Abstract

Cancer is a major health challenge. Globally, the estimated number of diagnosed cancer incidences is approximately 14.1 million people per year and a mortality rate of 8.2 million deaths per year. The primary objective was to develop robust predictive models to forecast prostate cancer incidences and identify significant trends and patterns that inform healthcare planning and interventions in Meru County Kenya using AutoRegressive Integrated Moving Averarge with exogeneous variable (ARIMAX) Models.The dataset used comprised historical records of prostate cancer incidences in Meru County. The data spaned from [Jan 2018] to [Nov 2023], providing a comprehensive overview of the trends over time. Additionally, exogenous variable age was included in the ARIMAX model to enhance the accuracy of the prostate cancer predictions. Data on the prevalence of prostate cancer was obtained from Meru Cancer Registry for 71 months.The ARIMAX model was fitted using the Box-Jenkins methodology which include four iterative steps that is model identification,parameter estimation, diagnostics and forecasting.The prostate cancer time series data was made stationary by differencing and log transformation. R programming (Version 4.3.3) software was used in the analysis. Further, given the highly sensitive nature of the forecast values, interpolated data from daily values to monthly values were used. The best models for the Prostate cancer incidences was ARIMAX (0,0,1). Majority of the Prostate cancer incidences were within the age group 70-79 years at 50.7%,ages 60-69 was 42.3% while 80-90 years was 7%. After log transformation and and differencing of the prostate cancer time series data the Augmented Dickey Fuller test was performed and the p-value was (0.01) which was less than the significance level of (0.05), the null hypothesis was rejected that the prostate cancer time series had a unit root. Therefore, there was sufficient evidence to conclude that the time series was stationary. Ljung-Box test checked for the presence of autocorrelation at multiple lags and a high p-value = 0.719 greater than 0.05 indicated that there is no significant autocorrelation remaining in the residuals, thus the ARIMAX model was adequate. The MA(1) coefficient was -0.9, which indicated strong short-term negative autocorrelation. A positive value of 0.587 suggested that as the external variable increases by one unit, the log-transformed and differenced prostate cancer monthly cases (lnPCa Monthlycases d1) were expected to increase by 0.5871 units, holding all else constant. Results show that the ARIMAX(0,0,1) model slightly outperformed the ARIMA (0,0,1) model.This study successfully modeled the trends of prostate cancer incidences in Meru County using ARIMAX models. The findings indicated a rising trend in incidences, with the

ARIMAX model providing
the most accurate forecasts by incorporating the external variable age.