

# Forecasting the Number of Outpatient Patient Visits Using ARIMA at Mitra Sehat Clinic: Case Study in Sukoharjo, Central Java Indonesia

Susilo Hariyanto

Department of Mathematics, Diponegoro University, Semarang 50275, Indonesia

Received June 10, 2022; Revised September 9, 2022; Accepted September 27, 2022

## Cite This Paper in the Following Citation Styles

(a): [1] Susilo Hariyanto , "Forecasting the Number of Outpatient Patient Visits Using ARIMA at Mitra Sehat Clinic: Case Study in Sukoharjo, Central Java Indonesia," *Universal Journal of Public Health*, Vol. 10, No. 5, pp. 482 - 491, 2022. DOI: 10.13189/ujph.2022.100506.

(b): Susilo Hariyanto (2022). *Forecasting the Number of Outpatient Patient Visits Using ARIMA at Mitra Sehat Clinic: Case Study in Sukoharjo, Central Java Indonesia*. *Universal Journal of Public Health*, 10(5), 482 - 491. DOI: 10.13189/ujph.2022.100506.

Copyright©2022 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

**Abstract** In the pandemic situation, Clinic Mitra Sehat Sukoharjo has become a forefront health facility to maintain the health of society. During the last three years, the number of patients visited has reached 142.792 people. With that large patient visitation, it is necessary to provide everything to serve patients well. This study will provide estimation data results that can be used by the Mitra Sehat Clinic, which previously only made preparations without data. This study will estimate the Number of Outpatient Patient Visits Using ARIMA at Mitra Sehat Clinic in the next few months. ARIMA or Autoregressive Integrated Moving Average is a forecasting method that will be used to predict the number of patient visits at the next time, so that in the future, Mitra Sehat Clinic can prepare matters related to serving the visiting patients optimally. The steps for using the ARIMA method are using data from the required sample from patients visiting in January 2019-December 2021, determining the type of time series data pattern, then conducting a stationarity test, determining the ARIMA model, calculating and analyzing the accuracy of the model used, and then forecasting the number of outpatient visits. And the best ARIMA model for this forecasting based on the calculation is (1, 0, 1) with an error value of 334183.

**Keywords** Arima, Forecasting, Patient Visits, Sukoharjo Indonesia

## 1. Introduction

As the human population increases and the condition nowadays, the world is already facing the pandemic era. Even statistically, some countries have started to show a decline in the transmission rate. Still, it does not make the situation back to normal directly like the condition before the pandemic era. For example, Indonesia is one of the countries which is also infected by Covid-19, even though, according to national statistical data, Indonesia has begun to experience a decline in Covid-19 transmission [1], and vaccinations are still being carried out to continue to minimize the transmission of Covid-19 and towards a new era of habituation with social activities that return to normal [2].

One of the crucial roles that are the responsibility of health facilities is to provide education, prevention, and care to the community at all times, including in this pandemic era where health facilities are at the forefront to play an active role in helping the country and the community in the process of solving these pandemic problems in terms of health [3]. In addition to focusing on the Covid-19 pandemic, Health Facilities also focus on public health in general, which is crucial in maintaining a controlled community condition in terms of health. Furthermore, everyday patients come to the Health Facility to do their health consultation with varying and not small amounts. Therefore, it is necessary to have special

attention from the Health Facilities to prepare for the fulfillment of facilities and to support service tools ranging from registration to drug prescriptions and other matters such as sick or healthy certificates. Therefore, the health facility must carry out preparatory activities to balance the patient visits who come to the health facility.

Based on that, it is necessary to forecast the number of patient arrivals to the public health facility so that in the future, the health facility can adequately provide services to patients who visit with predictions of patients who will visit before. That condition will be able to provide more well-prepared preparations compared to no data to prepare for things that will be faced in the future. The increase in the total human population also impacts increasing visits to health facilities. In this pandemic era, people will be more secure in their health by conducting health consultations at existing health facilities.

Indonesia is one of the countries affected by the Covid-19 pandemic, including Sukoharjo Regency in Central Java, Indonesia, where until now, there are still positive cases of Covid-19 being treated [4]. In Sukoharjo Regency, there is a health facility with the highest number of registered BPJS patients, known as Mitra Sehat Clinic [5], which is one of the mainstay health facilities in maintaining public health in Sukoharjo. With more than 18,000 registered patients [5], sound planning is needed to provide good service to patients who visit for health consultations at these health facilities.

This research aims to analyze and predict the number of patients visiting the Mitra Sehat Clinic to be used to determine policies and prepare health services for patients. In this research, forecasters use the ARIMA method with a case study in Sukoharjo Regency, Central Java, Indonesia, with a specific location at the Mitra Sehat Clinic Health Facility because it has the highest number of BPJS registered patients in the district area. This research is structured in 6 parts: (1) Introduction, in which the rationale of the research is briefly introduced; (2) Materials and Methods; (3) Case Study, where the health facility Mitra Sehat in Sukoharjo regency as a case problem is described; (4) Results and Discussion, which intensively analyzes the results; (5) Conclusions, which provides a summary of the main finding of this research.

## 2. Materials and Methods

### 2.1. Research Design

This study uses the Autoregressive Integrated Moving Average (ARIMA) forecasting method to forecast the number of patients visiting Mitra Sehat Clinic in Sukoharjo-Central Java, Indonesia.

