مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
Journal of the College of Basic Education Vol.32 (NO. 136) 2026, pp. 477-488

Forecasting Anemia Prevalence Among Women Of Reproductive Age In Iraq Using ARIMA

Azhar Kadhim Jbarah[1] **Suhad Ali Shaheed Al-Temimi** [2]
Hamid Saad Nour AL-Shammrty[3]
[1] Statistics Department, College of Administration and Economics- Mustansiriyah University, Baghdad-Iraq.
azkdfr_2017@uomustansiriyah.edu.iq
[2] Statistics Department, College of Administration and Economics- Mustansiriyah University, Baghdad-Iraq
dr.suhadali@uomustansiriyah.edu.iq
[3] College of Business Administration, AL-Bayan University, Baghdad-Iraq.

Abstract:
ANEMIA is one of the diseases that threaten global health indicators and mainly affects children and women of reproductive age (15-45) years. It also affects women of postpartum hemorrhage and during menstruation. In (this) research, women of childbearing age highlighted as an important period that affects the health of the fetus during pregnancy and has serious consequences for the child after birth. Data representing the number of women suffering from ANEMIA of reproductive age in Iraq was taken for the period (2000-2019). A time series model was used to analyze the study data, which is called the Autoregressive Integrated Moving Model (ARIMA), and the best model was chosen to predict the number of infections for the coming periods. The ARIMA model is distinguished by its flexibility and accuracy in the short-term forecasting process, and it has been widely used in many social, economic, health and other applications. Iraq witnessed a significant decrease in ANEMIA rates among women of childbearing age, as it reached (28.6) in 2019, after it was (39.5) in 2000. However, this indicator is still high compared to developed countries, so the forecasting process helps in taking Health measures necessary to reduce this threat. The ARIMA model (1,2,2) was chosen according to the AIC and HQC standards, and then the validity of the proposed model was examined when time changes. Through the MAPE test. After confirming the validity of the model, its parameters were estimated for use in predicting the number of ANEMIA cases in women of reproductive age for the period (2020-2028). The expected results indicate that there will be an increase in the number of infections during the coming period.

Keywords: ANEMIA among women in Iraq, ARIMA, ACF, PACF, forecasting.

1- Introduction:

¬One third of women from MYNCH are suffering from, which reflected negatively on the future of the region as there is significant negative effects on children being born to mother with suffer of ANEMIA during pregnancy, such as: immune weakness, inability to concentrate major peripheral SMI report 2012 – First version growth delay spreading premature birth and . When we are born, it cements the transfer of inequality from one generation to another. The most frequent physical manifestations of ANEMIA are "fatigue, asthenia and dyspnea" which in turn cause "depression, anxiety and decreased quality of life", and limit women's ability to carry out their daily activities or tasks at work. ANAEMIA in pregnancy is also linked with higher risk of maternal death, poor cognitive development and memory issues. (Yakubu,2022, p.3)

Time series analysis is primarily used to study data while accounting for the effect of time, and its importance lies in enabling researchers and planners to make informed predictions and decisions (Berhe & Box, 2019).

Forecasting techniques consist of two main components:

1. **Statistical prediction models**, such as Moving Average (MA), Exponential Smoothing (ES), Regression, and Autoregressive Integrated Moving Average (ARIMA) (R. Krispin, 2019).
2. **Artificial intelligence-based prediction techniques**, including Neural Networks (NN), Genetic Algorithms(GA) , Simulated Annealing (SA), Genetic Programming (GP), Classification, and hybrid methods.

The ARIMA model was extensively developed by George Box and Gwilym Jenkins in 1976. This model relies on probabilistic analysis of time series data to enable accurate forecasting of future values

2- Research objective

The research aims to predict the number of among women of reproductive age (15-45) by using "The Autoregressive Integrated Moving Average (ARIMA)" model to annually predict the number of cases of the disease in Iraq until 2030. The Autoregressive Integrated Moving Average (ARIMA) method is used in forecasting because it is characterized by its flexibility and accuracy in the short-term forecasting process, as it has been widely used in many applications, including social, economic, health, and others. Several models were proposed and a comparison was made of the predictive ability between these models to choose the best predictive model. Among these tests are the Akaike Information Criterion (AIC), and the Hannan-Quinn Criterion (HQC). (Box& R. Patil,2021, p.2).

