MODELLING THE PREVALENCE OF CHILDHOOD OBESITY USING THE INAR MODEL

KIPRUTO KIPLIMO EZRA - SCM223-0705/2020 RUGA

KIMANI EDWIN - SCM223-0092/2020 LINDA

MUYOKA LUPAO - SCM223-0144/2020 VICTOR

MUUO MUSYOKA - SCM223-0113/2020

A research project submitted to the Department of Statistics and Actuarial Science in the school of Mathematics and Physical Sciences in partial fulfilment for the degree of BSc.Statistics at Jomo Kenyatta University of Agriculture and Technology

DECLARATION

This research is our original work and has not been presented for the award of adegree in any university.

Name ………………………………………Reg.No …………………………..

Signature …………………………………..Date ………………………………

Name ………………………………………Reg.No …………………………..

Signature ………………………………….. Date ………………………………

Name ……………………………………….Reg.No …………………………..

Signature ………………………………….. Date ………………………………

Name ……………………………………….Reg.No …………………………..

Signature ………………………………….. Date ………………………………

The research proposal has been submitted with my approval as universitysupervisor

Name ……………………………………….

Signature ………………………………….. Date ………………………………

ACKNOWLEDGEMENT

We thank God for His grace and mercies throughout this study. He has been a loving and faithful God. Our sincerest gratitude also goes to our supervisor Dr. Walter Yodah for his guidance and support. It has been a pleasure working with him. Lastly, we want to thank our department (STACS), loving parents for theiremotional, social and financial support throughout this study. May God bless you.

TABLE OF CONTENTS

Contents

DECLARATION 2
ACKNOWLEDGEMENT 3
TABLE OF CONTENTS 3
ABSTRACT 6
CHAPTER ONE 8
1.1 Background of the Study 8
1.2 Statement of the Problem 9
1.3 Objectives 10
1.3.1 General Objective 10
1.3.2 Specific Objectives 10
1.4 Research Question 10
1.5 Assumptions 10a
1.6 Justification of the Study 10
1.7 Scope of the study 11
1.8 Limitations of the study 11
CHAPTER TWO 12
LITERATURE REVIEW 12
2.1 Introduction 12
2.2 Research Gap 18
CHAPTER THREE: METHODOLOGY 18
3.1 Introduction 19
3.2 Data collection 19
3.4 Sampling design 19
3.5 Model 20
3.6 Properties of the model 20
3.7 Parameter Estimation using Conditional Least Squares 21
3.8 Diagnostic model 22
3.9 Forecasting 22
3.10 Bayesian Forecasting Method 22
CHAPTER FOUR: RESEARCH RESULTS AND DISCUSSSIONS 24
4.1 Introduction 24
4.2 Descriptive statistics 24
4.3 Fitting INAR (1) Model 29
4.4 Properties of the Model 31

Diagnostic plot 32
Residual autocorrelation..33
Chi-square goodness of fit test..34
4.5 Forecasting..36
CHAPTER FIVE 40
SUMMARY, CONCLUSION AND RECOMMENDATIONS 40
5.1 Summary...41
5.2 Conclusions...42
5.3 Recommendations...42
APPENDICES...43
REFERENCES...47

LIST OF ABBREVIATIONS

INAR	Integer-valued Autoregressive
WHO	World Health Organization
CDC	Centers for Disease Control
BMI	Body Mass Index
KDHS	Kenya Demographic and Health Survey
SMB's	School Management Boards
OR	Odds Ratio
CI	Confidence Interval
NDC	Non-communicable diseases
SSA	Sub-Sahara Africa
CLS	Conditional Least Squares
ARMS	Adaptive Rejection Metropolis Samplin

ABSTRACT

The purpose of this study is to model and predict the prevalence of childhood obesity based on annual count time series data for the period 2001 to 2022. The study has used INAR process to model the count data. The PACF was plotted in order to determine the order of the count time series data hence First-order Integer- valued Autoregressive (INAR (1)) was used. The model was fitted and estimated parameters of mu and alpha were obtained. Diagnostic plot, residual analysis and chi-square test of goodness of fit were used to determine the properties of the fitted model. The results indicated that, the residuals followed a normal distribution and were uncorrelated hence the model was a good fit. Further, Bayesian forecasting method has been employed to forecast childhood obesity for the next five years. The predicted count of obese children shows an upward trend from 3914 to 4080 over the forecasted period of 5 years, indicating an expected rise. The model's performance was reasonably good with a Mean Absolute Percentage Error (MAPE)of 8.33.

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

Obesity is a health condition where an individual has abnormal or excessive fat accumulation in the adipose tissue which may impair a person’s health (World Health Organization, 1997). Childhood obesity is also defined as a Body Mass Index at or above the 95th percentile for children of the same sex and age according to the Centres for Disease Control, CDC (2009). To assess a child’s body weight, the BMI of the child has to be measured. BMI (Body Mass Index) is calculated using a child’s weight and height. A child’s classification of body weight is then established based on age and sex percentiles.

Childhood obesity has become a global pandemic in developed countries, with an alarming increase in prevalence over the past few decades. This could be attributedto the changes in lifestyle, for instance, many children seem to have shifted from being physically active to a sedentary lifestyle, with many changes in their diet. These changes have disproportionately affected children, making them vulnerable to the adverse health effects associated with obesity, such as cardiovascular diseases, diabetes, and social stigmatization.

However, despite the recognized importance of addressing childhood obesity, there is a scarcity of comprehensive studies that focus on its prevalence. This research aims to bridge this gap by employing a statistical modelling technique to analyse the prevalence of childhood obesity.

The findings from this study are expected to provide valuable insights for policymakers, healthcare professionals, and community stakeholders. Moreover,the research may serve as a model for similar studies in other countries facing comparable challenges, fostering a broader understanding of the global epidemicof childhood obesity.

1.2 Statement of the Problem

Despite the increasing awareness of childhood obesity in urban areas, there is a lack of comprehensive understanding regarding the simultaneous influence on the prevalence of childhood obesity. This knowledge gap hinders the development of targeted interventions and evidence-based policies to effectively combat childhood obesity in diverse environments. This research attempts to investigate commonness of childhood obesity in the US considering count time series data from 2001 to 2022. The fitting and validation of the INAR model has been conducted and the model has been used to forecast the prevalence of childhood obesity for the next five years (2023-2027).

1.3 Objectives

1.3.1 General Objective

The general objective of this study is to model the prevalence of childhood obesity.

1.3.2 Specific Objectives

1. Fitting INAR model on the prevalence of childhood obesity among children.
2. Properties of the fitted INAR model.
3. Forecasting the prevalence of childhood obesity among children.

1.4 Research Question

What is the current prevalence rate of childhood obesity among children?

1.5 Assumptions

1. We assume that observations are independent of each other. Implying that the weight status of one child does not influence the weight status of another child in the sample.

