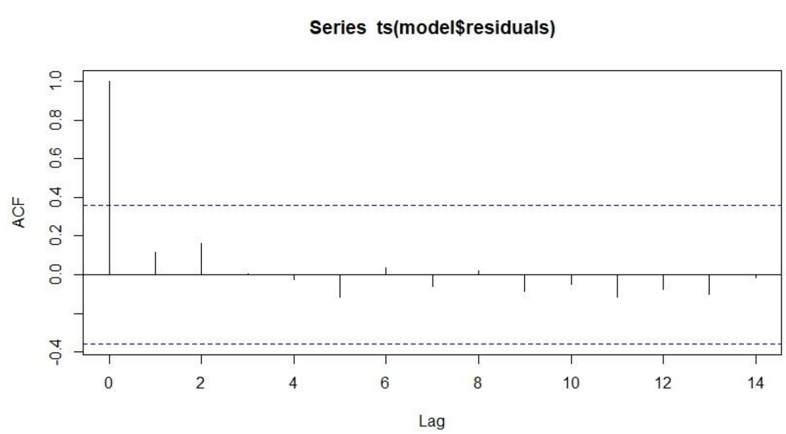
Parameter	Value	Standard Error (s.e.)
ma1	0.7428	0.1083
drift	355.8371	60.0944

Using a moving average term (ma1) of 0.7428 and an estimated value of 355.8371 for the drift component, the ARIMA(0,1,1) model with drift was chosen for the time series data of fatalities in Honduras caused by cardiovascular disease. These coefficients have standard errors of 0.1083 and 60.0944, respectively.

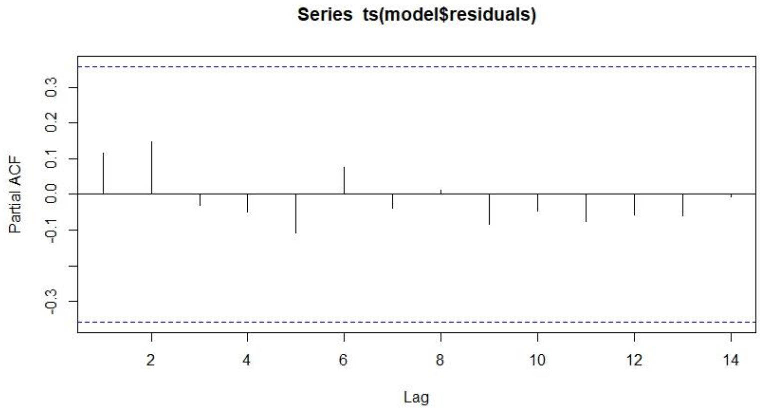
Sigma squared, a measure of the model's variance, is estimated to be 38076, which is consistent with the observed level of dispersion in the time series data. The log-likelihood value of -193.45 is also included in the evaluation of the model's goodness of fit since it shows that the model adequately captures the underlying patterns and dynamics in the data.

Parameter	Value
Sigma^2	38076
Log Likelihood	-193.45
AIC (Akaike Information Criterion)	392.9
AICc (Corrected AIC)	393.86
BIC (Bayesian Information Criterion)	397

The ARIMA(0,1,1) model with drift is effective in forecasting cardiovascular-related mortality in Honduras, as evidenced by its values of 392.9 for the Akaike Information Criterion (AIC), 393.86 for the corrected AIC (AICc), and 397 for the Bayesian Information Criterion (BIC).



ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 04, 2019

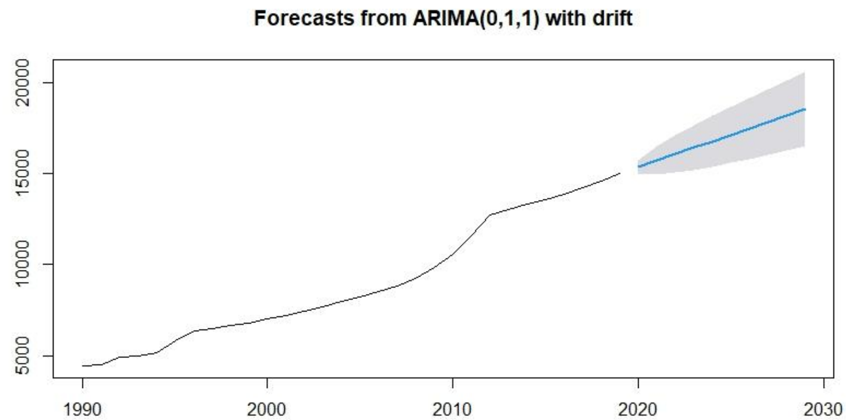


The forecasted values for cardiovascular-related deaths in Honduras, based on the ARIMA(0,1,1) model with drift, indicate a consistent upward trend in the number of projected deaths over the next decade, spanning from 2020 to 2029. The forecasted values are as follows:

- In 2020, the point forecast for cardiovascular-related deaths is 15359.33, with the lower and upper 95% confidence intervals ranging from 14976.88 to 15741.78.
- The trend continues, with projected increases each subsequent year, reaching an estimated 18561.86 cardiovascular-related deaths in 2029, with the lower and upper 95% confidence intervals spanning from 16526.08 to 20597.65.