3- Literature Reviews:

In a systematic review, Kasa et al. (2017) analyzed the prevalence of ANEMIA and determinants associated with the incidence in Ethiopia and concluded that about one-third of pregnant women in the country are anemic. (Kassa,2017,https://bmchematol.biomedcentral.com/articles/10.1186/s12878-017-0090-z) Berhe et al.(2019) diagnosed the prevalence of ANEMIA and factors associated with its prevalence in Adigrat General Hospital in Ethiopia. Multivariable logistic regression was applied. The results showed that the prevalence of ANEMIA among the pregnant women studied was 7.9%. (Box ,1975, p.70-79). In (2021), Damaris Kinyoki and others also studied geospatial estimates for the period (2000-2018) of the prevalence of ANEMIA in women of reproductive age (15-49 years) across 82 low- and middle-income countries (LMICs). Blood by severity and sub-national disparity analyzes were presented to provide a comprehensive view of inequalities in ANEMIA prevalence within these countries and to forecast progress towards achieving the WHO global nutrition target (WHO GNT) to reduce ANEMIA by half by 2030. (Kinyoki,2021, p. 1761–1782)

4- The methodology of building ARIMA models:

One of the main foundations of the ARIMA model is that it primarily relies on analyzing and forecasting univariate time series data, using the autoregressive integrated moving average model (ARIMA)". The ARIMA model predicts a value in a response time series as a linear combination of its past values, and bases the prediction on past errors (also called shocks) and the current and past values of other time series. The ARIMA approach was first published by "Box and Jenkins", and ARIMA models are generally referred to as Box-Jenkins models. (Kinyoki,2021, p. 1761–1782)

The general transfer function model used by the ARIMA procedure was discussed by "Box and Tiao (1975)". The ARIMA procedure provides a comprehensive set of tools for univariate time series model identification, parameter estimation, and forecasting, and provides significant flexibility in the types of ARIMA models that can be analyzed. The ARIMA procedure supports ARIMA seasonal, sub- and analytical models; Intervention or interrupted time series models; Multiple regression analysis with ARMA errors; It models rational transfer functions of any complexity.(Kassa,2017)

Let a time series $X_t, t = 1,2,\dots,n$, represent the objective is to select the best fitted ARIMA (p, d, q) model, so we can write ARIMA model as follow: (Al-Morshedy& Kassa,2021).

$$X_t = b_0 + b_1 X_{t-1} + b_2 X_{t-2} + \cdots + b_p X_{t-p} - w_1 e_{t-1} - w_2 e_{t-2} - \cdots - w_q e_{t-q} + e_t \quad (1)$$

Where X_t represents the predicted values of the variable, b_0 Constant, which represents the average of X_t over the period, b_p Autoregressive coefficients (AR model), w_q represents the error term coefficient or the moving average coefficient (MA model) and e_{t-1} the error term for the period before the current period. For example, of writing an equation with ARIMA model:

$$\left.\begin{array}{ll} X_t = b_0 + b_1 X_{t-1} - w_1 e_{t-1} + e_t & \rightarrow ARIMA(1.0.1) \\ X_t = b_0 + b_1 X_{t-1} + b_2 X_{t-2} - w_1 e_{t-1} + e_t & \rightarrow ARIMA(2,0,1) \\ X_t = b_0 + b_1 X_{t-1} - w_1 e_{t-1} - w_2 e_{t-2} + e_t & \rightarrow ARIMA(1,0,2) \end{array}\right\} \quad \ldots (2)$$

To building ARIMA models according to "Box and Jenkins" involve four stages:

Stage (1): "Identification Stage" included:

- The data is plotted to determine whether it is **seasonal** or **non-seasonal**..
- Perform the **Augmented Dickey-Fuller (ADF) test** to determine whether the data is stationary or not; if the data is not stationary, the **differentiation process** is applied.

