Forecasting HIV cases in Awka South (2025-2028)

Divine-Favour Ekemezie

December 2023

DATA PRESENTATION AND ANALYSIS

Data presentation and analysis will be comprehensively discussed in this section. Starting with presenting the data on HIV-positive cases in Awka South sourced secondarily from Amaku teaching hospital, Awka, to presenting relevant plots and forecasting the prevalence rate of HIV in Awka South, and hence drawing necessary conclusions.

The table below shows the HIV cases in Awka South from the year 2014 to 2023.

Table 1: HIV-positive cases in Awka South from the year 2014 to 2023

Year	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**	Average
2014	299	185	225	127	233	260	155	164	243	163	223	292	214.0833
2015	238	186	276	145	229	194	190	188	240	188	227	285	215.5
2016	252	195	291	138	229	208	158	172	170	205	266	307	215.9167
2017	266	170	250	166	222	254	152	273	175	177	247	304	221.3333
2018	302	171	245	155	264	180	161	210	164	198	246	305	216.75
2019	214	183	230	178	222	197	185	248	165	214	202	231	205.75
2020	173	168	204	187	191	191	180	180	210	178	170	176	184
2021	186	173	176	176	185	165	173	180	167	174	177	169	175.0833
2022	181	180	185	159	170	171	155	195	185	157	186	181	175.4167
2023	200	174	170	190	194	180	161	161	173	174	178	188	178.5833

Following the data are the plots of the yearly HIV cases of HIV alongside it's representation with the normal curve.

The graph of yearly average HIV cases from 2014 to 2023 shows a clear declining trend in the number of HIV-positive cases over time. HIV cases peaked around 2018, followed by a noticeable decline, particularly from 2019 onwards. This downward trend may reflect the impact of enhanced public health interventions, increased awareness, or improved treatment programs. The earlier portion of the dataset (2014 to 2018) shows some fluctuations, with

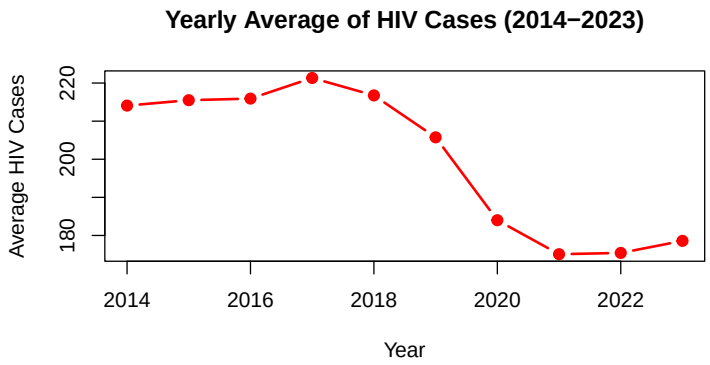

Figure 1: Yearly Average of HIV cases

periods of rising and stabilizing case averages. Overall, the trend from 2019 onwards suggests gradual progress in reducing HIV cases in the region.

The Histogram of Yearly Average HIV Cases from 2010-2019 plot shows that the average number of HIV cases fluctuated between approximately 150 and 200 over the years, indicating some variation in yearly cases, though the overall trend remains relatively consistent without drastic increases or decreases.

1 Data Modelling and Fitting

This section encompasses the subjection of the HIV data to normality tests to determine if the data is considered normal or not. Several methods will be employed such as Shapiro-Wilk, Anderson Darling, Crammer Von Mises, and Lilliefors (Kolmogorov's Smirnov) tests, and then in order to visualize how well the data fits a normal distribution, a Q-Q plot alongside a histogram showing the normal curve will be established.

The table below shows the results from these tests.

Table 2: Normality Test Results

Test	Result	p-value
Shapiro-Wilk	0.97	0.0208
Anderson Darling	0.86	0..237
Crammer Von Misses	0.12	0.0434
Kolmogorov Smirnov	0.08	0.0220

To assess the normality of the data, we test the following hypotheses:

Null hypothesis (H_0): There is no significant difference between the observed data and the generated normal distribution.
Alternative hypothesis (H_1): There is a significant difference between the observed data and the generated normal distribution.
Since the p-values for all the normality tests in Table 2 are less than 0.05, we will reject the null hypothesis. Thus, we conclude that there is a significant difference between the observed data and the normal distribution, indicating that the data does not have a normal distribution.

In this analysis, we will evaluate the accuracy of three models by employing three key metrics: Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). By comparing these metrics across the models, we aim to identify the most effective model for forecasting the trend of HIV cases in Awka South.

Table 3: Model Fitting of the HIV Data

Model	MAE	RMSE	MAPE
Linear	28.12	37.51	14.0
Quadratic	28.79	37.21	14.38
Exponential	27.95	37.76	13.63

Based on the results obtained, it suggests that the changes in HIV cases over time follow an exponential pattern rather than a linear or quadratic trend. This generally implies that the number of cases increases or decreases at a consistent percentage rate, rather than by a fixed amount in each period.

2 ARIMA Forecasting

To model the data using ARIMA, it's essential to first determine whether the time series is stationary or non-stationary. This is done by analyzing the graphs of the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF).

