PROFESSIONAL STATISTICIANS SOCIETY OF NIGERIA

THE BOOK OF ABSTRACT

9TH INTERNATIONAL CONFERENCE

THEME: ADVANCING STATISTICAL LITERACY IN AI APPLICATIONS: NIGERIA'S STRATEGY FOR DIGITAL TRANSFORMATION

VENUE: FEDERAL POLYTECHNIC, ADO-EKITI, EKITI STATE, NIGERIA

AUGUST, 2025

Table of Contents

CLASSIFICATION OF DIABETES MELLITUS INCIDENCES FOR KANO STATE, BY AGE, GENDER, DIABETIC CONDITIONS AND RESIDENTIAL STATUS: A DATA MINING APPROACH.

*Musa Uba Muhammad1, Sani Salihu Abubakar1, Ibrahim Abdullahi2 and Salihu Gambo3

1 Department of Statistics Aliko Dangote University of Science and Technology Wudil, Kano State 2 Faculty of Mathematical Sciences North-west University Kano 3 Department of Statistics School of Technology, Kano State Polytechnic Kano * Correspondence: musaubamuhammad@gmail.com

Tel.: +2348032126733

`Corresponding Author: Musa Muhammad`

A precise classification for diabetes mellitus (DM) decides the suitable treatment and management of its nuisance, particularly in emerging countries like Nigeria. This study intends to employ data mining procedures for classifying and identifying the identified diabetes cases by age, gender, diabetic conditions, and residential area in Kano state of Nigeria, based on real-life data collected from public hospitals in the State. A K-mean assessment was employed to cluster the instances, after 11 iterations the occurrences classified out of 3062: 2692 (87.9%) non-insulin dependent (NID), 181 (5.91%) insulin-dependent (IND), and 189 (6.17%) gestational diabetes (GTD). The total number of identified diabetes cases was 3062: 1395 males (45.56%) and 1667 females (54.44%). The greater incidence was found in females compared to males, and mostly in cities and towns, rather than in villages (36.5%, 34.2%, and 29.3%, respectively). The highest incidence among the age groups was in the age group 50–69 years, which brings about 43.9% of the total identified cases. Moreover, the NID situation had the highest cases (87.9%). These were the first findings of the classification of the incidence in the State, and the figures have been of greatest importance to the healthcare authorities and non-governmental organizations for the proper preparation and management of diabetes.

Keywords: age; gender; diabetics-condition; classification; K-means;

SINE GOMPERTZ EXPONENTIATED INVERSE RAYLEIGH DISTRIBUTION: PROPERTIES, ESTIMATION, AND APPLICATIONS

S.O. Bashiru

Department of Mathematics and Statistics, Confluence University of Science and Technology, Osara, Kogi State, Nigeria

`Corresponding Author: OMEIZA SULE`

This study introduces a novel extension of the Gompertz exponentiated inverse Rayleigh distribution using the sine family of distributions, termed the Sine Gompertz Exponentiated Inverse Rayleigh Distribution (SGEIRD). The proposed model enhances flexibility and improves modeling capability for real life data with complex behaviors. We derive key statistical properties, including moments, quantile function, and hazard rate. Four classical estimation methods—Maximum Likelihood Estimation (MLE), Least Squares Estimation (LSE), Weighted Least Squares Estimation (WLSE), and Cramér von Mises Estimation (CME)—are employed to estimate the model parameters. A comprehensive simulation study is conducted to evaluate the consistency and efficiency of the estimators. The practical utility of the SGEIRD is demonstrated through its application to two real life datasets, where it outperforms several existing distributions in terms of goodness of fit criteria. The results highlight the proposed model’s robustness and versatility in statistical modeling.

Keywords: Sine generator; Gompertz exponentiated inverse Rayleigh distribution; maximum likelihood estimation; simulation study; goodness of fit

HYBRID ESTIMATORS FOR SOLVING MULTICOLLINEARITY IN A GAUSSIAN LINEAR REGRESSION MODEL BASED ON RIDGE – PCA ESTIMATORS

R. E. Alabi; O. O. Ojo; O. O. Alabi; T. S. Fayose
Federal Polytechnic Ado-Ekiti
Federal University of Technology, Akure
Federal University of Technology, Akure
Federal Polytechnic Ado-Ekiti
Corresponding Author: Remilekun Alabi

This study addresses the persistent issue of multicollinearity in Gaussian linear regression models which often renders Ordinary Least Squares (OLS) estimators inefficient and unreliable. While Ridge Regression and Principal Component Analysis (PCA) are established methods for mitigating multicollinearity, both techniques exhibit limitations in bias reduction and interpretability respectively. To overcome these challenges, this research proposes new hybrid estimators that combine novel ridge parameters with PCA. Four new ridge parameters were developed and integrated with PCA to form Ridge – PCA estimators. The performance of these hybrid estimators was assessed through a Monte Carlo simulation under varying sample sizes, error variances and multicollinearity levels using the Mean Squared Error (MSE) criterion. A total of 21 estimators including existing OLS, Ridge, PCA, and Liu estimators were evaluated. The results reveal that the proposed Ridge – PCA estimator particularly the minimum version of Chand and Kibria 2 (2024) combined with PCA (PCARCK2MIN) consistently outperformed others across all scenarios. This estimator exhibited the lowest MSE demonstrating superior efficiency and robustness. The findings underscore the effectiveness of combining regularization and dimensionality reduction in handling multicollinearity and provide a promising framework for improving regression estimates in high – dimensional and correlated data environments.

Keywords: Multicollinearity, Ridge Estimator, Principal Component Estimator, Hybrid Estimators, MSE

COMPARATIVE EVALUATION OF MACHINE LEARNING ALGORITHMS FOR THREE-CLASS DIABETES CLASSIFICATION USING NHANES DATA

T.M. Ayeni; E.A. Odukaya; A.O. Ilesanmi
Northeastern University, Toronto, Canada
Ekiti State University, Ado-Ekiti
Corresponding Author: Taiwo Ayeni

Background: Early detection of diabetes and prediabetes is critical for preventing long-term complications, including cardiovascular disease, neuropathy, and nephropathy. Type 2 diabetes affects over 537 million adults globally, with nearly half remaining undiagnosed. Machine learning offers potential for improving classification accuracy, but comprehensive comparative evaluations across multiple algorithms for three-class diabetes status prediction remain limited, particularly studies focusing on minimal feature sets suitable for resource-constrained settings. Objective: This study evaluates and compares four machine learning algorithms, Multinomial Logistic Regression (MLR), Decision Trees, Random Forest, and XGBoost, for classifying individuals into normal, prediabetic, and diabetic categories using readily available clinical indicators that can be obtained in primary care settings. Methods: Using data from the National Health and Nutrition Examination Survey (NHANES) with 2,029 participants (62.0% normal, 26.7% prediabetic, 11.3% diabetic), we employed Recursive Feature Elimination with 10-fold cross-validation to identify six key predictors from 27 candidate variables: fasting glucose, age, diabetes history, insulin level, waist circumference, and systolic blood pressure. Models were trained on a 70/30 stratified split and evaluated using overall accuracy, Cohen's Kappa, class-specific sensitivity, specificity, positive predictive value, and balanced accuracy. Results: Random Forest achieved the highest performance with 78.1% accuracy and Cohen's Kappa of 0.471, significantly outperforming XGBoost (72.6%), MLR (73.0%), and Decision Trees (72.4%). All models demonstrated excellent specificity for diabetes detection (>97%), minimizing false positives. Normal class identification showed robust sensitivity (87-88%) across algorithms. However, all models showed consistent challenges in prediabetes classification, with sensitivity ranging from 38-42%, reflecting the transitional nature of this metabolic state where physiological markers overlap with both normal and diabetic populations. Conclusions: Our comparative evaluation validates Random Forest as the optimal algorithm for three-class diabetes classification using NHANES data. The high performance achieved with only six routine clinical features obtainable with basic equipment in primary healthcare facilities suggests potential applicability in diverse clinical settings, including those with limited laboratory infrastructure. Future work should validate these findings in diverse populations, explore techniques to improve prediabetes detection, and develop implementation strategies for real-world deployment.

Keywords: Diabetes classification; Random Forest; NHANES; Machine Learning Comparison; Prediabetes detection; Multi-lass Classification

EMPIRICAL INVESTIGATIONS ON THE IMPACT OF RENEWABLE ENERGY PRODUCTION AND CONSUMPTION ON FOREIGN DIRECT INVESTMENT (FDI) IN AFRICA

Alhaji Abdullahi Gwani; Usman Waziri; Abbas Babayaro; Sek Siok Kun; Salahuddeen Abdulwaheed Adabayo
Department of Mathematical Sciences, Faculty of Science, Saadu Zungur University Bauchi State, Nigeria
School of Mathematical Sciences, Universiti Sains Malaysia, Malaysia
Corresponding Author: Alhaji Gwani

Amidst African pursuit of sustainable economic development and the global shift towards decarbonisation, the nexus between renewable energy and Foreign Direct Investment (FDI) has emerged as a critical area of inquiry. This study empirically investigates the impact of renewable energy production and consumption on FDI inflows into African nations. Employing a panel dataset covering 49 African countries from 2005 to 2022, this paper utilises panel data regression models, including Fixed Effects and System Generalised Method of Moments (GMM) estimators, to control for endogeneity and country-specific heterogeneity. Key independent variables include renewable energy production and renewable energy consumption per capita. The model also controls for established macroeconomic and institutional determinants of FDI; other variables such as GDP growth were included. The results indicate that African countries' commitment to the green transition serves as a robust signal to foreign investors, mitigating perceived risks and emphasizing new investment opportunities. This study concludes that renewable energy development constitutes not only an environmental mandate but also a compelling economic strategy for attracting sustainable investment in Africa.In the context of Africa's pursuit of sustainable economic advancement and the global transition towards decarbonization, the interplay between renewable energy and Foreign Direct Investment (FDI) has emerged as a vital subject of investigation. This research empirically examines the influence of renewable energy production and consumption on FDI inflows into African nations. Utilizing a panel dataset encompassing 49 African countries from 2005 to 2022, this paper employs panel data regression models, including Fixed Effects and System Generalized Method of Moments (GMM) estimators, to address endogeneity and country-specific heterogeneity. Essential independent variables include renewable energy production and renewable energy consumption per capita. The model also accounts for established macroeconomic and institutional determinants of FDI; additional variables such as GDP growth were incorporated. The findings disclose a positive and statistically significant correlation between renewable energy production and FDI inflows. Both augmented renewable energy production capacity and elevated consumption levels are identified as significant determinants of FDI. The results indicate that African countries' commitment to the green transition serves as a robust signal to foreign investors, mitigating perceived risks and emphasizing new investment opportunities. This study concludes that renewable energy development constitutes not only an environmental mandate but also a compelling economic strategy for attracting sustainable investment in Africa.

Keywords: Foreign Direct Investment, Renewable Energy production, Renewable Energy Consumption; Sustainable Development; Panel Data Analysis; Green Growth

EXPLORING DETERMINANTS OF STUDENTS' PREPAREDNESS FOR HIGHER EDUCATION: THE MEDIATING ROLE OF INTEREST IN STUDYING STATISTICS

I.O. Ajao; O. Olaitan
Federal Polytechnic, Ado-Ekiti, Nigeria
Walter Sisulu University, Buffalo City Campus, East London 5241 South Africa
Corresponding Author: ISAAC AJAO

This study, titled Exploring Determinants of Students' Preparedness for Higher Education: The Mediating Role of Interest in Studying Statistics, investigates the factors influencing secondary school students' readiness for higher education, emphasizing the mediating effect of their interest in statistics. A total of 300 students from four public and private schools in Ado-Ekiti, Nigeria, participated in the survey conducted in April 2025. The validated questionnaire measured constructs such as statistical awareness, preparedness, parental influence, societal influence, mathematical anxiety, and interest in statistics. Reliability analysis showed acceptable to high internal consistency, with Cronbach's alpha ranging from 0.60 (Parental Influence) to 0.92 (Statistical Awareness). PLS-SEM results revealed that both parental influence (? = 0.28, $p < 0.001$) and societal influence (? = 0.26, $p < 0.001$) significantly predicted students' interest in studying statistics. Interest, in turn, strongly predicted their preparedness for higher education (? = 0.44, $p < 0.001$), highlighting its mediating role. However, statistical awareness and mathematical anxiety had minimal or non-significant direct effects. These findings suggest that enhancing students' interest—through supportive family, community, and school environments—can improve their readiness for academic advancement, particularly in statistics-related disciplines. The study recommends targeted interventions to strengthen these determinants for better educational outcomes.