1- After completing the differentiation process, the **ADF** value will be tested, and the differentiation process will continue until the data becomes stationary.(Iraq IQ& R. Krispin,2019, (http://apps.who.int/gho/data/node.main.1?lang=en))

-Determine the graphs of "the autocorrelation function (ACF) and partial autocorrelation function (PACF)": (to assist in the process of estimating the values of the parameters p and q in the ARIMA model (p, d, q). The sample Autocorrelations which is the plot of the sample Autocorrelations

$$r_p = \frac{\sum_{t=p+1}^{n}(X_t - \bar{X})(X_{t-p} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})} \quad \ldots (3)$$

Where (r_p) estimate the population autocorrelation (ρ_p) and ($\bar{X}$) is sample mean of ($X_1, X_2, \ldots, X_n$). The set of r_p defines the sample autocorrelation function. If these estimates do not decay towards zero, the process is not stationary and therefore it needs to be examined in first differences. And in case the process is stationary these estimates will provide useful information with respect to AR or MA components.

The "partial autocorrelation" coefficient φ_{pp} (population) indicates the autocorrelations between (X_t, X_{t-p}) , for given values $X_{t-1} - X_{t-2}, \ldots, X_{t-p+1}$. to find the values of φ_{pp} we need to construct two errors:

$$\left.\begin{array}{l} X_{t,t-1,t-2,\ldots,t-p+1} = X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \cdots - \varphi_{p-1} X_{t-p+1} \\ X_{t-p.t-k+1,t-p+2,\ldots,t-1} = X_{t-p} - \varphi_1 X_{t-p+1} - \varphi_2 X_{t-p+2} - \cdots - \varphi_{p-1} X_{t-1} \end{array}\right\} \quad \ldots (4)$$

Partial autocorrelations are calculated based on the two values in (2) and are defined as regular autocorrelations. The "partial autocorrelation" coefficients (φ_{pp}) are computed as follows:

$$\left.\begin{aligned} \varphi_{11} &= Corr(X_t, X_{t-1}) = Cov(X_t, X_{t-1})/Var(X_t) = \rho_1 \\ \varphi_{22} &= Corr[X_t - \varphi_1 X_{t-1}, X_{t-2} - \varphi_1 X_{t-1}) = [\rho_2 - \rho_1^2]/[1 - \rho_1^2] \\ \varphi_{33} &= Corr[X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2}, X_{t-3} - \varphi_1 X_{t-1} - \varphi_2 X_{t-2}) \dots and\ so\ on \end{aligned}\right\} \dots (5)$$

When X_t~ AR(p), that means that only the first p partial autocorrelations exist.

Stage 2: Model Parameter Estimation: At this stage, the smallest **AIC** value is determined, as shown in the following figure. (Berhe& Yakubu,2022)

$$AIC\ (p) = N\ ln(\sigma_a^2) + 2p \quad \dots (6)$$

Where p is the number of parameters in the model, N is the number of data, and σ_a^2 is the maximum likelihood estimate of σ_a^2. Another test to choosing best model is "Hannan–Quinn information criterion" is a criterion for model selection. It is an alternative to "Akaike information criterion" . It is given as:

$$HQC = -2L_{max} + 2p\ \ln(\ln(n)) \quad \dots (7)$$

L_{max} refers to log-likelihood function , (p) is the number of parameters and (n) is the number of observations.

Stage 3: In order to obtain the best model, and in the case of forecasting, the future period must be predicted using the training data and the forecasting data based on model (3).

Stage 4: Calculate Accuracy Rate: To measure the comparison of the prediction error with the actual value the Mean Absolute Percentage Error (MAPE) criterion is calculated. If the MAPE value is less than 10%, then the prediction is characterized by excellence performance, and Since the accuracy is moderate to good In the case where the MAPE value ranges between 10% and 30%. It is given as: (Box& Iraq IQ,1976)

$$MAPE = \frac{\sum_{t=1}^{n}\left|\left(\frac{A_t - F_t}{A_t}\right)\times 100\right|}{n} \quad \dots (8)$$

Where n is the value of the time period; A_t is the actual value; F_t is the expected value .

