SCIENTIFIC REPORT ON

TIME SERIES ANALYSIS OF PATIENT ATTENDANCE AT JUMUIA FRIENDS

HOSPITAL



Chart

Description automatically generated

Table of Contents

[1 Chapter one 3](#_Toc121943564)

[PROJECT OVERVIEW 3](#_Toc121943565)

[Statement of the problem and metrics 4](#_Toc121943566)

[1.3 Objectives of study 4](#_Toc121943567)

[1.4 Significance of study 4](#_Toc121943568)

[Chapter 2 5](#_Toc121943569)

[2.1 Data exploration and visualization. 5](#_Toc121943570)

[Testing for stationarity. 6](#_Toc121943571)

[SUMMARY OF THE MODEL 8](#_Toc121943572)

[Normality assumption 8](#_Toc121943573)

[**Forecasted data** 9](#_Toc121943574)

[Model summary. 11](#_Toc121943575)

[Chapter 3 METHODOLOGY 11](#_Toc121943576)

[3.1 DATA PREPROCESSING 11](#_Toc121943577)

[3.2 IMPLEMENTATION OF THE MODEL. 12](#_Toc121943578)

Chapter 4 Model evaluation and validation ……………………………………………………..13

Chapter 5 Conclusion……………………………………………………………………….........15

References……………………………………………………………………………….16

# 1 Chapter one

# PROJECT OVERVIEW

The project's main aim is to examine and forecast the number of outpatients attending Jumuia friends Hospital at a particular time. We want to build a model that best defines our data and use that model to forecast future Attendance rates.

Our study before WHO was about the usefulness of statistics in the Field of medicine. In this Field of medicine, statistics are also highly relied on for medicine discovery trials. Over the years, Jumuia Friends Hospital was sometimes helped by numerous patients that attended the facility. This was a result of poor planning for their service delivery. Jumuia Friends Hospital is a private hospital, and profit maximization depends on proper planning and service delivery.

Our project will help the management to have appropriate plans for the expected number of outpatients at a particular time. The Ministry of Health and the World Health Organization can also use the same technique to estimate the medical requirement and other resources for citizens at a specific time in the future.

# Statement of the problem and metrics

Poor planning on several nurses serving patients, i.e., numerous circumstances had been recorded where asset prerequisites and nursing services had been overwhelmed with the number of patients seeking assistance from the Hospital.

Other times Hospitals would set a high number of nurses and experience fewer patients.

The kind of data to be used here will be from patient attendance records. This study will use several outpatient attendance (dependent variable) data from 2018 to November 2022(independent variable). The availability of a dataset from the Hospital determined our sample size. Jumuia Friends Hospital started storing its records in a computer system in 2018. Records were kept manually (on paper) for the previous years, which is why we could not get attendance for the last years. The research will be based mainly on secondary data from all outpatient clinics in Jumuia Friends Hospital. The data will be obtained from the records found in Health Records Department at Jumuia Friends Hospital. MAPE will be used to evaluate error metrics.

# 

# 1.3 Objectives of the study

Main objective.

1. Modelling patient attendance at Jumuia Friends Hospital.

Specific objective.

2. Develop a model that best describes our data.

3. Forecasting attendance rate for at least one year.

# 1.4 Significance of the study

Time series analysis on patient attendance at Jumuia Friends Hospital will help the Hospital improve service delivery.

To maximize profit in the Hospital, management should have a model showing direct variation between the number of patients attending the facility and resource allocation by the Hospital to serve those patients.

This will be archived by; estimating the number of health workers to help patients during a specific period and estimating bedding capacities and other medical resources to serve patients at

a given period [4].

## Chapter 2

## 2.1 Data exploration and visualization.

The kind of data to be used here will be from patient attendance records. This study will use several outpatient attendance(dependent variable) data from 2018 to November 2022(independent variable). The availability of the dataset from the Hospital determined our sample size. Jumuia Friends Hospital started storing its records in a computer system in 2018.

Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec  
 2018 621 453 657 432 567 453 234 687 691 688 691 510  
 2019 534 718 732 738 789 610 631 619 551 618 639 621  
 2020 589 436 492 470 585 534 619 624 789 639 694 711  
 2021 834 591 414 614 501 572 619 408 562 449 397 393  
 2022 390 427 415 497 602 906 800 613 713 516 418

autoplot(k)

I have my data value k for simple analysis. Our data is already set in time series format. Let's look at the plot of our data.

Chart

Description automatically generated

Fig 2.1. Time series plot.

The above graph shows how the attendance number was distributed within the specified duration of time. This graph shows whether the data followed an inevitable trend before conducting any analysis.

## Testing for stationarity

Before carrying out a time series analysis, the first step is to test whether the data is

Stationary or not.

For this reason, the Augmented Dickey-Fuller test is used to test for stationarity in time series analysis.

The results of this test are shown below.

Augmented Dickey-Fuller Test  
   
 data: k  
 Dickey-Fuller = -3.798, Lag order = 3, p-value = 0.0246  
 alternative hypothesis: stationary

###our data is stationary

Our data is stationary because our p-value is less than a 5 percent significance level.

We can therefore continue to perform time series analysis.

The auto. Arima() function uses a combination of unit root test and depreciation of the A.I.C.

(Akaike information criterion), and MLE(maximum likelihood estimator) to obtain the ARIMA model. KPSS test is then used to determine the number of differences (d).

The values of p,d, and q are then chosen by minimizing the AICc.

kmodel=auto.arima(k, ic="aic", trace = TRUE)

##   
## ARIMA(2,0,2)(1,0,1)[12] with non-zero mean : Inf  
## ARIMA(0,0,0) with non-zero mean : 745.84  
## ARIMA(1,0,0)(1,0,0)[12] with non-zero mean : 736.8619  
## ARIMA(0,0,1)(0,0,1)[12] with non-zero mean : 739.3337  
## ARIMA(0,0,0) with zero mean : 923.4838  
## ARIMA(1,0,0) with non-zero mean : 734.8859  
## ARIMA(1,0,0)(0,0,1)[12] with non-zero mean : 736.8622  
## ARIMA(1,0,0)(1,0,1)[12] with non-zero mean : 738.8619  
## ARIMA(2,0,0) with non-zero mean : 736.5282  
## ARIMA(1,0,1) with non-zero mean : 736.5954  
## ARIMA(0,0,1) with non-zero mean : 737.4236  
## ARIMA(2,0,1) with non-zero mean : 738.3746  
## ARIMA(1,0,0) with zero mean : 753.8683  
##   
## Best model: ARIMA(1,0,0) with non-zero mean

Using the method of Akaike information criterion, the best ARIMA model would be of order (1,0,0).

You can also try getting the same model's values using the Bayesian method.

kmodel=auto.arima(k, ic="bic", trace = TRUE)

##   
## ARIMA(2,0,2)(1,0,1)[12] with non-zero mean : Inf  
## ARIMA(0,0,0) with non-zero mean : 749.9951  
## ARIMA(1,0,0)(1,0,0)[12] with non-zero mean : 745.1721  
## ARIMA(0,0,1)(0,0,1)[12] with non-zero mean : 747.6438  
## ARIMA(0,0,0) with zero mean : 925.5614  
## ARIMA(1,0,0) with non-zero mean : 741.1186  
## ARIMA(1,0,0)(0,0,1)[12] with non-zero mean : 745.1724  
## ARIMA(1,0,0)(1,0,1)[12] with non-zero mean : 749.2496  
## ARIMA(2,0,0) with non-zero mean : 744.8383  
## ARIMA(1,0,1) with non-zero mean : 744.9055  
## ARIMA(0,0,1) with non-zero mean : 743.6562  
## ARIMA(2,0,1) with non-zero mean : 748.7623  
## ARIMA(1,0,0) with zero mean : 758.0234  
##   
## Best model: ARIMA(1,0,0) with non-zero mean

We can see that even using Bayesian method; the model does not change. I can affirm using the ARIMA order model (1,0,0) for my analysis.