1.6 Justification of the Study

The findings of this study can be useful to different stakeholders such as;

The Ministry of Health, Ministry of Education and School Management Boards (SMBs) can offer insights on the level of childhood obesity prevalence.

Quality Assurance and Standards Officers and Health Officers may use the findings of the study to create awareness in general on how the schools can be involved in assessing and monitoring children's weight and how to reduce obesity among children through involving parents in the awareness creation on childhood obesity.

Understanding the prevalence of childhood obesity helps in designing targeted interventions. Policymakers can use the study's findings to develop and implement policies that promote healthier environments in urban centres.

1.7 Scope of the study

The scope of study of prevalence of childhood obesity is a topic of interest in many countries. Our study mainly focuses on children in the US at the ages of 5-12 yearsof age using the US population in a given time period. The majority of the conducted studies were cross-sectional surveys and the majority focused on both male and female participants and we are going to have a study on children obesity in the ages stated above.

1.8 Limitations of the study

a) Data Source: Using a dataset from an external source (e.g. PubMed) might limit the generalizability of findings, as it represents a specific population or might lack diversity, potentially affecting the applicability of results to broader demographics.

b) Dataset Completeness and Quality: Inherent limitations in the dataset, such as missing or incomplete data, inaccuracies, or biases, might affect the reliability and validity of the analysis and subsequent conclusions.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter presents a review of related literature and studies concerning the prevalence of childhood obesity between the ages of 5-12 years of age in the US.

With the current statistics globally 39 million children under the age of 5 were overweight or obese in 2020. Over 340 million children and adolescents aged 5-19 were overweight or obese in 2016.

The 2019 data from the National Health and Nutrition Examination Survey showthat the prevalence of obesity among US children and adolescents was 18.5% in 2015-2016. Overall, the prevalence of obesity among adolescents (12-19 years; 20.6%) and school-aged children (6-11 years; 18.4%) was higher than among preschool-aged children (2-5 years; 13.9%). School-aged boys (20.4%) had a higher prevalence of obesity than preschool-aged boys (14.3%). Adolescent girls(20.9%) had a higher prevalence of obesity than preschool-aged girls (13.5%; Moreover, the rates of obesity have been steadily rising from 1999-2000 through2015-2016.

According to Ahmad et al, twenty-five percent of children in the US are overweight and 11% are obese. About 70% of obese adolescents grow up to become obese adults; 80% of adolescents aged 10 to 14 years, 25% of children younger than the age of 5 years, and 50% of children aged 6 to 9 years with obesity are at risk of remaining adults with obesity.

From the US National Health and Nutrition Examination Survey from 1999 to 2018, 35,907 children aged 2–19 with body mass index (BMI) data were included.Obesity and severe obesity were defined as BMI ≥95th percentile and ≥120% of 95th percentile of US Centers for Disease Control and Prevention growth charts, respectively. Trends in the prevalence of obesity and subgroup analyses accordingto socioeconomic factors and language used in the interview were analyzed.

Results***:*** The prevalence of obesity and severe obesity increased from 14.7 [95% confidence interval: 12.9–17.0]% to 19.2 [17.2–21.0]% and 3.9 [2.9–5.0]% to 6.1 [4.8–8.0]% in 1999–2018, respectively ($p = 0.001$ and $p = 0.014$, respectively). In2017–2018, the prevalence of obesity among children from Spanish-speaking households was 24.4 [22.4–27.0]%, higher than children from English-speaking households ($p = 0.027$).

Withrow D, Alter DA. The economic burden of obesity worldwide: a systematic review of the direct costs of obesity. Obese Rev. 2010. DOI: 0.1111/j.1467-789X.2009.00712.x. The author had in the last decade, the prevalence of obesity has increased significantly in populations worldwide. A less dramatic, but equallyimportant increase has been seen in our knowledge of its effects on health and theburden it places on healthcare systems. This systematic review aims to assess the current published literature on the direct costs associated with obesity. A computerized search of English language articles published between 1990 and June 2009 yielded 32 articles suitable for review. Based on these articles, obesitywas estimated to account for between 0.7% and 2.8% of a country's total healthcare expenditures. Furthermore, obese individuals were found to have medical costs that were approximately 30% greater than their normal weight peers.

According to the National Institutes of Health (NIH) childhood obesity research,nearly 1 in 5 children have obesity. Children with obesity are more likely to develop other serious health problems, including heart disease and type 2 diabetes. They are also more likely to suffer from anxiety, depression, and low self-esteem. Obesity affects children from different backgrounds differently.

About 1 in 4 Hispanic and non-Hispanic Black children have obesity. This is a challenge for parents, because addressing their child's weight often means makinglifestyle changes for the whole family.

Childhood obesity is a serious problem in the United States, putting children and adolescents at risk for poor health. Obesity prevalence among children and adolescents is still too high. For children and adolescents aged 2-19 years in 2017- 2020: The prevalence of obesity was 19.7% and affected about 14.7 million children and adolescents. Obesity prevalence was 12.7% among 2- to 5-year-olds, 20.7% among 6- to 11-year-olds, and 22.2% among 12- to 19-year-olds.

Childhood obesity is also more common among certain populations. Obesity prevalence was 26.2% among Hispanic children, 24.8% among non-Hispanic Black children, 16.6% among non-Hispanic White children, and 9.0% among non- Hispanic Asian children. Obesity-related conditions include high blood pressure, high cholesterol, type 2 diabetes, breathing problems such as asthma and sleep apnea, and joint problems. Childhood obesity is a serious health problem in the United States where 1 in 5 children and adolescents are affected. Some groups of children are more affected than others, but all children are at risk of gaining weight that is higher than what isconsidered healthy. Obesity is complex. Many factors can contribute to excess weight gain including behaviour, genetics and taking certain medications. But societal and communityfactors also matter: child care and school environment,neighbourhood design, With the current statistics globally 39 million children under the age of 5 wereoverweight or obese in 2020. Over 340 million children and adolescents aged 5-19were overweight or obese in 2016. Obesity is preventable.

Body mass index (BMI) is an anthropometric index of weight and height that is calculated by dividing a person's weight (in kilograms) by the square of their height (in meters). Because children and teens are growing, the ranges of height, weight, and BMI vary by age and sex. As a result, BMI values need to be expressed relativeto other children of the same sex and age.

The CDC Growth Charts display sex-specific BMI-for-age percentile curves and can be used to monitor the growth of children and teens aged 2-19 years.

