BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF STATISTICS AND MATHEMATICS

STATISTICAL ANALYSIS OF PREVALENCE OF DIABETES IN ZIMBABWE.

BY

BYBIT PANASHE MUTUDA

A DISSERTAION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF THE BACHELOR OF SCIENCE HONOURS DEGREE IN STATISTICS AND FINANCIAL MATHEMATICS

JUNE 2024

APPROVAL FORM

The undersigned certify that they have read and recommended to the Bindura University of Science Education the acceptance of a dissertation entitled “STATISTICAL ANALYSIS OF PREVALENCE OF DIABETES IN ZIMBABWE," submitted in partial fulfillment of the requirements of the Bachelor of Science (Honours) Degree in Statistics and Financial Mathematics.

STUDENT

Bybit Panashe Mutuda	…………………	07/06/2024
(B202085B)	Signature	Date

Certified by:

Ms. J.C. Pagan’a	………………………	10/06/2024.
Supervisor	Signature	Date

Dr.M. Magodora	…………………………..	
Chairperson	Signature	Date

RELEASE FORM

NAME OF STUDENT: BYBIT PANASHE MUTUDA

REGISTRATION NUMBER: B202085B

DISSERTATION TITLE: STATISTICAL ANALYSIS OF PREVALENCE OF DIABETES IN ZIMBABWE

YEAR GRANTED: 2024

I hereby granted permission for my academic dissertation, to be reproduced and made available for public access within the premises of the Bindura University of Science Education Library and Department of Statistics and Mathematics, for academic and research purposes only. I retain copyright ownership of my dissertation. Any copies produced as a result of this permission must not be sold, lent, or distributed to external parties without my prior written approval.

Signature of Student ……………………………

Date Signed **07/06/2024**

Permanent Address: House number 898

Muguta Extension

Epworth, Harare

DEDICATION

I dedicated this research project to my lovely mother and father, who always supported me in my academic pursuits and installed in me a love of learning from a young age. I want to express my gratitude to my best friend, Tinotenda Meza, for being a source of encouragement and support throughout the process. Finally, this research is dedicated to my mentor, Ms. J.C Pagan'a, who taught me the importance of hardworking and perseverance.

ACKNOLEDGEMENT

First of all, thanks go to the Almighty God, without everything being impossible. I would like to express my gratitude and appreciation for the courage and opportunity of the memorable fruitful years at Bindura University of Science Education. Through his grace and wisdom, I was able to learn and understand the education that was given to me.

My special gratification goes to my supervisor, Ms. J.C Pagan'a, for her constant encouragement, follow-up, patience, and guidance, which will assist me in my life. Without her continuous assistance and dedication throughout the entire project, the completion of this dissertation would not have been possible. Another expression of gratitude was extended to my colleagues as well, who made this learning process easier through their unwavering support and the enriching experiences we shared during this program in statistics and financial mathematics.

The author wanted to express her sincere gratitude to the department of Statistics and Mathematics at Bindura University of Science Education for providing her with the invaluable academic knowledge and abilities she desperately required to complete the course, as well as the assistance and direction provided by her academic lecturers.

Special thanks to my loving parents for the invaluable support (moral), caring, motivation, and sacrifice they have made for me during these difficult times to get to this point in my life. I would also like to thank the rest of my family and friends for the support and encouragement they gave me during this time. God bless you all.

ABSTRACT

Zimbabwe is facing a growing public health challenge due to the increasing prevalence and incidence of diabetes. Diabetes is one of the leading causes of death and complications, such as morbidity. The aim of this study was to conduct a comprehensive statistical analysis of prevalence of diabetes in Zimbabwe focusing on understanding the prevalence and risk factors of the disease. A study utilized historical data from medical records of 768 diabetic patients obtained from Chitungwiza Central Hospital in Zimbabwe. Collected data was for the period 2014 to 2018. This study employed a cross-sectional research design. The collected data was analysed to examine the prevalence of diabetes across different age groups and gender, to identify significant predictors of diabetes and quantify their respective impact. Furthermore, descriptive, inferential, correlation and logistic regression analyses were conducted, and the results presented in tables and graphs. According to the study, age, blood sugar, and BMI were the main indicators of diabetes in adult Zimbabweans. Through the use of a confusion matrix, the accuracy of the model was determined to be 76.3% and AUC curve indicated 82% accuracy. This study provides valuable insights into the statistical trends and relationships of diabetes in Zimbabwe. The study highlighted the importance of monitoring and controlling key risk factors to mitigate the growing burden of diabetes in Zimbabwe. This research recommended that healthcare providers should prioritize regular monitoring of blood sugar, BMI, and age to identify high-risk patients. Additionally, patients with diabetes should adhere to treatment plans and attend regular follow-up appointments to manage their condition effectively. Furthermore, the government should develop and implement a comprehensive national strategy with specific goals, timelines, and measurable indicators involving key stakeholders such as healthcare professionals, NGOs and patients with diabetes to address diabetes prevention, management, and control.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADA- American Diabetes Association

AUC-ROC- Area Under the Receiver Operating Characteristic

BMI- Body mass index

DALY - Disability-adjusted life year

DM- Diabetes Mellitus

CDC- Centres for Disease Control and Prevention

GBD- Global Burden of Disease

IDDM- Insulin dependent diabetes mellitus

PAHO- Pan American Health Organisation

IDF- International Diabetes Federation

DREAM-Diabetes Remission and Medical Trial

HBM- Health Belief Model

NGO- Non-Government Organisation

NPR- National Population Registry

T1DM- Type 1 Diabetes Mellitus

T2DM- Type 2 Diabetes Mellitus

VIF- Variance Inflation Factor

WHO- World Health Organisation

ZDHS- Zimbabwe Demographic and Health Survey

ZIMSTAT- Zimbabwe National Statistics Agency

ZNHS- Zimbabwe National Health Survey

CHAPTER 1

INTRODUCTION

1.0 Introduction

Diabetes is a chronic and crippling illness that has grown to be a significant public health issue in Zimbabwe, putting a significant financial strain on patients, families, and the healthcare system. Given that the nation has a higher prevalence of diabetes than the world average, creating efficient management and control plans requires a thorough grasp of the disease's epidemiology, causes, and effects. In order to give important insights that might guide evidence-based therapies and policy decisions, this project intends to provide a thorough statistical analysis of diabetes in Zimbabwe, looking at the disease's distribution, trends, and connections with different risk factors.

1.1 Background of the study

Millions of individuals worldwide suffer from diabetes mellitus (DM), a serious chronic illness whose consequences continue to be a major source of worry (WHO, 2021). Because of anomalies in how the body processes and stores nutrients like glucose, lipids, and proteins, it is regarded as a metabolic condition (ADA, 2020). Ninety-nine percent of adults with diabetes have type 2 diabetes mellitus, and five to ten percent of those with diabetes have type 1 diabetes (CDC, 2023).By 2045, 629 million more people are expected to have diabetes mellitus, with emerging nations experiencing the largest increase (IDF, 2022). The World Health Organization (2005) states that diabetes is expected to rise two to three times more prevalent than it is now in Africa and Asia, the two regions with the largest potential increase in diabetes mellitus in the future.

According to Daniel et al. (2011), there is a growing global population with diabetes, projected to affect 285–350 million individuals. According to estimates from the International Diabetes Federation (IDF), 382 million people worldwide have diabetes, translating to an 8.3% global prevalence (International Diabetes Federation, 2013). Diabetes is the eighth most common cause of death worldwide, contributing 8.4% of the world's all-cause mortality (International Diabetes Federation, 2013). Diabetes is one of the major causes of death worldwide. Pan American Health Organization (PAHO) 2023 states that in high-income nations, the premature mortality rate due to diabetes mellitus decreased from 2000 to 2010 but then increased in 2010–2016, while in lower-middle-income countries, the premature mortality rate due to diabetes mellitus increased across both periods.

Most diabetics in low- and middle-income nations are between the ages of 45 and 64 (King, H. R.E. et al., 1998). On the other hand, most diabetics in affluent nations are under 64 years old. According to estimates, there will be more than 82 million individuals with diabetes under 64 in low- and middle-income countries by 2030, and more than 48 million in high-income countries (Wild S et al., 2004). According to Bjork S et al. (2003), the age groups that are most economically productive are expected to see the most growth in the number of diabetics. According to reports, diabetes mellitus ranks five out of the ten most prevalent disorders in Zimbabwe (Mudiay TK, 1997). In Zimbabwe, the prevalence of diabetes tripled overall between 1990 and 1997, rising from 150 to 155 cases per 100,000 persons. Zimbabwe National Health Profiles (1996–1998) show that there were 5114 new cases among those aged 15 and over in 1998 compared to 2734 instances in 1996, an increase of 87% above the number of cases that were recorded (Mufunda J., 2006).

Zimbabwe currently holds the fourth-highest prevalence incidence of diabetes among African nations. According to WHO (2016), the majority of diabetes deaths occur in people under the age of 70. It is estimated that 80% of diabetics reside in low- and middle-income nations. The socially disadvantaged are the most susceptible to the disease in any given nation. By the end of 2045, there will likely be 82 million Africans living with diabetes mellitus, up from 39 million in 2017 (WHO, 2016). Now recognized as a pandemic, the illness is primarily impacting impoverished populations in sub-Saharan Africa and is spreading quickly in developing nations (Mbanya JCN, 2015; Whiting DR, 2016). As the population ages and moves from a traditional healthy and active lifestyle to a modern sedentary, stressful lifestyle and overconsumption of energy-dense foods linked to rapid urbanization and westernization, the prevalence of diabetes in African communities is rising (Hjelm K, Mufunda E, 2012).

Due to its ability to impact every organ in the body and result in a wide range of symptoms, diabetes mellitus is referred to as a silent killer disease. In the year 2014, 8.5% of persons who were 18 years of age or older were diagnosed with diabetes mellitus. In the year 2019, diabetes mellitus accounted for 1.5 million fatalities directly, with diabetes mellitus occurring before the age of 70 years accounting for 48% of all deaths (WHO, 2023). According to the Institute of Health Metrics and Evaluation (2020), diabetes mellitus also contributed to 460 thousand fatalities from kidney disease, and elevated blood glucose levels are responsible for 20% of deaths from cardiovascular disease.

Diabetes mellitus also contributes significantly to renal failure, heart attacks, strokes, impotence, difficult-to-heal wounds, lower limb amputation, and decreased vision (blindness) (WHO, 2023). According to the World Health Organization (2023), diabetes mellitus is also commonly characterized by blurred vision, fatigue, polydipsia (a lot of drinking), polyuria (frequent nighttime urination), polyphagia (a lot of eating), increased appetite but rapid weight loss of up to 5–10 kg in two to four weeks, and tiredness easily. With the highest prevalence of diabetes mellitus and the lowest ability to prevent and treat it, low- and middle-income nations bear the brunt of the disease's effects (WHO, 2021).