### 2.2. Sample and Population

The target populations in this study were patients who

visited Mitra Sehat Clinic either with complaints of illness or only for health consultations at Mitra Sehat Clinic. Therefore, the sample was patients visiting with those criteria from January 2019-December 2021.

### 2.3. Data Collection Methods

The data used in this research were obtained from secondary data from patients' visitation in Mitra Sehat Clinic from January 2019-December 2021 [6] accessed on the Pcare BPJS web page [6,7].

### 2.4. Data Analysis Methods

A time series is a series of observations taken based on time sequences. Between adjacent and correlated observations, it is said that in a time series, each observation taken from a variable is correlated with the variable itself at the previous time [13]. The methods to analyze the data used to forecast the number of outpatient visits for the next months using the Autoregressive Integrated Moving Average (ARIMA) method. Compared to the other models between Autoregressive Model, Moving Average, ARMA and ARIMA, ARIMA is the best prediction model for forecasting daily trends or monthly cases in this paper [11]. Using this model, we could estimate the daily number of confirmed cases or patient visitation for the next month [11]. For forecasting a time series, ARIMA modeling is one of the best modeling techniques. ARIMA models are always represented with the help of some parameters, and the model is expressed as ARIMA (p, d, q). Here, p stands for the order of auto-regression, d signifies the degree of trend difference, and q is the order of moving average [12]. The first step is to determine the pattern type of the data before conducting advanced analysis. The first stage is an inspection of the data's pattern with plotting data to see trends and patterns in data [15].

The stationary of the data also needs to be tested, which is already stationary or not to variance and mean or average. Data can be stationary in the average if the fluctuations in the data are around a constant average value, independent of the time and variance of these fluctuations [13]. And data can be said to be stationary in variance if the data structure from time to time has a constant or constant fluctuation [13]. The stationarity of the data can be obtained with time series plot graphs and plots of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). Data is said to be stationary if it is not affected by the change in time [8]. If the data is not stationary with variance, it is necessary to transform until the rounded value is 1.00. If the value of the rounded value or lambda ( $\lambda$ ) is more than or equal to 1, the data is already stationary in variance [15]. If the data is not stationary to the average, it is necessary to do the differencing process, which is a process to eliminate trend elements and seasonal trends. After doing level one

differencing, the data needs to be retested whether it is stationary to the average by looking at the graph of the Autocorrelation Function (ACF) plot and Partial Autocorrelation Function (PACF). If not, then level differencing is necessary, going to the second level until the data is stationary. The stationarity of ACF and PACF plots is also obtained from the lag outside the red line. If only one lag outside, it can be said that the data is stationer [15]. If the data is stationary, the next step is to identify temporary models based on ACF for MA values, PACF for AR values, and differencing for the value of  $d$  to determine the best forecasting model and measure accuracy by looking at the small error value.

The ARIMA model is said to be significant and feasible if it has a final P-value estimate of the parameter is below the error tolerance limit ( $\alpha$ ) 5% or 0.05 [14,15] and the Ljung-Box P-value is above the error tolerance limit ( $\alpha$ ) 5% or 0.05 [15]. In addition, based on the significance test, the best model can be determined with the smallest error value from each resulting model [15].

The ARIMA method is used for short-term forecasting because it has accurate accuracy [9]. Autoregressive Integrated Moving Average (ARIMA) is a model that completely ignores the independent variables in forecasting. The values used by ARIMA for forecasting are past and present values of the dependent variable to produce accurate short-term forecasts [10]. The assumptions that must be met to use this method are data stationarity and error, which is white noise or not autocorrelated and normally distributed. White noise is a form of random variables that are not mutually correlated. The white noise process is determined by a constant average [13]. The residual model is said to be normally distributed if the residual plot probability result exceeds the error tolerance limit ( $\alpha$ ) of 5% or 0.05 [15].

### 3. Case Study

Mitra Sehat Clinic, located in Sukoharjo, Central Java, Indonesia, has the highest number of registered BPJS patients. With more than 18,000 registered patients [5], sound planning is needed to provide good service to patients who visit for health consultations at these health facilities. Based on that, Mitra Sehat Sukoharjo is highlighted in this study. With that high number of patients, so it is necessary to have a good plan to maintain optimal servicing to the patients. And this study can help

analyze which is a good plan the Clinic will prepare. This study will show the forecasting data for the Clinic and the prediction of patient visitation to the Clinic based on the data in the past. The finding of this study may help the policy maker in the Clinic to make a good policy or plan based on their understanding of their perspective to serve the patients in the future.

## 4. Results and Discussion

### 4.1. Secondary Data

The number of patient visits in Mitra Sehat in the range of 2019 – 2021 decreased in 2020 and increased again in 2021. Forecasting is necessary to see an increasing trend for patient visits in Mitra Sehat. Mitra Sehat can use the forecast result to plan and run more specific programs based on these problems. If the forecasting results indicate a predictive amount, then Mitra Sehat can create a program that aims to make a good plan to give the optimal services to the Patient. The ARIMA method will use these data from 2019 – 2021 to predict the number of patient visits to Mitra Sehat Clinic.