BMI categories for children and teens are based on sex- and age-specific BMI percentiles, whereas BMI categories for adults are based on BMI only. Child BMI categories and their corresponding sex- and age-specific BMI percentiles are in the following table:

TABLE 1

BMI Category	BMI Range
Underweight	Less than the 5th percentile
Healthy Weight	5th percentile to less than the 85th percentile
Overweight	85th percentile to less than the 95th percentile
Obesity	95th percentile or greater
Severe Obesity	120% of the 95th percentile or greater OR 35 kg/m2 or greater

For example, a 10-year-old boy who is 54.5 inches tall (50th percentile for height) and weighs 96 pounds would have a BMI of 22.5 kg/m2, placing him at the 96th percentile for his age and sex, which is in the obesity category. This means that the boy's BMI is greater than the BMIs of 96% of 10-year-old boys in the reference population.

An expanded definition of severe obesity is used by the American Academy of Pediatrics (AAP):

- Class 2 Obesity: BMI ≥120% to <140% of the 95th percentile or BMI ≥35to <40 kg/m2
- Class 3 Obesity: BMI ≥140% of the 95th percentile or BMI ≥40 kg/m2

For example, a 15-year-old girl who is 63.7 inches tall (50th percentile for height) and weighs 210 pounds would have a BMI of 36.4 kg/m2, placing her at the 99th percentile for her age and sex. Her BMI is 129% of the 95th percentile, which is class 2 obesity based on the expanded definition of severe obesity.

Having a high BMI-for-age percentile is associated with clinical risk factors for cardiovascular disease, including high cholesterol and high blood pressure3, and other chronic conditions. In 2023, the AAP released the Clinical Practice Guideline for the Evaluation and Treatment of Children and Adolescents With Obesity to inform paediatric healthcare providers about the standard of care for youth with overweight and obesity and related comorbidities.

access to healthy and affordable foods and beverages, and access to safe and convenient places for physical activity affect our ability to make healthy choices.

Although variations in inclusion/exclusion criteria, reporting methods and included costs varied widely between the studies, a lack of examination of how and why theexcess costs were being accrued appeared to be a commonality between most studies. Accordingly, future studies must better explore how costs accrue among obese populations, in order to best facilitate health and social policy interventions.

2.2 Research Gap

The studies that have been reviewed have proved that despite the several empiricalpieces of evidence indicating the health implications of obesity, there is no study which has been done for children aged 5-12 years. Also in the reviewed studies there is no research which has used the INAR model to model the prevalence of childhood obesity. The reviewed studies have modelled obesity and factors contributing to obesity while in our research we are going to deal with prevalence only. Therefore, this research project has come up with better methodology to model the prevalence and forecast childhood obesity in ages of 5-12 years.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

This chapter elaborates the methodology that used to achieve the research objectives stated in chapter 1 sequentially. In this section INAR (1) model was used. In all the plots, graphics and calculations Rx64 4.3.2 statistical program was used.

3.2 Data collection

For the study, data was from a secondary online database with abstracts of medical articles, hosted by US National Library of Medicine (PubMed)

https://pubmed.ncbi.nlm.nih.gov/19990112/

3.3 Sampling design

Data used in this study is from a secondary data source. This data was sampledfrom various regions in US to be representative of the whole country. For the purposes of this research, focus was on children aged 5-12 years.

3.4 Model

For studying the prevalence of childhood obesity over time, a time series statisticalmodel of count data can provide valuable insights into trends, patterns, and potential forecasting. If such time series was modelled using ARIMA (Autoregressive Moving Average), we would get a continuous number for forecast value. The First- Order Integer-Valued Autoregressive (INAR (1)) model was used

$$a^{\circ} Z = \sum_{i=1}^{Z} B_i$$

Where B_i is a series of variable random iid, B_i independent with Z and

$\Pr(B_{i-1}) = 1 - \Pr((B_{i=0}) = a)$

Model fitting

Let Z be a non –negative integer. The process: $Z_t; t = 0,1,2, \dots$

was defined as INAR (1) if it satisfies the equation:

$Z_t = a^{\circ} Z_{t-1} + \varepsilon_t \dots \dots \dots \dots \dots 1$

The interpretation of INAR (1) model is that the process at time t that is Z_t, is theis the summation of the survivors at $t - 1$ that can survive until t with probabilityof surviving α and the objects which entered the system in the time interval $(t - 1, t)$ which denoted by ε_t.

3.5 Characteristics of the model

In this section mean, variance, auto covariance, autocorrelation, and partial autocorrelation was calculated. The autocorrelation and partial autocorrelation function can be considered as specification tools of the model. Expected functionfor model INAR (1) can be written as:

• Conditional expectation of Z_t, given

$$Z_{t-1}: E(Z_t|Z_{t-1}) = \alpha Z_{t-1} + \mu_s$$

• Unconditional expectation of Z_t:

$$E(Z_t) = a^t\ E(Z_0) + \mu_s \sum_{j=0}^{t-1} a^j$$

• Variance of Z_t

$$Var(Z_t) = a^{2t} + Var(Z_o) + (1-\alpha) \sum_{j=1}^{t} a^{2j-1}\ (Z_{t-j}) + \sigma^2{\varepsilon} \sum_{j=1}^{t} a^{2(j-1)}$$

• Auto covariance of Z_t:

$$Cov\ (Z_t, (Z_{t-k}) = a_k\ Var(Z_t)$$

• Autocorrelation of Z_t:

$$Corr\ ((Z_t, (Z_{t-k}) = a_k$$

The magnitude of autocorrelation function (ACF) decreases exponentially as the number of lags k increases. The partial autocorrelation function (PACF) was obtained Because ACF of INAR (1) looks like ACF of AR (1). Therefore, the only specification of INAR (1) has PACF significance on lag 1.

3.6 Parameter Estimation using Conditional Least Squares (CLS)Method:

The CLS method is finding parameter estimation by minimizing sum square ofdifference between $\boldsymbol{Z_t}$ and conditional expectation of $\boldsymbol{Z_t}$, given $\boldsymbol{Z_{t-1}}$ With the assumption $s_{\boldsymbol{t}}$~Poisson (λ) and given time series until $\boldsymbol{t} = \boldsymbol{n}$ we obtain that

$$\hat{\alpha} = \frac{\sum_{t=1}^{n} Z_t Z_{t-1} - \hat{\lambda} \sum_{t=1}^{n} Z_{t-1}}{\sum_{t=1}^{n} Z_{t-1}^2}$$

$$\hat{\lambda} = \frac{1}{n}\left(\sum_{t=1}^{n} Z_t - \hat{\alpha} \sum_{t=1}^{n} Z_{t-1} \right).$$

3.7 Diagnostic model

After obtaining parameter estimation of model INAR (1), we need to diagnose themodel. Residual of fitted model does not have correlation each other. Residual formodel INAR (1) is;

$$\boldsymbol{r_t} = \boldsymbol{Z_t} - a\boldsymbol{Z_{t-1}} - \boldsymbol{\lambda}.$$

If the model is adequate, then we plot standardized residual scatter around a zero horizontal level with no trends

3.8 Forecasting.

The conditional expectation concept to forecast INAR (1) model, would give us continuous number forecast value. Therefore, Bayesian forecasting method was used in order to get the non-negative integer forecast value