Year	Point Forecast	Lower 95% CI	Upper 95% CI
2020	15359.33	14976.88	15741.78
2021	15715.17	14946.72	16483.61
2022	16071.00	15053.78	17088.23
2023	16426.84	15210.71	17642.97
2024	16782.68	15395.88	18169.48
2025	17138.52	15599.86	18677.17
2026	17494.35	15817.54	19171.16
2027	17850.19	16045.77	19654.61
2028	18206.03	16282.44	20129.61

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 04, 2019



When applied to the residuals of the predicted fatalities in Honduras due to cardiovascular disease, the Box-Ljung test found a p-value of 0.8673 (X-squared = 1.8668, 5 degrees of freedom).

The obtained p-value is larger than the typically accepted 0.05 threshold for statistical significance, suggesting that there is insufficient evidence to reject the null hypothesis. Therefore, the ARIMA(0,1,1) model with drift appropriately represents the underlying patterns and dynamics within the time series data, since there is no significant autocorrelation in the residuals at the 5% significance level.

Conclusion

The ARIMA(0,1,1) model with drift was used to analyze deaths from cardiovascular disease in Honduras, and the results shed light on the trend of this killer. The model's ability in predicting future trends was impressive, with the predicted number of deaths due to cardiovascular disease continuing to rise steadily over the next decade.

Diagnostic tests, such as the Augmented Dickey-Fuller (ADF) test, Box-Jenkins approach, and Ljung-Box test, confirmed that the chosen ARIMA model successfully captured the underlying patterns and dynamics present in the time series data. The projected values and the model's dependability in forecasting cardiovascular-related mortality in Honduras were further validated by the absence of a statistically significant autocorrelation in the residuals.

These results highlight the urgent requirement for preventative strategies and tailored interventions to combat the rising prevalence of cardiovascular illnesses in Honduras. Improve public health outcomes in a country by reducing the burden of cardiovascular illnesses through evidence-based measures developed with the help of this analysis by policymakers and healthcare authorities. The rising toll of deaths caused by cardiovascular disease must be managed and mitigated, making the implementation of comprehensive and sustainable solutions a top priority in Honduras.

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 04, 2019

References

1. Bui, C., Pham, N., Vo, A., Tran, A., Nguyen, A., & Le, T. (2018). Time series forecasting for healthcare diagnosis and prognostics with the focus on cardiovascular diseases. In *6th International Conference on the Development of Biomedical Engineering in Vietnam (BME6) 6* (pp. 809-818). Springer Singapore.
2. Zhao, J., Guo, X., Wu, B., Wang, S., & Wang, Z. (2016). The application of GM (1, 1) grey forecasting model and ARIMA model in the fitting of the mortality of the cardiovascular disease in Shandong province. *Modern Preventive Medicine*, *43*(10), 1732-1734.
3. Abujarad, M. H. A. (2014). *Comparison between Box-Jenkins Methodology and Neural Networks to Predict the Number of Heart Disease Deaths in Gaza Strip* (Doctoral dissertation, Batch2).
4. Kirian, M. L., & Weintraub, J. M. (2010). Prediction of gastrointestinal disease with over-the-counter diarrheal remedy sales records in the San Francisco Bay Area. *BMC medical informatics and decision making*, *10*, 1-9.
5. Wang, Y., & Gu, J. (2015, August). A hybrid prediction model applied to diarrhea time series. In *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)* (pp. 1096-1102). IEEE.
6. Kam, H. J., Choi, S., Cho, J. P., Min, Y. G., & Park, R. W. (2010). Acute diarrheal syndromic surveillance. *Applied Clinical Informatics*, *1*(02), 79-95.
7. Ahasan, M. N. (2019). *Modeling of Climatic Index on Infectious Diarrheal Disease* (Doctoral dissertation, University of Rajshahi).
8. Yan, L., Wang, H., Zhang, X., Li, M. Y., & He, J. (2017). Impact of meteorological factors on the incidence of bacillary dysentery in Beijing, China: a time series analysis (1970-2012). *PLoS One*, *12*(8), e0182937.
9. Wang, Y., & Gu, J. (2015). A Novel Hybrid Approach for Diarrhea Prediction. In *SEKE* (pp. 168-173).
10. Porter, C. K. (2011). *Time Series Evaluation of Childhood Diarrhea in Abu Homos, Egypt* (Doctoral dissertation, The George Washington University).
11. Wang, Y., Gu, J., Zhou, Z., & Wang, Z. (2015). Diarrhoea outpatient visits prediction based on time series decomposition and multi-local predictor fusion. *Knowledge-Based Systems*, *88*, 12-23.
12. Weisent, J., Seaver, W., Odoi, A., & Rohrbach, B. (2010). Comparison of three time-series models for predicting campylobacteriosis risk. *Epidemiology & Infection*, *138*(6), 898-906.
13. Anokye, R., Acheampong, E., Owusu, I., & Isaac Obeng, E. (2018). Time series analysis of malaria in Kumasi: Using ARIMA models to forecast future incidence. *Cogent social sciences*, *4*(1), 1461544.