5- Results and Discussion:

For the purpose of determining the best model for estimating the number of annual cases of ANEMIA for women of reproductive age in Iraq. Study data were obtained based on World Health Organization data (2025) (World Health Organization, https://data.worldbank.org/indicator/SH.PRG.ANEM)

We draw the data series representing the incidence of "ANEMIA" for women of reproductive age recorded in Iraq for the years (2000-2019), in order to

identify the behavior of the time series and its initial characteristics, as shown in Figure (1), which represents the drawing of the series:

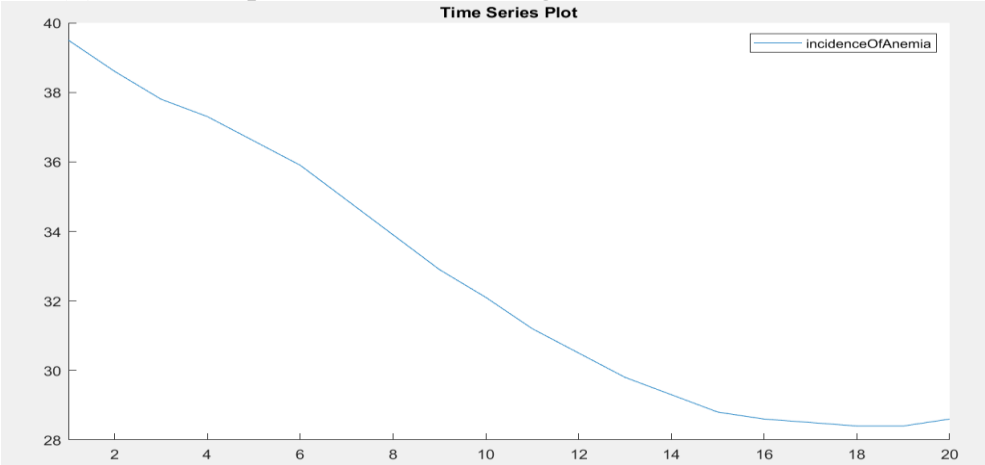
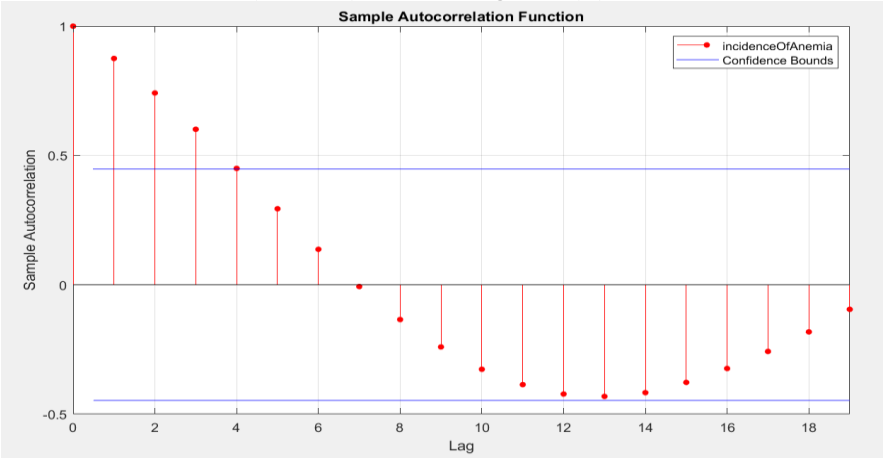


Figure (1) shows the annual time series of ANEMIA cases for women of reproductive age in Iraq for the years (2000-2019)

It is clear from the figure (1) that the time series is unstable due to the presence of concavities and convexities, which indicate the presence of a general reaction component and that the series is unstable. For more detail, we draw both the "autocorrelation function" (ACF) and the "partial autocorrelation function" (PACF), as in Figure (2).



مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
Journal of the College of Basic Education Vol.32 (NO. 136) 2026, pp. 477-488

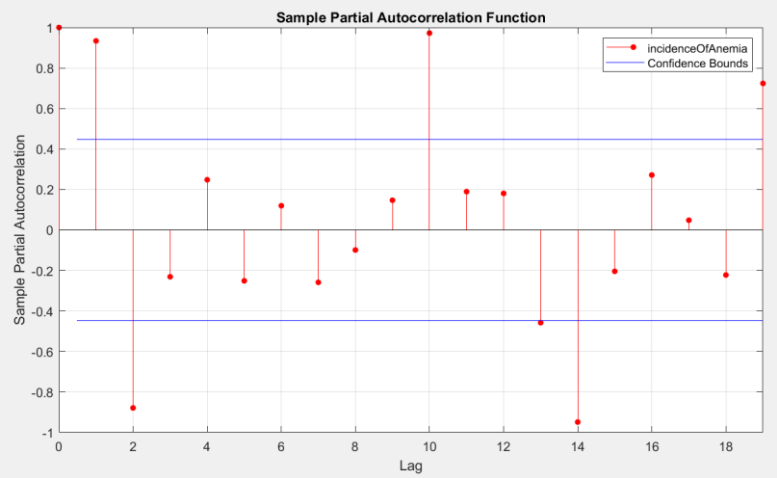


Figure (2) shows a plot of both the "autocorrelation function" and the "partial autocorrelation function" for the incidence of ANEMIA among women of childbearing age.