### 4.2. Data Plotting

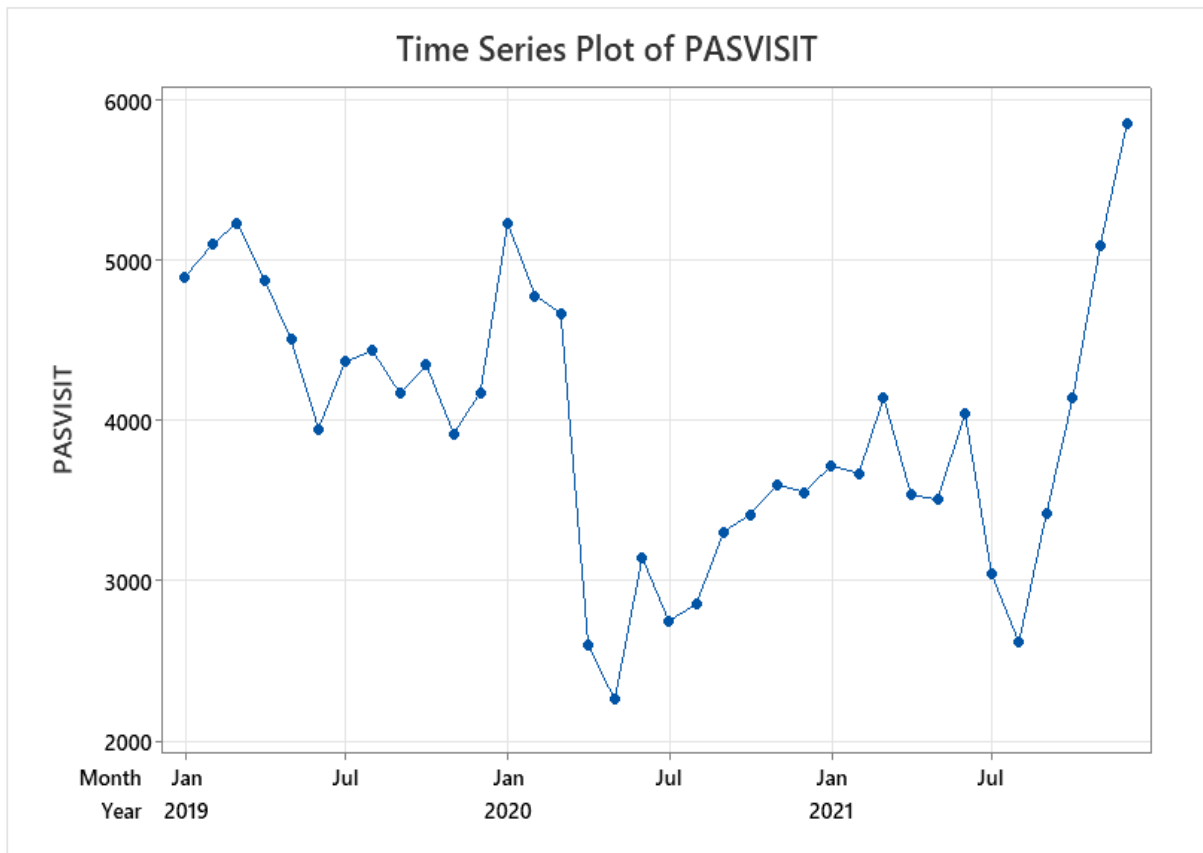
The requirement in using the ARIMA model in analyzing data is to determine that the analyzed data is in a stationary condition. One way to determine the stationarity of the data is by plotting the data. Figure 1 shows that the data is not fluctuating. So further analysis of seasonal patterns is necessary, to determine whether the data is stationary in variance.

### 4.3. Stationary Test of Variance and Means

Figure 2 shows that the rounded value on that graph is already 1. Hence the data is already stationer in variance. Therefore, the value 1 obtained without performing transformations on the original data due to the original data already stationer in variance with the index of rounded value is 1. Apart from observing Box-Cox plots, data stationarity can also be carried out with the results of testing the graph of the auto-correlation function and auto-partial correlation with indicators of whether the data has been stationary in the mean. The following results are the graph autocorrelation plot and partial autocorrelation.

**Table 1.** Number of Patient Visits 2019-2021

No	Month	2019	2020	2021
1	January	4.892	5.226	3.716
2	February	5.096	4.776	3.667
3	March	5.229	4.666	4.139
4	April	4.874	2.594	3.535
5	May	4.502	2.255	3.503
6	June	3.945	3.143	4.042
7	July	4.363	2.747	3.040
8	August	4.433	2.849	2.613
9	September	4.168	3.302	3.417
10	October	4.343	3.409	4.141
11	November	3.914	3.595	5.091
12	December	4.173	3.550	5.846
<b>Total</b>		53.932	42.112	46.750