The current way of life and ancestral genes designed for a hunter-gatherer lifestyle collide to cause diabetes mellitus (Zimmet P., 2000). Reduction in life expectancy, elevated mortality, and morbidity linked to the emergence of costly complications are the primary effects of diabetes mellitus, and they will therefore place a significant financial strain on both individuals and society. For people with diabetes, their families, and governments, the condition and its effects entail a substantial financial and economic cost (ADA, 2017). The American Diabetes Association (2017) states that the expenses of medication, medical treatment, and missed productivity due to disease can put a heavy burden on healthcare systems and governments when it comes to managing diabetes.

Furthermore, the course of diabetes mellitus is mostly defined by the patient's self-management, which is influenced by personal and cultural views on health and illness (Hjelm K. and Mafunda

E., 2012). Blood glucose monitoring, frequent exercise, taking medicine, and eating a balanced diet are examples of self-management practices. When it comes to assisting patients in acquiring the abilities, mindsets, and information required for self-management, health practitioners bear a significant duty. Diabetes can have a variety of effects on regular social contacts. The patient has to limit the kinds and quantities of food they eat; they may also need to take medicine while they are participating in social activities and check their blood sugar levels at certain times of the day (Brezo J. et al., 2006).

In order to address this growing health issue, global leaders have committed to lowering the burden of diabetes mellitus, designating it as one of four priority non-communicable diseases (NCDs). Taking into account the global targets for non-communicable diseases for 2025 and the Sustainable Development Goals for 2030, member states have set the lofty goal of reducing premature mortality from non-communicable diseases, including diabetes mellitus, by one-third, achieving universal health coverage, and providing affordable essential medicines (1981, 2021).

1.2 Statement of the problem

The rising incidence and prevalence of diabetes present a serious public health challenge for Zimbabwe. One of the main causes of mortality and consequences, including morbidity, is diabetes. Furthermore, the economic toll that diabetes mellitus and its complications have on individuals, families, and governments is substantial. As the number of people with diabetes rises, there is a need for a deeper comprehension of the disease's effects. As a result, patients must develop the necessary attitudes, abilities, and knowledge to successfully manage their diabetes. This study conducts a thorough statistical analysis of diabetes in Zimbabwe, offering important insights that help guide healthcare policies and actions to better manage and prevent diabetes in Zimbabwe.

1.3 Aim of the study

The goal of this study was to develop a statistical model and analysis of prevalence of diabetes in Zimbabwe in order to find better methods for managing and preventing the disease.

1.4 Research objectives

The overall objectives of this study are:

- To identify the risk factors associated with the development of diabetes in Zimbabwe.
- To determine the prevalence of diabetes, examining variations across different age groups, blood sugar and BMI.
- To explore the relationship between various risk factors and the likelihood of developing diabetes.
- To develop a logistic regression model to predict the likelihood of developing diabetes in Zimbabwean adults.

1.5 Research questions

- What are the primary risk factors associated with the development of diabetes in Zimbabwe?
- How does the prevalence of diabetes vary across different age groups, blood sugar and BMI?
- Which risk factors are more strongly associated with an increased likelihood of diabetes diagnosis and how do various risk factors contribute to the likelihood of developing diabetes?
- How can a logistic model predict diabetes diagnosis based on the significant predictors?

1.6 Assumptions

Certain presumptions had to be made in order to conduct research in the first place. The presumptions that have been made are listed below:

• The research techniques are adequate and valid; the data and information used in the study were accurate and dependable.

• The study was carried out in accordance with pertinent ethical principles and laws

• The sample size selected was a genuine representative of the target population.

1.7 Significance of the study

This research was expected to benefit the students, university, and government of Zimbabwe.

Student

This research benefitted the student by bridging the gap between theory and practice by allowing the researcher to link theory with practice as well as developing research skills. Participating in this research on diabetes in Zimbabwe equipped the researcher with valuable skills, knowledge, and experience that can shape her future careers. This study provided opportunities to foster personal growth, professional growth, and self-discovery (self-confidence, responsibility, and sense of independence).

University

This study was added to the existing literature on diabetes in Zimbabwe, providing a valuable resource for future research as it advanced the understanding of The materials provided in this study are useful to other researchers who would like to carry out further studies on this topic in the future. This study contributes to the existing body of knowledge on diabetes in Zimbabwe. By addressing research gaps and new insights, it can guide future research endeavors and stimulate further investigations into specific aspects of diabetes management in Zimbabwe.

Government

This study was useful to the government when formulating policies and regulations related to diabetes mellitus management and control based on the results of the study. This research findings provide valuable insights that can help the government provide strategies that effectively address the burden of diabetes in Zimbabwe. Moreover, this study can assist the government in resource allocation, planning equitable healthcare services, and implementing some preventative measures specific needs of the people in Zimbabwe.

1.8 Scope of the study

The study aimed to determine a statistical model of prevalence of diabetes in Zimbabwe. Additionally, the analysis focused on individuals aged 21–65 who have been diagnosed with diabetes from 2014 to 2018. The statistical analysis techniques used in this study include descriptive analysis, inferential analysis, and logistic regression analysis. R Studio was used for statistical analysis and modelling of diabetes in Zimbabwe. This research used quantitative data. Additionally, a cross-sectional study was employed. The determinants used in this study are demographic factors (age, gender), lifestyle factors (BMI), and physiological or metabolic factors (blood sugar, blood pressure, cholesterol).

1.9 Limitations of the study

Some obstacles and restrictions came up while the research was being conducted. The following are some limitations of the research:

- One of the challenges the researcher faced was the study's restricted budget. She was forced to pay for stationery and internet services out of her own pocket.

- The hospital was hesitant to give the researcher access to some material that they regarded as private and sensitive, but the researcher managed the study's limited funding wisely and made sure it was successful. The researcher did, however, reassure them that their information would remain private. Knowing that their identity would be protected made them feel more at ease to supply the necessary information.

Definition of terms

Diabetes is a chronic illness that arises from insufficient insulin production by the pancreas or from the body's inability to utilize the insulin that is produced (WHO, 2023).

Diabetes type 1 (also known as insulin-dependent diabetes) is a chronic illness in which the pancreas fails to produce any insulin at all (ADA, 2021).

Type 2 diabetes is a long-term medical condition that arises from an inability of the body to appropriately utilize or generate insulin, the hormone that is accountable for controlling blood glucose levels (ADA, 2023).

Incidence is the number of new cases that occur during a given time period (CDC, 2021).

Prevalence is the percentage of a population that is found to have a condition at a specific point in time (WHO, 2021).

1.10 Summary

In conclusion, this chapter has introduced the research on statistical analysis of prevalence of diabetes in Zimbabwe, highlighting its significance and potential impact. The research questions and objectives have been stated, providing a direction for the study. Additionally, the next chapter will review relevant literature on diabetes.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

The chapter highlighted the theoretical literature and previous research studies that are related to this research study. This chapter provided several contributions by different authors. The chapter also provided the conceptual model and empirical literature of the previous research on the analysis of prevalence of diabetes in Zimbabwe. This chapter represented a review of the concept from prior research.

2.1 Theoretical literature

According to Bandura (2006), the social cognitive model is a theory that suggests that people's health behaviors, including eating habits and physical activity, are influenced by their environment and the people around them. People's beliefs about their own abilities and self-efficacy can influence their behavior, and this means that people in Zimbabwe who believe they can effectively manage their diabetes might be more likely to take steps to manage their condition. In addition to that, people's self-regulation can also influence their behavior, which means people who are able to set goals and monitor their progress in managing their diabetes might be more successful in doing so. Social cognitive theory has been used to help people change their behaviors and is often used to develop a number of public health interventions related to diabetes to promote healthy behavior. One group-based intervention that employs social cognitive theory to encourage physical activity and healthy eating among individuals with type 2 diabetes is the DREAM (Diabetes Remission and Medication Trial) program. Research has demonstrated that this program enhances participant quality of life and lowers depression. Research conducted in 2019 by Crandall and Doljanac demonstrated that diabetes management behaviors like blood glucose monitoring, medication adherence, and healthy eating may be explained and predicted by social cognition theory.

Social cognitive theory provides some insights into the diabetes health situation in Zimbabwe. Firstly, this theory highlights the importance of personal factors like knowledge and attitudes about diabetes, which are likely to be influenced by cultural beliefs and social norms. Furthermore, a number of research have examined the connection between social support and diabetes control using this idea. For instance, Li et al.'s (2018) study discovered a link between better diabetes outcomes and social support. The social cognitive model adopts a comprehensive perspective, taking into account the interplay between behavioral, environmental, and personal elements. It can be applied to evaluate the success of treatments and offer suggestions for enhancements. Furthermore, this theory takes into account how emotions play a part in managing diabetes. Emotions have been demonstrated to be a significant influence in behavioral change and to lower the incidence of diabetes-related comorbidities including kidney and cardiovascular disease. The social cognitive model, however, is not necessarily applicable to all cultures and social groupings, and putting it into practice can be challenging because it takes a lot of time and money to gather information and evaluate the variables that affect behavior.

Another theory that is related to diabetes is the biopsychosocial model. The biopsychosocial model is a framework that suggests that biological, psychological, and social factors all play a role in diabetes and any other chronic diseases (Rush A., and Stevens A., 2013). According to Li et al. (2019), the biopsychosocial model suggests that biological factors (such as genetic, hormonal, and lifestyle factors); psychological factors (such as stress, coping strategies, and health beliefs) and social factors (such as family support, social support, and socioeconomic status) are all interconnected and that they all influence each other. In addition, this model proposes that all of these factors have an influence on diabetes management and outcomes. Ryan et al. (2016) used a biopsychosocial framework to predict diabetes self-management and found that the model had significant predictive power.

Public health treatments have been developed using the biopsychosocial model in a variety of ways (Nagoya and Russell, 2018). For example, it has been applied to create interventions aimed at resolving socioeconomic factors of health, like homelessness, food insecurity, and poverty. It has also been used to develop interventions that aim to improve access to care, as well as interventions that focus on health literacy and education. Moreover, it has been used to develop interventions that target specific populations, such as ethnic minorities or people with low socioeconomic status. This model is used to help researchers, public health professionals, and policymakers understand the impact of different interventions on diabetes outcomes.