3.9 Bayesian Forecasting Method

The idea of Bayesian forecasting method is based on the two terms of INAR (1) model are variable random. Hence, the conjugate prior of α is $\beta(a, b)$ while the conjugate prior of λ is

$$\pi(\alpha \mid \lambda, Z) \propto \alpha^{a-1}(1-\alpha)^{b-1} \prod_{t=2}^{n} \sum_{i=0}^{M_t} \frac{\lambda^{Z_t - i}}{(Z_t - i)!} \binom{Z_{t-1}}{i} \alpha^{i} (1-\alpha)^{Z_t - i},$$

With the full conditional posterior distribution of Λ written as:

$$\pi(\lambda \mid \alpha, Z) \propto \lambda^{c-1} \exp(-(d+n-1)\lambda) \prod_{t=2}^{n} \sum_{i=0}^{M_t} \frac{\lambda^{Z_t - i}}{(Z_t - i)!} \binom{Z_{t-1}}{i} \alpha^{i} (1-\alpha)^{Z_t - i}.$$

The predictive posterior Z_{n+h} given Z_n is complicated. Hence we cannot find the solution using the standard method of Bayesian. The algorithm of Bayesian forecasting can be used to find the forecasting value. The algorithm of Bayesian forecasting written as follows;

1) Finding estimation parameter of model INAR (1)

2) Defining m as how much the iterations that will be performed and S_n as the sequences of proceeds of a collection (α, λ) in every iteration.

3) Doing Adaptive Rejection Metropolis Sampling (ARMS) in Gibbs sampling procedures.

4) Sampling μ~$Uniform$ (0, 1).

5) Finding non-negative integers.

6) Obtaining Z_{n+hi}.

The Algorithm ARMS in Gibbs Sampling written as follows:

1. Initializing m as number of iteration, S_n as the collection that contain sequences of α that has been generated, and α CLS as the estimation of α that has been obtained from CLS method.
2. Sampling a* from the probability posterior sampling.
3. Sampling μ~$Uniform$ (0, 1).
4. If $u > \pi\ (\alpha|\lambda\ Z)/\ exp\ (\alpha|\lambda\ Z)$, then a* enter $S_n, n \leftarrow n + 1$, and back to step 2. If not, then go to next step.
5. Sampling μ~$Uniform$ (0, 1)

CHAPTER FOUR

RESEARCH RESULTS AND DISCUSSSIONS

4.1 Introduction

This chapter showcases the study's results and their interpretations. It also highlights the performance metrics. This illustrates the reliability and viability of the model, ultimately leading to well-supported conclusion.

Rx64 4.3.2 statistical program has been mainly utilized to obtain these results.

4.2 Descriptive Statistics

The study uses time series count data in the period January 2001 to December 2022. Data on obese children were obtained from PubMed. The figure below shows the distribution of obese children from 2001 to 2022, the ages range from 5 years to 12years.

Table 4.1

Year	Count
2022	3872
2021	4103
2020	3833
2019	3581
2018	3518
2017	3158
2016	3276
2015	3115
2014	2813
2013	2233
2012	1704
2011	1471
2010	1268
2009	1073
2008	962
2007	796
2006	697
2005	586
2004	457
2003	327
2002	302
2001	240

Table 4.2 below shows the summary statistics of the main variables that have been included in the model. These include the mean, median, maximum, minimum quantiles and standard deviation of the annual data observed. Mean is used to locate the centre of the relative frequency distribution. Additionally, standard deviation gives the spread or dispersion in a series.

Table 4.2 Summary statistics

	Year		Count
Min	2001	Min	240.0
1st Qu	2006	1st Qu	721.8
Median	2012	Median	1587.5
Mean	2012	Mean	1972.0
3rd Qu	2017	3rd Qu	3246.5
Max	2022	Max	4103.0

The dataset on childhood obesity exhibits considerable variation in the count variable, with values ranging from a minimum of 240 to a maximum of 4103. The average count of the obese children across the entire time series is approximately 1972, with a standard deviation of around 1382. This indicates that, on average, the count tends to be around 1972, but individual counts can deviate from this mean by roughly 1382 units. The median count is 1587.5, suggesting a right-skewed distribution where the mean is pulled upwards by relatively higher count values. The interquartile range, spanning from the first quartile (721.8) to the third quartile(3246.5), further highlights the variability in the dataset. This analysis provides insights into the central tendency, variability and distribution of the count variable in the dataset, aiding in understanding its temporal patterns and trends.

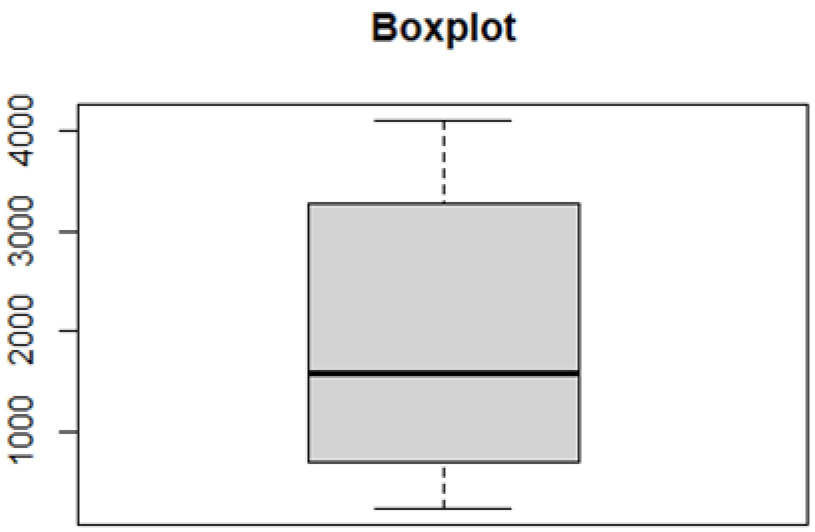

Figure 1

The boxplot shows that the number of obese children is right- skewed, with more groups having a lower number and a smaller number of groups having a higher number. The data is spread out above and below the median, and there are no outliers.