The figure 2 indicates that the HIV case data exhibits short-term dependencies, with recent values (like those from the previous month) being strongly correlated with the current values. This is useful for fitting autoregressive (AR) models, as the PACF helps determine the appropriate number of lags to include, likely 1 or 2 in this case. It suggests that the current month's HIV cases can be effectively forecasted using data from recent months, while contributions from more distant periods are less significant for prediction. The ARIMA forecasting was conducted using the statistical tool R, which can be seen in the appendix.

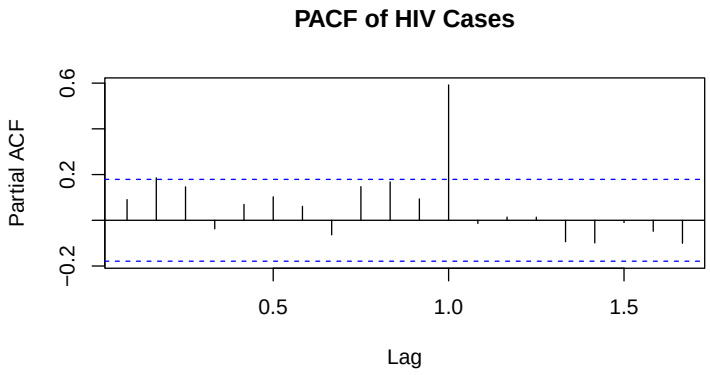

Figure 2: PACF of Differenced Series

Table 4

Month	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan-24	188.0946	150.89369	225.2956	131.2007	244.9886
Feb-24	168.5304	131.32951	205.7314	111.63652	225.4244
Mar-24	168.4366	131.23566	205.6375	111.54267	225.3305
Apr-24	175.4453	138.2444	212.6463	118.55141	232.3393
May-24	181.8713	144.67038	219.0723	124.97739	238.7653
Jun-24	170.8123	133.61133	208.0132	113.91834	227.7062
Jul-24	154.4695	117.26855	191.6704	97.57556	211.3634
Aug-24	164.2346	127.03362	201.4355	107.34063	221.1285
Sep-24	169.2949	132.09392	206.4958	112.40093	226.1888
Oct-24	164.1168	126.91591	201.3178	107.22292	221.0108
Nov-24	173.2468	136.04581	210.4477	116.35282	230.1407
Dec-24	178.9609	141.75999	216.1619	122.067	235.8549
Jan-25	183.5666	138.51296	228.6203	114.66298	252.4703
Feb-25	164.0024	118.94877	209.0561	95.0988	232.9061
Mar-25	163.9086	118.85493	208.9623	95.00495	232.8123
Apr-25	170.9173	125.86366	215.971	102.01369	239.821
May-25	177.3433	132.28965	222.397	108.43967	246.247
Jun-25	166.2843	121.2306	211.3379	97.38062	235.1879
Jul-25	149.9415	104.88781	194.9952	81.03784	218.8451
Aug-25	159.7066	114.65289	204.7602	90.80292	228.6102
Sep-25	164.7669	119.71318	209.8205	95.86321	233.6705
Oct-25	159.5888	114.53517	204.6425	90.6852	228.4925
Nov-25	168.7188	123.66508	213.7724	99.8151	237.6224
Dec-25	174.4329	129.37925	219.4866	105.52928	243.3366
Jan-26	179.0386	127.31091	230.7664	99.92791	258.1494
Feb-26	159.4744	107.74672	211.2022	80.36372	238.5852
Mar-26	159.3806	107.65288	211.1083	80.26988	238.4913
Apr-26	166.3893	114.66161	218.1171	87.27861	245.5001
May-26	172.8153	121.0876	224.543	93.7046	251.926
Jun-26	161.7563	110.02855	213.484	82.64555	240.867
Jul-26	145.4135	93.68577	197.1412	66.30276	224.5242
Aug-26	155.1786	103.45084	206.9063	76.06784	234.2893
Sep-26	160.2389	108.51113	211.9666	81.12813	239.3496
Oct-26	155.0608	103.33312	206.7886	75.95012	234.1716
Nov-26	164.1908	112.46303	215.9185	85.08003	243.3015
Dec-26	169.9049	118.1772	221.6326	90.7942	249.0157

The table 4 above represents predicted HIV-positive case numbers from January 2024 to December 2026, derived from historical data spanning 2014 to 2023 and utilizing an ARIMA (0,1,1) model for projection. The point forecast provides the projected number of HIV-positive cases for each month, representing the model's most accurate estimate based on historical trends and patterns. The point forecast for January 2024 is 188.09, indicating that the model predicts around 188 HIV-positive cases for that month. By December 2026, the forecast drops to 169.90, suggesting a gradual decline in cases over time compared to previous years.

The confidence intervals as shown in the table above are 80% and 90%. These intervals depict the range at which the true value will fall. Lo 80 and Hi 80: This shows the 80% confidence interval, which simply explains that there is an 80% chance that the actual number of HIV-positive cases will fall within this range.

- For January, the 80% confidence interval is 150.89 to 225.30, indicating there is an 80% chance the actual number of cases will fall between 151 and 225.
- By the end of 2026, the 80% confidence interval is 118.18 to 221.63.

Lo 95 and Hi 95 represent the 95% confidence interval, indicating a wider range where the actual values are expected to fall with 95% confidence.

- The 95% confidence interval for January 2024 ranges from 131.20 to 244.99, indicating that with 95% confidence, the actual number of cases is expected to fall between 131 and 245.
- By December 2026, the 95% confidence interval ranges from 90.79 to 249.02, reflecting a broader range of uncertainty in the model's prediction.