Furthermore, this model is used to develop and test culturally tailored interventions that take into account the unique cultural and social context of diabetes in different countries and communities. In addition to that, the biopsychosocial model helps to reduce health disparities in diabetes by identifying and addressing the unique barriers that different groups of people face, helps to build capacity among local communities to better manage diabetes, and is also used to inform policy decisions about diabetes prevention and management. However, it can be difficult to apply in practice as it requires a lot of time and resources to collect data on all of the different factors involved. The biopsychosocial model can be difficult to translate into concrete actions and interventions. Another limitation is that it may not be as useful for understanding the experience of people with type 1 diabetes, as it is more focused on factors that contribute to type 2 diabetes. Despite these drawbacks, the model has a lot of potential for improving diabetes outcomes.

Another idea related to diabetes is the social determinants of health. According to WHO (2022) and CDC (2020), social determinants of health are the circumstances in the settings in which individuals are born, live, learn, work, play, worship, and age and have an impact on a variety of health, functional, and quality-of-life outcomes and hazards. The distribution of resources, power, and money at the international, national, and local levels also has a role in shaping these factors. Furthermore, as per the CDC (2020), there exist other categories of social determinants of health, such as economic stability, educational accessibility and quality, built environment and neighborhood, and social and communal context.

Social determinants of health model suggest that diabetes mellitus is influenced by factors beyond individual control, such as poverty, social inequality, and access to healthcare (Lochner and O'Campo, 2019). The social environment can have a significant impact on how individuals manage their diabetes and the level of support they receive (Schoenberg et al., 2018). Neal, Mellor, and Nettleton (2018) state that healthier eating and physical activity options, together with other interventions aimed at addressing the socioeconomic determinants of diabetes, can lead to better outcomes for those with the disease. Diabetic management and its course can be greatly influenced by social determinants of health. For instance, having little access to healthcare and having a low socioeconomic position can both raise the risk of diabetes and make it more difficult to treat the disease. Further variables that may negatively affect diabetes care include food insecurity, a lack of mobility, and social isolation. A study by Balfour et al. (2018) found that socioeconomic position (SEP) was a strong predictor of diabetes in lower-income countries. This study also found that low SEP was associated with poorer glycemic control, higher rates of complications, and higher diabetes-related mortality.

Health is shaped by politics and power, as recognized by the social determinants of health approach. Transport, food systems, and land use policies, for instance, can all significantly affect diabetes rates. The model also acknowledges that maintaining health is a major objective and that everyone, regardless of background, should have the chance to live a healthy life. Public health interventions for diabetes can be developed using the social determinants of health approach in a number of ways. Firstly, it is the use of community-based interventions that aim to improve access to healthy foods, increase physical activity, and provide social support for individuals with diabetes (Nagoya and Russell, 2018). Secondly, it is the use of policy interventions that aim to address socioeconomic inequalities, improve educational opportunities, and provide affordable housing options. These measures can help lower the prevalence of diabetes and enhance community health overall. However, because it can be challenging to distinguish between the effects of individual interventions, measuring the influence of interventions on health outcomes can be challenging. Furthermore, the model doesn't always account for how personal agency and decision-making influence health outcomes.

Health belief model is another theory (HBM). An analytical and predictive paradigm for health behaviors is the Health Belief Model. Diabetes is among the many health conditions for which it has been used (Geboers et al., 2017; White, 2019). According to this model, there are several elements that might impact health behaviors, including the perception of one's vulnerability to illness, the severity of the sickness, the advantages of taking action, and the perception of obstacles to taking action (Glanz, Rimer, and Viswanath, 2008). Enhancing health outcomes can be achieved by developing and customizing public health interventions to the requirements of the target community based on an understanding of these factors. Patients with type 2 diabetes in the Netherlands were the subject of a study by Bax In a study focusing on type 2 diabetes patients in the Netherlands,(Bax ,2017) discovered a substantial correlation between medication adherence and perceived advantages as well as hurdles. This means that individuals who believed that there were benefits to taking their medication and who perceived fewer barriers to taking their medication were more likely to take their medication as prescribed. Bax et al. (2017) also found that perceived benefits such as prevention of complications and a better quality of life were important motivators for diabetes management.

The Health Belief Model (HBM) offers a theoretical framework for comprehending behaviors associated to health, which makes it beneficial (Kalra et al., 2021). This model is a helpful resource for scholars and practitioners since it is straightforward and simple to comprehend. HBM offers a precise framework for recognizing and addressing important beliefs that affect actions linked to health. Furthermore, this approach has been effectively applied to the development of therapies targeted at enhancing diabetes control and lowering problems associated with diabetes. It has been used, for example, to create interventions to enhance medication adherence in African American diabetic patients (Meisner et al., 2020) and to enhance diabetes self-management in individuals with low literacy (Vallejo et al., 2019). Key beliefs discovered by the HBM were the focus of these interventions. These beliefs included those regarding the gravity of diabetes, the advantages of taking medicine, and the obstacles to doing so. Those who lack motivation to modify their behavior or who do not possess a strong feeling of self-efficacy may find this paradigm unsuitable. Additionally, the HBM may not be as effective for individuals with more complex diabetes-related issues, such as those with co-morbid conditions or those who have difficulty adhering to a treatment plan. The socio-ecological model is another theory that is often used in diabetes management. This model acknowledges that diabetes management is influenced by a complex web of factors that is flexible and adaptable to changing both individual and population-level factors (Kruk et al., 2018). It acknowledges that managing diabetes involves a variety of social, environmental, and individual factors, including information, attitudes, and actions. In addition to that, the socioecological model for diabetes management suggests that interventions are tailored to the individual, target multiple levels of influence, and are culturally sensitive (Hales et al., 2018). The interventions to improve diabetes care consider the individual, community, and environmental levels of influence and also take into account factors such as cultural beliefs, access to resources, and social support (McCulloch et al., 2019). According to a study of diabetic therapies based on the model, glycemic control and quality of life—two outcomes connected to diabetes—were improved (McEwen et al., 2014).

Glanz in 2015, suggested that the socioecological model is a helpful framework for understanding how different levels of the social environment interact to influence health and disease and for developing interventions to address complex public health issues (Glanz, 2015). A study by Sperl-Hillen et al. (2018) used this model to develop an intervention to improve diabetes care in rural communities. In this study, the authors identified the social, economic, and environmental factors that influenced diabetes care in rural communities and used this information to develop a community-based intervention. Additionally, the model acknowledges the complexity and multifaceted nature of diabetes, which is significant because it can assist in identifying and addressing the major obstacles to efficient management and intervention. It also acknowledges the significance of cultural factors in the management of diabetes. Finally, the model provides a framework for organizing and implementing community-based interventions for diabetes management. However, this model can be difficult to apply in practice as it requires a comprehensive assessment of individual, community, and environmental factors. Moreover, the model does not always consider the individual's readiness for change, which can be an important factor in diabetes management.

Moreover, logistic regression model is a frequently employed statistical technique for predicting the risk factors connected to diabetes, according to Hosmer et al. (2017). Based on one or more independent factors, a binary or categorical outcome variable's probability can be predicted using the statistical modeling technique known as logistic regression. It's a kind of regression analysis, however logistic regression is used when the dependent variable is dichotomous, meaning it can only have two possible values, such as 0 or 1, true or false, as opposed to linear regression, which is used for continuous outcomes. A multitude of assumptions are made by logistic regression, such as the absence of multicollinearity, the presence of significant outliers, the independence and identical distribution (IID) of the observations, and a linear relationship between the logic of the outcome variable and the independent variables (Hosmer et al., 2017). These presumptions aid in ensuring the accuracy of the results and the validity of the model. It simulates the link between the dependent variable's log-odds and the independent variables.

The logistic regression model can be written mathematically as:

$$\text{Logit}(P) = \log(\text{Odds}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \qquad \text{Equation 2.1}$$

Where:

- β_0 is the intercept term.
- β_1 to β_n are the coefficients associated with the independent variables X_1, X_2, ..., X_n.
- Log is the natural logarithm function.

The odds of an event are defined as:

$$\text{Odds} = \frac{P}{1-P} \qquad \text{Equation 2.2}$$

Where:

- P is the probability of an event.
- 1-P is the probability of an event not occurring.

OR given the odds of an event, the probability of an event can be computed by:

$$P = \frac{\text{Odds}}{1+\text{Odds}} \qquad \text{Equation 2.3}$$

$$P = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n)}{1+\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n)} \qquad \text{Equation 2.4}$$

For estimating diabetes risk and locating possible risk factors, logistic regression is a useful technique. Logistic regression was employed by Ladefoged et al. (2015) to determine risk factors for type 1 and type 2 diabetes in a Danish population-based cohort. The risk of both forms of diabetes was found to be influenced by shared and non-shared environmental factors, which were modeled using logistic regression by Ladefoged et al. (2015). By emphasizing the higher risk of diabetes and offering advice on how to lower this risk, the findings of logistic regression analysis can be utilized to create precise and unambiguous communications for high-risk populations.

One of the most flexible tools for assessing a variety of data is logistic regression, which can handle both continuous and categorical predictor variables. Furthermore, easy-to-read metrics of the association between the independent factors and the likelihood of the outcome are provided via logistic regression coefficients and odds ratios. It can also be helpful to make forecasts and guide decision-making because this model provides a direct assessment of the outcome's probability. In order to help in prediction and decision-making, logistic regression provides a direct assessment of the likelihood of an outcome. Still, estimates that are skewed may result from breaking the logistic regression assumptions. Furthermore, relationships between predictor variables—which can be challenging to recognize and model—may be sensitive to in logistic regression.

2.2 Empirical evidence

Nyika et al. (2016) used a cross-sectional design to investigate the prevalence and distribution of diabetes in Zimbabwe. The Zimbabwe National Health Survey (ZNHS), which was carried out in 2013 and 2014 by the Zimbabwe National Statistics Agency (ZIMSTAT), provided the researchers with the data they needed. A multistage, cluster, stratified, and sampling design was employed in Zimbabwe to choose a nationally representative sample of homes. Standardized questionnaires were used in the survey, and interviewers with training gathered data. Nyika et al. (2016) employed multivariable logistic regression and descriptive analysis as statistical analytic techniques. The prevalence of diabetes was estimated and its distribution across different demographic and socioeconomic factors was examined using descriptive analysis. In order to evaluate the relationships between diabetes and other risk variables while accounting for possible confounders, multivariable logistic regression was utilized. As a result, the researchers were able to identify the variables were independently linked to diabetes.

The research by Nyika et al. (2016) found that 8.3% of Zimbabweans had diabetes. Furthermore, those who were obese, had higher socioeconomic position, and higher education levels also had higher rates of diabetes. The frequency rose with age and was more common in women than in males. Age and educational attainment were found to be the next most significant predictors of diabetes, after obesity, according to multivariable logistic regression. Furthermore, the study demonstrates that there is no substantial correlation between alcohol and tobacco usage and diabetes. Furthermore, 11.3% of the population in the nation has diabetes, with metropolitan areas having the highest frequency.