The figures 2 and 3 below shows the visualization of the event with a bar graphand a line plot respectively:

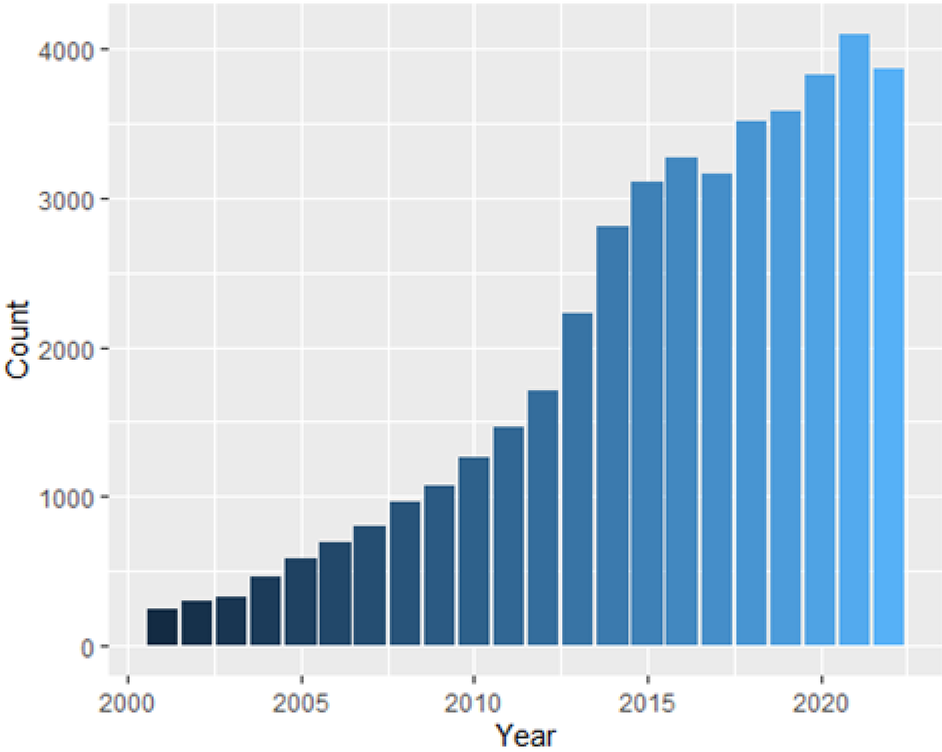

Figure 2

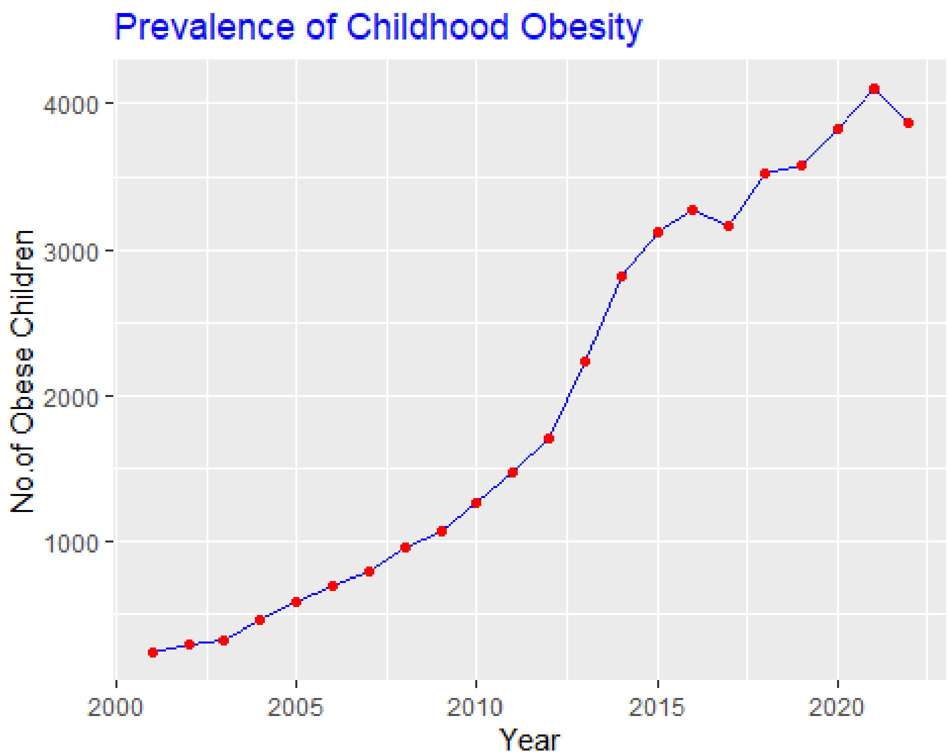

Figure 3

There is an increasing trend in the number of childhood obesity cases over thistime period. This number appears to steadily increase over the next two decades.

4.3 Fitting INAR model

Figure 4 below shows the partial autocorrelation function of the count time seriesdata.

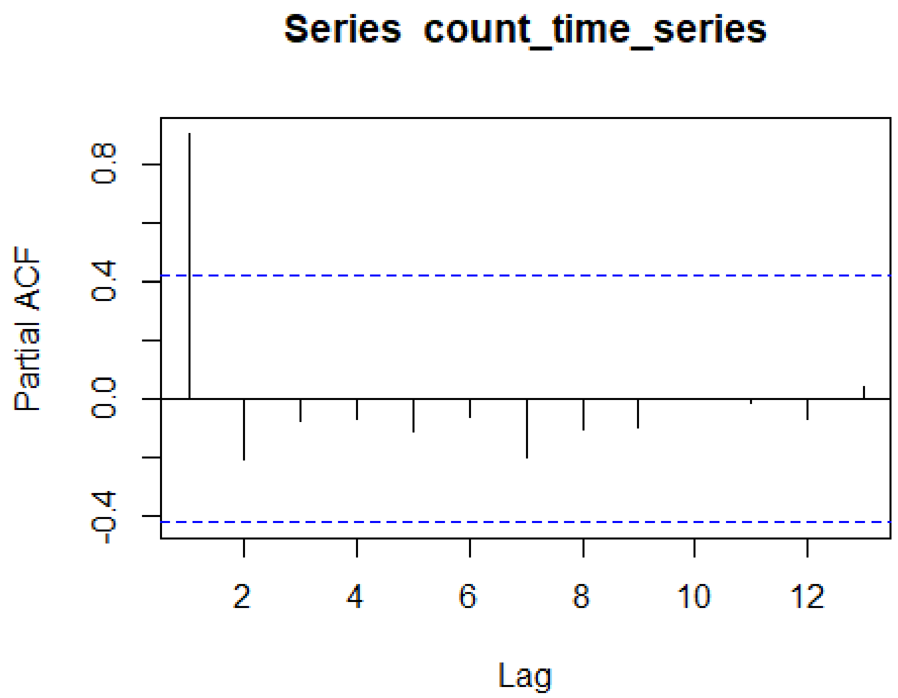

Figure 4

The spikes at lags greater than 1 appear to be within the confidence bounds (the horizontal blue lines), which means they are not statistically significant. This suggests that there is no significant correlation between the time series and its values at lags greater than 1.

The PACF plot above enables us to understand the characteristics, patterns, and dependencies present in the data. There is a significant spike at lag 1, which means there is a strong correlation between the time series and its value at lag 1(one period before). This is consistent with an INAR (1) model.

This suggests the appropriate model for fitting the count data on childhood obesity is INAR (1) model.