## SUMMARY OF THE MODEL

## Series: k   
## ARIMA(1,0,0) with non-zero mean   
##   
## Coefficients:  
## ar1 mean  
## 0.4462 579.9879  
## s.e. 0.1166 26.9854  
##   
## sigma^2 estimated as 14000: log likelihood=-364.44  
## AIC=734.89 AICc=735.32 BIC=741.12

From the above summary, we have the attributes for our model.

## 

## Normality assumption

Residuals are used to test for normality in time series analysis. The histogram for residuals should be bell-shaped (symmetrical) in nature. This shape defines normal distribution; our data followed a normal distribution for our case.

Chart, histogram

Description automatically generated

A standard tool for graphing is the histogram. It summarizes continuous or discrete data measured on an interval scale. It is frequently used to provide a convenient manner of illustration for the main characteristics of the data distribution. When working with enormous data sets, it is also helpful (greater than 100 observations). Any abnormal findings (outliers) or data gaps can be found with its assistance.

## **Forecasted data**

The data below was forecasted for 20 months from December 2022 at the 95 percent level.

myforecast=forecast(kmodel, level = c(95), h=20)  
my forecast

## Point Forecast Lo 95 Hi 95  
## Dec 2022 507.7081 275.7982 739.6179  
## Jan 2023 547.7363 293.7872 801.6854  
## Feb 2023 565.5971 307.4845 823.7096  
## Mar 2023 573.5666 314.6331 832.5001  
## Apr 2023 577.1227 318.0260 836.2193  
## May 2023 578.7094 319.5803 837.8385  
## Jun 2023 579.4174 320.2819 838.5530  
## Jul 2023 579.7333 320.5965 838.8702  
## Aug 2023 579.8743 320.7372 839.0114  
## Sep 2023 579.9372 320.8000 839.0744  
## Oct 2023 579.9653 320.8281 839.1025  
## Nov 2023 579.9778 320.8406 839.1150  
## Dec 2023 579.9834 320.8462 839.1206  
## Jan 2024 579.9859 320.8487 839.1231  
## Feb 2024 579.9870 320.8498 839.1242  
## Mar 2024 579.9875 320.8503 839.1247  
## Apr 2024 579.9877 320.8505 839.1249  
## May 2024 579.9878 320.8506 839.1250  
## Jun 2024 579.9879 320.8507 839.1250  
## Jul 2024 579.9879 320.8507 839.1251

Hospital management can use the above-forecasted data to do estimation on

Resources allocation on a future monthly basis.

**We forecasted graph data.**

Chart, histogram

Description automatically generated

The above graph shows a graphical extension for the forecasted data attached to the actual data used to form our model.

## Model summary.

## Forecast method: ARIMA(1,0,0) with non-zero mean  
##   
## Model Information:  
## Series: k   
## ARIMA(1,0,0) with non-zero mean   
##   
## Coefficients:  
## ar1 mean  
## 0.4462 579.9879  
## s.e. 0.1166 26.9854  
##   
## sigma^2 estimated as 14000: log likelihood=-364.44  
## AIC=734.89 AICc=735.32 BIC=741.12  
##   
## Error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -0.3809331 116.3007 95.50874 -4.891252 18.28741 0.6171173  
## ACF1  
## Training set -0.0281456

### Chapter 3 METHODOLOGY

### 3.1 DATA PREPROCESSING

Data preprocessing includes changing crude information to comprehensive framed informational collections with the goal that information mining investigation can be applied.

Raw information is frequently fragmented and has conflicting organizing. The sufficiency or insufficiency of information preparation has an immediate connection with the progress of any undertaking that includes information analysis.

Preprocessing includes the two-information approval and information ascription. The objective of information approval is to survey whether the information being referred to is complete and precise.

The objective of information ascription is to address blunders and information missing qualities - physically or naturally through business process automation (B.P.A.) programming.