Madebwe et al. (2017) conducted a second study in which they looked at the factors and prevalence of diabetes in hypertensive individuals in Harare, Zimbabwe. This study employed a cross-sectional design situated in a hospital and used data from the 2013 Zimbabwe Demographic and Health Survey (ZDHS). The study involved 910 patients in all. Adult patients who were diagnosed with hypertension at the hospital and were at least eighteen years old made up the study population. Participants' blood was also drawn in order to test their blood glucose levels and other biochemical markers. The data analysis was done using SPSS software. The data contained clinical test results, demographic data, and medical history information. The methods used for statistical analysis included logistic regression, chi-square tests, and descriptive statistics. Significant risk factors for diabetes were found using logistic regression once other variables were taken into account.

The study's conclusions showed that 19.8% of hypertension patients in Harare also had diabetes. According to Madebwe et al. (2017), the prevalence of diabetes mellitus was higher in females than in males and in urban versus rural areas. Furthermore, compared to people with secondary or higher education, the prevalence of diabetes was greater in people with no education or only a primary education. It was also higher in people without a job compared to people with a job. Compared to lower age groups, the prevalence of diabetes was greater in people 60 years of age and older. Additionally, it was shown that obesity, age, and a family history of diabetes were important risk factors for the disease. Furthermore, the study demonstrated that diabetes has a detrimental impact on patients' quality of life, with diabetics expressing higher levels of weariness, worry, and despair than non-diabetics. The management and prevention of diabetes in Zimbabwe will be significantly impacted by these findings.

Nyabadza et al. (2017) reviewed the statistics on diabetes and the factors that contribute to poor glucose tolerance in Zimbabwe. Reviewing the prevalence and contributing factors to diabetes and impaired glucose tolerance in Zimbabwe was another goal of the investigation. The Zimbabwe Demographic and Health Survey (ZDHS) was carried out in 2012, and that data was used. In a survey, one member of each household is interviewed after being chosen in two stages using a cluster sampling design. Impaired glucose tolerance and the prevalence of diabetes were estimated using a logistic regression model.

According to the findings of Nyabadza et al.'s (2017) investigation, 6.3% of people had diabetes and 9.9% had impaired glucose tolerance. Furthermore, men were more likely than women to have diabetes and impaired glucose tolerance, respectively. According to this study, persons who live in urban areas have higher odds of having diabetes or impaired glucose tolerance than people who live in rural areas. Those who are obese and have higher levels of education also have higher odds of having these conditions. Physical activity, smoking status, and alcohol use did not appear to be major factors of diabetes or impaired glucose tolerance. Furthermore, this analysis offered crucial details regarding the prevalence of diabetes and low glucose tolerance in Zimbabwe.

Zhang et al. (2017) carried out a meta-analysis of research on the prevalence of diabetes in low- and middle-income nations in another study. Zhang et al. (2017) used data from national surveys that the International Diabetes Federation and the World Health Organization carried out in low- and middle-income nations. The 48 nations where the research from Africa, Asia, and Latin America were conducted and published between 2000 and 2016 included the meta-analysis. Additionally, the data was examined in this study according to the nation, economic level, geographic area, and urban or rural location. The prevalence of diabetes in low- and middle-income nations was estimated by researchers using a random-effects model. The model facilitated the consideration of the variations among the research. To evaluate how reliable the findings were, researchers also performed sensitivity analyses. Moreover, publication bias was evaluated using Egger's test and funnel plots.

Kampfumhukutu et al. (2017) conducted a second study titled "Time Series Analysis of Diabetes Mellitus Prevalence in Zimbabwe, 2001–2012." Data from three nationwide surveys carried out by the Zimbabwe nationwide Statistics Agency in Zimbabwe were gathered by Kampfumhukutu et al. (2017). The Zimbabwe Demographic and Health Survey (ZDHS) was conducted in three different years: in 2001–2002, 2005–2006, and 2010–2012. The study employed a multistage stratified random sampling technique to choose roughly 7,000 families from the total population. Time series analysis, vector autoregression (VAR), and autocorrelation were the statistical analysis techniques applied.

Zimbabwe saw a rise in the prevalence of diabetes from 2.5% in 2001–2002 to 6.4% in 2010–2012, according to Kampfumhukutu et al. (2017). Diabetes was shown to be more common in urban regions than in rural ones, with a substantial difference in prevalence. The findings indicated that diabetes was more common in older age groups and in females. The study also discovered a relationship between the prevalence of diabetes and levels of physical activity and body mass index. Additionally, Kampfumhukutu et al. (2017) discovered that those with greater wealth and educational attainment had higher rates of diabetes prevalence. Furthermore, those who had never smoked and those who abstained from alcohol consumption had reduced diabetes prevalence rates. According to the research, those who consumed a lot of fruits and vegetables had a reduced prevalence of diabetes. This study demonstrated a relationship between the prevalence of diabetes and variables like food, physical activity, and body mass index. Researchers recommended that encouraging healthy lifestyle choices like exercise and a balanced diet should be the main goal of treatments to lower the prevalence of diabetes in Zimbabwe.

Shaw et al. (2017) examined a different paper titled "Diabetes: A Silent Pandemic." In order to inform preventative and treatment initiatives, the study set out to assess the global burden of diabetes and the risk factors associated with it. The World Health Organization, the World Bank, the United Nations, published literature, and national statistical offices were among the many sources of data that Shaw et al. (2017) consulted. Each piece of data was meticulously chosen and verified to guarantee its accuracy. Regression analysis and multivariate analysis were employed by Shaw et al. (2017). The link between the variables was predicted using regression analysis. By doing this, we can assess the strength of the link and account for other factors that might have an impact on it. A comparable technique that enables us to examine several variables at once is multivariate analysis. This is significant since diabetes is linked to numerous factors, and we must comprehend how these factors interact. The human capital approach was another technique employed by Shaw et al. (2017) to determine the expenses associated with diabetes. This method accounts for both direct costs—like the price of care—and indirect costs—like the price of lost production brought on by a disability or early death. Diabetes-related illness burden was also quantified using disability-adjusted life year (DALY).

The study's conclusions indicate that as obesity rates have risen over time, so too has the incidence of diabetes. Furthermore, the study discovered that while there is a high association between diabetes and obesity, it is not as strong as it is for some other risk variables like gender and age. Furthermore, the study found that the incidence of diabetes differs by nation, with the highest frequency found in the Middle East, Africa, and Southeast Asia. According to the report, diabetes ranked as the seventh most common cause of death globally in 2013. Additionally, this study demonstrated that diabetes was the primary cause of lower limb amputation, kidney failure, and blindness. According to disability-adjusted life years, diabetes was found to be significantly related with a lower quality of life. Furthermore, women are more likely than males to get diabetes, and this risk rises with age. According to this study, those who have a family history of diabetes are more likely to develop the illness themselves. Furthermore, the data showed that those with poor socioeconomic level, African Americans, and residents of rural areas had higher rates of diabetes. In their final paragraph, they urge additional action to combat diabetes, including improved diabetes management, prevention, and treatment.

An additional investigation conducted by Tanyanyiwa et al. (2019) looked at the ten-year trends in diabetes mellitus incidence and mortality in Zimbabwe. The study's objectives were to comprehend diabetes mellitus's incidence and death rates in Zimbabwe and to spot any possible trends or discrepancies. Data from the National Population Registry (NPR) and the Zimbabwe Demographic and Health Survey (ZDHS) were used in this analysis. The ZNHS is a nationally representative household survey that gathers information from the NPR for the years 2008–2016 as well as data on diabetes from 2008–2009 and 2015–2016. Tanyanyiwa et al. (2019) employed multivariable regression analysis and descriptive statistics. Descriptive statistics were also utilized to determine any demographic differences in the incidence and mortality of diabetes, as well as to explain the trends and patterns of diabetes in Zimbabwe. The relationship between diabetes and other variables, including age, sex, place of residence, and education level, was investigated using multivariable regression analysis. Regression analysis separated the impact of diabetes on mortality by adjusting for any confounding variables, such as comorbidities

The results showed that diabetes was the sixth most common cause of death in Zimbabwe and that the mortality rate from the disease was higher than anticipated. The incidence of diabetes rose from 0.79 per 1,000 persons in 2008–2009 to 1.24 per 1,000 people in 2015–2016, according to this report. Both the incidence and the prevalence of the condition were higher in urban than rural regions and among males, respectively. Furthermore, the research revealed that the diabetes-related death rate stayed mostly consistent during the study period, averaging 2.3 per 1,000 people in 2008–2009 and 2.5 per 1,000 people in 2015–2016. The review had limitations, including the fact that the fact that the data from the ZNHS is self-reported, which may result in under-reporting or over-reporting of diabetes.

McGavock et al. (2021) conducted an additional study to investigate the diabetes prevalence estimates for 2019 in the global, regional, and national contexts, as well as projections for 2030 and 2045. An invaluable resource for international diabetes research, the International Diabetes Federation Diabetes Atlas has data from more than 190 nations and territories (McGavock et al., 2021). Data on diabetes prevalence, mortality, and risk factors can be found in Atlas, a global repository. Furthermore, Atlas offered mortality and prevalence estimates for diabetes for the years 2030 and 2045. Data was collected from multiple sources, including national surveys, hospital registries, and population health surveys. Statistical analysis methods used by McGavock et al. (2021) were based on modeling techniques. Both the diabetes prevalence and future trends were projected using a Bayesian hierarchical model. In this model, factors like age, sex, the incidence of diabetes in a given nation, and variations in the incidence across time were included. The model also accounted for underreporting of diabetes and diabetes-related mortality. Researchers used Monte Carlo simulations to generate estimates and projections for each country and region. These estimates and projections were then used to generate summary statistics for the global, regional, and national levels.

The findings of McGavock et al.'s study from 2021 demonstrated that diabetes has become more commonplace worldwide over time, with the Middle East, North Africa, and Asia having the greatest prevalence rates. Furthermore, the results indicated that future estimates for the prevalence of diabetes were expected to rise, with the biggest increases expected in Asia and Africa. According to this research, diabetes was identified as a leading cause of death globally in 2019 (about 4.2 million deaths), with significant regional and national differences in both diabetes mortality and prevalence. A number of additional diabetes-related data were also examined by McGavock et al. in 2021.For instance, the authors discovered that persons with lower levels of education had a higher frequency of diabetes than those living in rural areas. Furthermore, researchers discovered that a greater proportion of individuals with diabetes had diabetes than those with other chronic illnesses like obesity and hypertension, and that the majority of diabetics are ignorant of their diagnosis. The financial burden of diabetes is also substantial, encompassing indirect expenses like impairment and missed productivity as well as direct costs like medication and hospital stays. Overall, the study by McGavock et al. (2021) paints a complete picture of the impact of diabetes and emphasizes how crucial it is to fund preventive efforts and provide treatment accessibility for those who already have the condition in order to lessen its effects.