Keywords: Statistical Preparedness; Interest in Statistics; Parental Influence; Societal Influence; Secondary Education

ADVANCING STATISTICAL LITERACY WITH THE ZERO-INFLATED POISSON-EXPONENTIAL DISTRIBUTION: A ROBUST MODEL FOR COMPLEX COUNT DATA

N. Z. Rufai; I, Suleman; W. B. Yahya; M. A. Umar
Federal Polytechnic Oko, Anambra state
University of Maiduguri, Maiduguri
University of Ilorin, Ilorin, Nigeria
Bayero University Kano
Corresponding Author: UMAR MUHAMMAD

In this paper, a new member of the Poisson family of distributions called the Zero-Inflated Poisson-Exponential (ZIPEx) distribution is derived and introduced by compounding the Zero-Inflated Poisson (ZIP) distribution with the Poisson-Exponential distribution (PED). The ZIPEx distribution, a special case of the more general Poisson-Exponential-Gamma family, is designed to model count data exhibiting both overdispersion and excess zeros—features commonly found in real-world datasets across health, social, economic, and environmental domains in Nigeria. Theoretical properties of the ZIPEx distribution are studied, and its parameters are estimated using the Maximum Likelihood Estimation (MLE) method. The performance of ZIPEx is compared with several related classical and zero-inflated distributions (Poisson, Negative-Binomial, Poisson-Exponential, Poisson-Lindley, Poisson Exponential-Gamma, Zero-Inflated Poisson, Zero-Inflated Negative-Binomial, and Zero-Inflated Poisson-Lindley) using various goodness-of-fit (GoF) criteria including the log-likelihood, Akaike Information Criterion (AIC), Corrected AIC (AICc), and Bayesian Information Criterion (BIC) on multiple published real-life count datasets. Results show that the ZIPEx distribution consistently provides a better fit and greater modeling efficiency than the existing alternatives, making it a robust and credible option for modeling count data where precision and interpretability are critical. This contributes meaningfully to the advancement of statistical literacy in AI applications by equipping analysts and data scientists with a more flexible and accurate statistical model. In the context of Nigeria's digital transformation, the ZIPEx distribution offers a powerful tool for improving AI-driven decision-making in domains such as healthcare, agriculture, education, and public policy, where data are often sparse, noisy, and complex. The work underscores the importance of context-aware statistical modeling in building trustworthy, inclusive, and locally relevant AI systems.

Keywords: Zero Inflated Distributions; Count Data Modelling; Poisson-Exponential Distribution; Goodness-of-fit; Statistical Modelling for AI Applications

PREDICTIVE MODELLING OF DIABETES RISK USING MACHINE LEARNING TECHNIQUES

Wale-Orojo, Oluwaseun A.; Beckley, Oluwamayowa D.; Olayiwola Olaniyi M.
Department of Statistics, College of Physical Sciences, Federal University of Agriculture, Abeokuta, Nigeria
Corresponding Author: Oluwaseun Wale-Orojo

The rising global prevalence of diabetes presents a critical public health challenge, particularly in resource-constrained environments where early diagnosis is often delayed. This study leverages supervised machine learning techniques—Logistic Regression and Random Forest—to develop predictive models for assessing diabetes risk using a secondary clinical dataset of 1,000 anonymized patient records. The dataset includes demographic and biochemical markers such as age, gender, HbA1c, BMI, cholesterol, and triglycerides. Rigorous data preprocessing steps, including outlier detection, log transformation, and multicollinearity checks, were implemented to optimize model accuracy and robustness. Model training and evaluation were performed using a 70:30 train-test split within the R statistical environment. The Logistic Regression model, refined through backward stepwise elimination, achieved an accuracy of 97.18% and an AUC of 0.99, highlighting HbA1c, BMI, and triglycerides as key predictors. In contrast, the Random Forest model demonstrated superior performance, attaining an accuracy of 98.33% and an AUC of 1.00, with a stronger balance between sensitivity (98.42%) and specificity (97.83%). Both models confirmed the critical influence of HbA1c and lipid profiles in diabetes classification. This research underscores the potential of predictive analytics in augmenting early detection strategies and informing clinical decision-making, advocating for the integration of machine learning into diabetes screening frameworks to improve health outcomes.

Keywords: Diabetes Prediction; Machine Learning; Logistic Regression; Random Forest; Clinical Decision Support

PREDICTION OF MULTIVARIATE CRYPTOCURRENCY RETURNS VOLATILITY WITH MGARCH AND DEEP LEARNING TECHNIQUES

S. Umar; G. M., Oyeyemi; A. Bolakale; A. Murtala
Department of Mathematics and Statistics, Umaru Ali Shinkafi Polytechnic, Sokoto
Department of Statistics, University of Ilorin
Corresponding Author: Gafar Oyeyemi

The increasing integration of cryptocurrencies into global financial systems has intensified the need for robust methods to model and forecast their volatility dynamics. This study evaluates the performance of the Dynamic Conditional Correlation Multivariate GARCH (DCC-MGARCH) model and Long Short-Term Memory (LSTM) networks using daily return data from major cryptocurrencies between 2018 and 2025. The return series exhibit stylized features such as volatility clustering, excess kurtosis and time-varying dependencies. The DCC-MGARCH framework successfully models conditional heteroskedasticity and dynamic correlations, notably the strong co-movement between leading digital assets. However, the LSTM model demonstrates superior accuracy in forecasting return volatility, as reflected in consistently lower error metrics across both in-sample and out-of-sample evaluations. These results highlight the advantages of deep learning networks in modeling complex volatility structures and provide empirical support for their application in financial decision-making, particularly within high-volatility environments such as digital asset markets.

Keywords: Cryptocurrency; DCC; Deep Learning; LSTM; MGARCH; Volatility Forecasting

EVALUATING THE IMPACT OF WATER POLLUTANTS ON HUMAN HEALTH: A ROBUST REGRESSION MODELING APPROACH

K. A. Odior; R. E. Emudiaga; B. E. Omokaro
Department of Statistics, Delta State Polytechnic, Otefe-Oghara
Corresponding Author: Rukevwe Emudiaga

This study evaluates the robustness and efficiency of five estimators; Ordinary Least Squares (OLS), Weighted Least Squares (WLS), Generalized Least Squares (GLS), MM Estimator, and Theil-Sen Estimator, in modeling environmental health data under varying sample sizes and contamination levels. Simulated datasets were designed to reflect real-world characteristics commonly found in environmental health studies, using predictors such as lead (Pb), cadmium (Cd), and water pH, with hemoglobin concentration (g/dL) as the response variable. Each simulation was conducted across five sample sizes (n = 25, 50, 75, 100, 1000), both under clean data conditions and under three contamination scenarios where 2%, 5%, and 10% of the observations were replaced with extreme outliers. Model performance was assessed using R^2, Root Mean Square Error (RMSE), and information criteria (AIC and BIC). The results reveal that while OLS and GLS perform well under clean data, they degrade rapidly with outliers. Theil-Sen and MM estimators offer modest robustness under contamination. However, WLS consistently outperforms all other methods in both clean and contaminated conditions, demonstrating superior predictive accuracy, resilience to outliers, and efficient handling of structural complexity. The findings underscore WLS as the most reliable method for fitting models in environmental health research where data irregularities are common.

Keywords: Robust Regression; Environmental Health; Outlier Contamination; Weighted Least Squares; Simulation Study

ENHANCING LINEAR REGRESSION ROBUSTNESS: A NOVEL APPROACH TO MITIGATE HETEROSCEDASTICITY AND OUTLIER INFLUENCE

T. S. Fayose; K. Ayinde; O. O. Alabi; O. F. Fayose; R. E. Alabi
Federal Polytechnic Ado Ekiti
Northwest Missouri State University USA
Federal University of Technology Akure
Federal Polytechnic Ado Ekiti
Federal Polytechnic Ado Ekiti
Corresponding Author: Taiwo Fayose

Linear regression is a fundamental statistical technique used for modeling the relationship between an output variable and one or more covariates. The most prevalent method of estimating the parameters in a linear regression model is the Ordinary Least Squares (OLS) Estimator. The OLS estimator is assumed best when all assumptions are satisfied. When any of these assumptions fail, there is need for other estimators. Kernel regression mixed with some ridge estimators as novel weighted estimators have been proposed as alternative techniques to improve estimation accuracy. However, both methods are non – robust to outliers. In previous studies, the M – estimator and S – estimator have been used in combination with the ridge estimators to address both correlated regressors and outliers. In this paper, we introduce new estimator by combining three estimators to form LQS – Weighted – Ridge estimator to address both heteroscedasticity and outliers simultaneously. Our simulation results demonstrate that the proposed technique performs favorably compared to existing methods using Mean Square Error (MSE) as the criterion to evaluate the performances of the new and existing estimators.

Keywords: Linear Regression Model; Kernel; Heteroscedasticity; LQS Estimator; Outlier; Monte Carlo

THE INFLUENCE OF PERSONALITY TRAITS ON THE ACADEMIC PERFORMANCE OF UNDERGRADUATE STUDENTS AT FUNAAB

Wale-Orojo, Oluwaseun A.; Emmanuel Oluwole A.
Department of Statistics, College of Physical Sciences, Federal University of Agriculture, Abeokuta, Nigeria
Corresponding Author: Oluwaseun Wale-Orojo

This study investigates the influence of personality traits on the academic performance of undergraduate students at the Federal University of Agriculture, Abeokuta (FUNAAB). Anchored on the Five-Factor Model of personality—Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism—the research employed a descriptive survey design. A total of 100 students were selected using stratified random sampling from various colleges within the university. Data were collected using a validated questionnaire and students' Cumulative Grade Point Averages (CGPAs) as academic performance indicators. Descriptive statistics revealed that 64% of respondents scored high in Conscientiousness, while 55% demonstrated high levels of Openness. Inferential analysis using Pearson correlation showed a significant positive relationship between Conscientiousness and academic performance ($r = 0.63$, $p < 0.01$), and between Openness and academic performance ($r = 0.51$, $p < 0.01$). Neuroticism exhibited a weak but negative correlation ($r = -0.29$, $p < 0.05$). Multiple regression analysis revealed that Conscientiousness alone accounted for 39.7% of the variance in academic performance ($R^2 = 0.397$), while the full model explained 58.4% of the total variance ($R^2 = 0.584$). The study concludes that personality traits—particularly Conscientiousness and Openness—significantly influence students' academic achievement. It recommends the integration of personality assessments into university counseling programs and curriculum design to foster personalized learning strategies and support student development.

Keywords: Personality Traits; Academic Performance; Five-Factor Model; FUNAAB; Regression Analysis

GEOSTATISTICAL MODELLING OF THE COMORBIDITY OF LYMPHATIC FILARIASIS AND MALARIA.

E.E Ukwajunor; F. Eshofonie; A.O Egonmwan; P Osatohanmwen
Department of Statistics, University of Lagos, Lagos State, Nigeria
Department of Statistics,Federal University of Technology, Akure, Nigeria
Mathematical, Computational and Statistical Biology Research Group (MCSB). Department of Mathematics, University of Benin, Benin City, Nigeria
Faculty of Economics and management, Free University of Bolzano, Italy
Corresponding Author: Eunice Ukwajunor

Lymphatic filariasis (LF) and malaria are two of the most burdensome mosquito-borne diseases in sub-Saharan Africa, often co-endemic in areas where Anopheles mosquitoes serve as the primary transmission vector. Un derstanding their spatial comorbidity is vital for designing efficient, integrated disease control strategies. In this study, we applied a shared component geostatistical model to jointly assess the spatial distribution of LF and malaria across Nigeria. Malaria prevalence data, obtained from the 2015 and 2020 Nigeria Malaria Indicator Surveys, were modeled using the beta distribution to account for the bounded and moderately skewed nature of the data. LF prevalence data, sourced from the ESPEN database spanning 2000 to 2022, were modeled using the gamma distribution, which appropriately captured its positively skewed distribution. A latent spatial effect common to both diseases was introduced through a zero-mean Gaussian Random Field with a Mat´ ern covari ance structure, enabling the identification of shared and disease-specific spatial risks. The results revealed that malaria prevalence was highest in the northwestern states, while LF prevalence showed localized clustering in central and southern states such as Nasarawa, Kogi, Ondo, Ekiti, Edo, and Benue. Although the overall spatial overlap between the two diseases was limited, these states exhibited relatively higher combined prevalence, high lighting them as potential targets for integrated surveillance and intervention. These findings underscore the spatial heterogeneity of malaria and LF transmission in Nigeria and demonstrate the value of shared component modeling in informing geographically targeted and cost-effective public health strategies.

Keywords: Spatial comorbidity, Lymphatic filariasis, Malaria, Shared component model, Nigeria, Geostatis tical modeling, Anopheles mosquitoes

PREDICTIVE MODELING OF STOCK PRICES USING SUPERVISED MACHINE LEARNING

Tolulope Olayemi James; B. Shehu
Abdullahi Fodio University of Science and Technology, Aliero, Kebbi State, Nigeria
Corresponding Author: TOLULOPE JAMES

The prediction of stock prices remains a significant challenge in financial analytics due to market volatility and complex data dynamics. This study applies machine learning techniques, including Decision Tree, Support Vector Machine (SVM), and Random Forest and developed predictive models for stock price forecasting using historical stock market data from January 2022 to May 2025. The dataset was partitioned into 70% training and 30% testing and training subsets to assess model performance. Each model was evaluated based on key regression performance metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2), and Mean Absolute Percentage Error (MAPE). The results revealed that the Random Forest model outperformed the others, achieving the lowest RMSE (2.7857), MAE (1.8876), and MAPE (1.60%), with the highest R^2 value (0.9988), indicating excellent predictive accuracy. The SVM model followed closely with an RMSE of 4.3903 and R^2 of 0.9970, while the Decision Tree model performed the least effectively with an RMSE of 13.5222 and MAPE of 13.08%. These findings suggest that ensemble methods like Random Forest offer superior performance in modeling and predicting stock prices over simpler decision-based or kernel-based algorithms. This study recommends the integration of robust machine learning models in financial forecasting for enhanced investment decision-making.

Keywords: Stock price prediction, machine learning, random forest, support vector machine (SVM), decision tree.