Regarding this description, we didn't process the data because we obtained it directly from the hospital database. I chose complete and accurate data for this study.

### 3.2 IMPLEMENTATION OF THE MODEL.

The process of using the model is quite simple.

**Identification of the model**-This step finds the appropriate values for p,d, and q. The method used to find these values is plotting the Autocorrelation function (A.D.F.) and partial autocorrelation function(PACF)[7]. The value of p is determined by line spikes that get outside the limits in a partial autocorrelation function plot (PACF)[1]. The points that reach outside the A.D.F. plot give the value of q, which is the order of autoregressive (A.R.). The value of d is simply several differences until our data was stationary. The story of A.C.F. and PACF helps obtain possible orders of ARIMA (p,d,q) models. ALL possible orderings of the model are listed down.

**Estimation** (Estimation of equation, estimation of coefficient) -

This step estimates the parameters of A.R. and M.A. included in the model. Method

The maximum likelihood technique is used where the equation is solved by nonlinear

function maximization. Backcasting is then used to obtain estimates of the initial residuals.

**Diagnostic checking**-In this step, we check whether the model fits our data. The main components to check here are the residuals. To validate the model, residuals of the estimated equation should be a white noise process, that is, without any correlation error. To check whether there is autocorrelation between lags, we use Box-test and Ljung-test to test the residual autocorrelation at a 5 percent significance level. If our p-value is less than 0.05, then there are autocorrelation errors. Our p-value should always be more significant than 0.05 whenever we increase the number of lags to claim the absence of autocorrelation between lags.

This implies that you should not get the value of p, which is less than 0.05

whenever we vary lag order in these tests. Suppose more than one ARIMA model is validated for the same time series. In that case, we can choose the best corresponding with the minimum value of the Akaike Information Criterion (A.I.C.)[2].

**Forecasting**

After checking the model, the next step is forecasting for the period needed by the researcher. This is done in R using the forecast function when you specify the confidence level and duration of estimation.

During implementation, one of the staff should have installed the correct software and understands statistical theories whenever the panel inquires about the same report. Otherwise, it's difficult to estimate without a statistical package, especially for large datasets like the one we are using in our study.

#### Chapter 4

#### 4.1 MODEL EVALUATION AND VALIDATION.

We used Box-test to validate the model.

The main operation here is to test whether there is autocorrelation between lags. We start with a small number of lags, say 2, and increase them, observing the p-value until the value is more than 0.05; then, your data is not having any autocorrelation.

Box-Ljung test

data: myforecast$resid

X-squared = 5.0113, df = 4, p-value = 0.2861

> Box.test(myforecast$resid, lag = 10, type = "Ljung-Box")

Box-Ljung test

data: myforecast$resid

X-squared = 8.6154, df = 10, p-value = 0.5680

> Box.test(myforecast$resid, lag = 20, type = "Ljung-Box")

Box-Ljung test

data: myforecast$resid

X-squared = 15.833, df = 20, p-value = 0.7269

> Box.test(myforecast$resid, lag = 30, type = "Ljung-Box")

Box-Ljung test

data: myforecast$resid

X-squared = 23.062, df = 30, p-value = 0.8127

> Box.test(myforecast$resid, lag = 35, type = "Ljung-Box")

Box-Ljung test

data: myforecast$resid

X-squared = 30.884, df = 35, p-value = 0.6672

There was no autocorrelation between lags as we increased the number of lags.

Therefore, our analysis was correct.

#### 4.2 Justification of the study

Proper service delivery by the Hospital will rely on future attendance rate estimates of patients. Therefore, our model will help the Hospital's overall management and policy enforcement toward service delivery to its patients (patients).

The outpatient visit pattern in Jumuia friends Hospital requires proper preparations in terms of personnel and logistics made in advance by the hospital management.

Therefore, the study findings would help the Hospital Management adequately prepare for the expected number of prospective patients.