2.3 Research gap

Many studies have been conducted worldwide on the incidence and prevalence of diabetes, risk factors for the disease, its complications, and the economic effects of diabetes, to name a few. However, very little research has been conducted in Zimbabwe, and little is still known about how these factors may differ amongst different regions of the country. This knowledge gap offers a chance for a cross-sectional study to examine the association between the prevalence of diabetes in Chitungwiza, a particular region within the Zimbabwean population, and demographic factors (age, gender), lifestyle factors (BMI), physiological or metabolic factors (blood sugar, blood pressure, cholesterol).

Prior studies using smaller sample sizes and shorter time spans have looked at the risk factors and prevalence of diabetes in Zimbabwe. A sizable dataset spanning five years was employed in this study, which gave important new information about the dynamics of diabetes and how it is managed in the Zimbabwean population. This research study is significant since it aided in the decision-making process for enhancing health systems and the delivery of healthcare services for the government, legislators, and several other relevant parties. Because of this, the research study was significant because it was conducted in Zimbabwe.

2.4 Conceptual model

A conceptual model, as defined by Humphreys and Kausel (2018), is a condensed depiction of a concept or circumstance that emphasizes the salient characteristics or analytical concepts. This study's conceptual model offers a framework for examining the connection between risk variables and the prevalence of diabetes in Chitungwiza, Zimbabwe. The link between the independent factors (age, blood pressure, cholesterol, blood sugar, body mass index, and gender) and the dependent variable was the main emphasis of the conceptual model. A model emphasized the salient characteristics and guiding principles of the investigation, such as the contribution of lifestyle, health-related, and demographic factors to diabetes risk. Researchers and policymakers were able to gain a better understanding of diabetes and its determinants by using a conceptual framework to illustrate the links between various variables and their effects on diabetes risk. The creation of research questions, hypotheses, and study designs was led by a conceptual framework, which made sure that the study was concentrated on the most crucial elements and how they interacted.

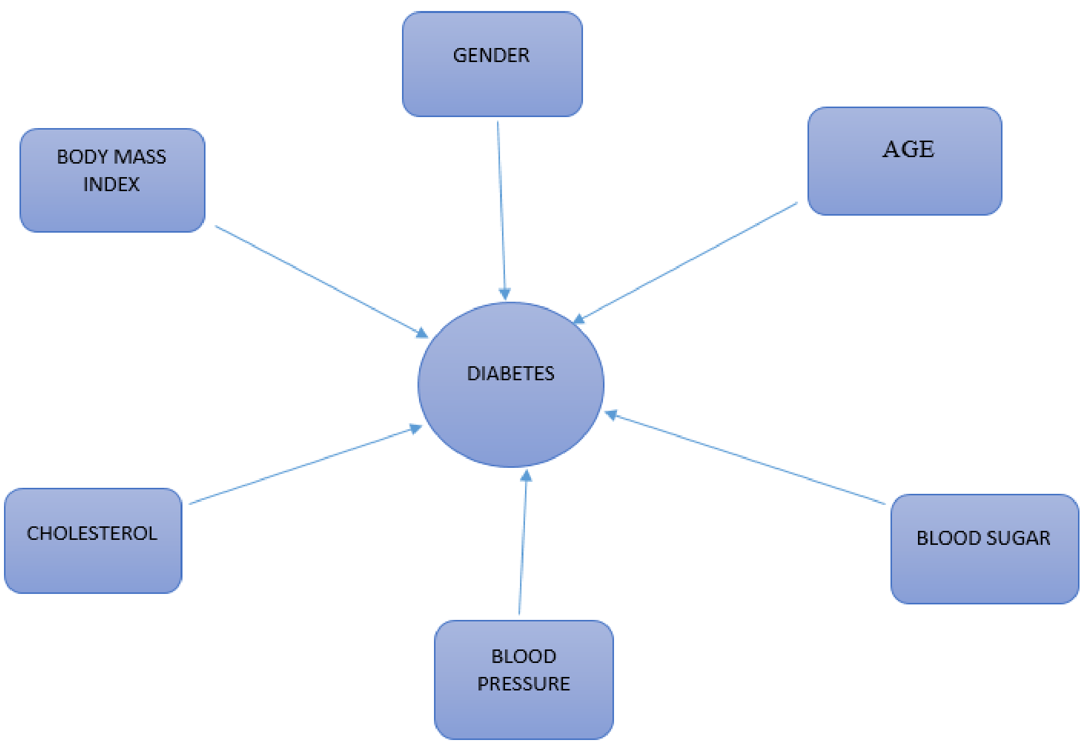

Figure 2. 1: Conceptual Model

2.5 Summary

This chapter focused on a review of different authors' analyses on diabetes. The chapter reviewed the relevant literature, including the theoretical frameworks supporting the conceptual frameworks and the empirical evidence of diabetes. A literature review helped to review the research gap, which the researcher wanted to close with the current research. The research procedure, data collection and analysis methods are covered in the next chapter.

CHAPTER 3

RESEARCH METHODOLOGY

3.0 Introduction

The study approach for the statistical analysis of prevalence of diabetes in Zimbabwe was the main topic of this chapter. This chapter's goal was to give a thorough and understandable explanation of the research techniques that were employed to achieve the goals. This chapter also covered the target population, sample techniques, research instruments, data analysis techniques, research design, and ethical considerations. The methods used for data collecting and presentation were also covered in this chapter.

3.1 Research design

A research design, as defined by MC Combe et al. (2019), is a comprehensive plan that addresses a number of research questions related to the preparation, coordination, and conduct of the study. Because it addressed the purpose and objectives of the research and offered a framework for responding to the research questions, the study design served as a guide for this investigation. Additionally, it served as a road map for the researcher as they carried out this investigation. This study used a quantitative research approach, focusing mainly on the analysis of numerical data related to diabetes prevalence, diabetes type, and risk factors. This research design was used because it was relatively easy to conduct since the data was collected at one point in time.

3.2 Data sources

Using secondary data, from Chitungwiza Central Hospital in Zimbabwe's Demographic Health Information System (DHIS 2) was used to create statistical models of diabetes. The collected data provided information on diagnosed cases of diabetes. Secondary data sources are less expensive and less time-consuming because the information is already public and can be readily checked from reliable sources. It made it possible for the researcher to have sufficient and relevant data on a few variables. However, because of mistakes in the data brought about by human tampering, they might include biased information. The utility of the secondary data may be limited by its inaccuracy, incompleteness, or the presence of outliers, which can limit its usefulness.

3.3 Target population and sampling

This study's target group comprised all adult patients at Chitungwiza Central Hospital, a significant public health facility in Chitungwiza, Zimbabwe, aged 21 and above, who were diagnosed with diabetes during a five-year period (2014–2018). Furthermore, 768 patients' worth of data were taken from the electronic medical records (EMR) of Chitungwiza Central Hospital. To make sure the sample was representative, the researcher used a stratified random sampling procedure.

Using stratified random sample and a backward elimination selection strategy for a logistic regression model, the researcher stratified the population according to pertinent factors such age, blood pressure, blood sugar, cholesterol, and BMI. Then, random sampling was done within each stratum to make sure that people from different subgroups are sufficiently represented in the sample. The backward elimination strategy was used to choose the most pertinent independent variables for the logistic regression model after the sampling procedure. By employing an iterative technique, the model is constructed using a representative sample that precisely represents the distribution of independent variables among various strata. Through an efficient combination of backward elimination and stratified random sampling, the investigator improved the logistic regression model's dependability while taking into consideration the wide range of features found in the population being studied.

3.4 Research Instrument

To gather information on instances of diabetes that have been diagnosed, a thorough assessment of medical and electronic health records was carried out for this study. Data extraction, cleaning, anonymization, and validation from the electronic health records system were all done using statistical tools.

3.5 Methods of data collection

Data from the Chitungwiza Central Hospital electronic medical records (EMR) system were used in this study. The electronic medical records were searched for demographic data on diabetes patients, including age, gender, lifestyle factors (BMI), and physiological or metabolic data (blood sugar, blood pressure, cholesterol). Furthermore, the fact that medical records are gathered by qualified medical professionals, which helps to minimize potential bias in the data, contributes to their broad reputation as a trustworthy source of information. Additionally, by using medical data, the researcher was able to see a lot of patients over an extended period of time. In order to enhance the EMR data, the researcher procured pertinent secondary sources, including government publications, journal articles, and healthcare recommendations concerning diabetes management and risk factors in Zimbabwe. In order to create a predictive model for diabetes risk and create evidence-based suggestions for lowering the incidence of diabetes, this secondary data offered more context on risk variables.

3.6 Description of variables and expected relationship

3.6.1 Description of Variables

To create a model, the researcher looked at diabetes cases at Chitungwiza Central Hospital. The study's independent variables include blood pressure, cholesterol, glucose level, age, body mass index, and blood pressure. Using a logistic regression model, the researcher ascertained the association between diabetes and the independent factors.

The author selected the following variables to apply the regression procedure:

Y: The response variable is the diabetes, which is binary (dichotomous) in nature, representing diabetes or no diabetes.

X: The explanatory variables; all variables are nominal.

X_1: Blood Sugar

X_2: Blood pressure

X_3: Body mass index

X_4: Cholesterol

X_5: Age

X_6: Gender

The relationship among the five independent variables and diabetes was assumed to be statistical.

Variable (s)	**Description**	**Data source**	**Expected Sign**
Diabetes	Diabetes is a binary variable where 0 represents a patient with diabetes and 1 represents a patient with no diabetes.	DHIS 2	Dependent Variable
Blood sugar	Blood sugar is measures in milligram per deciliter (mg/dL).	DHIS 2	Positive
Blood pressure	Diagnostic blood pressure is measured in millimeters of mercury (mmHg).	DHIS 2	Uncertain
Body mass index	Body mass index is measured in kilogram per square meters (kg/m^2).	DHIS 2	Positive
Cholesterol	Cholesterol is measures in milligram per deciliter (mg/dL).	DHIS 2	Uncertain
Age	Age of adult people with diabetes is 21 years and above.	DHIS 2	Positive
Gender	Diabetes is a binary variable where 0 represents female and 1 represents male.	DHIS 2	Uncertain

Table 3.1: Variables and Expected signs

3.7 Data Presentation and analysis procedures

First, data formatting and cleaning were done in Microsoft Excel using the Chitungwiza Central Hospital extracted data. The researcher did this by removing any records from the final analysis dataset that lacked essential variables. Statistical analysis was conducted by importing the cleaned data into R Studio. A predictive model of diabetes in Zimbabwe was developed by utilizing a logistic regression model to ascertain the statistical link between diabetes and the related risk variables.