We notice from Figure (2) that many of the ACF coefficients are outside the confidence Interval at the 95% level, as well as some of the PACF coefficients, and this is evidence of the instability of series. To test the series' s non-stationarity, we use the "Augmented Dickey – Fuller" test, as shown in the table (1):

Table (1): Augmented Dickey – Fuller test

Test name	Test-statistics	Critical Value	Significant level	P-value
Augmented Dickey – Fuller	-0.3446	-7.1698	0.05	0.6609

It is clear from the test results in Table (1) that the absolute value of the test statistic is less than the critical value of Augmented Dickey – Fuller test at a significance level (0.05). we conclude by not rejecting the null hypothesis and rejecting the alternative hypothesis that indicates the stability of the time series, meaning that the series is unstable.

H_0: The series is unstable.

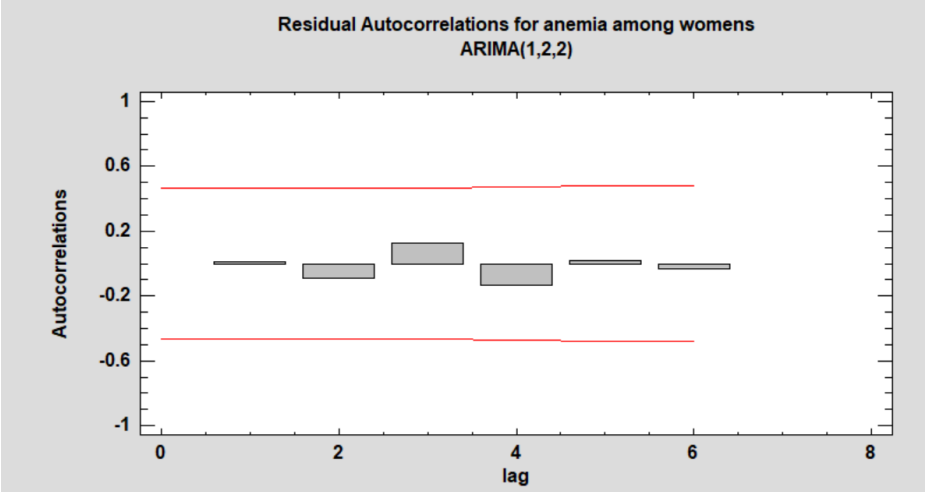
H_1: The series is stable.

The lags of the MA model and AR model are determined by acf and pacf, respectively.By testing the autocorrelation function (ACF) and partial autocorrelation coefficient (PAC),the result shows ACF decrease rapidly with increase of delay,this might indicate that time series is nonstationary and need to be fitted(see Fig2). We draw on the time series data and ADF test to investigate model fit. We introduce more than one ARIMA models in order to stabilize the series and select the best fitted model by choosing an

ARIMA model with minimum AIC. Finally, we obtain moving average of signed error (MASE) criteria for each proposed model to check the consistency between the predicted error and the actual value. From the analysis, ARIMA (1,2,2) is considered as a best model since has stable time series based on ACF and PACF both and also gives minimum AIC and HQC values. The estimates of the drift function and noise term are presented in Table 2.

Table (2): the results for the proposed models

Model	MAPE	AIC	HQC
(1,2,2)	0.369713	-3.644	-3.61485
(2,2,1)	0.358959	-3.61806	-3.5889
(0,2,0)	0.400446	-3.56372	-3.56372
(1,1,0)	0.396039	-3.53208	-3.52236
(2,2,0)	0.394664	-3.52066	-3.50122