GRAPHICAL ASSESSMENT OF THE PREDICTION VARIANCE OF RESPONSE SURFACE DESIGNS FOR FITTING THE MULTIVARIATE LINEAR REGRESSION MODELS

N. C. Umelo-Ibemere; K. E. Anyiam
Federal University of Technology, Owerri
Corresponding Author: NGOZI UMELO-IBEMERE

Many experiments in science and technology require simultaneous measurement of more than one response function for each setting of the input variables. Usually, the multiple responses are approximated using multivariate model and appropriate experimental designs for fitting the model are required in modeling and optimizing such experiments. Most times, the responses are correlated therefore, multivariate methods are needed to obtain the prediction variance-covariance matrix. Often, it is challenging to make a choice on the appropriate design to use. Scaled prediction variance is a common measure of design performance and fraction of design space (FDS) plot is a graphical display that uses the scaled prediction variance (SPV) to evaluate the performance of designs. FDS plots are used to assess the prediction variance of the central composite design (CCD) and three types of extended central composite design (ECCD) for fitting a multivariate linear regression model. The designs are chosen such that the variances and covariances of the predicted responses are constant on spheres centered at the origin. From the plots, none of the designs is consistently superior all over the design region. However, type II ECCD and CCD have more stable prediction variances than the other designs.

Keywords: Extended central composite design; Fraction of design space; Multivariate linear model; Multi-response rotatability; Scaled prediction variance

ON THE DEVELOPMENT OF A PARAMETRIC PH MODEL USING A WEIBULL-EXPONENTIAL DISTRIBUTION BASELINE HAZARD FUNCTION

Abdulfatai Lawal ; Olalekan Akanji Bello ; Salihu M. Zakari
Department of Mathematics and Statistics Kaduna Polytechnic Kaduna
Department of Statistics Osun State University, Osogbo
Corresponding Author: Abdulfatai Lawal

This study proposes a parametric proportional hazards (PH) model, referred to as the Weibull-Exponential Proportional Hazards (WEPH) model, which incorporates the Weibull-Exponential distribution as the baseline hazard function, thereby extending the traditional Cox PH framework. The motivation for this development arises not only from the need for modeling flexible hazard shapes but also from the need to improve model adequacy in survival analysis, where more accurate time-to-event modeling can significantly impact decision-making and interpretation. The proposed model is evaluated using two distinct datasets: historical child mortality, and clinical survival data of breast cancer patients. Comparative analyses based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) reveal that the new Weibull-Exponential-based PH model consistently outperforms the Cox PH model. This superior performance is attributed to the parametric nature of the baseline hazard, which allows for better representation of survival dynamics and yields more efficient estimates. These results underscore the potential of the proposed model as a robust and clinically meaningful alternative for parametric survival analysis in both historical demographic and medical research settings.

Keywords: Weibull-Exponential Distribution, Parametric Proportional Hazards Model, Survival Analysis, Cox PH Model Comparison, Historical and Clinical Survival Data

MODIFIED A CLASS OF RATIO–TYPE ESTIMATORS FOR FINITE POPULATION MEAN WITH TWO AUXILIARY VARIABLES.

K.A MUHAMMED; A .A ADEWARA
Department of Statistics University of Ilorin
Corresponding Author: KAMALDEEN MUHAMMED

In sample survey, the uses of auxiliary information incorporated with the study variable has played a significant role in an estimation stag, for population mean. Several authors have placed greater emphasis on the use of a single auxiliary variable (x) in estimating the study variable (y) and suggested a large number of the estimators along with their properties. In this study, two auxiliary variables were used with optimum value of weight assigned to each auxiliary variable with study variable, to modify four ratio-type estimators The Bias and Mean Square Error of the modified estimators were obtained up to the first order approximation, and the percentage relative efficiency of modified estimators over some existing estimators were established. An empirical study was carried out, and the results from the analysis shows that, the modified estimators were performed efficiently than the existing estimators considered in this study, having attained minimum Mean Square Errors.

Keywords: Keyword: population mean; auxiliary variable; weight; Bias and mean square error.

GENERALIZED ALMOST UNBIASED ESTIMATOR OF FINITE POPULATION VARIANCE UNDER STRATIFIED RANDOM SAMPLING

I. Muhammad; G. M. Oyeyemi, A. Bolakale and A. Murtala
aDepartment of Statistics, Binyaminu Usman Polytechnic, Hadejia, Nigeria bDepartment of Statistics, University of Ilorin, Ilorin, Nigeria
Corresponding Author: BOLAKALE ABDULHAMEED

This study developed a generalized almost unbiased estimator for obtaining finite population variance under stratified random sampling. The generalized estimators were developed using the approach of linear combination. The approach of almost unbiased was utilized in deriving the theoretical properties of the generalized estimators and were derived along with the expressions of the filtration parameters. The efficiency conditions of the proposed estimator over some existing population variance estimators were obtained theoretically. The performance of the proposed estimator was evaluated empirically using four real life datasets. Based on the criteria of bias, mean square error and percentage relative efficiency, the developed estimator performed better with minimum values of bias and mean square error, and higher value of percentage relative efficiency. Therefore, the generalized almost unbiased estimator can be used to estimate the variations of various phenomenon in stratified random sampling.

Keywords: Estimator, variance, unbiased, mean square error, efficiency.

LOG-LINEAR AND BINARY LOGISTIC REGRESSION MODELLING OF CARDIOVASCULAR DISEASE RISK FACTORS AND ITS ASSOCIATED SYMPTOMS IN NORTHWESTERN NIGERIA

I. O. Aliyu; O. M. Adetutu
Department of Statistics, Federal University of Technology Minna
Corresponding Author: Ilyasu Aliyu

Cardiovascular diseases encompass a diverse group of medical conditions that impact the heart and blood vessels, particularly arteries and veins. These conditions can have severe and sometimes life-threatening consequences, making them a significant global health concern and a leading cause of morbidity and mortality worldwide. Our study focused on the Northwestern region of Nigeria, where we precisely collected a total of 1,632 responses via a Google form from a specific cohort comprising healthcare personnel and patients over a period extending from August 15, 2024 to September 15, 2024. The study involved a multifaceted analysis, which including estimation of the prevalence rate and quantify the significant risk factors associated with cardiovascular diseases, determination of the associations between reported symptoms and the likelihood of cardiovascular disease diagnosis and assessment of the overall model fit and predictive accuracy of the logistic regression model. In order to achieve research aim, log-linear and binary logistic regression models were used. Research findings revealed key risk factors includes: Hypertension, High Cholesterol, Smoking, Alcohol Consumption, Unhealthy Diet, Physical Inactivity, Obesity and Overweight, and Diabetes admitted by 21.1, 15.2, 12.4, 8.2, 11.8, 12.1, 10.3, and 8.8 percent of the respondents. These findings provide valuable insight into the prevalence of risk factors in the Northwestern Nigerian context. Logistic regression evidently suggested age, gender, symptoms, risk factors, and family history as predictors of cardiovascular diseases; however, the variable "state" prove statistically insignificant and did not impact the model outcome. On classification, results of the logistic regression model performed well in predicting residing environments, achieving an overall accuracy of 83.9%. When dealing with log-linear modelling, interaction between age and gender was statistically significant, but the interaction between age and residing site was statistically insignificant when controlling for other variables (p-value of 0.064, which is not below the typical 0.05 threshold), this is an indication that place where you reside have no impact or relationship with age of a person towards cardiovascular disease, assess the model's fit, we reported the Nagelkerke pseudo R-squared, which indicates that approximately 53.5% of the variation in the dependent variable is explained by the model.

Keywords: log-linear; logistics; risk factors; morbidity; mortality

THE EFFECT OF HETEROSCEDASTICITY AND OUTLIERS ON VARIOUS ESTIMATION TECHNIQUES IN SIMULTANEOUS EQUATIONS MODELS

A. Bolakale, A. M. Kolawole, M. O. Babatunde
Department of Statistics, University of Ilorin, Ilorin, Nigeria.
Corresponding Author: Babatunde Mariam

The Gauss-Markov assumptions regarding equal variances of the error term and the absence of extreme values in any linear model can adversely affect not only the reliability of a model but also the significance of the contributing variables. Therefore, this study examines how heteroscedasticity and outliers influence the performance of different estimation techniques in simultaneous equation models utilizing data characterized by non-constant error variance and leveraged-endogenous variables. Various estimators, including Two-Stage Least Squares (2SLS), Three-Stage Least Squares (3SLS), Indirect Least Squares (ILS), Ordinary Least Squares (OLS), Seemingly Unrelated Regression (SUR), and Weighted Two-Stage Least Squares (W2SLS), were evaluated under different sample sizes (20, 50, 100) and levels of outliers (10%, 20%, and 30%). The findings indicate that both heteroscedasticity and outliers have a notable impact on the accuracy and efficiency of parameter estimates, with increased sample sizes typically enhancing performance, while outliers consistently diminish the model fit across all estimation techniques. This research offers important information for practitioners addressing simultaneous equation models amidst model irregularities.

Keywords: Simultaneous equation models, heteroscedasticity, outliers, 2SLS, 3SLS, OLS, SUR, W2SLS, ILS

CONVERTING A 2-WAY CLASSIFICATION MODEL INTO A MACHINE LEARNING LINEAR REGRESSION MODEL

N. I. Badmus, R. K. Ogundeji and B. W. Akinkunmi
University of Lagos, Akoka, Nigeria(1 & 2) and Federal School of Statistics, Ibadan, Nigeria.
Corresponding Author: Idowu Badmus

The study presents a way of formulating a multiple linear model from a 2-way classification called Completely Randomized Block Designs (CRBD). This classification consists of treatment and block known as features (input), and the formation is done by taking rows and columns elements as the response variable (target or output). In addition, the number of n-trials, slope, predictors, and regression parameters were obtained within the system. The exploratory data analysis (graphical representation), such as histograms, boxplots, and QQ-plots, was conducted to ascertain the normality of the target variable and for proper selection of the error term distribution. Therefore, the data used for illustration is on the dosage that produces the crop yield for cotton. Four fields divided into six plots, with dosages of 5, 10, 15, 20, 25, and 30 units of herbicide assigned to the plots at random, are considered. Furthermore, results reveal that the machine learning linear model is better used than a 2-way classification model. This is determined by some metric values, including: MAE, MSE, RMSE, MAPE, MBE, and RAE = (0.8913, 1.5801, 1.2570, 32.6469%, -0.0000, and 0.1964).

Keywords: Boxplot, Block, herbicide, input, Output, QQ-plot, Randomized, Treatment

PARAMETER ESTIMATION IN DYNAMIC PANEL DATA MODELS USING A HIERARCHICAL BAYESIAN FRAMEWORK

Oladimeji, L. A.; Garba, M. K. and Adeshola, A. D.
Ladoke Akintola University of Technology, Ogbomoso, Nigeria
University of Ilorin, Ilorin, Nigeria
Kwara State University, Malete, Nigeria
Corresponding Author: Mohammed Garba

This study develops a hierarchical Bayesian framework for dynamic panel data models to cope with the challenges of heterogeneity and unequal variances regularly encountered in empirical research. Lagged dependent variables are incorporated into dynamic panel information models, which take individual precise effects and temporal dynamics into account. However, conventional estimation methods, such as the Generalized Method of Moments (GMM) and Ordinary Least Squares (OLS), often fail to accommodate heterogeneity thoroughly, main to biased and inconsistent parameter estimates. Developing a robust hierarchical Bayesian model for dynamic panel data and assessing its overall performance under various panel configurations, as determined by the scale of cross-sectional devices (N) and time durations (T), are the primary goals of this study. The framework provides great flexibility in handling complex heterogeneity and diverse panel systems by using Markov Chain Monte Carlo (MCMC) techniques to derive posterior estimates. Through substantial simulation experiments, the proposed hierarchical Bayesian estimator demonstrates superior overall performance throughout situations where $N<T$, $N=T$, and $N>T$. The method's robustness is demonstrated by the results, which show consistent decreases in numerical standard error and posterior values that closely match real parameter values. Sensitivity analyses also demonstrate the significance of prior specification, showing that notably informative priors improve estimation accuracy. This work addresses key limitations of traditional and current Bayesian methods, providing a practical framework for modeling dynamic panel data. Its applications extend to various empirical contexts, making it a treasured tool for researchers handling datasets with heterogeneity and unequal error variances.

Keywords: Dynamic Panel data, Heterogeneity, Lagged Dependent Variables, Parameter, Hierarchical Bayesian, Simulation, Covariates

A PREDICTIVE STATISTICAL FRAMEWORK FOR ASSESSING AIR POLLUTION EFFECTS ON CROP YIELD USING MACHINE LEARNING

T. E. Faniyi; Y. Yisa; H. Shehu
Federal University of Technology, Minna
Federal University of Technology, Minna
Niger State Polytechnic, Zungeru
Corresponding Author: Toyin Faniyi

This study explores the predictive power of machine learning models in assessing air quality impacts on cassava yield in Osara, Kogi State, Nigeria, using data collected from the OpenWeatherMap API between January 2023 and April 2025. Five classifiers, Random Forest, Decision Tree, Support Vector Machine, K-Nearest Neighbor, and Ordinal Logistic Regression, were trained on pollutant indicators such as PM?.?, NO?, NH?, and others. The Random Forest and Decision Tree models achieved outstanding performance, with 100% classification accuracy, precision, and recall, even with reduced feature sets. These results align with earlier studies emphasizing the importance of ensemble and rule-based learners for environmental modeling. Feature importance analysis revealed that NH? and PM?.? are the most influential predictors of air quality-induced yield stress in cassava. The models' robustness was confirmed through 5-fold cross-validation and normalized confusion matrices. These findings support the integration of machine learning into environmental early warning systems for agriculture, especially in data-scarce, rural communities. This study provides evidence-based guidance for deploying low-cost, sensor-based forecasting tools to safeguard cassava production and inform agricultural policy and intervention planning in pollution-prone regions.