3.7.1 Diagnostic Tests

1. Descriptive statistics

A collection of methods for summarizing and describing a data set is known as descriptive statistics (Field, 2014). Descriptive statistics were employed by the researcher to generate an overview of the variables' properties. Field (2014) states that measures of variability like range and measures of central tendency like mean and median were included in descriptive statistics.

2. Multicollinearity Test

When two or more predictor variables in a multiple regression model have a high degree of correlation, this is referred to as multicollinearity, according to Bowers, D. (2020). Regression coefficient interpretation and the validity of statistical tests are hampered by multicollinearity. Additionally, it can cause issues for multiple regression models, making it more challenging to distinguish between the impacts of the variables and producing estimates of the regression coefficients that are unstable, deceptive, or untrustworthy (Bowers, D. 2020). To accurately isolate and estimate the individual effects of each variable in a multiple logistics regression model, it is ideal for the predictor variables to be as independent of one another as feasible. Estimating the distinct effects of each predictor variable gets challenging when there is a significant correlation between two or more of them. To verify if the data were multicollinear, the researcher employed a correlation matrix and a variance inflation factor.

3.7.2 Analytical Model

3.7.2.1 Dependent Variable

The dependent variable is dichotomous, meaning that the variable has two categories, which are patients with either diabetes no diabetes for all the participants received from Chitungwiza Central Hospital.

The logistic regression model used a logit function to model the relationship between independent variables and the log-odds of the dependent variable.

3.7.2.2 Model fitting and variable selection

Model fitting and variable selection are closely related concepts in statistical analysis, including logistic regression (Bernardo, 2022). The process of fitting a model involved a stage called variable selection, which established variables that were significant to the model. Additionally, the goal of model fitting was to identify the best model that could explain the relationship between the predictor variables and the outcome variables (Hosmer and Lemeshow, 2021). This procedure was carried out to guarantee that the final model has a minimum variance and fits the data perfectly. The model with less variation had more predictive power than one with greater variance.

The likelihood ratio (LR) test is a method used for variable selection in logistic regression based on the logit function. Variable selection process applied to a model is derived from the theoretical logit model. A logit function is defined as the natural logarithm of the odds ratio, which is the probability of the event occurring (for example, having no diabetes) compared to the probability of the event not occurring (for example, having diabetes). The logit function is unbounded due to the range of the logit function. As P approaches 0, the logit (P) approaches negative infinity, and when P approaches 1, the logit (P) approaches positive infinity. A link function that connects the probabilities is called the logit. The study model's explanatory variables comprised both continuous and categorical data.

The proposed empirical model for this study is a logistic regression model for diabetes, mathematically represented as:

$$\text{Logit}(Y) = \beta_0+\beta_1X_1+\beta_2 X_2+ \beta_3 X_3+\beta_4 X_4+\beta_5X_5 +\beta_6X_6 \qquad \text{Equation 3.1}$$

Where:

X_i for i = {blood sugar, blood pressure, cholesterol, body mass index, age and gender} = {1, 2, 3, 4, 5,6}.

Logit (P) is a logit transformation of the probability P of having no diabetes.

β_0 is the intercept term.

β_1 to β_6 are the coefficients (log-odds ratios) associated with each independent variable.

The best model of fit in binary logistic regression was extremely unlikely to emerge. Additionally, the researcher used backward elimination for variable selection in order to identify the optimal set of predictor variables that still provided a good model fit, and it also helped to reduce the risk of overfitting the model. The researcher started with the full model, then assessed the significance of each variable by examining the p-values of the coefficients for each independent variable in the full model. At every stage of the modeling process, the researcher eliminated variables that were less significant and created a new model with significant values from the previous model. When all the less significant variables are removed, it results in a model where all the remaining variables have a $p\text{-value}<0.05$. This indicated that the final model is statistically significant and is the most appropriate model.

3.7.2.3 Assumptions of logistic regression

1. Binary outcome

The logistic regression model assumed that the dependent variable was dichotomous in nature. It means that the variable must take only two possible values, for example, 0 or 1.

2. Linearity of independent variables and log odds

Logit/log odds and the linearity of independent variables are assumed in logistic regression. This model suggests that the log odds of the outcome variable are a linear combination of the predictor variables, but it does not require a linear relationship between the dependent and independent variables. The logistic regression equation was estimated after the researcher evaluated the linearity of the data using scatterplots. Plotting the standardized residuals as a function of the standardized expected values allowed the researcher to analyze the residual plots after the equation was computed.

3. The independence of observation

It was essential that they observe each other independently in order to build a logistic regression model. As a result, neither repeated measurements nor matched data should be the source of any consistent correlations or dependencies between the observations. The Durbin-Watson test was used by the researcher to confirm the error term's independence.

4. The absence of multicollinearity

The independent variables in a logistic regression analysis must have little to no multicollinearity. To put it another way, multicollinearity happens when two or more independent variables have such strong of a correlation with one another that a regression model cannot use them to produce distinct or independent information (Brodeur et al., 2021). In order to check for multicollinearity between variables, the researcher employed the correlation matrix and Variance Inflation Factor (VIF). In a regression model, each independent variable has a numerical value called the VIF. A criterion that is universally accepted for identifying multicollinearity based only on variance indicator product (VIF) does not exist. A VIF of five to ten, on the other hand, is often considered to imply a troublesome level of multicollinearity. No multicollinearity exists between the predictor variable and the other independent variables when the VIF is 1. An increasing degree of multicollinearity is indicated by a VIF larger than 1.

5. Homoscedasticity

Logistic regression relies heavily on the premise of homoscedasticity. For all values of the predictors, this implies that the variance of the errors must be identical. Inaccurate p-values and biased estimations result from not meeting this assumption. To test for homoscedasticity, the researcher made a plot of the standardized residuals against the anticipated dependent values.

3.7.3 Model Validation

To determine whether or whether the model was appropriate, the validation analysis procedure was carried out (Wallenstein, 2020). Several tests that were fitted to the logistic regression model were also used to validate the model. Below was a discussion of the tests pertaining to the model's goodness of fit and the significance of the estimated parameters.

3.7.3.1 Wald test

Under some circumstances, the Wald statistic—which is the ratio of the regression coefficient's square to its standard error—asymptotically has a chi-square distribution (Brennan, 2021). After estimating the model parameters, the researcher employed the Wald test to see whether the regression coefficients were significant.

The hypotheses are as follows:

H_0: $\beta_j=\beta_1=0$

H_1: $\beta_j=\beta_1\neq 0$....where j = 1, 2,..., k

W has a Chi-square distribution in the big sample under H_0, with degrees of freedom matching the number of limitations.

The Wald test formula is as follows:

$$W=\frac{\beta^2{}_j}{SE^2{}_{\beta j}} \qquad \text{Equation 3.2}$$

3.7.3.2 Likelihood ratio (LR) test

A likelihood ratio test was employed to evaluate the logistic regression model's overall significance. Likelihood ratio test refers to two times the logarithm of the ratio of the likelihood functions of two distinct models evaluated at their maximum likelihood estimations (MLEs). Furthermore, based on the maximum likelihood estimations of the entire model, L (full), the likelihood ratio test presupposes the existence of r_1 variables in the model under examination. The degree of freedom of this test's chi-square distribution asymptotically equals the total number of parameters measured in the two models.

The likelihood ratio test is defined as:

$$LR= -2\, [\ln\{L(\text{reduced})\} - \ln\{L(\text{full})\}] \qquad \text{Equation 3.3}$$

3.7.3.2 Hosmer-Lemeshow goodness of fit test

One of the most crucial components of a logistic regression model for determining whether or not the model fits well is its goodness-of-fit (Gupta, S.N. and Gupta, M.M, 2016). A statistical measure that contrasts the estimated and actual occurrences of events in subgroups with equal expected probabilities is the Hosmer and Lemeshow goodness of fit test. By computing the Pearson Chi-square statistic from the (2xg) table of observed and predicted frequencies, where g is the number of groups, the Hosmer-Lemeshow goodness of fit test was obtained. In order to achieve a minimal difference in goodness of fit, this test assessed the similarity be

The definition of the Hosmer-Lemeshow goodness-of-fit test is:

$$\chi^2 HL= \sum_{i=1}^{g} \frac{(O_i - N_i E_i)^2}{N_i E_i (1-E_i)} \sim \chi^2 g-2 \qquad \text{Equation 3.4}$$

Where:

O_i= number of occurrences that were seen in the i^{th} group.

N_i= how many participants there were in i^{th} group.

E_i= the mean calculated likelihood of an occurrence in the i^{th} group.

g= the total count of the groups.

The test was conducted at a 5% level of significance.

3.8 Ethical considerations

In doing this study, ethical considerations were crucial. The researcher followed ethical norms and took precautions to safeguard patient confidentiality. The privacy and confidentiality of patients' personal information as well as health-related information were strictly safeguarded throughout the study, and this research conformed to all applicable ethical guidelines. To avoid unwanted access or exposure, all of the data was anonymized and kept in a secure location. Additionally, the data was handled solely for research purposes and in compliance with existing data protection legislation. In order to guarantee adherence to the pertinent regulations, the researcher secured the required approvals and consents from the relevant authorities.

3.9 Summary

The study's methodology was presented in this chapter. This chapter also covered the definition of variables, expected relations, diagnostic tests, analytical model testing, model validation tests, and ethical considerations. Next chapter will focus on data presentation, analysis, and discussion of the research study's findings.

CHAPTER 4

DATA PRESENTATION, ANALYSIS AND DISCUSSION

4.0 Introduction

The study's findings are presented in this chapter, which also analyzes and interprets the information gathered about prevalence of diabetes in Zimbabwe. The data presentation, data analysis, and discussion of findings parts make up the three sections of the chapter. While the analysis portion uses inferential statistics to find important links and trends, the presentation section presents the descriptive statistics and visualizations. In the discussion part, the results are interpreted, the patterns, trends, and links found in the data are discussed, and the findings are connected to the goals of the research and the literature review.