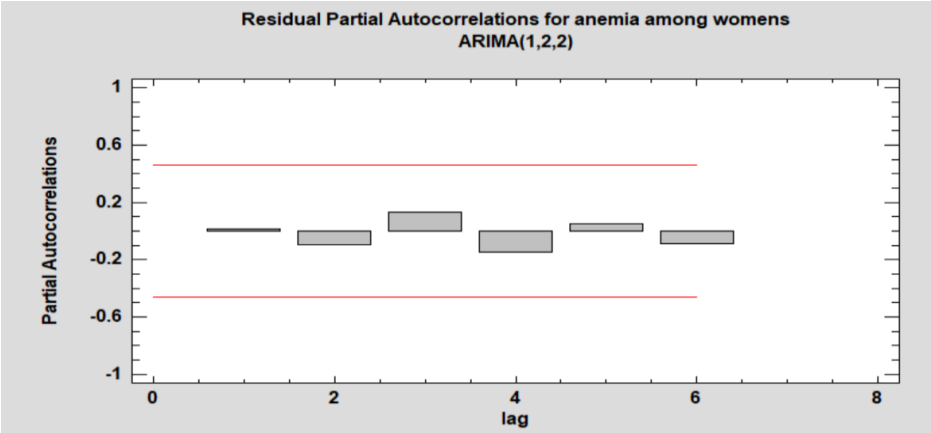


Figure (3) shows a plot of ACF and PACF after taking the second difference

It is shown in the figure below that all ACF coefficients, as well as the PACF coefficients, fall within the confidence limits. The model parameters were then estimated, and by applying the maximum likelihood method to the time series data, the ARIMA (1,2,2) model was obtained, as shown below.:

Table (3): Estimation results of the proposed model

Parameter	Estimate	Stnd. Error	t	P-value
AR(1)	-0.928044	0.136339	-6.80688	0.000006
MA(1)	-1.20267	0.220107	-5.464	0.000065
MA(2)	-0.508468	0.220632	-2.3046	0.035904

It is shown from the table below that the model is statistically significant, as the p-value is less than 0.05. Therefore, to predict cases of anemia among women of reproductive age, predictions were made using the ARIMA (1,2,2) model, as shown below

Table (4): predict cases of ANEMIA among women of reproductive age (2020-2028)

		Lower 95.0%	Upper 95.0%
Period	Forecast	Limit	Limit
2020	28.8553	28.5363	29.1743
2021	29.1139	28.3213	29.9066
2022	29.3695	27.9204	30.8186
2023	29.6279	27.4503	31.8056
2024	29.8837	26.8532	32.9142
2025	30.1419	26.1973	34.0865
2026	30.3978	25.4406	35.3551
2027	30.6559	24.6309	36.6808
2028	30.912	23.7362	38.0878

From Table 4, it can be noted that there is a slight increase in the number of ANEMIA cases expected during the coming period. The table also includes the expected lower and upper limits at a 95% significance level, as the expected upper limit for infections reached (38), and Figure (4) shows this.