Keywords: Airquality, Air-pollutant, Machine Learning, Agricultural product, yield loss

A GENERALIZED MIRRA DISTRIBUTION WITH APPLICATIONS IN ROBUST AI ALGORITHM DEVELOPMENT

M. B. Aliyu; I. Suleman; O. A. Oyegoke; M. A. Umar
Modibbo Adama University, Yola
University of Maiduguri, Maiduguri
Osun State Polytechnic, Iree, Osun State
Bayero University, Kano
Corresponding Author: UMAR MUHAMMAD

This paper introduces a novel probability distribution, termed the Generalized Mirra Distribution (GMD), which unifies and extends the classical Mirra and Xgamma distributions. The proposed distribution is constructed to capture a wider range of data behaviors—including skewness, kurtosis, and heavy-tailedness—that are commonly encountered in real-world datasets, particularly in the context of artificial intelligence (AI) and machine learning. Some key statistical properties of the GMD were derived, including its probability density function, cumulative distribution function, moments, hazard and survival functions. Its parameters were estimated via the Maximum likelihood method. The flexibility of the GMD allows it to model complex data structures more effectively than its baseline counterparts. To demonstrate its practical relevance, the GMD is applied to real-world datasets typical in AI applications, such as model residuals and anomaly scores. Comparative analyses based on goodness-of-fit criteria revealed that the GMD outperforms traditional distributions in robustness and accuracy. Furthermore, the integration of the GMD into robust AI algorithm development was discussed—highlighting its role in uncertainty quantification, robust loss modeling, and synthetic data generation for underrepresented events. The results established the Generalized Mirra Distribution as a powerful and versatile statistical tool for enhancing the reliability and interpretability of modern AI systems.

Keywords: Mirra and Xgamma distributions; Robust Statistical Modeling; Artificial Intelligence (AI); Heavy-Tailed and Skewed Data

NOTES ON THE DISTRIBUTIONS OF ZERO INFLATED PROCESSES

W. B. Yahya; M. A. Umar
1Department of Statistics, University of Ilorin, Nigeria
2Department of Mathematical Sciences, Bayero University, Kano, Nigeria
Corresponding Author: Waheed Babatunde Yahya

Count data analyses are mostly performed using the Poisson model, but it cannot model some data with over- or under-dispersion. Although, its compound models arise as alternative to analyze datasets with different level of dispersion, typically, sufficient conditions are given for such models to apply. These conditions are rarely satisfied completely in some instances, especially when large number of zeros are observed. As a result, a number of work-s focused on proposing methods that modeled and give quite satisfactory fit to these datasets. In this Paper, some of these distributions were studied and applied to real-life datasets. Their Goodness-of-fit has been discussed based on some Information Criteria. These were achieved by estimating the parameters of the distributions using the real-life datasets considered. The distributions fit the datasets satisfactorily. On the basis of which it is concluded that the distributions can serve as important alternatives to real-life count data modeling especially when excess zeros are observed.

Keywords: Count data; Zero-Inflated; Poisson; Exponential-Gamma; Goodness-of-fit.

RESPONSE ENVELOPE FOR EFFICIENT MULTIVARIATE LINEAR REGRESSIONS WITH INFORMATION COMPLEXITY

A. N. Emmanuel; K. O. Olorede; W. B. Yayah
Kwara State University, Malete Nigeria
University of Ilorin, Ilorin, Nigeria
Corresponding Author: Abraham Emmanuel

This study explores Sufficient Dimension Reduction (SDR) as a method for extracting insights from high-dimensional data. The aim is to evaluate the effectiveness of the response envelope model in reducing standard errors in multivariate linear regression. A novel variant of the Information Complexity (ICOMP) criterion, called ICOMP (IFIM), is introduced to enhance the performance of the response envelopes. Benchmark datasets, such as the Berkeley Guidance Study and the cattle weight study, are used to compare the envelope model with traditional regression. Efficiency gains and optimal dimension selection using ICOMP (IFIM) are assessed against Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Results show reduced standard errors, indicating enhanced efficiency of the envelope model. In the Berkeley Guidance Study, differences in height between boys and girls at ages 12, 13, and 14 are highlighted. The cattle weight data reveals distinct treatment effects from week 12. This underscores the envelope’s model value in uncovering subtle distinctions. Overall, this research demonstrates the effectiveness of the response envelope model for dimension reduction and coefficient estimation, particularly in big data analyses.

Keywords: Multivariate linear regression; Dimension reduction; Sufficient Dimension reduction; Envelope model; Information Complexity

ON THE OLS ESTIMATION OF ZERO-ORDER INTEGRATED MACROECONOMIC TIME SERIES: A METHODOLOGICAL REVIEW

S. B. Akanni; M. K. Garba; S. A. Adeosun; K. T. Amzat
Department of Mathematical Sciences, Crescent University, Abeokuta, Nigeria
Department of Statistics, University of Ilorin, Nigeria
Department of Mathematical Sciences, Crescent University, Abeokuta, Nigeria
Department of Statistics, University of Ibadan, Nigeria
Corresponding Author: Mohammed Garba

In most real-life scenarios of empirical settings, the dynamics of related multivariate time series are usually examined using standard cointegration and non-cointegration techniques, including the Vector Error Correction Model (VECM), Autoregressive Distributed Lag (ARDL) model, and Vector Autoregressive (VAR) model. This is because most related macroeconomic time series are either stationary processes of order one {I (1)s} or mixtures of both zero-order integrated {I(0)s} and I(1)s. However, there are occasional scenarios in which related multivariate time series are zero-order integrated processes and still fulfill the key assumptions of the Ordinary Least Squares (OLS) estimators. This work establishes such a scenario by examining these key assumptions of the OLS techniques of zero-order integrated related macroeconomic time series using Monte Carlo Simulation (MCS) and real-life macroeconomic time series data. Pretests of the simulated data using time series plots and stationarity tests confirmed the stationarity of all simulated series , , and . Based on the Absolute Bias (AB) and Mean Squared Error (MSE) metrics, the regression coefficients converge to their respective true values as sample sizes increase from 50 to 1000, which shows improvement in estimation accuracy. Results from diagnostic checking revealed that the model is free from autocorrelation, heteroscedasticity, multicollinearity, and endogeneity. The study found that across all simulations, violations of key OLS assumptions are minimal, with no evidence of multicollinearity or endogeneity. These findings support the reliability of OLS for stationary macroeconomic time series, particularly at larger sample sizes.

Keywords: Zero-order integrated macroeconomic time series; Ordinary Least Squares; Monte Carlo Simulation; Autocorrelation; Heteroscedasticity

ENHANCED DIAGNOSTIC TECHNIQUE FOR DETECTING OUTLIERS IN TWO-WAY RANDOM EFFECTS PANEL DATA MODELS

A. T. Ridwanullahi; M. K. Garba; and W. B. Yahya
Department of Statistics, University of Ilorin, Nigeria
Corresponding Author: Mohammed Garba

Detecting outliers in two-way random effects panel data models is crucial due to their extensive application in fields such as economics, healthcare, and environmental studies, where undetected anomalies can distort statistical inferences and mislead decision-making. Traditional diagnostic techniques were developed for simpler models and fail to account for the dual variation across units and time, leading to inefficiencies in identifying influential observations in complex datasets. Additionally, the presence of autoregressive moving average (ARMA) processes and heteroscedastic errors further complicates the detection of outliers, as standard residual-based diagnostics become unreliable in such settings. This study incorporated the techniques of autoregressive moving average (ARMA) processes into the diagnostic framework to account for time-dependent correlations in residuals and introduced heteroscedasticity-aware influence diagnostics that adjusted for variations in error variance across units and time. With several Monte Carlo experiments, the study was able to develop an enhanced diagnostic procedure for the detection of outliers in panel data models with two-way random effects.

Keywords: Panel data; Outlier detection; two-way random effect; heteroscedasticity; Monte Carlo Experiment

THE EFFECT OF CONTRACEPTIVE USAGE AMONG MARRIED AND UNMARRIED WOMEN IN FEDERAL POLYTECHNIC, ADO EKITI.

D.F. Olorunfemi; A.O. Ilugbusi; E.O. Kayode
Department of Statistics, Federal Polytechnic Ado-Ekiti
Department of Statistics, Federal Polytechnic Ado-Ekiti
Department of Statistics, University of Ibadan
Corresponding Author: Deborah Olorunfemi

Adolescents is an effort being implemented throughout Nigeria to enhance voluntary contraceptive utilization among unmarried women. Utilizing baseline data from evaluation research, we delineated the characteristics of sexuality, fertility, and contraceptive usage among married and unmarried individuals in Nigeria. A cross-sectional baseline survey of married and unmarried women was undertaken at Federal Polytechnic Ado-Ekiti. A clustered sampling methodology was employed. We assessed the effects and drawbacks of contraceptive use in this demographic utilizing the chi-square test. Results from 211 interviewed women indicate that the majority of respondents concurred that contraception is utilized to prevent unwanted pregnancies. Hypothesis testing demonstrates that the use of contraceptives has a significant effect, and the study further corroborates this finding. The utilization of contraceptives presents disadvantages for both married and unmarried women, which are; the insertion and removal of IUDs can induce pain, and there is a possibility of displacement, which may result in depression and vaginal irritation, among other issues. Sexual and reproductive health education, encompassing counseling for gender equality, safer sex practices, and the prevention of unintended pregnancy, should be delivered to women through a synthesis of traditional and contemporary perspectives on sexuality. Research interventions to establish sexual and reproductive health education programs, encompassing contraception counseling, should be conducted.

Keywords: : Utilization of Contraceptives; Married and Unmarried women; Sexual and Reproductive health; Chi-square test

DEVELOPMENT OF A HIGH-ORDER MULTISTEP BLOCK METHOD FOR THE NUMERICAL INTEGRATION OF COUPLED SYSTEMS OF ORDINARY DIFFERENTIAL EQUATIONS

E.O. Omole; K. Edaogbogun; D.F. Olorunfemi
Department of Physical Sciences (Mathematics), Landmark University, Omu-Aran
Department of Basic Medical Science, College of Health Sciences and Technology, Ijero Ekiti
Department of Statistics, Federal Polytechnic, Ado-Ekiti
Corresponding Author: Deborah Olorunfemi

This paper presents the development and analysis of a high-order multistep block method tailored for the efficient numerical integration of coupled systems of ordinary differential equations (ODEs). The proposed method employs a combination of interpolation and collocation strategies to derive a simultaneous scheme capable of solving multiple solution components over a block of points within each integration step. This approach enables parallel computation, enhances stability properties, and improves the accuracy of the numerical solution, especially for stiff and nonlinear systems. The method is tested on a variety of systems including biological compartmental models and physical dynamical systems to validate its performance. Comparative studies against classical single-step and linear multistep methods highlight its superiority in terms of convergence rate, error reduction, and computational efficiency. Rigorous stability and consistency analysis is provided to support the theoretical underpinnings of the method. The proposed technique offers a robust and efficient tool for simulating large-scale and complex systems of differential equations arising in science and engineering.

Keywords: Multistep block method; Coupled systems of ODEs; Numerical integration; Stiff systems; High-order accuracy

APPLICATION OF GAMMA-EXPONENTIAL DISTRIBUTION IN ANALYZING LIFETIME DATA RELATED TO RELIEF DURATIONS OF PATIENTS RECEIVING ANALGESICS

K.P. Ajewole; D.F. Olorunfemi; A.O. Ilugbusi
Department of Physical Sciences, Mathematics Programme, Landmark University, Omu-Aran
Department of Statistics, Federal Polytechnic, Ado-Ekiti
Department of Statistics, Federal Polytechnic, Ado-Ekiti
Corresponding Author: Deborah Olorunfemi

When evaluating and predicting real events, the significance of statistical distributions must be considered. Numerous statistical distributions have been employed by various scholars to examine real-world data. The primary objective of this study was to utilize the newly formulated Gamma-Exponential distribution to model empirical data. The usability and performance of the Gamma-Exponential distribution were assessed by comparing it with other commonly utilized statistical distributions. The findings indicate that the Gamma-Exponential distribution surpassed the other distributions regarding adequacy and fit.

Keywords: Gamma-Exponential distribution; Analgesic; AIC; BIC; Log-Likelihood

A BAYESIAN AGE-STRUCTURED SEIR MODEL WITH VACCINATION AND CONTACT HETEROGENEITY FOR DIPHTHERIA TRANSMISSION DYNAMICS IN NIGERIA

I. Y. Ibrahim; H. M. Salihu; H. A. Murtala; A. A. Abdullahi; M. A. Abbas; A. I. Yusuf; O. D. Adeyemi
Kano Independent Research Centre Trust (KIRCT) and Bayero University Kano
Kano Independent Research Centre Trust (KIRCT)
Kano Independent Research Centre Trust (KIRCT)
Kano Independent Research Centre Trust (KIRCT)
Bayero University Kano
4Office of the Commissioner of Health, Kano State Ministry of Health, Nigeria
Federal Ministry of Health, Nigeria
Corresponding Author: Ibrahim Ibrahim

Diphtheria remains a significant public health threat in Nigeria, particularly due to suboptimal vaccination coverage and age-dependent transmission dynamics. This study presents a Bayesian age-structured Susceptible-Exposed-Infectious-Recovered (SEIR) model to explore the transmission dynamics of diphtheria, explicitly incorporating vaccination status and contact heterogeneity across age groups. Using Nigerian surveillance data and age-specific contact matrices, we calibrated the model via Markov Chain Monte Carlo (MCMC) simulations to estimate key epidemiological parameters. The Bayesian framework allows for uncertainty quantification and robust inference in the presence of noisy or incomplete data. Our results highlight the differential impact of vaccination coverage across age strata and underscore the role of school-aged children in sustaining transmission. The estimated effective reproduction number (R?) indicates ongoing transmission potential in several age groups, even under current vaccination levels. The model provides a valuable decision-support tool for public health policymakers to evaluate age-targeted interventions and optimize vaccination strategies to curb future outbreaks in Nigeria.