4.1 Descriptive statistics

Insightful patterns in health measures for the Zimbabwean population were shown by the summary statistics. In addition, the slightly elevated mean blood sugar level of 121.6 would point to a higher chance of hyperglycemia or diabetes. The mean blood pressure of 73.92, on the other hand, is within the normal range and suggests a generally healthy blood pressure profile. Nonetheless, the average Body Mass Index (BMI) of 31.83 indicates an obese population, indicating a high incidence of obesity and a major risk factor for diabetes and other health problems. Furthermore, the average cholesterol level of 186 was somewhat higher than normal, which may suggest an increased risk of cardiovascular disease. Finally, the population's mean age of 33.14 indicates that it is comparatively young, which could account for some of the reported health indicators.

```
  Blood.Sugar     Blood.Pressure        BMI          Cholesterol        Age       
 Min.   : 44.0   Min.   :60.00   Min.   :18.20   Min.   :150   Min.   :21.00  
 1st Qu.: 99.0   1st Qu.:66.00   1st Qu.:27.40   1st Qu.:170   1st Qu.:24.00  
 Median :117.0   Median :72.50   Median :32.00   Median :185   Median :29.00  
 Mean   :121.6   Mean   :73.92   Mean   :31.82   Mean   :186   Mean   :33.14  
 3rd Qu.:140.2   3rd Qu.:80.00   3rd Qu.:36.00   3rd Qu.:200   3rd Qu.:41.00  
 Max.   :199.0   Max.   :95.00   Max.   :44.60   Max.   :235   Max.   :65.00  
```