Table 4.3 Parameter estimates

	alpha	mu
EM estimates	0.855	124.866

The output above shows the estimated parameters of the model:

alpha: This is the autoregressive coefficient. The value is 0.855. In an INAR (1) model, alpha determines the impact of the previous count on the current count.

An alpha close to 1 indicates a strong positive dependence. The current count is heavily influenced by the previous count.

In this case, an alpha of 0.855 suggests a moderately strong positive dependence. The current count is likely to be higher if the previous count was high.

mu: This represents the estimated conditional mean of the count process. It's the average count expected for a given observation, considering the influence of the previous observation through the autoregressive parameter.

The fitted INAR (1) model suggests that the current number of obese children is positively influenced by the previous count (autoregressive effect).

4.4 Properties of the Fitted Model

Diagnostic plot.

The plot displays the relationship between predicted counts and observed (actual) counts over multiple time periods.

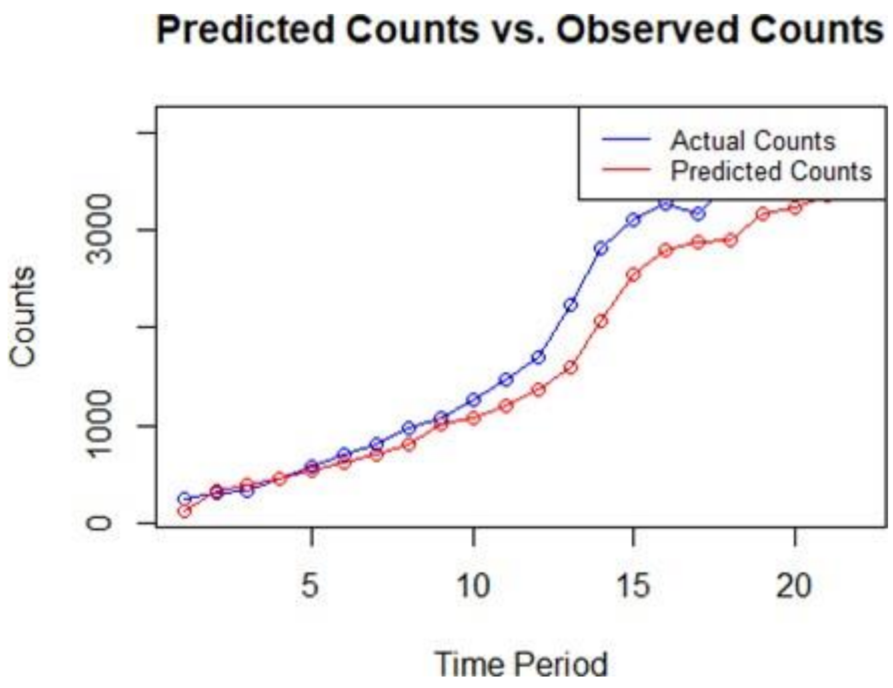

Figure 5

The plot displays the relationship between predicted counts and observed (actual) counts over multiple time periods. Both the predicted counts (represented by thered line) and the actual counts (represented by the blue line) follow an increasing trend over time. However, there are slight deviations between the predicted and observed values at certain time points.

In general, the predicted counts and actual counts show a strong positive correlation, suggesting that the predictive model used is relatively accurate in estimating the counts over the given time periods.

Autocorrelation

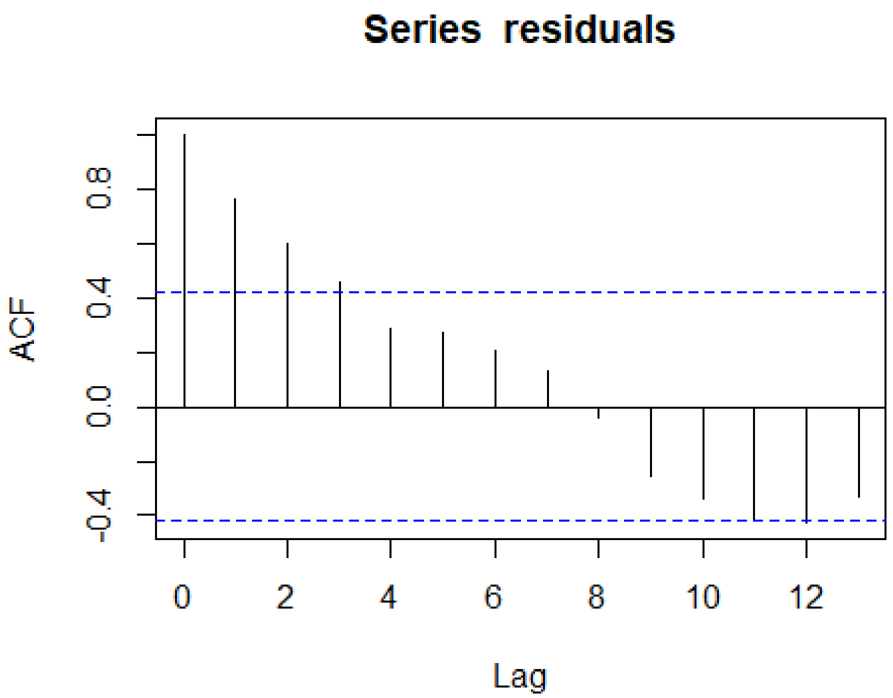

Figure 6

This plot above displays the series residuals, which represent the differencesbetween the observed values and the predicted values from a model or time series analysis, across different lags or time periods. The ACF indicates there is no significant correlation between the residuals at different lags. This suggests that themodel is capturing all of the important information in the data.

Chi-square goodness of fit test

Null hypothesis

Observed and predicted counts are not statistically different.

Alternative hypothesis

Observed and predicted counts are statistically different.

Table 4.4 Chi-square goodness of fit test

X-squared	df	p-value
462	441	02362

A high p-value > 0.05 suggests we fail to reject the null hypothesis. In this case, with a p-value of 0.2362, we don't have strong evidence to reject the null hypothesis at a significance level of 0.05..

The Chi-squared test result suggests that the observed counts and the predicted counts in the model are not statistically different at a significance level of 0.05.

This suggests the model captures the overall distribution of the data well.