Keywords: Diphtheria, SEIR model, Bayesian inference, Age-structured modeling, Vaccination, Contact heterogeneity, Markov Chain Monte Carlo (MCMC), Nigeria, Infectious disease modeling, Public health.

DEVELOPMENT OF LOCALLY AND BAYESIAN D-OPTIMAL DESIGNS FOR MULTINOMIAL LOGISTIC RESPONSES

I. Y. Ibrahim; Abubakar Yahaya; Tasiu Musa; Yahaya Zakari
Bayero University Kano And Ahmadu Bello University Zaria
Ahmadu Bello University Zaria
Ahmadu Bello University Zaria
Ahmadu Bello University Zaria
Corresponding Author: Ibrahim Ibrahim

This research focuses on developing Locally and Bayesian D-optimal designs for multinomial logistic regression models, which are vital in experiments involving categorical responses, especially in clinical trials. Multinomial logistic models often depend on unknown parameters, posing challenges for experimental design. Locally D-optimal designs address this by assuming fixed parameter values, but their efficiency diminishes when parameter uncertainty is high. To overcome this, Bayesian D-optimal designs integrate prior knowledge, offering more robust solutions. Using the baseline-category logit model with non-proportional odds assumptions, this study constructs the Fisher Information Matrix (FIM) for multinomial responses with four categories. Locally D-optimal designs were developed under various model structures—including one covariate, two covariates, and interaction terms—using both grid search and the Imperialist Competitive Algorithm (ICA). These methods aimed to maximize the determinant of the FIM, thereby improving parameter estimation efficiency. The General Equivalence Theorem was employed to validate optimality, and simulation studies were conducted to assess design performance. Bayesian D-optimal designs were then constructed using prior distributions such as Normal, Skew-Normal, Student's t, and Uniform. Results showed that designs using Normal priors were more efficient and required fewer subjects than their locally optimal or non-optimal counterparts. Comparative analysis revealed that wider design spaces generally yielded more efficient designs. This research provides practical methodologies for optimizing experimental design in the presence of parameter uncertainty and limited resources. The findings highlight the advantages of incorporating prior knowledge and advanced search algorithms to enhance the precision and efficiency of categorical response models in scientific research.

Keywords: Diphtheria, SEIR model, Bayesian inference, Age-structured modeling, Vaccination, Contact heterogeneity, Markov Chain Monte Carlo (MCMC), Nigeria, Infectious disease modeling, Public health.

A NOTE ON LEHMAN TYPE-2 TOP-LEONE TYPE-2 INVERSE RAYLEIGH DISTRIBUTION AND APPLICATION TO NECK CANCER DATA

Ajayi B.; Omosigho D. O.; Adefolarin D. A.
Federal Polytechnic Ado Ekiti, Ekiti State
Federal Polytechnic Ado Ekiti, Ekiti State
Federal Polytechnic Ado Ekiti, Ekiti State.
Corresponding Author: Bamidele Ajayi

In this article we develop a new four-parameters model called the Lehman type II Top-Leone Fréchet (LT-2TLT-2IR) distribution which exhibits non-monotone hazard rate. Many models such as Lehman type II Fréchet, Type II Top-Leone Fréchet, Generalized Exponentiated Inverse Rayleigh, and Inverse Rayleigh distribution are sub models. Some of its properties including moment, reliability, moment generating function, Incomplete moments, and hazard rate are investigated. The method of maximum likelihood is proposed to estimate the model parameters. Moreover, we give the advantage of the LT-2TLT-2IR distribution by an application using two real datasets.

Keywords: moments; moment generating function; non-monotone; incomplete moments

A GENERALIZED FRÉCHET PROMOTION TIME CURE MODEL AND ITS FITNESS TO CANCER DATA

Ajayi B.; Omosigho D. O.; Adefolarin D. A.
Federal Polytechnic Ado Ekiti, Ekiti State
Federal Polytechnic Ado Ekiti, Ekiti State
Federal Polytechnic Ado Ekiti, Ekiti State.
Corresponding Author: Bamidele Ajayi

The cure models based on standard probability distributions like exponential, Lomax, Gumbel, Weibull, lognormal, Gompertz, Fréchet gamma, are often used to analyze survival data from cancer clinical trials with long-term survivors. Sometimes, the data is simple to handle for statistical analysis, and the standard cure models fit them very well, however, most often the data are complex and the standard cure models is unable to fit them reasonably well. In this article, we develop a novel generalized Fréchet promotion time cure model and illustrate its fitness to gastric cancer data by three different methods. The generalized Fréchet distribution is as simple as the generalized Exponential distribution and is not computationally intensive as the generalized beta distribution. One detailed real-life data application is provided for demonstration and comparison purposes.

Keywords: standard probability distribution; Generalized Fréchet distribution; cure model; clinical trials

INTEGRATING ROBUST NONLINEAR REGRESSION AND AI TECHNIQUES FOR PREDICTIVE ANALYTICS IN NIGERIAN AGRICULTURE

B. S. Yakura; A. H. Bala; K. A. Manju; I. Mohammed
Department of General Studies Education, Federal College of Education, Yola, Nigeria
1Department of General Studies Education, Federal College of Education, Yola, Nigeria
1Department of General Studies Education, Federal College of Education, Yola, Nigeria
1Department of General Studies Education, Federal College of Education, Yola, Nigeria
Corresponding Author: Shiwaye Bassa

Agriculture in Nigeria faces increasing risks from climate variability, pest infestations, and erratic rainfall, leading to unstable crop yields and food insecurity. Predictive models can help mitigate these risks, but data quality issues such as outliers, missing values, and nonlinear trends often undermine the performance of conventional forecasting methods. This study proposes a hybrid framework that integrates robust nonlinear regression via M-estimation with Long Short-Term Memory (LSTM) neural networks to enhance early warning systems. The robust regression layer filters outliers and captures complex trend dynamics, while the LSTM component models sequential climatic variations to predict future crop yields. When applied to multi-year datasets across Northern Nigeria, the hybrid model achieved a 27% reduction in RMSE and a 19% improvement in MAPE compared to traditional forecasting approaches. This integrated approach improves accuracy and resilience to data irregularities, offering a scalable solution for agricultural planners and policymakers. The model supports Nigeria’s ongoing efforts in digital transformation, food security, and climate-resilient agriculture.

Keywords: Robust regression; LSTM; nonlinear modeling; agricultural forecasting; climate smart agriculture

PROPORTION LOG-PRODUCT-TYPE ESTIMATOR OF FINITE POPULATION MEAN

J. O. Muili; A. Audu; A. A. Adewara
Skyline University Nigeria Kano
Usmanu Danfodiyo University, Sokoto
University of Ilorin
Corresponding Author: Jamiu Muili

The population mean is a crucial parameter in numerous statistical studies and simple random sampling offers a clear method for estimating it. This research contributes to the existing knowledge in the field of sample surveys on mean estimators by using information on an auxiliary attribute. A new and efficient estimator is proposed for estimating the finite population mean. The expressions for the bias and mean square error (MSE) of the proposed estimator were derived up to the first degree of approximation using Taylor's series method. The efficiency conditions of the proposed estimator are proven and established. Numerical comparisons demonstrate that the proposed estimator outperforms the existing estimators considered in this study.

Keywords: Auxiliary Attribute; Population Proportion; Product Estimator; Mean Square Error (MSE).

MODELING THE RELATIONSHIP BETWEEN SOCIAL DETERMINANTS AND COVID-19 OUTCOMES USING MULTIVARIATE REGRESSION IN NORTH-EASTERN REGION OF NIGERIA.

Y. M Strong; D. Jesse
Plateau State College of Health Technology Pankshin, Nigeria
Corresponding Author: STRONG YUSUF

Abstract This study presents a retrospective analysis of COVID-19 outbreak in North Eastern Nigeria between 2020 and 2022, examining the relationship between social determinants and pandemic outcomes across six states: Adamawa, Bauchi, Borno, Gombe, Taraba, and Yobe. Using secondary data from the World Health Organization (WHO), the study employs descriptive statistics, correlation analysis, and multiple linear regression, to explore geographical dispersion, transmission trends, and the effectiveness of public health interventions. Findings revealed significant regional disparities in testing, confirmed cases, and active cases, with high coefficients of variation and pronounced positive skewness, indicating inconsistent disease surveillance and unequal health access. Correlation analysis confirmed strong linear associations between confirmed cases, recoveries, and deaths. Notably, public health measures such as testing and isolation were statistically significant predictors of COVID-19 outcomes ($p < 0.001$). However, transmission modalities and disease severity were statistically uniform across states. The study concludes that social and geographic factors significantly influenced outbreak patterns and highlights the critical role of consistent testing, targeted intervention, and localized policy response in managing future pandemics. These findings provide essential insights for data-driven public health planning and resource allocation in Nigeria's vulnerable regions.

Keywords: Keywords: COVID-19, Social Determinants, North Eastern Nigeria, Regression Analysis, Geospatial Epidemiology

PRODUCTION-WEIGHTED FRACTIONAL ARIMA MODELLING OF CRUDE OIL PRICE DYNAMICS

1Afolayan, R.B., 1Garba, M. K., 1Gatta, N. F., 1Omekam, I. V., 2Akanni, S. B., 1Babatunde, M.O
1Department of Statistics, Faculty of Physical Sciences, University of Ilorin, Ilorin, Nigeria.
2Department of Mathematical Sciences, College of Natural and Applied Sciences, Crescent University, Abeokuta, Ogun State, Nigeria
Corresponding Author: RAZAQ AFOLAYAN

Crude oil prices are a critical global commodity that significantly influences other economic sectors. In Nigeria, where over 70% of the Gross Domestic Product (GDP) is derived from oil exports, fluctuations in global crude oil prices have a profound impact on the pricing of goods and services. Due to the volatility and economic importance of oil prices, numerous studies have explored their dynamic behaviour, often employing Box-Jenkins methodology (ARIMA models) to forecast future values based solely on historical patterns. However, many of these studies rely exclusively on univariate time series models, focusing on autoregressive patterns and overall trends without considering key explanatory variables that may drive price changes. This study addresses that gap by examining the effect of oil production volume on crude oil price behaviour. The crude oil price series was transformed using oil production volume, and the Hurst exponent was computed to assess the persistence and memory structure of the transformed series. The estimated Hurst exponent of 0.76 for the transformed data indicates strong long-memory characteristics, implying that shocks to the series have a persistent impact. Consequently, an Autoregressive Fractionally Integrated Moving Average (ARFIMA) model was fitted to the data. The performance of ARFIMA models with and without transformation was evaluated using Root Mean Square Error (RMSE). The findings suggest that incorporating oil production volume improves the accuracy of crude oil price modelling, particularly for long-memory processes.

Keywords: Hurst Exponent; ARFIMA; ARIMA; GDP ; and Production-Weighted

THE PERFORMANCE OF SOME EXTREME VALUES DETECTION METHODS IN MULTIVARIATE MULTIPLE LINEAR REGRESSION

A. Bolakale; F. M. Ayoade and N. F. Gatta
Department of Statistics, University of Ilorin
Corresponding Author: Nusirat Gatta

Outlier detection is a critical aspect of multivariate linear regression analysis, as the presence of outliers can significantly distort parameter estimates and reduce model reliability. This research conducted a simulation-based comparative study of various outlier detection methods in multivariate multiple regression. Samples of sizes 20, 60, and 100 were generated from a multivariate normal distribution, each contaminated with 5%, 10%, 15%, and 20% outliers. Across 1000 replications, seven techniques—DFFITS, Cook's Distance, Standardized Residuals, Leverage, Principal Component Analysis (PCA), Mahalanobis Distance, and Robust Mahalanobis Distance—were evaluated for their ability to detect outliers in dependent variables Y1, Y2, and their combination. Each method is assessed based on its sensitivity, computational efficiency, and effectiveness in identifying influential observations under various simulated data conditions. The findings revealed that no method consistently outperformed others under all conditions. Outlier detection effectiveness varied with sample size, contamination level, and data structure. Standardized Residuals proved the most stable and interpretable across general cases. While Residuals and Cook's Distance worked best at low contamination levels, their performance weakened with larger sample sizes. Robust Mahalanobis Distance showed superior robustness under high contamination, despite being overly sensitive in some instances. Mahalanobis Distance and PCA offered more balanced, conservative detection, especially when minimizing false positives or addressing multivariate complexities. The findings offer practical guidance for researchers and analysts in selecting appropriate outlier detection techniques tailored to their specific data scenarios.

Keywords: Outliers; multivariate regression; over identification; under identification; ,

COMPARISON OF PREDICTION VARIANCE PERFORMANCE FOR VARIATIONS OF RESPONSE SURFACE DESIGNS IN THE PRESENCE OF MISSING OBSERVATIONS

Yisa Yakubu; Usman Abubakar; Oluwakemi Balikis Yekeen
Department of Statistics, School of Physical Sciences, Federal University of Technology, Minna, Nigeria
Corresponding Author: Yisa Yakubu

The primary aim of the researcher when constructing an experiment is to choose a design that allows for minimum variation between the predicted and observed responses through a best fitted model. In this work, precision of estimates of predicted response vector and optimality status for full central composite and box-behnken designs (CCDs and BBDs) for 3 to 6 factors are evaluated under different center point replicates. Then, precision of estimates of model parameter vector and predicted response vector for face-centered, spherical, rotatable, and orthogonal CCD variations (FCCD, SCCD, RCCD, and OCCD, respectively) for 2 to 5 factors are evaluated when a single observation of factorial, axial, and center run is missing. Prediction variance properties and distribution of these CCD variations throughout the design space are tracked and compared using Fraction of Design Space (FDS) plots. The findings show that, under two center run replicates, a five-factor BBD is G-optimal (G-efficiency of 100%); for 3 to 5 factors, the precision of the predicted response estimates from each of the CCD variations was observed to be robust to a single missing observation of the center runs. The FDS plots show that the missing factorial run observation caused the largest damage throughout the entire design space.