**Figure 1.** Time Series plot of Patient Visits 2019 – 2021

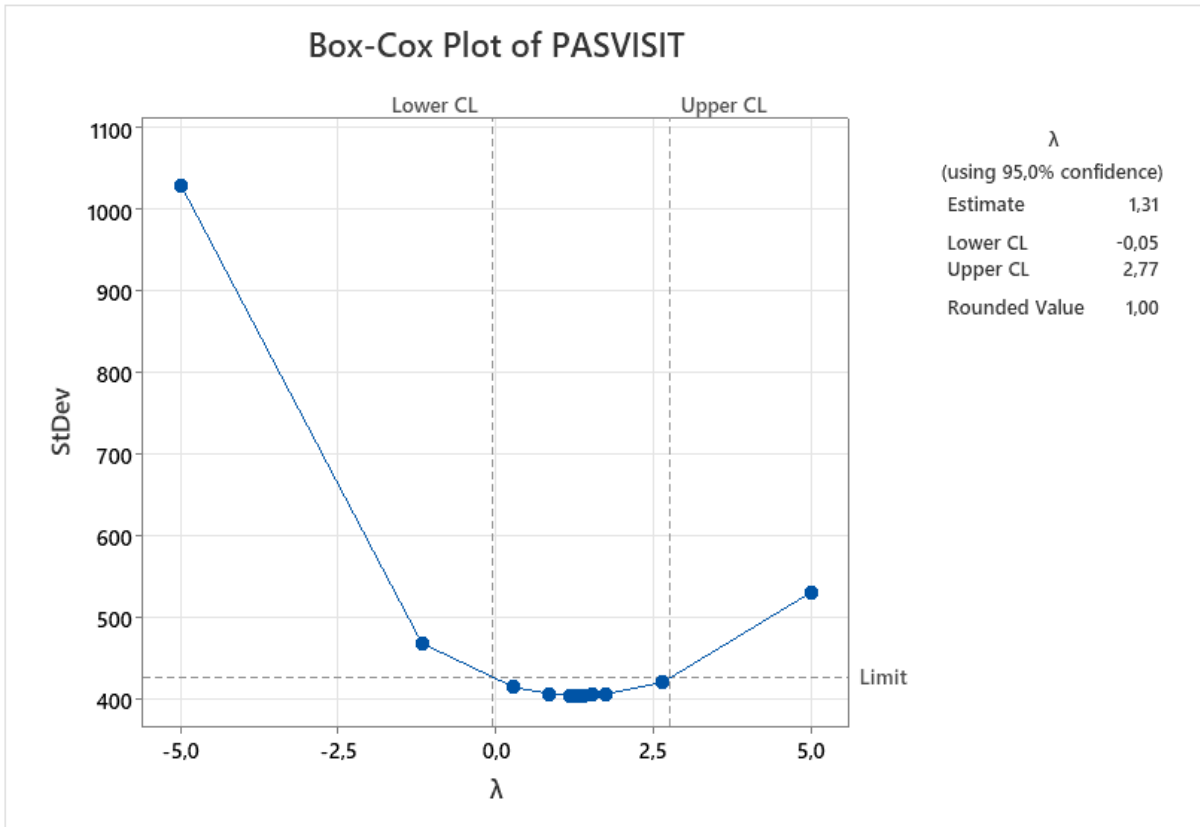


Figure 2. Box-Cox Plot Data Transformation

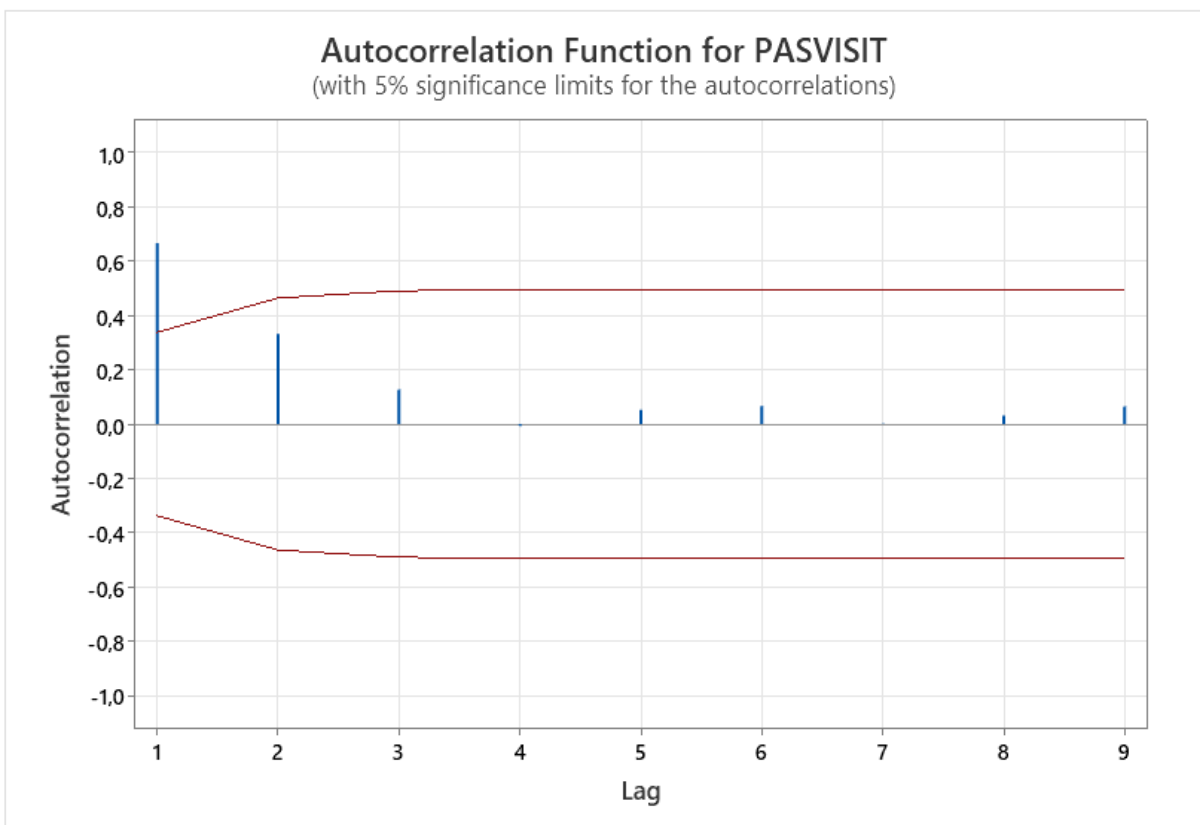


Figure 3. ACF Graph Patient Data Transformation

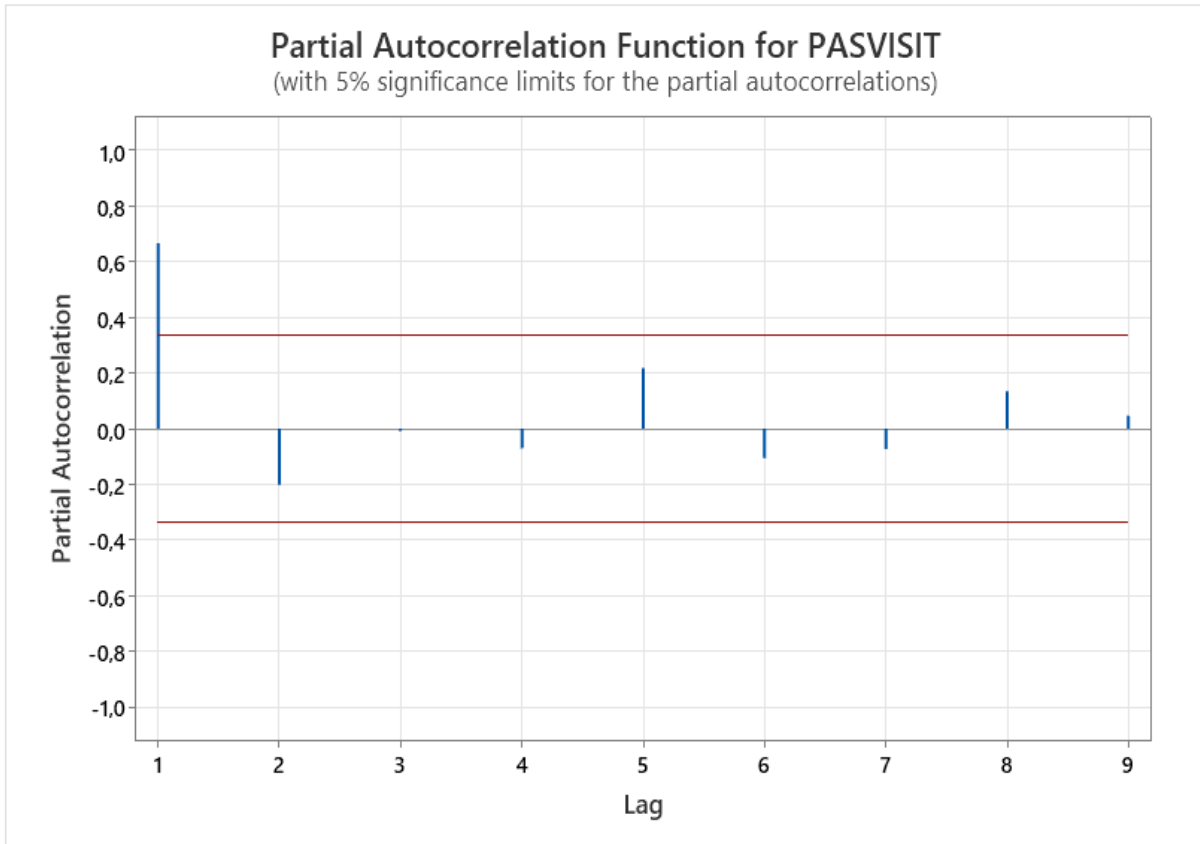


Figure 4. PACF Graph Patient Data Transformation

ACF and PACF plots in Figure 3 and 4 shows that the data transformation already stationer in means due to the coefficient auto-correlation and partial auto-correlation which is outside the red line (Bartlett) is only one coefficient it is lag one which outside on the red line, and the data graph is not too dies down, so there is no need to do the differencing process. So to determine the parameter model using ACF and PACF plots form the data transformation only. Temporary models which can be used to do advanced testing are ARIMA (1,0,1), ARIMA (0,0,1), and ARIMA (1,0,0).

**4.4. ARIMA Models Estimation**

The ARIMA model is said to be significant and feasible if it has a P-value final estimate of the parameter is below the error tolerance limit ( $\alpha$ ) 5% or 0.05 and the Ljung-Box P-value is above the error tolerance limit ( $\alpha$ ) 5% or 0.05. In addition, based on the significance test, the best model can be determined with the smallest error value from each

resulting model.

The temporary model is based on parameters testing with P-Value as Table 2.

The results of the analysis show that there are two ARIMA models with different significant parameters, namely (1, 0, 0) and (1, 0, 1), because it has a P-Value Final Estimates of Parameters value below the error tolerance limit ( $\alpha$ ) 5% or 0.05. And also, the Ljung-Box P-value is above the tolerance limit error ( $\alpha$ ) 5% or 0.05.