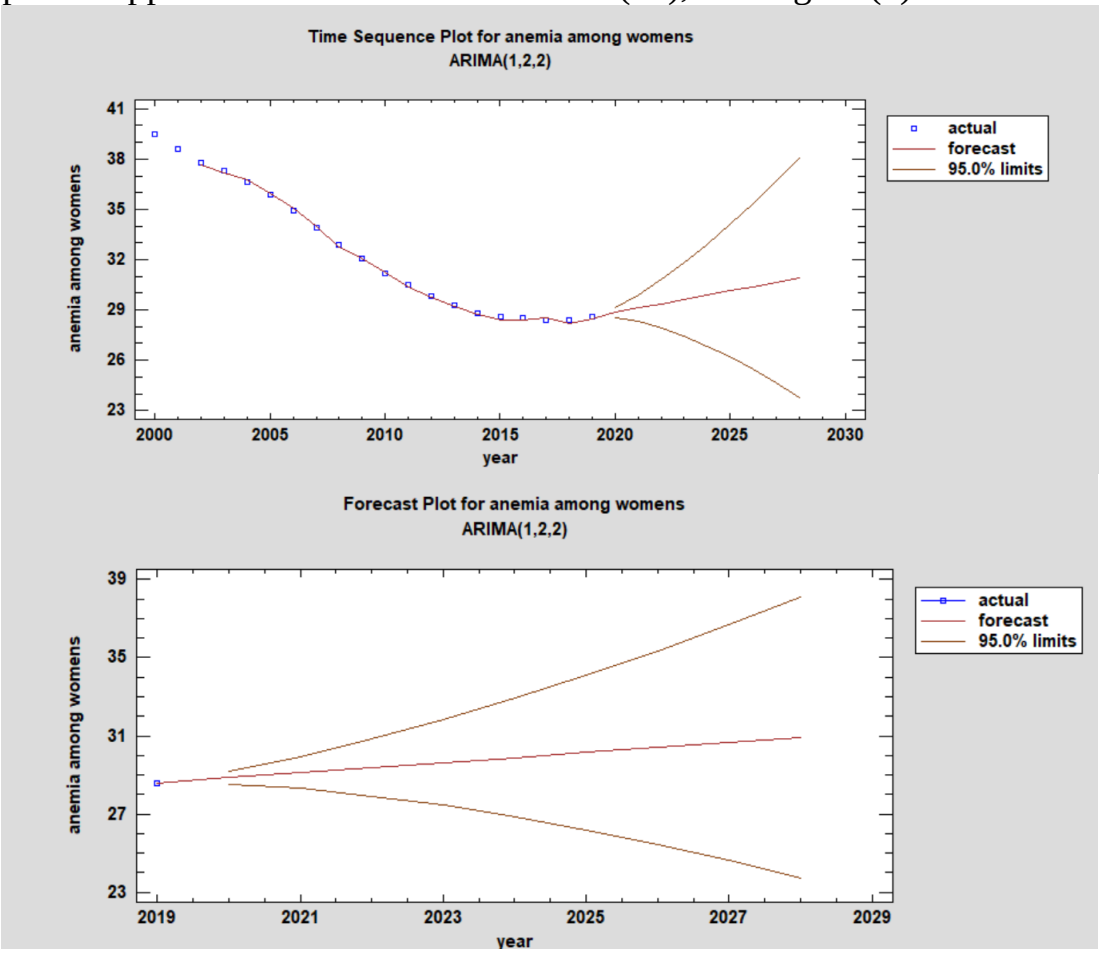


Figure (4) shows a plot of actual value of time series and forecast values for ANEMIA among women of reproductive age (2020-2028) at limit 95%

6- Conclusion:

The ARIMA model is a valid way to estimate the value of data over time as both ARIMA model and SARIMA model is workable maximum for 50 observations only which means you also face one issue in it which is a limitation. As the study aim is to choose models for forecast number of cases 0/6 ANEMIA which include women's across fecundity period in Iraq, ARIMA was focused as regards to our data. By the estimated shown time series model based on data of incidence of ANEMIA among women in reproductive ages, it is fitted for AR (1) and MA (2). MAPE value was also estimated to test the validation of proposed model with time change.

Therefore, the predicted values for the number of "ANEMIA" (the dependent variable) cases among women in reproductive age group during 2020-2028 were estimated after estimation of parameters of suggested model. It was mentioned that there would be more cases in the future.

7- References:

2- Al-Morshedy, Karrar Haider Hussein. (2021); "Diagnosis and estimation of seasonal time-series models with practical application". A thesis submitted to the council of the college of Administration &Economics\ University of Karbala.

3- Berhe ,B. et al, (2019); "Prevalence of ANEMIA and associated factors among pregnant women in Adigrat General Hospital, Tigrai, northern Ethiopia, 2018". BMC Res Notes 12:310 https://doi.org/10.1186/s13104-019-4347-4.

4- Box, G.E.P. and Jenkins, G.M. (1976), Time Series Analysis: Forecasting and Control, San Francisco: Holden-Day.

5- Box, G.E.P. and Tiao, G.C. (1975), "Intervention Analysis with Applications to Economic and Environmental Problems," JASA, 70, 70-79.

6- Iraq IQ: Prevalence of ANEMIA among Women of Reproductive Age: % of Women Aged 15-49, (http://apps.who.int/gho/data/node.main.1?lang=en).

7- Kassa et a,(2017); "Prevalence and determinants of ANEMIA among pregnant women in Ethiopia; a systematic review and meta-analysis ".College of health Sciences, Debre Markos University, Debre Markos, Ethiopia ,BMC Hematology 17:17. https://bmchematol.biomedcentral.com/articles/10.1186/s12878-017-0090-z .

8- Kinyoki, D. and et. al. ,(2021); "ANEMIA prevalence in women of reproductive age in low- and middle-income countries between 2000 and 2018". Nature Medicine | VOL 27 | October 2021 | 1761–1782 | www.nature.com/naturemedicine.