This is likely to help them make advanced plans in terms of staffing and logistical requirements for better service delivery to the satisfaction and expectations of their customers.

Again, the Ministry of Health can adapt the same model for nationwide forecasting of both outpatients and improve the model to capture data for inpatients.

The level of accuracy in our model is very high.

MAPE value of 18.28741 is enough to conclude that our model deviated with only 18% of the actual estimate.

I decided to use the ARIMA model over other models for many reasons.

Comparison of the ARIMA model with other models used in time series.

|  |  |
| --- | --- |
| ARIMA MODEL | MOVING AVERAGE |
| It combines both autoregressive  with a moving average. | The moving average relies on the standard of  data per specified time duration. |
| ARIMA model has a fixed structure  and is specifically built for time series  (sequential) data | The structure is unfixed. |

##### CHAPTER 5. CONCLUSION

##### 5.1 REFLECTION

The researcher achieved his goals of modeling and forecasting future data from the study findings. The purpose of attaining the best plans for any organization is well explained in its mission and vision. Any organization that plans to succeed in its operations should try to model their data to improve accuracy in its future operations. Basing this argument on our study, Jumuia Friends Hospital is liked by many people according to the attendance rate and consistency observed in the data.

There is one problematic aspect of estimating patients based on the type of disease. Remember, our model only covers attendance rates, not patients suffering from different diseases.

My model would not be necessary if management concentrates on service for one disease.

##### 5.2 Improvement

Based on the research findings, I recommend hospitals to.

1) Use time series analysis to predict because it's to identify the pattern in your data.

2) Use auto. Arima function and the best criteria for selecting the model to avoid overfitting and underfitting the model.

We recommend Jumuia hospital always rely on the estimated number of patients attendance at least 95 percent level; this will help the hospital management improve their resource allocation to patients at any time.

I suggest to any other researcher who will research using the ARIMA model to increase the sample size, try to increase the number of predictors, and perform the A.D.F. test for testing stationarity. Remember, if the residuals of your data will not be normally distributed, then your estimates will be wrong. If your data is not normally distributed, try transforming it using logarithms.

###### References

[1] Avril Coghlan. : A little book on R For Time Series release 0.2, January 2014.

[2] Akaike, H.: Fitting autoregressive models for prediction, Ann. Inst. Stat.Math., 21, 243-247, 1969.

[3] Box G.E.P. and G.M. Jenkins, Time Series Analysis, Forecasting, and Control, Oakland, CA: Holden-Day. 1970

[4] Dr. Gavin Shaddick, using R (with applications in time series analysis), January 2004.

[5] Oleg Nedadic, Walter Zucchini, Time Series Analysis with R-Part I, 2004.

[6] Chen, S. M., and Hsu C. C. 2004: New Method to Forecast Enrollments Using Fuzzy Time

Series. International Journal Applied Science and Engineering, 2004. 2, 3: 234-244.

[7] Zhang J. G. 2007, Looking Ahead of the Curve: "ARIM" Modeling Approach to Enrollment Forecasting, Association for Institutional Research," Ju" e 4, 2007.

[8] The Analysis of Time Series: An Introduction by Chris Chat\_eld

[9] Dickey, D., and W. Fuller (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root, Journal of the American Statistical Association, 74, 427-431.

[10] Introduction to Time Series Analysis and Forecasting by Douglas C. Montgomery, Cheryl L. Jennings, and Murat Kulahci

[11] World health report 2002. Reducing risks, promoting healthy life. Geneva, World Health Organization, 2002.

[12] Danaei G, Lawes C.M.M., Vander Hoorn S, Murray C.J.L., Ezzat M. Global and regional

mortality from ischemic heart disease and stroke attributable to higher-than-optimal blood glucose concentration: comparative risk assessment. Lancet, 2006, 368:16511659.

[13] Baan R, Straif K, Grosse Y, Secretan B, El G.F., Bouvard V, et al. Carcinogenicity of some

aromatic amines, organic dyes, and related exposures. Lancet Oncology, 2008, 9:322323.