And then result of residual test results as the Table 3.

The best model is ARIMA (1,0,1). Then it is necessary to verify the model with the residual test consisting of white noise and normality test using Kolmogorov-Smirnov. The residual model is said to be normally distributed if the result of the residual plot probability is more than the error tolerance limit ( $\alpha$ ) of 5% or 0.05.

The residual graph of the ARIMA model (1,0,1) is as Figure 4.

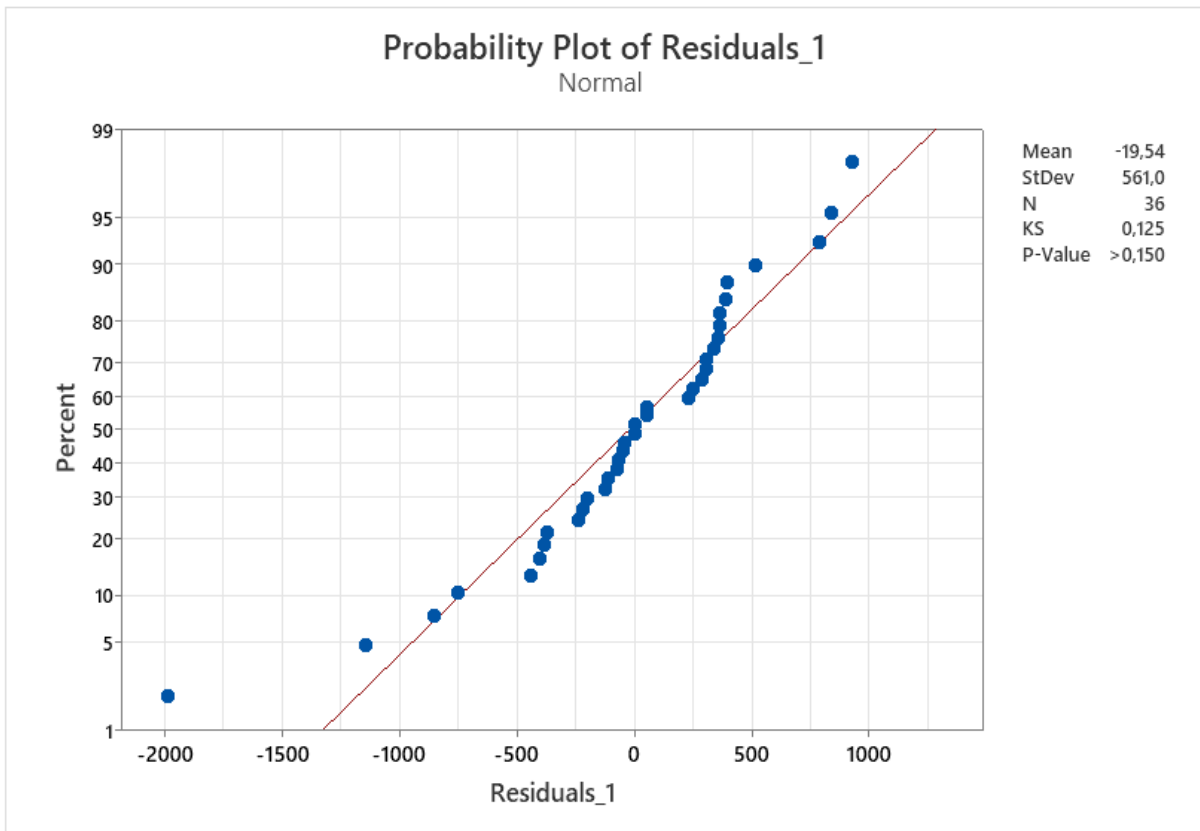
**Table 2.** P-Value Analysis Report

ARIMA Model	P-Value Final Estimates of Parameters				P-Value Ljung-Box	
	AR 1	AR 2	MA 1	MA 2	12	24
1,0,1	0,002	-	0,052	-	0,536	0,524
Transform 0,0,1	-	-	0,000	0,000	0,472	0,038
1,0,0	0,000	-	-	-	0,707	0,453

ARIMA Model	Decision
1,0,1	Significant
Transform 0,0,1	Rejected
1,0,0	Significant

**Table 3.** Residuals and Error Testing

Models	MS	Probability Plot of Residuals	Decision
Transformation 1,0,1	334183	> 0,150	Significant
Transformation 1,0,0	346994	> 0,150	Rejected