9- R. Krispin, Hands-On Time Series Analysis with R: Perform time series analysis and forecasting using R. Packt Publishing Ltd, 2019, https://books.google.co.id/books?hl=id&lr=&id=tTmbDwAAQBAJ.

10- R. Patil, D. M. Nagaraj, B. S. Polisgowdar, and S. Rathod, "Forecasting Potential Evapotranspiration for Raichur District Using Seasonal ARIMA Model," Mausam, vol. 73, no. 2, pp. 433–440, 2022, https://doi.org/10.54302/mausam.v73i2.5488.

11- World Health Organization, Global Health Observatory Data Repository/World Health Statistics. https://data.worldbank.org/indicator/SH.PRG.ANEM.

12- Yakubu, A. U., Saputra, A. P. M., (2022); "Time Series Model Analysis Using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) for E-wallet Transactions during a Pandemic". International Journal of Global Operations Research Vol. 3, No. 3, pp. 80-85.

التنبؤ بمعدلات فقر الدم (الأنيميا) بين النساء في سنّ الإنجاب في العراق: باستخدام نموذج ARIMA

ازهار كاظم جبارة [1]، سهاد علي شهيد التميمي [2]، حامد سعد نور الشمرتي [3]
[1]قسم الاحصاء ، كلية الادارة والاقتصاد ،الجامعة المستنصرية ، بغداد/العراق
azkdfr_2017@uomustansiriyah.edu.iq
[2]قسم الاحصاء ، كلية الادارة والاقتصاد ،الجامعة المستنصرية ، بغداد/العراق
dr.suhadali@uomustansiriyah.edu.iq
[3]كلية ادارة الاعمال ،جامعة البيان،بغداد/العراق

مستخلص البحث:

يعدّ فقر الدم (الأنيميا) أحد الأمراض التي تُهدِّد المؤشرات الصحية العالمية، ويؤثر بشكل رئيسي في الأطفال والنساء في سنّ الإنجاب (15–45 سنة). كما يؤثر في النساء في حالات نزف ما بعد الولادة وخلال فترة الحيض. في هذا البحث، تم تسليط الضوء على النساء في سنّ الإنجاب بوصفها مرحلة مهمة تؤثر في صحة الجنين أثناء الحمل ولها آثار خطيرة على الطفل بعد الولادة. تم الاعتماد على بيانات تمثل عدد النساء المصابات بفقر الدم من سنّ الإنجاب في العراق خلال الفترة (2000–2019). واستُخدم أحد نماذج السلاسل الزمنية لتحليل بيانات الدراسة، وهو نموذج الانحدار الذاتي والمتوسطات المتحركة المتكامل(ARIMA) ، حيث تم اختيار أفضل نموذج للتنبؤ بعدد الإصابات خلال الفترات القادمة. ويتميّز نموذج ARIMA بالمرونة والدقة في التنبؤ قصير الأجل، وقد استُخدم على نطاق واسع في العديد من التطبيقات الاجتماعية والاقتصادية والصحية وغيرها.شهد العراق انخفاضًا ملحوظًا في معدلات فقر الدم بين النساء في سنّ الإنجاب، إذ بلغت (28.6) عام 2019 بعد أن كانت (39.5) عام 2000. ومع ذلك، لا يزال هذا المؤشر مرتفعًا مقارنة بالدول المتقدمة، لذا فإن عملية التنبؤ تساعد في اتخاذ الإجراءات الصحية اللازمة للحد من هذا التهديد.

وبناءً على نتائج تحديد نموذج السلسلة الزمنية باستخدام مخططات الارتباط الذاتي (ACF) والارتباط الذاتي الجزئي (PACF) لبيانات الدراسة، تم اختيار نموذج ARIMA (1,2,2) وفقًا لمعياري AIC وHQC، ثم جرى فحص صلاحية النموذج المقترح مع تغير الزمن من خلال اختبار MAPE. وبعد التأكد من صلاحية النموذج، تم تقدير معلماته لاستخدامها في التنبؤ بعدد حالات فقر الدم لدى النساء في سنّ الإنجاب للفترة (2020–2028). وتشير النتائج المتوقعة إلى حدوث زيادة في عدد الإصابات خلال الفترة القادمة.