Noor International Conference for Science and Technology (2NICST2020) Baghdad , Iraq

COVID-19 Cases in Iraq; Forecasting Incidents Using Box - Jenkins ARIMA Model

Hadeel I. Mustafa
Computer Information System Department
University of Basra
Basra, Iraq
hadeelismu@gmail.com

Noor Y. Fareed
Pharmaceutics Department
University of Basra
Basra , Iraq
lanayo.ly@gmail.com

Abstract: The pandemic outbreak of COVID-19 created panic all over the world. The mathematical principle in developing forecasting models aims to predict the number of future infections is considered crucial at this stage. The present investigation aims to analyze the time series using the Box-Jenkins method (Diagnostic, The Estimate, and selection, Forecasting) to find the best ARIMA model (Autoregressive Integrated Moving Average) for predicting the numbers of people infected with Covid-19 disease in Iraq. The data used were collected in the period between 1 –March and 31- July. The results showed that the appropriate forecasting model is ARIMA (2,1,5). Depending on this model, they predict the numbers of those infected with COVID-19 daily and for thirty days. Predictive values are consistent with original series values, indicating the efficiency of the model.

Keywords: ARIMA, Box-Jenkins, Minitab programming, Iraq, COVID-19.

I. Introduction

The quick spread and the highly contagious nature of COVID -19 created a serious crisis worldwide [1]. The absence of specific treatment for this decrease further raises the concerns of the public [2]. Therefore, world governments utilize all the possible measures to prevent the infection and decrease the disease's devastating outcomes [3].

A forecast is a quantitative, probabilistic statement about an unobserved event, outcome, or trend. Its surrounding uncertainty, conditional on previously observed data [4]. Time-series analysis is a powerful tool of forecasting, in which a mathematical model is established according to the regularity and trend of the observed historical values with time [8]. Box–Jenkins model is an autoregressive integrated moving average (ARIMA) model and is the most common time series prediction model [14][15].

The application of the mathematical principle of forecasting to predict the number of future infections relying on existing numbers has been used on many occasions [5],[7]. Forecasting would provide decision-makers with the necessary information required to prepare health care.

In COVID-19, several studies were conducted to predict the future burden utilizing the time series approach. For instance, the ARIMA model was used to forecast COVID-19 future infection in Nigeria[18]. Similarly, Researchers in China created a time series ARIMA model for new COVI-19 cases incidence and death [19]. Also, curve estimation models, Box-Jenkins and Brown/Holt linear exponential

Smoothing methods were used to forecast COVID-19 cases in eight different countries[20].

The present investigation aims to postulate the suitable forecasting model for the COVID-19 outbreak in Iraq by modeling actual data using (Box- Jenkin) models by the Minitab program.

II. The Methodology Of Research

Time series is defined as a series of recorded values (observations) for a specific phenomenon in limited periods [9]. It represents a historical record over time, and under the influence of economic, social, and environmental factors [10]. The time series is considered stable if the variable values' deviation from the mean values is zero or converts to decay. The stable time series has fixed arithmetic mean, and its variance and co-variations are constant over time, [11] i.e.

$$E(Y_t) = E(Y_{t+K}) = \mu \quad (1)$$

$$\text{Var}(Y_t) = E[Y_t - E(Y_t)]^2 = \text{Var}(Y_{t+k}) = E[Y_{t+k} - E(Y_{t+k})]^2 = \gamma(0) = \sigma^2 < \infty \quad (2)$$

$$\text{Cov}(Y_t, Y_{t+k}) = E[(Y_t - \mu)(Y_{t-k} - \mu)] = \text{Cov}(Y_{t-k}, Y_{t-k+s}) = \gamma(k) \quad (3)$$

The Autocorrelation Function(ACF) means that some variables or observations are related to each other during a specific time series period, which are important in clarifying the time series's characteristics. The mathematical formula for the Autocorrelation function as in "(4)":

$$\rho_K = \frac{\text{COV}(Y_t, Y_{t+k})}{\sigma_Y^2} \quad , t = 1,2,\ldots,N \ , \ k = 0,1,2, \quad (4)$$

The series is stable if it has an Autocorrelation function equal to or close to zero. Meaning, the lower the auto-correlations, the higher k, while for the unstable series, the differences are taken to them and to different degrees to convert them to stable [12]. Model stability is one of the important steps to implement the Box- Jenkins package [13]. To obtain a stable series, we use W_t as a separate series as follows:

$$W_t = \nabla^1 Y_t = Y_t - Y_{t-1} \quad , t = 2,3,\ldots,N \quad (5)$$