Keywords: Variations of central composite designs, Prediction variance performance, Missing observations, Fraction of design space, Relative efficiency

BAYESIAN VECTOR INTEGER GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY(INGARCH) MODEL FOR ZERO-INFLATED COUNT DATA

Amali V. E., Nwaosu. S.C., Kuhe A.D., and Udomouh . E. F.
Joseph Sarwuan Tarka University Makurdi,
Corresponding Author: Victoria Ochanya/ Ojile

This study employed Bayesian time series models to capture the complex dynamics of COVID-19 data, characterized by overdispersion, zero inflation, and cross-variable dependencies. Daily COVID-19 data on new cases, death cases, and recoveries in Nigeria from March 2020 to May 2023 were obtained from the National Centre for Disease Control (NCDC). The research applies for linear and log-linear specifications of the Vector Zero-Inflated Generalized Poisson (ZIGP-INGARCH) and Vector Zero-Inflated Negative Binomial (ZINB-INGARCH) models. Model fit and adequacy were checked through residual diagnostics, AIC/BIC model selection, and MCMC convergence checks. The results show that the log-linear ZINB-INGARCH(1,1) model outperforms alternatives with COVID-19 data, with significant autoregressive and moving average effects observed in New Cases, strong overdispersion in Death cases, and notable cross-effects from Recoveries to New Cases. Zero-inflation was highest in Death Rate (?? = 0.45), reflecting frequent zero counts. The study confirms the importance of incorporating both zero-inflation and overdispersion in multivariate epidemic modeling. It contributes a flexible Bayesian framework for accurately capturing public health care dynamics, offering valuable insights for pandemic surveillance, forecasting, and policy response.

Keywords: Bayesian inference; ZIGP-INGARCH; ZINB COVID-19, multivariate time series

STATISTICAL ANALYSIS ON THE NUMBER OF TREATMENTS FOR SOME SELECTED CHILDREN ILLNESS IN NIGERIA

N.A. Alao; S.A. Suleiman; G.A. Yusuf; A.O. Adeoye
Kwara State Polytechnic ilorin
Federal Polytechnic Offa
Corresponding Author: Sikiru Suleiman

This paper examined the number of treatments on some selected common children illness across the six regional areas in Nigeria using a secondary data obtained from the Nigeria Demographic Health Survey (NDHS) 2023-2024 key indicators. Using descriptive statistics and multivariate techniques, from the analysis the results reveals that the North west region have the highest average number of treatments for all the children illness (FEVER (285.4), ARI (47.1), and DIARH (306.8)) considered under the study while the South west have the least average number for ARI (1.0) and FEVER (42.7); and South south for DIARH (24.7), respectively. Also, the average number of children illness treatment is statistically not the same across the six regions in Nigeria. Finally, the overall average number of treatments for FEVER (107.8) is the highest follow by DIARH (97.6), then ARI (14.9), respectively. This implies that Nigeria children are mostly affected and treated for FEVER illness.

Keywords: Treatment; Children; Illness; Nigeria-regions; MANOVA;

MARKOV CHAIN MODELING OF THE IMPACT OF CLIMATE CHANGE ON GROWTH OF SELECTED CROPS IN MAKURDI, NIGERIA

V. ADAH; J. A. IKUGHUR; S. C. NWAOSU; E. F. UDOUMOH
Joseph Sarwuan Tarka University, Makurdi, Nigeria
Corresponding Author: Victoria Adah

Crop growth is impacted by multiple climatic variables, such as rainfall, temperature, humidity, and solar radiation, of which rainfall and temperature are key. Modeling the impact of these key climatic variables is of paramount interest. This study discretized rainfall (mm) and temperature (°C) amounts using indicator function and combined these indicator functions with the logical “AND” and “OR” to obtain a sequence of climatic conditions, favourable and unfavourable states, deemed suitable for the growth of crops. The sequence of the climatic states was investigated for the Markov property of dependency. Subsequently, Markov chain model was fit to this sequence on the growth of yam and pearl millet crops in Markudi, Nigeria. The parameters of the Markov chain model were estimated using the maximum likelihood estimation (MLE) and Bayesian techniques. The estimated transition probabilities have a higher likelihood of unfavourable climatic states of 0.6207 and 0.5773, respectively, for the growth of the yam and pearl millet crops when the logical “AND” is used. However, the logical “OR” has higher probability of favourability of 0.7151 and 0.7619, respectively, for the growth of the yam and pearl millet crops. The Bayes factor (BF) revealed that the informative prior is a better model when compared with the non-informative priors. The log-likelihood of the Bayesian estimates with the informative prior provides a good fit than the maximum likelihood estimate (MLE). The climate change impact (CCI) indicator > 50% revealed that climate change adversely affects the growth of the yam and pearl millet crops. The mean recurrent time of 3 years for yam and 2 years for pearl millet provided the favorable period for adaptation. The study established that farmers in Makurdi should invest less in yam and pearl millet during the unfavourable period and more during the favourable period as an adaptation strategy.

Keywords: Rainfall, Temperature, indicator function, Logical connectives, Markov chain

VOLATILITY MODELLING OF BRITISH POUNDS, EURO AND NAIRA VERSUS UNITED STATE DOLLAR EXCHANGE RATE

Maruf Ariyo Raheem; Mfon Etim UDO
Abiola Ajimobi Technical University, Ibadan
University of Ibadan, Ibadan
Corresponding Author: MARUF RAHEEM

Studying volatility characterizing financial assets is an important issue of interest in risk management not only as a barometer for making informed investment decisions, but for understanding the robustness of a country economy. This paper studied the exchange rate of US-Dollar against three major currencies namely British Pound, Euro and Naira. GARCH-type models such as Standard GARCH, Integrated GARCH, Asymmetric Power ARCH and Exponential GARCH models were explored to investigate the pattern of volatility characterizing each of the currencies using generalized and student t errors distribution. Information criteria, ARCH-LM test and Ljung-Box test were applied to identify the best-fitted models, make decision for heteroskedasticity and serial correlation for all the considered models. The finding revealed that IGARCH (1,1) and EGARCH (1,2) are the best fitted models in series of US-Dollar/British Pound. In US-Dollar to Euro series, EGARCH (1,1) was found to be the best model for both student-t and generalized error distribution. For US-Dollar to Naira, EGARCH (1,2) outperformed other models with respect to the two error distributions. The forecasts performance accuracy of daily volatility was observed for the three series using forecasts plot. The results show that IGARCH (1,1) model with the generalized error distribution has stronger forecast ability compared to EGARCH (1,2) in the case of US-Dollar against British Pound. For the US-Dollar to Euro and Naira, EGARCH (1,1) and EGARCH(1,2), with the student-t innovations outperformed same candidate models with generalized error distributions. Based on the findings, we recommend that generalized error distribution be applied in predicting the future rate return volatility of the US-Dollar/British Pound exchange. And that rate return volatility of US-Dollar to Euro and Naira exchange rate of the future is better with student t error distributions

Keywords: Volatility, Foreign exchange, Heteroskedasticity, GARCH-family model, Error distribution

EVALUATING THE BEST-FITTED DISTRIBUTION FOR OVER-DISPERSED TIME-TO-COUNT DATA: A COMPARATIVE STATISTICAL MODELLING OF MURDER SUSPECTS ARRESTED IN AKWA IBOM STATE, NIGERIA

Bassey, A. E. ; RAHEEM, M.A; Ezugwu, V.O.
University of Uyo, Uyo
Abiola Ajimobi Technical University, Ibadan
University of Uyo, Uyo
Corresponding Author: MARUF RAHEEM

Many discrete distributions have been proposed for modelling crime, however commonly used discrete distribution including Geometric, Poisson and Pascal have limited applicability. To contributes towards tackling this limitation, this research is motivated by the properties of a one-parameter discrete Lindley distribution in handling over-dispersion and balances complexities in discrete data. The research models the number of suspects arrested for murder in Akwa Ibom State between 2018 and 2023 using Geometric, Pascal and a one-parameter discrete Lindley distribution, and then using the relative ?^2 goodness of fit, compares the performance of these models. The arrest data were obtained from the Akwa Ibom State Police command and Power Bi visualization software was used in visualizing the properties of the data. Aggregating the data into quarters shows that 32.03% of murder suspects were arrested at 1st quarter, 23.53% of the suspects arrested at 2nd quarter, 26.8% and 17.65% of the suspects were arrested at the 3rd and 4th quarters respectively within the period under study. The discrete Lindley distribution with rel-?^2=1.22, was found to fit the data better than the Pascal distribution with rel-?^2=1.25 and Geometric distribution with rel-?^2=1.86. Thus, number of suspects arrested for murder cases in Akwa Ibom State, follows a one-parameter discrete Lindley distribution. Apart from providing valuable insights for decision-makers, aiding in law enforcement activities, crime trend analysis and forecasting future occurrences of arrest. Utilizing the discrete Lindley distribution could lead to improved crime prevention strategies and contribute to the field of criminology

Keywords: Geometric distribution, Pascal distribution, Lindley distribution, Murder cases, Suspects

VALIDATING THE HAUSA TRANSLATIONS OF THE ACCEPTABILITY AND FEASIBILITY OF INTERVENTION MEASURES (AIM AND FIM)

MK Khalid. I. Y. Ibrahim; H. A. Murtala; H. M. Salihu. Muhammad Saluhu
Kano Independent Research Centre Trust
Corresponding Author: MUSTAPHA KHALID

This study assessed the reliability and validity of Hausa-translated versions of the Acceptability of Intervention Measure (AIM) and the Feasibility of Intervention Measure (FIM). Internal consistency was evaluated using Cronbach’s alpha, yielding values of 0.77 for AIM and 0.73 for FIM, indicating acceptable to good reliability. Confirmatory Factor Analysis (CFA) supported the hypothesized two-factor structure, with model fit indices indicating adequate fit: $?^2(19) = 78.7$, $p < 0.001$; CFI = 0.926; TLI = 0.891; RMSEA = 0.089 (90% CI: 0.069–0.110); and SRMR = 0.051. All standardized factor loadings were statistically significant ($p < 0.001$) and above the acceptable threshold of 0.50, providing evidence for construct validity. Inter-rater reliability, assessed using Cohen’s Kappa between AIM and FIM items, ranged from 0.15 to 0.31, reflecting fair to moderate agreement. These findings suggest that the Hausa-translated AIM and FIM scales are both reliable and valid tools for measuring the acceptability and feasibility of interventions among Hausa-speaking populations.

Keywords: Acceptability, Feasibility, Reliability, Validity, Confirmatory Factor Analysis, Internal Consistency,

DYNAMIC OPTIMAL TREATMENT STRATEGIES FOR CHRONIC DISEASE MANAGEMENT USING GENETIC ALGORITHM-REINFORCEMENT LEARNING IN NON-STATIONARY SETTING

R.F. Moronkola; W.B. Yahya
University of Ilorin
Corresponding Author: Roseline Moronkola

Precision medicine aims to deliver tailored treatment plans based on individual patient's characteristics, thereby enhancing healthcare outcomes. Mobile health technologies have become pivotal to this vision by enabling real-time monitoring and adaptive interventions especially for chronic diseases such as cancer, and diabetes which typically requires long time management rather than cure. The dynamic and individualized nature of chronic disease progression present significant challenges in designing optimal treatment strategies, particularly under changing conditions over time. This research addresses the limitations of current Reinforcement Learning (RL) approaches a sequence of decision rules tailored to individual patient characteristics and evolving health states. RL aims to maximize long-term patient outcomes by adapting treatment strategies over time. It often assumes patient characteristic remain stationary over time and also relies on finite treatment stages. Such assumptions often lead to suboptimal treatment strategies when modeling the complex, evolving nature of chronic disease progression. To overcome this challenge, I propose a novel methodological framework that integrates Genetic Algorithm with Reinforcement Learning (GA-RL) to estimate dynamic, personalized treatment policies, in non-stationary Markov decision process (MDP) Models, particularly over infinite horizons with both discrete and continuous actions. The objectives of this research work include (i) developing a multi-objective GA-RL model for disease classification and treatment strategy optimization; (ii) evaluating the model's performance through extensive simulation study; (iii) benchmarking the proposed method over state-of-the-art techniques; (iv) validating the model on real world clinical dataset. The expected outcome of this research is a robust, scalable algorithm which is able to generate adaptive and personalized treatment policies that evolve with patient-specific conditions over time. It is anticipated that the proposed method will demonstrate advantages in terms of sample efficiency, computational tractability, and treatment policy adaptability, thereby contributing meaningfully to the development of intelligent, patient-centered healthcare systems.

Keywords: Precision Medicine; Chronic Disease; Reinforcement Learning; Genetic Algorithm; Non-stationary MDP; Personalized treatment.