The model is adequate, then the plot of standardized residual scatter around a zero horizontal level with no trends

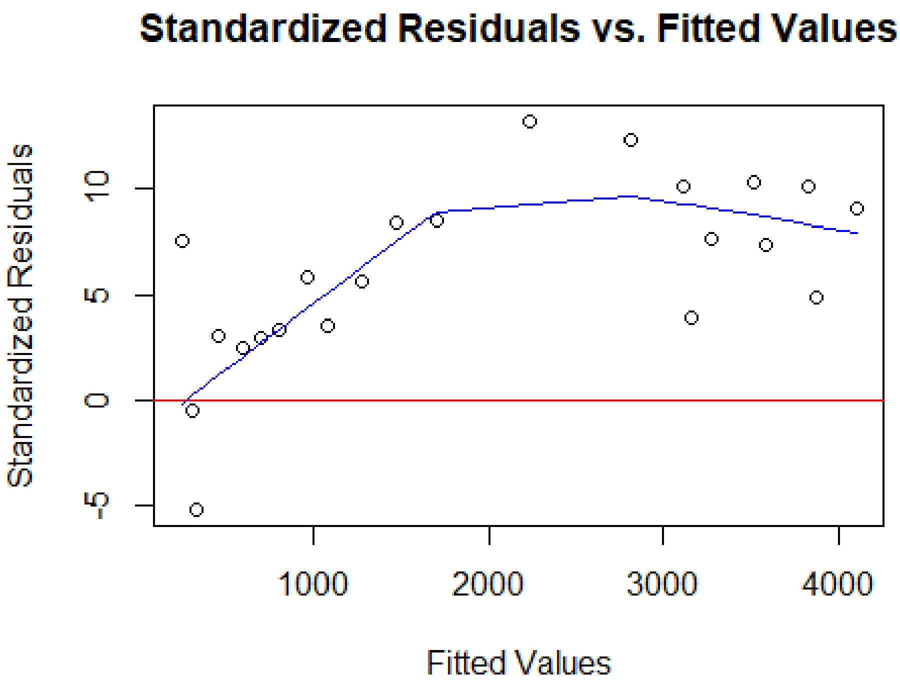
Standardized Residuals vs. Fitted Values
Standardized Residuals
-5
0
5
10
1000
2000
3000
4000
Fitted Values

Figure 7

Scatter around zero: The standardized residuals are randomly scattered around the horizontal zero line. This indicates that there's no strong evidence of systematic biasin the residuals.

No horizontal or curved bands: There are no clear horizontal or curved bands (no trend) in the plot. This suggests that the variance of the standardized residuals is likely consistent across the fitted values.

4.5 Forecasting.

To investigate the last objective, Bayesian forecasting method with MCMC simulation was used. Model parameters have been sampled based on the observed data and then uses those parameters to generate probabilistic forecasts for future counts of obese children. According to the results in table below, the forecasted number of obese children suggest an upward trend over the forecastedperiod, starting at 3914 and gradually increasing to 4080. This indicates an expected rise in number of obese children.

Table 4.7

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
23	3914	3347	4480	3047	4780
24	3955	2971	4939	2451	5460
25	3997	2560	5433	1799	6194
26	4038	2102	5974	1077	6999
27	4080	1596	6564	281	7879

Table 4.8 Forecasted values

Year	Forecast
2023	3914
2024	3955
2025	3997
2026	4038
2027	4080

Plot of forecasted count of obese children.

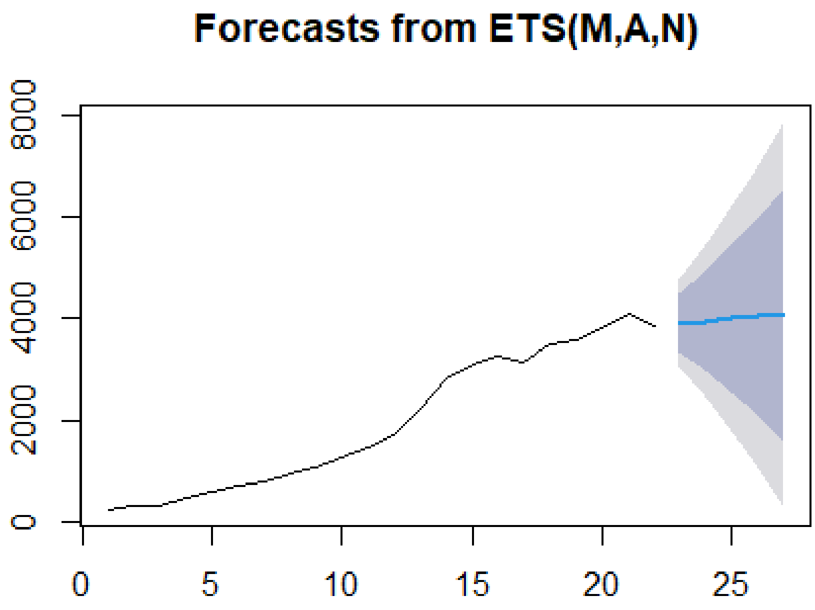

Figure 8

Figure 8 visualizes the expected future prevalence of childhood obesity based on the patterns identified in the historical time series data.

Point forecast (line): The solid line represents the point forecasts for the childhood obesity counts at future time points. These are the expected values of the counts based on the fitted INAR (1) model.

Prediction intervals (shaded area): The shaded area around the forecast line represents the prediction intervals. These intervals show the range of plausible values for the future counts with a certain level of confidence (e.g., 95% prediction interval). The wider the interval, the greater the uncertainty around the forecast. The prediction intervals provide a sense of how much the actual counts might deviatefrom the point forecasts.

The forecast line indicates the predicted trend in childhood obesity counts over thefuture time points, it suggests a predicted rise in child obesity cases.

Uncertainty: The width of the prediction intervals reflects the uncertainty associated with the forecasts.

Here are some reasons for uncertainty:

- Randomness inherent in count data (Poisson distribution)
- Limitations of the INAR (1) model (may not capture all factors affecting childhood obesity)
- Forecasting into the future (inherently uncertain)

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 Summary

This study specifically sought to model and forecast the prevalence of childhood obesity based on annual time series data for the period 2001 to 2022. Descriptive statistics offered a nuanced understanding of variable characteristics, including central tendencies and distribution features. The PACF was plotted in order to determine the order of the count time series data hence First-order Integer-valued Autoregressive (INAR (1)) was used to model the data.

The first objective was to fit INAR model on the prevalence of childhood obesity among children aged 5-12 years. The model was fitted and estimated parameters ofmu and alpha.

The second objective was to check properties of the fitted INAR model. Diagnosticplot, residual analysis and chi-square test of goodness of fit were used to determine the properties.

The last objective, was to forecast the prevalence of childhood obesity. Bayesian forecasting method was employed to for forecast childhood obesity for the next fiveyears. The predicted count of obese children shows an upward trend from 3914 to 4080 over the forecasted period of 5 years, indicating an expected rise.

5.2 Conclusions

The findings of this study clearly indicate that there is a projected increase in childhood obesity prevalence. It also suggests a potential increase in childhood obesity, highlighting a significant public health concern. In general, the findings of this study emphasize the critical need for proactive measures to address childhoodobesity.