**Figure 5.** Residual Graph Model (1,0,1)

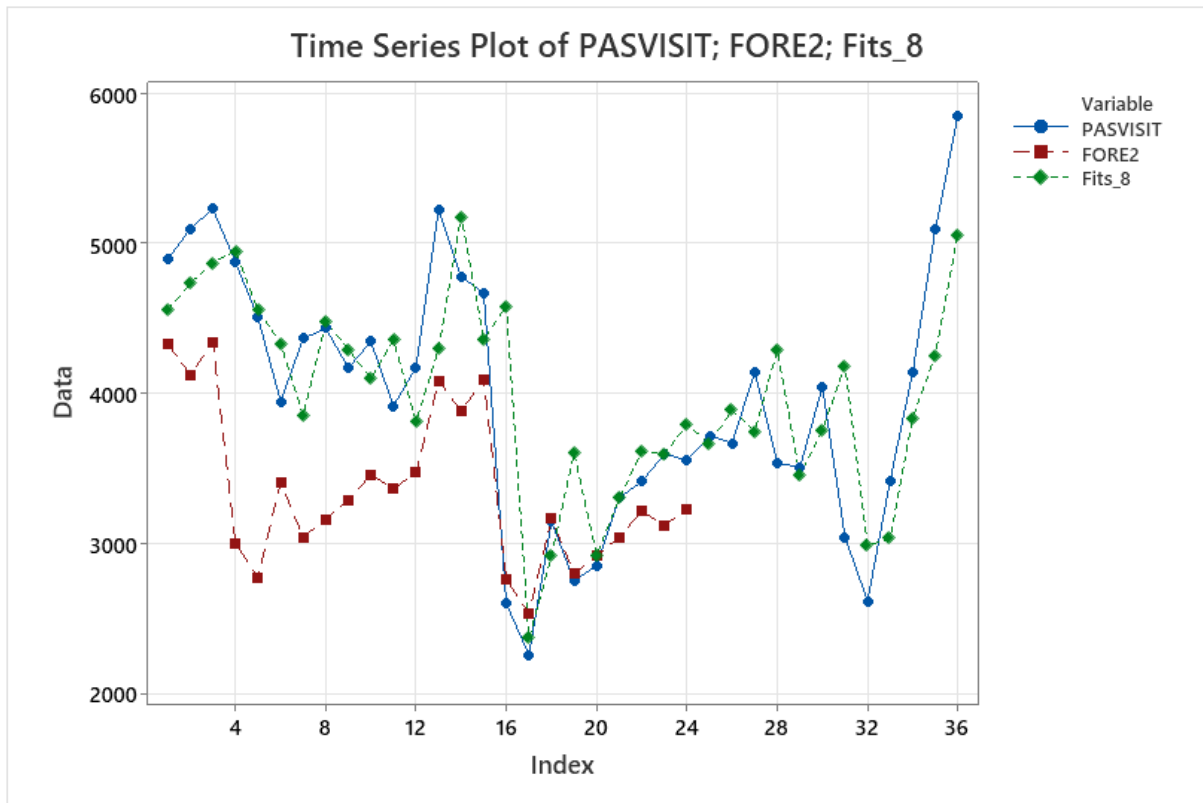


Figure 6. Time Series Plot Actual data, Fits, and Fore

Based on Table 3 and Figure 4 above, it is obtained that the P-Value for the residual ARIMA model (1,0,1) is worth  $> 0.150$ , which means that the value exceeds the value of the error tolerance limit ( $\alpha$ ) 5% or 0.05 in other words, the residuals from the ARIMA (1, 0, 1) model are already normal. From Table 3, It is also found that the ARIMA (1, 0, 1) model has a relatively small error value of 334183. So, in the end, it was determined that the best model for forecasting using the ARIMA method for this data is the ARIMA (1, 0, 1) model. The graph of the fits and fore is as Figure 6.

**4.5. Forecasting**

The data to be forecasted is the original data. In this case, it is 36 data on the number of patients visiting Mitra Sehat Clinic. The ARIMA model (1,0,1) with the error value (MS) 334183 is to be used as the best ARIMA model for forecasting the original data. After the forecasting process for 12 months later, the results are obtained.

Based on Table 4, it can be seen that the highest forecast is in January with 5507.27, and the lowest is at the end of December 2022 with 4116.47. In the actual column, the actual patient visits in the first three months of 2022 are also presented with results that are still within the tolerance threshold.

Table 4. Forecasting Results

Forecasting for 12 Months				
Period	Forecast	95% Limits		Actual
		Lower	Upper	
37	5507,27	4373,99	6640,55	5269,00
38	4978,19	3354,40	6601,98	5608,00
39	4649,30	2871,86	6426,74	4746,00
40	4444,86	2611,50	6278,22	
41	4317,77	2463,25	6172,29	
42	4238,77	2376,14	6101,40	
43	4189,66	2323,91	6055,42	
44	4159,14	2292,17	6026,10	
45	4140,16	2272,73	6007,59	
46	4128,36	2260,76	5995,97	
47	4121,03	2253,35	5988,71	
48	4116,47	2248,77	5984,18	

From the forecast results, an estimated error calculation will be carried out on the value of the forecast results to the estimated actual data value that will occur in the 12 months. The error calculation result of the actual data estimate is as follows.



**Table 5.** Error value of forecast data and actual data estimate

Actual Data Estimate	MAPE	MAD	MSD
Linear Trend Model	2,0	95,9	12414,8
Quadratic Trend Model	0,408	19,521	514,404
Growth Curve Model	1,8	85,3	10197,0
S-Curve Trend Model	0,0769	3,8713	95,8036

## 5. Conclusions

Health issues in the current era have become an essential topic of discussion with various existing issues, so derivative health problems are also significant. In this research concerning Forecasting the Number of Outpatient Patient Visits Using ARIMA at Mitra Sehat Clinic, it can be concluded that based on the forecasting results, the number of patient visits in 2022 will increase with various underlying factors, potentially be able to exceed the previous three years. Therefore, the ARIMA forecasting method must choose the best model with the smallest error value (MS). In the search to get the best model, several tests are needed, such as the test stationarity, residual test, and normality test. After finding the best model, it can be used to estimate the number of visiting patients in that Clinic.