A PANEL DATA MODELING OF THE IMPACT OF SELECTED DEMOGRAPHIC INDICATORS ON HUMAN DEVELOPMENT IN NIGERIA

Aniah-Betiang Elizabeth Ishagba; David Adugh Kuhe; Tersoo Uba; Onuche Peter
Department of Statistics, Joseph Sarwuan Tarka University, Makurdi
Corresponding Author: Elizabeth Aniah-Betiang

This study examines the impact of selected demographic indicators on human development in Nigeria using panel data from 1995 to 2024. Human Development Index (HDI) serves as the dependent variable, with explanatory variables including life expectancy at birth (LEB), population growth (PG), fertility rate (FR), maternal mortality rate (MMR), urbanization rate (UR), population density (PD), female labour force participation rate (FLP), dependency ratio (DR), and youth population (YP). Data for the analysis were obtained from reputable national and international sources including the National Bureau of Statistics (NBS), World Bank, Nigeria Demographic and Health Survey (NDHSS), Harmonized National Living Standard Survey (HNLSS), and National Population Commission (NPC). The study employs descriptive statistics and normality measures, panel unit root tests using Levin-Lin-Chu and Im-Pesaran-Shin methods, Johansen Fisher panel cointegration test, Dynamic Fully Modified Ordinary Least Squares (FMOLS) estimation, Panel Vector Error Correction Model (PVECM) and Dumitrescu-Hurlin panel causality test as methods of analysis. The descriptive statistics show high variability and non-normality in the data, especially for youth population, population density, and maternal health. Unit root tests (LLC and IPS) indicate that all variables are non-stationary at levels but become stationary after first differencing, confirming they are integrated of order one (I(1)). The Johansen Fisher panel cointegration test confirms the existence of long-run stable equilibrium relationships among HDI and the demographic indicators, justifying the application of long-run estimation techniques. The Fully Modified Ordinary Least Squares (FMOLS) estimation shows that life expectancy, urbanization, female labour participation, and youth population exert significant positive effects on HDI, while fertility, population growth, maternal mortality, dependency ratio, and population density negatively affect it. The Panel Vector Error Correction Model (PVECM) further confirms a stable long-run adjustment mechanism, with about 78% of disequilibrium corrected annually. Short-run dynamics indicate that health, fertility, and labour-related factors are key determinants of HDI fluctuations. The Dumitrescu-Hurlin panel Granger causality test further reveals both unidirectional and bidirectional causal relationships. Life expectancy, youth population, and female labour force participation exhibit feedback causality with HDI, while population growth, fertility rate, maternal mortality, dependency ratio, and urbanization unidirectionally Granger-cause HDI. No significant causal relationship is found between HDI and population density in either direction. The study recommends that policymakers should prioritize health system strengthening, reproductive health services, inclusive urban planning, female economic empowerment, and youth development to enhance human development outcomes in Nigeria.

Keywords: Demographic Indicators; Human Development Index; Panel Vector Error Correction Model; Panel Granger Causality; Nigerian Geopolitical Regions.

EVALUATING THE PERFORMANCES OF AN ESTIMATOR IN SUCCESSIVE SAMPLING BY COMPARING MATCHED AND UNMATCHED SAMPLES

D. T. Ailobhio; J. A. Ikughur; S. C. Nwoasu; T. Uba
1Department of Statistics, Federal University of Lafia, Nasarawa State, Nigeria
2,3,4Department of Statistics, Joseph Sarwuan Tarka University, Makurdi, Benue State, Nigeria
Corresponding Author: Titilayo Ailobhio

Abstract Successive sampling monitors changes over time in repeated surveys or observational studies by using both matched and unmatched samples across occasions. This study comparatively analyzes the estimator's performances under two conditions, when unmatched sample is greater than matched sample ($u > u$) and when unmatched unit is less than matched units ($u < m$) in successive sampling structure, focusing on the proposed estimator of (Ailobhio et al., 2025). This study evaluates the estimator's efficiency in measuring finite population parameter under the two conditions judging from their Mean Square Error (MSE) and coefficient of variation (CV). Using a synthetic population and a real dataset, results shows the estimators performances under this two conditions, with the estimators being more prosperous when unmatched sample is greater than matched sample ($u > m$) and the estimator by Ailobhio et al. 2025, showing competitive efficiency, outperforming most of the study estimators in most sample scenarios with increases efficiency at large sample sizes. These findings explain the estimator's robustness under different sample configurations, providing empirical insights and methodological improvements for survey practitioners.

Keywords: Matched and Unmatched Sample; Successive Sample; Mean Square Error; Coefficients of Variation

SURVIVAL ANALYSIS OF DIABETES PATIENTS: A RETROSPECTIVE STUDY AT DELTA STATE UNIVERSITY TEACHING HOSPITAL, ABRAKA

Alaba Akinleye Obabire, Ogehenefejiro Akpodamure, Isaac Oluwaseyi Ajao, Alademomi Aladesuyi,
1,2Department of Statistics, Federal Polytechnic, Orogun, Delta State, 3,4Department of Statistics, Federal Polytechnic, Ado-Ekiti, Ekiti State
Corresponding Author: Alaba Obabire

This study investigates the survival dynamics of diabetic patients through the application of survival analysis techniques, with a focus on identifying the prognostic influence of selected covariates. Utilizing retrospective data from 550 patients diagnosed with diabetes and treated at Delta State University Teaching Hospital (DELSUTH) between 2021 and 2024, we employed the Kaplan–Meier estimator to evaluate survival probabilities and the log-rank test to assess differences in survival distributions across demographic and clinical subgroups. To model the relationship between covariates and time-to-event outcomes, the Cox proportional hazards model was applied. Results indicate no statistically significant difference in survival between male and female patients ($p > 0.05$), although males exhibited a slightly higher hazard ratio. In contrast, the year of admission emerged as a significant predictor of survival ($p < 0.05$), with patients admitted in later years demonstrating better survival outcomes. These findings suggest temporal improvements in clinical management or healthcare delivery. The study underscores the importance of temporal factors in patient prognosis and supports the continued use of survival models in evaluating longitudinal health outcomes in chronic disease populations.

Keywords: Kaplan–Meier estimator, log-rank test, Cox proportional hazards model, Patients, survival models

COMPARATIVE ANALYSIS OF HYBRID FUZZY ARIMA MODEL WITH ADJUSTED FUZZY NUMBER (FARIMA-AFN) AND CLASSICAL ARIMA MODEL FORECASTS

M. A. Bakawu; A. B. Tarajo; A. Malik
Mai Idris Alooma Polytechnic Geidam
The Federal Polytechnic Damaturu
Mai Idris Alooma Polytechnic Geidam
Corresponding Author: Maigana Bakawu

Model performance evaluation is one of the key phase in predictive modelling analytics for ensuring clarity, efficiency and effectiveness in decision-making. In this paper, a novel hybrid fuzzy ARIMA and classical ARIMA forecasts are compared on the basis of interval widths and error measurement values. The results indicate that the predicted intervals generated by the Hybrid Fuzzy Autoregressive Integrated Moving Average Model with Adjusted Fuzzy Number (FARIMA-AFN) are significantly narrower than the ARIMA at ninety-five percent confidence intervals. In addition, the width of the forecasted intervals obtained from the FARIMA-AFN model is 0.064 which is significantly lower compared to 1.295 obtained by the ARIMA model (95% confidence intervals). Further, the FARIMA-AFN achieves minimal mean absolute error (MAE) values in both lower and upper bounds predictions. This implies the FARIMA-AFN model could provide better forecasts for small historical data than the ARIMA model's minimum data requirement for quick decision-making.

Keywords: ARIMA; FARIMA-AFN; Model performance evaluation; Predictive modelling

COMPARATIVE ANALYSIS OF MACHINE LEARNING AND HYPERBOLIC GROWTH MODELS FOR URBAN VEGETATION COVER SIMULATION IN LAGOS, NIGERIA

S. O. Oyamakin
University of Ibadan, Nigeria
Corresponding Author: Oluwafemi Oyamakin

Urban vegetation plays a crucial role in mitigating climate change effects, improving air quality, and enhancing the livability of megacities. This study presents a comprehensive comparative analysis of machine learning algorithms and hyperbolic growth models for simulating vegetation cover dynamics in Lagos, Nigeria's largest metropolitan area. We developed and evaluated five modeling approaches: Random Forest, Support Vector Machine, Gradient Boosting Machine, Neural Networks, Generalized Additive Models, and hyperbolic growth models adapted from Oyamakin et al.'s framework. Using synthetic data representing realistic Lagos environmental conditions across 20 urban zones over 15 years, we assessed model performance for five vegetation types under varying urban density scenarios. The study provides insights into optimal modeling approaches for urban vegetation management and future green infrastructure planning in rapidly growing African megacities.

Keywords: Urban vegetation, Lagos, machine learning, hyperbolic growth models, urban forestry, green infrastructure

SPARSE SUFFICIENT DIMENSION REDUCTION METHODS FOR MICROARRAYS WITH APPLICATION TO CENSORED SURVIVAL DATA

K. O. Olorede ; R. O. Oladoja; M. Babalola
Kwara State University, Nigeria
Texas Tech University, USA
KPMG Ireland
Corresponding Author: Kabir Olorede

Recent advances in biotechnology and genomics have demonstrated the potential of DNA microarray data for predicting clinical phenotypes, including tumour types, drug responses, and patient survival. While tumour classification has gained traction in statistical machine learning, less attention has been given to survival outcomes, especially under right-censored conditions. This study focuses on modelling censored survival time using gene expression data through sufficient dimension reduction (SDR) and information-theoretic feature selection. We propose a novel, unified estimation strategy that integrates regression-based SDR, information complexity measures, covariance hybridisation, and shrinkage estimation to identify sparse, interpretable predictors. These sufficient predictors retain essential phenotypic information and enable accurate prediction of survival outcomes. The method is compatible with existing SDR techniques like sliced inverse regression (SIR) and sliced average variance estimation (SAVE). This study presents a novel SSIR-CoxPH approach for high-dimensional regression in censored survival data. Its effectiveness was evaluated on a diffuse large B-cell lymphoma (DLBCL) dataset comprising 240 patients and 7,399 genes. The method effectively distinguishes risk groups and shows strong predictive performance, as confirmed by survival analysis and ROC results. It holds promises for identifying key predictive genes and improving patient survival prediction.

Keywords: Dimension reduction; Information complexity criterion (ICOMP); Survival analysis; Shrinkage sparse estimator; Sufficient dimension reduction

ASSESSING THE IMPACT OF ARTIFICIAL INTELLIGENCE ON MALARIA SURVEILLANCE IN OSUN STATE, NIGERIA: A STATISTICAL ANALYSIS OF PUBLIC HEALTH DECISION-MAKING

S. A. Ajala; A.O. Oyedepo; O.R. Apanaku; L.L. Oke ;S.A. Fabunmi ; Y.O. Oderemi.
Samson Ade Ajala, Department of Statistics, Osun State Polytechnic, Iree Oyedepo A. O., Department of Statistics, Igbajo Polytechnic, Igbajo Apanaku O. R., Department of Computer Science, Igbajo Polytechnic, Igbajo Oke, Lydia Layo, General Studies Department, Igbajo Polytechnic, Igbajo Fabunmi S. A., Department of Computer Science, Igbajo Polytechnic, Igbajo Dr. Oderemi Y. O., Department of Science Laboratory Technology, Igbajo Polytechnic, Igbajo
Corresponding Author: Adeola Ajala

The rise of Artificial Intelligence (AI) has opened new frontiers in disease surveillance and public health intervention. This study evaluates the role and effectiveness of AI in malaria surveillance across Osun State, Nigeria, with a focus on enhancing public health decision-making through data-driven approaches. By employing statistical models and data analytics tools, we examined patterns of malaria incidence, detection speed, and response efficiency before and after the integration of AIenabled surveillance systems. Data were collected from selected healthcare institutions and analyzed using timeseries analysis, logistic regression, and geospatial mapping techniques. The results demonstrate a significant improvement in early detection, spatial targeting of highrisk areas, and faster response to outbreaks. Furthermore, stakeholders reported increased confidence in health planning due to predictive insights generated by AI tools. The study concludes that AI, when integrated with statistical surveillance frameworks, can significantly transform malaria control strategies and serve as a sustainable model for digital public health innovation in Nigeria and similar contexts. Artificial Intelligence (AI), a key driver of the Fourth Industrial Revolution, offers powerful tools for enhancing public health surveillance, particularly in resource-constrained settings. AI techniques—ranging from machine learning and neural networks to natural language processing—can detect patterns in large datasets, make predictions, and automate decisions, offering new ways to track disease transmission and forecast outbreaks.Osun State, located in southwestern Nigeria, is characterized by a mix of urban and rural communities, with varying levels of access to healthcare and data infrastructure. In this study, we assess the role of AI in improving malaria surveillance in Osun State by integrating AI tools into statistical monitoring and public health decision-making frameworks. We aim to identify the impact of AI-driven models on disease detection, trend analysis, and intervention efficiency, using rigorous statistical methods to provide empirical evidence.This paper contributes to the growing body of knowledge on AI applications in public health, while offering practical policy.

Keywords: Artificial Intelligence, Malaria Surveillance, Public Health, Statistical Analysis, Osun State, Nigeria, Disease Mapping, Decision-Making, Predictive Modeling

DEVELOPING A ROBUST AND CAUSAL EXPLAINABLE AI (RC-XAI) TO ENHANCE MODEL TRANSPARENCY

B.S Rotimi;A.S Oke;A.E Omotoye;A.Adeleye
Adekunle Ajasin University, Akungba Akoko,Ondo State, Nigeria
Corresponding Author: BUNMI ROTIMI

Explainable Artificial Intelligence (XAI) has become essential for increasing the transparency and trustworthiness of complex machine learning models. However, many existing XAI techniques, such as SHAP, LIME, and Integrated Gradients, rely on post-hoc correlation-based approximations that are often sensitive to small perturbations and also fail to capture the causal relationships behind model decisions. This can lead to either misleading, unstable, or incomplete explanations. Therefore, to address these limitations, we propose a novel XAI framework named Robust and Causal Explainability (RC-XAI). This framework integrates causal inference techniques with robustness-aware attribution methods. We apply both structural causal modelling (SCM) and perturbation-resilient attribution techniques to construct stable and causally grounded explanations. We demonstrated the effectiveness of RC-XAI across several datasets and discovered that RC-XAI shows significant improvements in explanation stability, fidelity to model reasoning, and alignment with causal relationships. This framework advances the goal of making AI systems not only interpretable but also trustworthy and reliable for real-world deployment.