5.3 Recommendations

- Public health initiatives: Implement and strengthen programs that promote healthy lifestyles for children. This could include encouraging physical activity, providing access to nutritious food options, and promoting healthy weight management practices.
- Further research: More in-depth studies can be to identify the specific factors driving the projected rise in childhood obesity. This could involve investigating socioeconomic factors, dietary habits, physical activity levels, and access to healthcare. Develop targeted interventions based on the identified drivers.
- The Integer-Values Autoregressive model can be employed by countries like Kenya to monitor the prevalence of childhood obesity.

Code Snippets

```
library(ZINAR1)

library(ggplot2)
library(tscount)
data<-read.csv("C:\\Users\\RAINMAKER\\Desktop\\PUBMED2.csv")
data

## Descriptive analysis
summary(data)

count_time_series<-ts(data$Count)
count_time_series

# Mean of a variable
mean(count_time_series)

# Standard deviation of a variable
sd(count_time_series)


pacf(count_time_series)


#Data visualization
# Boxplot
boxplot(count_time_series, main = "Boxplot")
#barplot
ggplot(data,aes(x=Year,y=Count,fill=Year))+geom_bar(show.legend =
FALSE,stat="identity")
#line plot
ggplot(data=data, aes(x = Year, y = Count)) +
  geom_line(color="blue")+geom_point(color="red")+
  labs(title = "Prevalence of Childhood Obesity",
       x = "Year",
       y = "No.of Obese Children")+theme(plot.title = element_text(color = "blue"))


#FITTING INAR(1) MODEL
```

```
pacf(count_time_series)
inar1_model<-EST_ZINAR(count_time_series,init = NULL,tol = 1e-05,iter =
5000,model="inar",innovation="Po",desc = FALSE)

##PROPERTIES OF THE MODEL

mu <- 124.866

alpha <- 0.855
year <- 2001:2022
counts <-
c(240,302,327,457,586,697,796,962,1073,1268,1471,1704,2233,2813,3115,3276,3158,3
518,3581,3833,4103,3872)
counts

lambda <- numeric(length(counts))
lambda[1] <- mu # Initial lambda value
for (i in 2:length(counts)) {
  lambda[i] <- mu + alpha * counts[i-1]
}
predicted_counts <- rpois(length(counts), lambda)
predicted_counts

residuals <- counts - predicted_counts
residuals

acf(residuals)
# Plot predicted counts vs. actual counts
plot(1:length(counts), counts, type = "o", col = "blue", ylim = range(c(counts,
predicted_counts)),
     xlab = "Time Period", ylab = "Counts", main = "Predicted Counts vs. Observed
Counts")
lines(1:length(predicted_counts), predicted_counts, type = "o", col = "red")
legend("topright", legend = c("Actual Counts ", "Predicted Counts"), col = c("blue",
"red"), lty = 1, cex = 0.8)

# Calculate standardized residuals
std_residuals <- residuals / sqrt(counts)

# Create a plot of standardized residuals vs. fitted values
plot(counts, std_residuals,
     xlab = "Fitted Values", ylab = "Standardized Residuals",
     main = "Standardized Residuals vs. Fitted Values")
```

```
abline(h = 0, col = "red") # Add a horizontal line at zero

# Check for trends using LOESS smoother
lines(lowess(counts, std_residuals), col = "blue")

# Chi-square goodness-of-fit test
chisq_test <- chisq.test(counts,predicted_counts)
print(chisq_test)


#Bayesian Forecasting
library(MASS)

alpha <- 0.855
m <- 1000
Sn <- matrix(NA, nrow = m, ncol = 2)
# 2 columns for mu and alpha
# Helper function to calculate the log-likelihood of the model
log_likelihood <- function(mu, alpha, counts) {
  lambda <- mu + alpha * c(0, counts[-length(counts)])
  sum(dpois(counts, lambda, log = TRUE))
}
log_likelihood(mu,alpha,counts)

## [1] -879.648
```

```
# Define proposal distribution for alpha
proposal_alpha <- function(alpha, sigma) {
rnorm(1, alpha, 0.1)
        }
        proposal_alpha(alpha,sigma)


        # Set initial values and tuning parameters
        alpha_current <- alpha
        sigma_alpha <- 0.1



        for (i in 1:m) {
          # Sample u ~ Uniform(0, 1)
          u <- runif(1)

          # Sample alpha using ARMS
          alpha_proposed <- proposal_alpha(alpha_current, sigma_alpha)
          log_ratio <- log_likelihood(mu, alpha_proposed, counts) - log_likelihood(mu,
        alpha_current, counts)

          if (log(u) < log_ratio) {
            alpha_current <- alpha_proposed
          }

          # Store parameter values
          Sn[i, ] <- c(mu, alpha_current)
        }

        future_years <- 2023:2028 # Example future years
        future_years

        ## [1] 2023 2024 2025 2026 2027 2028

        forecasts <- matrix(NA, nrow = length(future_years), ncol = 2) # Store forecasts
        colnames(forecasts) <- c("Year", "Forecasted Count")

        for (j in 1:length(future_years)) {
          lambda_forecast <- Sn[, 1] + Sn[, 2] * counts[length(counts)] # Forecasted lambda
        using last observed count
          forecasts[j, ] <- c(future_years[j], rpois(1, lambda_forecast[length(lambda_forecast)]))
        }
        print(forecasts)

        plot(forecasts)
```

REFERENCES

1. Hales CM, Carroll MD, Fryar CD, Ogden CL. *Prevalence of Obesity AmongAdults and Youth: United States, 2015-2016*. NCHS Data Brief, No. 288. Hyattsville, MD: National Center for Health Statistics; 2017

2. Ahmad QI, Ahmad CB, Ahmad SM. Childhood obesity. *Indian J Endocrinol Metab*. 2010;14:19-25.

3. Childhood obesity, prevalence and prevention. Dehghan M, Akhtar-Danesh N, Merchant AT. Nutr J. 2005;4:24.
4. G. Box and G. Jenkins. Time Series Analysis: Forecasting and Control. First edition, 1970.
5. Y. Cui and R.B. Lund. A new look at the time series of counts. Biometrika, 96:781–792, 2009.
6. P. J. Brockwell and R. A. Davis. Time Series: Theory and Methods. SpringerSeries in Statistics. Springer-Verlag, New York, second edition, 1991.
7. Obesity and insulin resistance in children. Marcovecchio ML, Mohn A,Chiarelli F. J Pediatr Gastroenterology Nutr. 2010;51 Suppl 3:0–50
8. Cross-national comparison of childhood obesity: the epidemic and the relationship between obesity and socioeconomic status. Wang Y. Int J Epidermal. 2001;30:1129–1136.

9. Obesity in India: the nation's weight. Kalra S, Unnikrishnan AG. J Med Nutr Nutraceut. 2012;1:37.
10. Journal of Time Series Analysis, 34:206–220, 2013
11. Withrow D, Alter DA. The economic burden of obesity worldwide: a systematic review of the direct costs of obesity. Obes Rev. 2010. DOI: 10.1111/j.1467-789X.2009.00712.x.