With the potential increase in patients visit, advice that can be given to the Clinic is to prepare all kinds of tools and needs to provide services to these patients. So with this data, the Clinic can have a basis for preparing the quantity of equipment needed and maintaining the quality of service to remain optimal.

## Appendix A

A list of abbreviations used in this paper is shown below:

Abbreviation	Meaning
BPJS	Badan Penyelenggara Jaminan Sosial (Indonesian Health Social Security Agency)
ARIMA	Auto-Regressive Integrated Moving Average
ACF; PACF	Autocorrelation Function; Partial Autocorrelation Function
AR; MA	Auto-regression; Moving Average
MS	Minimum Mean Square Error
MAPE	Mean Absolute Percentage Error
MAD	Mean Absolute Deviation
MSD	Meanok as Squared Deviation

## REFERENCES

- [1] Ministry of Health Republic of Indonesia, "Peta Sebaran Covid-19.", Online available from <https://covid19.go.id/peta-sebaran> (accessed on 24 March 2022)
- [2] Ministry of Health Republic of Indonesia, "Vaksinasi Covid-19.", Online available from <https://vaksin.kemkes.go.id/#/vaccines> (accessed on 24 March 2022)
- [3] Ministry of Health Republic of Indonesia, "Apresiasi FKTP Berjuang Di Garda Terdepan, Kemenkes Adakan Jambore Virtual Bagi Tenaga Kesehatan di FKTP dalam rangkaian HKN.", Online available from <https://sehatnegeriku.kemkes.go.id/baca/umum/20201114/0035693/apresiasi-fktp-berjuang-di-garda-terdepan-kemenkes-adakan-jambore-virtual-bagi-tenaga-kesehatan-di-fktp-dalam-rangkaian-hkn/> (accessed on 10 February 2022)
- [4] Sukoharjo Government General Hospital, "Monitoring Data COVID-19 Kabupaten Sukoharjo.", Online available from <https://corona.sukoharjokab.go.id/> (Accessed on 10 February 2022)
- [5] Indonesian Health Social Security Agency (BPJS), "Jumlah Peserta Terdaftar.", Online available from <https://pcarejkn.bpjs-kesehatan.go.id/eclaim/ViewRekapPstPerPpk> (Accessed on 1 February 2022)
- [6] Indonesian Health Social Security Agency (BPJS), "Daftar Kunjungan Peserta Sakit." Online available from <https://pcarejkn.bpjs-kesehatan.go.id/eclaim/ViewKunjunganSakitByPpkTgl> (Accessed on 1 February 2022)
- [7] Indonesian Health Social Security Agency (BPJS), "Daftar Kunjungan Peserta Sehat." Online available from <https://pcarejkn.bpjs-kesehatan.go.id/eclaim/ViewKunjunganSehatByPpkTgl> (Accessed on 1 February 2022)
- [8] Montgomery D. C., Jennings C. L., Kulahci M., "Time Series Analysis And Forecasting: Statistics Background For Forecasting," 2nd ed, John Wiley & Sons Inc., 2015, pp. 25–41.
- [9] Fadhilah Abd. Razak, Mahendran Shitan, Amir H. Hashim, Izhah Z. Abidin, "Load Forecasting Using Time Series Models.", *Jurnal Kejuruteraan*, pp. 53-62, 2009. DOI: [dx.doi.org/10.17576/jkukm-2009-21-06](https://doi.org/10.17576/jkukm-2009-21-06)
- [10] Nany Salwa, Nidya Tatsara, Ridha Amalia, Aja Fatimah Zohra, "Peramalan Harga Bitcoin Menggunakan Metode ARIMA (Autoregressive Integrated Moving Average).", *Journal of Data Analysis*, Vol. 1, No. 1, pp. 21–31, 2018. DOI: <https://doi.org/10.24815/jda.v1i1.11874>
- [11] Saleh I Alzahrani, Ibrahim A Aljamaan, Ebrahim A Al-Fakih, "Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions." *J Infect Public Health*, National Library of Medicine, pp. 914-919, 2020. DOI: [10.1016/j.jiph.2020.06.001](https://doi.org/10.1016/j.jiph.2020.06.001)
- [12] Hiteshi Tandon, Prabhat Ranjan, Tanmoy Chakraborty, Vandana Suhag, "Coronavirus (COVID-19): ARIMA based time-series analysis to forecast near future.", *arXiv Cornell University*, 2020. DOI: <https://doi.org/10.48550/arXiv.2004.07859>
- [13] Willian W.S. Wei, "Time Series Analysis, Univariate and Multivariate Methods," 2nd ed, Pearson Addison Wesley, 2006.

- [14] Ridha Maya Faza Lubis, Zakarias Situmorang, Rika Rosnelly, "Autoregressive Integrated Moving Average (ARIMA-Box Jenkins) Pada Peramalan Komoditas Cabai Merah di Indonesia", *Jurnal Media Informatika Budidarma*, pp. 487-489, 2021. DOI: 10.30865/mib.v5i2.2927
- [15] Muhammad Bintang Pamungkas, Arief Wibowo, "Aplikasi Metode Arima Box-Jenkins Untuk Meramalkan Kasus Dbd Di Provinsi Jawa Timur", *The Indonesian Journal of Public Health*, pp. 183-186, 2018. DOI: <https://doi.org/10.20473/i-jph.v13i2.2018.183-196>