Keywords: Structural Causal Model;Perturbation-resilient attribution;stability;Fidelity;Casual relationships

EXPONENTIAL RATIO -TYPE ESTIMATORS FOR FINITE POPULATION MEAN USING TWO AUXILIARY VARIABLES IN TWO PHASE SAMPLE

Muhammed K. A, Adewara A. A
University of Ilorin, university of Ilorin
Corresponding Author: KAMALDEEN MUHAMMED

Abstract: This paper proposes a class of exponential ratio-type estimators for estimating the finite population mean using two auxiliary variables under a double sampling framework. The use of auxiliary information is known to improve the efficiency of estimators, especially in survey sampling. In the proposed methodology, exponential forms are developed by incorporating information from both auxiliary variables, one of which is observed in a large preliminary (first-phase) sample and the other in a smaller subsample (second-phase). Expressions for the bias and mean squared error (MSE) of the proposed estimators are derived up to the first order of approximation. The efficiency conditions under which the proposed estimators outperform existing ones are also discussed. A numerical illustration based on real data is provided to demonstrate the performance and applicability of the proposed estimators.

Keywords: Population mean, auxillary variables,Bias and mean square error.

FORECASTING MALARIA AND TYPHOID FEVER CASES USING DECOMPOSITION METHOD, HOLT-WINTER'S METHOD, AND SARIMA METHOD

Adeniyi Oyewole Ogunmola, Yunusa Namale Jibo
Department of Statistics, Faculty of Physical Sciences, Federal University Wukari
Corresponding Author: Adeniyi Ogunmola

This study investigates the forecasting accuracy of three univariate time series models, Decomposition, Winter's and Seasonal autoregressive integrated moving average (SARIMA) method to predict monthly malaria and typhoid fever cases from January 2008 to December 2024. The data obtained from Federal Medical Centre Jalingo were analyzed. The study applied time series decomposition method, Holt-Winter's method, and SARIMA method to compare their forecasting accuracy using Root Mean Square Error (RMSE) as the selection criterion. The results for malaria fever cases revealed that SARIMA (0,0,1) × (1,1,2) model outperformed Decomposition and Holt-Winter's methods with the lowest RMSE, indicating its superior accuracy in forecasting monthly malaria fever cases. The results for typhoid fever cases revealed that SARIMA (4,0,0) × (1,0,0) best fits and predict the data more than the Decomposition, and the two Holt-Winter's model. The monthly malaria fever indices indicates that the month of April to August has a higher malaria fever cases compared to the annual average. The highest index is in the month of June accounting for 92% increase over the annual average. Also, The monthly Typhoid fever indices indicates that the month of April, May and July to October has higher typhoid fever cases compared to the annual average. The highest index is in the month of August accounting for nearly 6.6% increase over the annual average.

Keywords: Malaria Fever, Typhoid Fever, Time Series, Decomposition Method, Holt-Winter Method, SARIMA Method, .

MACHINE LEARNING-DRIVEN AIR QUALITY CLASSIFICATION IN ABUJA METROPOLIS

Oladapo M. Oladoja; Elizabeth A. Oladejo; Taiwo M. Adegoke ; W.B. Yahya
Department of Mathematics and Statistics, Abiola Ajimobi Technical University, Ibadan, Nigeria
Department of Statistics, University of Ilorin, Kwara State, Nigeria
Corresponding Author: Taiwo Adegoke

Air pollutants pose a significant threat to human health worldwide, particularly in urban areas. Nine out of ten humans breathe in polluted air, of which Particulate Matter (PM2.5) is most significant. Abuja, the capital of Nigeria, lacks an effective predictive model that can provide real-time air quality forecasts, leaving residents vulnerable to exposure because there is no early warning system in place to help mitigate risks. Traditional statistical approaches are not able to process the data in a convenient and flexible way, especially when it comes to high-dimensional data. Recent advances in artificial intelligence and machine learning have shown promising results in predicting air pollution levels. For this study, the selection of features, training, and assessment of different types of models with the usage of eight classification algorithms namely, Multinomial Logistic Regression (MLogR), Random Forest (RF), K-Nearest Neighbors (KNN), Neural Networks (NN), Naive Bayes (NB), XGBoost (XGB), Support Vector Machine (SVM), and Decision Tree (DT), were considered. Abuja air quality data and other environmental variables are used for validation. The presented superiority of ensemble methods (RF: 77.84% accuracy, 0.7509 F1 score) over the traditional parametric ones justifies the use of advanced methods of machine learning to work with complex environmental data with non-linear associations and seasonality. The determination of the best hyperparameter values (especially the mtry=2, the 5-neuron architecture with the 1e-04 regularisation and c=1.0 with sigma=0.3253 value of hyperparameters in RF, NN, and SVM) allows the practitioner who plans to incorporate a similar solution to air quality monitoring to make adequate recommendations. This study can be added to a list of works devoted to the use of machine learning in environmental science, proving that innovative approaches to the development of ensemble techniques are instrumental in modelling intricate meteorological-pollutant relationships. The study gives the quantitative data on the extreme seasonal fluctuation of air quality, where the concentration of PM2.5 during the dry season is 4-5 times higher than in the rainy season. Researchers must continue studying patterns on a sub-seasonal scale, such as diurnal patterns, weekly and seasonal patterns, and daily seasonal transitions.

Keywords: Multinomial Logistic Regression; Random Forest; Particulate Matter; Environmental Data; Air Pollution.

YIELD PREDICTION IN YAMS USING REDUCED-DIMENSION MACHINE LEARNING FRAMEWORKS

Taiwo M. Adegoke; David O. Oyeleke; Oladapo M. Oladoja; W.B. Yahya
Department of Mathematics and Statistics, Abiola Ajimobi Technical University, Ibadan, Nigeria
Department of Statistics, University of Ilorin, Kwara State, Nigeria
Corresponding Author: Taiwo Adegoke

Yam (Dioscorea spp.) is one of the most important staple foods in West Africa which is crucial to food security. Yam yield is difficult to predict because of many complicated factors in the genetic and environmental factor. This study evaluates the effectiveness of the Principal Component Analysis (PCA) to enhance the usage of the machine learning in the prediction of yield. These assessments were based on 4 models (Linear Regression, Random Forest, Support Vector Regression (SVR) and XGBoost) with and without the use of PCA on 720 phenotypic datasets collected from International Institute of Tropical Agriculture (IITA), Ibadan. It was found that both XGBoost model and Random Forest model improved significantly with the application of PCA, with a maximum test R2 of 0.706 and lowest RMSE of 0.762 of the Random Forest model, which is therefore the best performing model in this study. SVR and Linear Regression had very small benefits of PCA. Such results indicate the potential of PCA followed by ensemble learning techniques, especially Random Forest, to closely predict yields as an additive value to the inconveniences of conducting an entire breeding cycle or conducting a large-scale test.

Keywords: Yam yield prediction, Principal Component Analysis; Machine Learning; Random Forest; XGBoost, Feature transformation; Agricultural analytics

EXCHANGE RATE VOLATILITY AND POLICY TRANSITIONS IN NIGERIA (1960–2025): A HISTORICAL AND EMPIRICAL ANALYSIS.

Ayokunle Oluwatosin Ilugbusi;Isaac Oluwaseyi Ajao;Alaba Akinleye Obabire 1,2Department of Statistics, Federal Polytechnic, Ado-Ekiti, Ekiti State 3Department of Statistics, Federal Polytechnic, Orogun, Delta State. 3 ilugbusi_ao@fedpolyado.edu.ng, ajao_io@fedpolyado.edu.ng, obabire.akinleye@fepo.edu.ng
Federal Polytechnic Ado Ekiti, Ekiti State.
Corresponding Author: AYOKUNLE ILUGBUSI

Abstract: This study investigates the historical trajectory and volatility of Nigeria's exchange rate from 1960 to 2025, aiming to uncover structural patterns, policy impacts, and macroeconomic implications. Exchange rate trends serve as indicators of economic stability and policy direction, especially in developing economies reliant on global trade. Using annual exchange rate data, time series techniques and volatility models were applied to examine percentage changes and regime shifts over six decades. Findings reveal three key phases: a stable fixed-parity regime (1960–1985); a volatile transition period (1986–1998) influenced by the Structural Adjustment Program and repeated devaluations; and an extended era of volatility (1999–2025), marked by external shocks such as the 2008 global financial crisis, the 2016 recession, COVID-19, and the 2023–2024 currency unification policy. The exchange rate surged from ?0.71/$ in 1960 to over ?1,560/$ in 2025, with a cumulative change exceeding 200,000%, reflecting severe currency depreciation. Volatility clustering was particularly evident during periods of policy uncertainty and economic shocks, including a 321% spike in 1999 and an unprecedented ?845 jump in 2024. These fluctuations complicate inflation control, deter foreign investment, and undermine long-term economic planning. The study underscores the need for credible exchange rate management, consistent monetary policy, and economic diversification. By incorporating advanced volatility models such as GARCH, this research contributes empirical insights into exchange rate behavior in resource-dependent economies and advocates for deeper exploration of structural factors driving currency instability in Nigeria's financial system. Keywords: Exchange Rate Volatility, Nigerian Naira, Monetary Policy, Structural Adjustment Program (SAP), Currency Devaluation, Time Series Analysis

Keywords: Keywords: Exchange Rate Volatility, Nigerian Naira, Monetary Policy, Structural Adjustment Program (SAP), Currency Devaluation, Time Series Analysis

EXPONENTIAL-GAMMA-RAYLEIGH DISTRIBUTION OF EG-X FAMILY OF DISTRIBUTIONS WITH APPLICATIONS TO ENVIRONMENTAL DATA.

Agbona A. A.; Alabi P.; Alade F. A.; Mubarak S.;Abdulazeez I. A.
Department of Statistics Federal Polytechnic Ede, Osun State
Corresponding Author: Anthony Agbona

This paper introduces a new distribution called Exponential_Gamma_Rayleigh distribution (EGRD) derived from Exponential_gamma – X family of distributions and aims to evaluating the performance of the new distribution with selected comparators distributions (models) using environmental data on exceedance flood and rainfall. Statistical properties of EGRD like moments, moment generating function, reliability function, skewness and kurtosis are derived and investigated. Parameters of the new distribution are obtained using the method of Maximum likelihood estimates. We compared the fits of the new EGRD with some existing comparators models of Exponential-gamma developed by Ogunwale et al (2019), Rayleigh, Gamma and Exponential distributions using best fit selection criteria of the log-likelihood function, Akaike information criterion (AIC) and Bayesian information criterion (BIC) It found that the new Exponential –gamma-Rayleigh distribution performed better than the comparators distributions in terms of flexibility, adaptability, and precision in analyzing data on environment issues .

Keywords: : Exponential_Gamma-X family; model selection criteria;Moments; Moments generating function; Reliability function; Maximum likelihood

CAN GROK AI GENERATE AND PERFORM POLITICAL DATA ANALYTICS EFFECTIVELY? EVIDENCE FROM AN EMPIRICAL STUDY

P. C. Okwumabua; S. A. Ogunmodede; O. O. Bamigbala; H. A. Ikedinma
Department of Political Science Obafemi Awolowo University, Ile-Ife
Department of Political Science University of Ilesa
Department of Peace Studies and Conflict Resolution University of Ilesa
Department of Political Science Obafemi Awolowo University, Ile-Ife
Corresponding Author: Paul Okwumabua

This study investigates the performance of Grok AI (version 3), a large language model (LLM) developed by Elon Musk's X AI, in generating and performing political data analytics. With the increasing integration of artificial intelligence into political science research, the ability of AI tools to generate, analyze, and interpret political datasets has become a compelling and innovative area of inquiry. Using an empirical approach, the study examines whether Grok AI founded in 2023 can effectively generate structured political data from textual prompts and conduct analytical tasks such as descriptive statistics, trend identification, and comparative analysis. The research draws on a case study involving the deaths of African presidents and prime ministers between 1960 and 2025. Grok AI was used to extract and structure data from online reports and news sources. A total of 36 leaders from 29 African countries were analyzed based on variables such as place, cause, and age at death, as well as whether death occurred in office or abroad. Key findings include that 56% (20 out of 36) of the leaders died while in office, and approximately 70% of these individuals were between 60 and 83 years old. Notably, 58% of the leaders also died while receiving medical treatment abroad, with 62% of these deaths occurring in Europe. Furthermore, analytics assisted by Grok AI revealed that 53% of the cases lacked an officially disclosed cause of death, highlighting persistent secrecy surrounding the health of African leaders. The results suggest that while Grok AI can effectively generate, summarize, and analyze politically relevant datasets with reasonable accuracy and speed, it falls short in data visualization and requires cross-checking due to occasional factual errors. Despite these limitations, the study highlights the potential of conversational AI like Grok to drive innovation in political research methodology and enhance statistical literacy in AI applications.

Keywords: Grok AI, Conversational AI, Analytics, Datasets, Political Data