Figure 4. 1: Descriptive Statistics

4.2 Diagnostic Test

4.2.1 Outlier detection

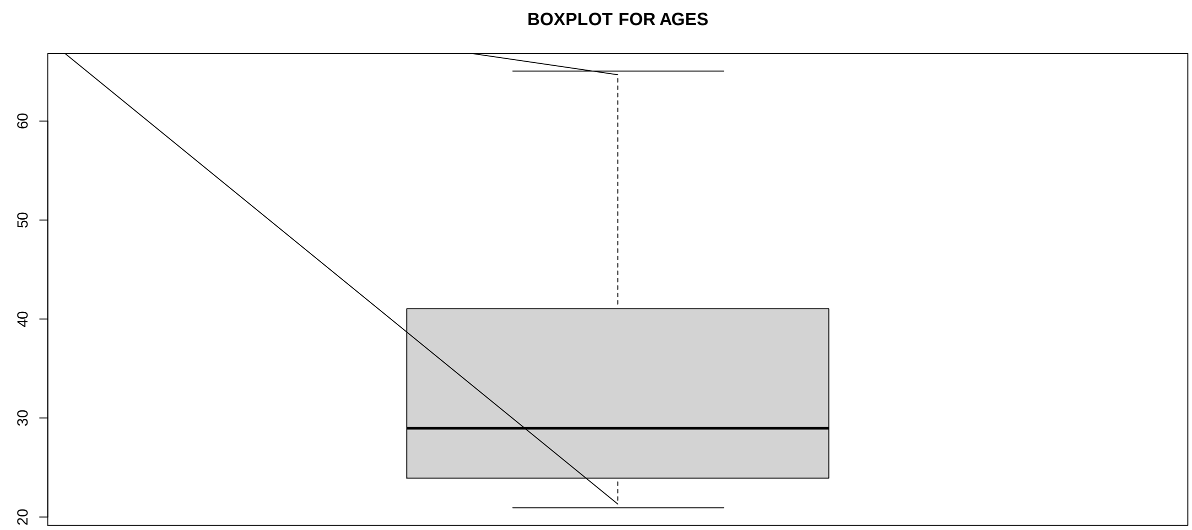

Figure 4. 2: Boxplot for Ages

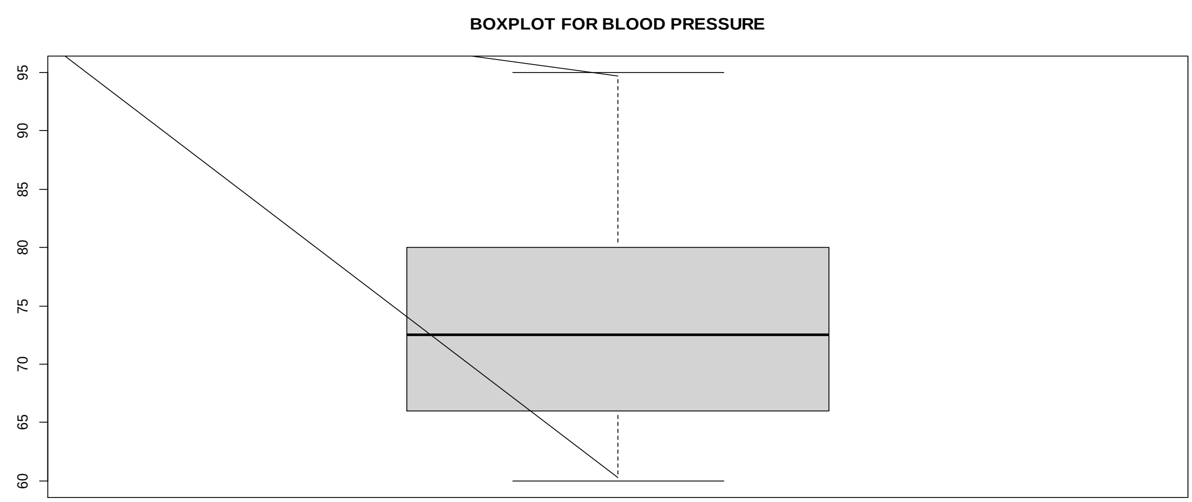

Figure 4. 3: Boxplot for Blood Pressure

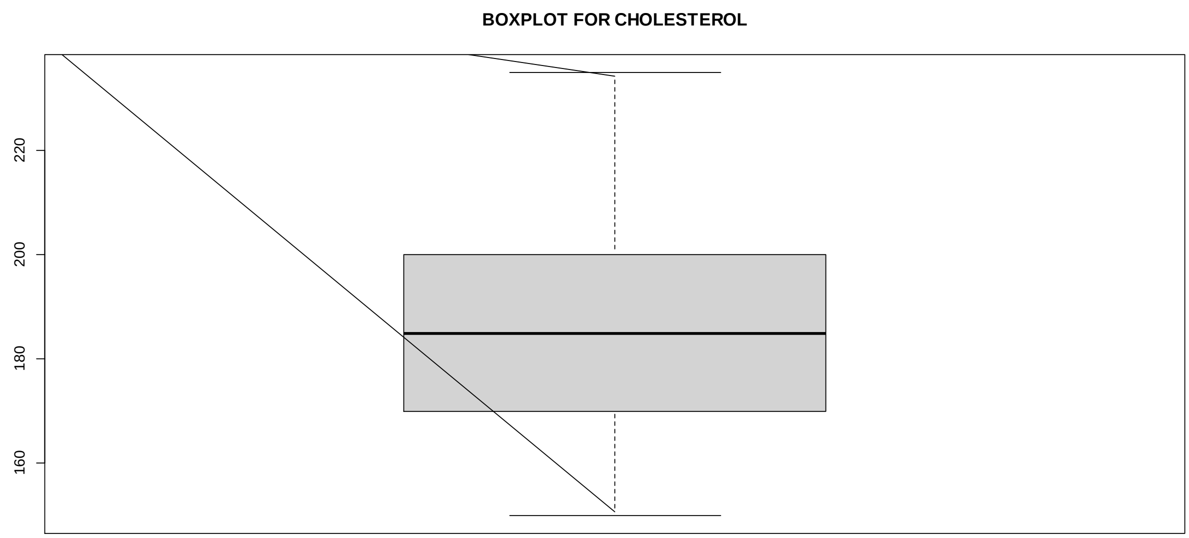

Figure 4. 4: Boxplot for Cholesterol

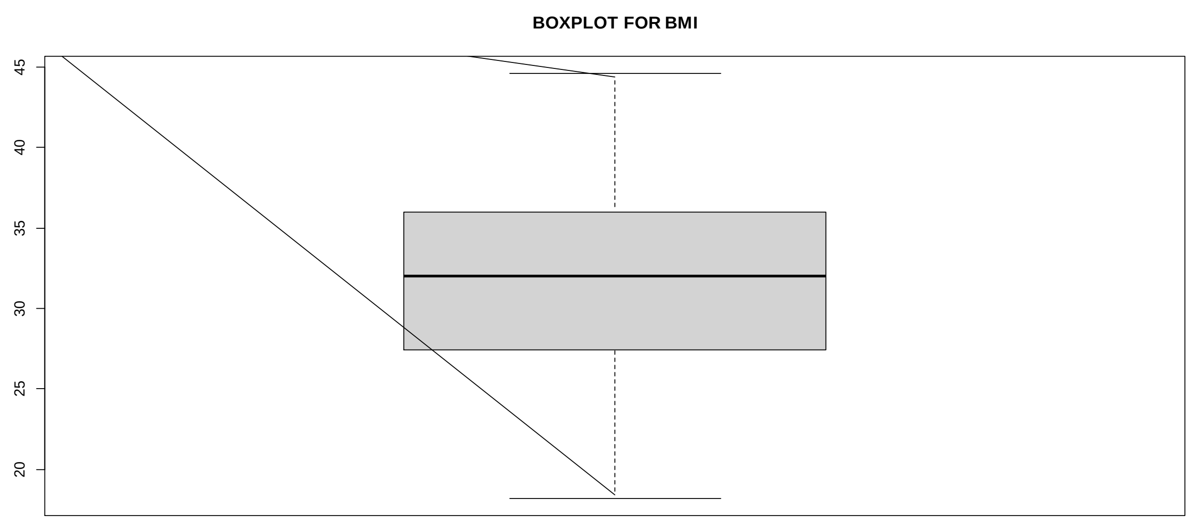

Figure 4. 5: Boxplot for BMI

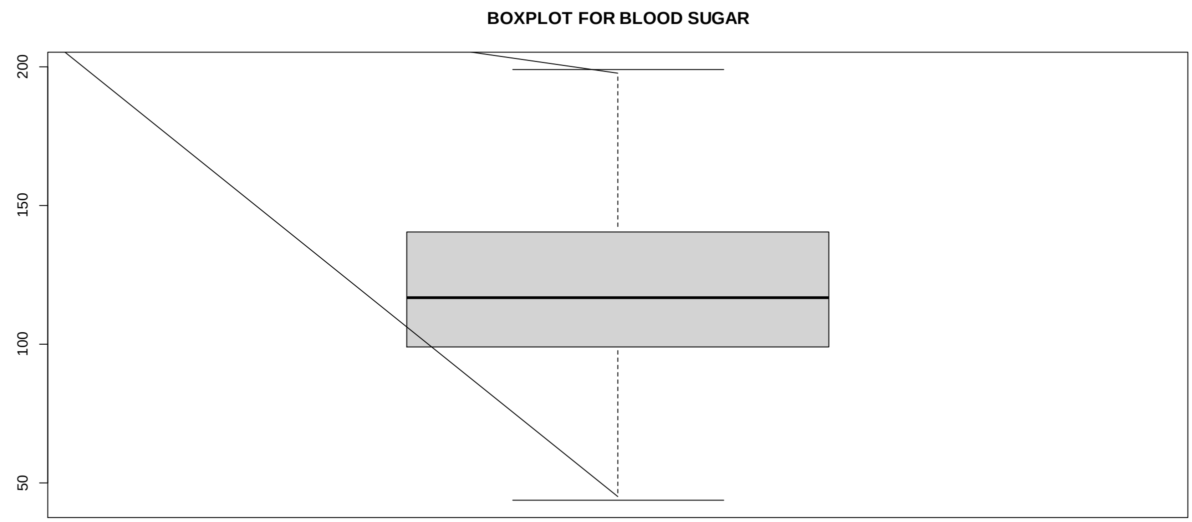

Figure 4. 6: Boxplot for Blood Sugar

An examination of the summary statistics and box plots revealed insightful trends in the health metrics of the Zimbabwean population. The absence of outliers in all box plots indicated a consistent data distribution, free from extreme values that could skew the analysis. In addition to that, the interquartile ranges (IQRs) suggested the moderate variability in blood sugar (IQR = 41.2), BMI (IQR = 8.6), and cholesterol (IQR = 20) levels, while blood pressure (IQR = 14) and age (IQR = 17) exhibit relatively less variability. These statistics provide a solid foundation for further statistical analysis and modeling, enabling a deeper understanding of diabetes in Zimbabwe. The boxplots above showed that there are no outliers in age, BMI, blood sugar, blood pressure, or cholesterol.

4.2.2 Multicollinearity Test

The magnitude and direction of the linear link between two variables are measured by a correlation coefficient. It gave a numerical representation of the degree to which two variables move in tandem in a linear connection. From -1 to 1, the correlation coefficient is the range. A perfect positive linear relationship is represented by a value of 1, a perfect negative linear relationship by a value of -1, and no linear relationship by a value of 0. As a result, stronger linear relationships are indicated by correlation coefficient values that are closer to 1 or -1, whereas weaker linear correlations are indicated by values that are closer to 0.

Age and blood pressure were found to have a moderate link, according to the highest positive correlation (0.279145555), while age and cholesterol were found to have a weak association, according to the highest negative correlation (-0.0019813307). These findings clearly show that the correlations are low, suggesting that there is little to no link between the variables. Consequently, multicollinearity does not exist, indicating that the independent variables are independent

	Blood.Sugar	Blood.Pressure	BMI	Cholesterol	Age	Gender
Blood.Sugar	1.000000000	0.2024160127	0.24038775	0.0440736643	0.269207259	0.001691073
Blood.Pressure	0.202416013	1.0000000000	0.19390945	0.0000814975	0.279194556	0.012757918
BMI	0.240387754	0.1939094514	1.00000000	-0.0183309290	0.041469166	-0.038207710
Cholesterol	0.044073664	0.0000814975	-0.01833093	1.0000000000	-0.001981331	-0.035341552
Age	0.269207259	0.2791945555	0.04146917	-0.0019813307	1.000000000	0.025619226
Gender	0.001691073	0.0127579181	-0.03820771	-0.0353415520	0.025619226	1.000000000

Figure 4. 7: Correlation Matrix

4.2.2.1 Variance Inflation Factor (VIF)

Blood.Sugar	Blood.Pressure	BMI	Cholesterol	Age	Gender
1.032630	1.116786	1.054691	1.006788	1.123622	1.005295

Figure 4. 8: Variance Inflation Factor Results

The findings of the multicollinearity test, as shown by the values of the Variance Inflation Factor (VIF) results in (figure 4.8), demonstrated that there is not much correlation between the independent variables in the regression analysis. The blood pressure, cholesterol, age, gender, blood sugar, and BMI all have VIF values of 1.032630, 1.116786, 1.054691, 1.006788, 1.123622, and 1.005295, respectively. All of these values are below 5, which denotes low to extremely low levels of multicollinearity. According to this, it is likely that the regression coefficients will not increase as a result of correlations between the independent variables, but rather will remain steady. The findings supported the dependability of the statistical model and facilitate the derivation of significant inferences regarding the correlations among the variables.

4.2.3 Durbin-Watson Test

The Durbin-Watson test findings, which are displayed in the figure below, confirmed to the study's observations' independence, which is a necessary precondition for accurate statistical analysis. The researcher came to the conclusion that there was no significant autocorrelation between the residuals, implying that each data point contributed independently and uniquely to the study, with a DW statistic of 1.9666 and a p-value of 0.3253. This finding verified the independence assumption, enabling the researcher to move forward with confidence in the regression analysis and result interpretation. This means that, in the context of the diabetes research being conducted in Zimbabwe, the observations reflect unique and unconnected measures, allowing the researcher to meaningfully infer associations between variables and make well-informed judgments based on the data.

```
        Durbin-Watson test

data:  model
DW = 1.9666, p-value = 0.3253
alternative hypothesis: true autocorrelation is greater than 0
```

Figure 4. 9: Durbin-Watson Test

4.2.4 Standardized residuals

This study used the scatterplot below to check for homoscedasticity and linearity of independent variables and log odds. The residual scatterplot provided a visual diagnostic for the logistic regression model's assumptions. A random scatter of residuals around the horizontal axis (residuals =0) indicated homoscedasticity, suggesting constant variance across the range of predicted values. Furthermore, the plot was utilized to evaluate the independent variables' linearity, with a random scatter signifying linearity. Furthermore, the residuals, which show the variation between the observed and expected log odds of diabetes, enabled the researcher to evaluate how well the model predicted outcomes.

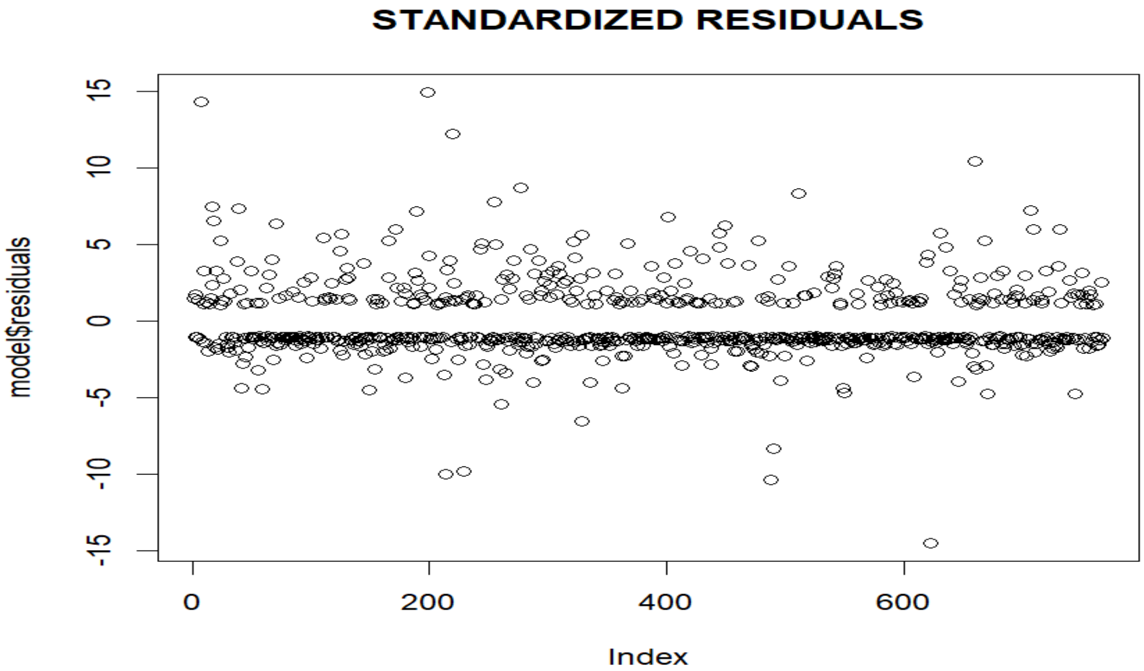

Figure 4. 10: Plot of Standardized Residuals

4.3 Model Output

4.3.1 Model selection

The researcher used the backward elimination model selection method to identify the most influential predictors for a binary outcome. This iterative approach began with a model containing all potential predictors and systematically removes variables that fail to contribute significantly to the model's predictive power. At each step, the variables with the highest p-value were eliminated, and the model's performance was re-evaluated. The process continued until only statistically significant predictors remained in the final model. Backward elimination helped the researcher streamline the model by retaining only the most relevant variables, thereby enhancing its interpretability and reducing the risk of overfitting. A full model with all the predictors was demonstrated in the table below.

```
Call:
glm(formula = Outcome ~ Blood.Sugar + Blood.Pressure + BMI + 
    Cholesterol + Age + Gender, family = "binomial", data = DiabetesData)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)    
(Intercept)    -9.673619   1.254198  -7.713 1.23e-14 ***
Blood.Sugar     0.035042   0.003474  10.087  < 2e-16 ***
Blood.Pressure -0.002640   0.010371  -0.255    0.799    
BMI             0.104959   0.017104   6.137 8.43e-10 ***
Cholesterol     0.001989   0.004458   0.446    0.655    
Age             0.032225   0.008225   3.918 8.93e-05 ***
Gender         -0.044857   0.183563  -0.244    0.807    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 993.48  on 767  degrees of freedom
Residual deviance: 741.34  on 761  degrees of freedom
AIC: 755.34

Number of Fisher Scoring iterations: 5
```

Figure 4. 11: Full Model

It was clear from the results in the above figure that there were multiple predictors in the whole model, each of which had a distinct p-value. Low p-values (less than 0.05) for age, blood sugar, and BMI show a statistically significant association with diabetes. Conversely, blood pressure, cholesterol and gender have higher p-values, suggesting a weaker statistical significance in relation to diabetes. The null and residual deviance values, along with the AIC, provide insights into the model's goodness of fit, with lower deviance and AIC values indicating a better fit. Additionally, the number of Fisher Scoring iterations, in this case 5, reflects the iterative process used to optimize the model's parameters.

The researcher eliminated blood pressure, cholesterol and gender, which are statistically insignificant in relation to diabetes since they have p-values greater than 0.05. The figure below showed the reduced model with the significant variables.

```
Call:
glm(formula = Outcome ~ Blood.Sugar + BMI + Age, family = "binomial",
    data = DiabetesData)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -9.461537   0.761073 -12.432  < 2e-16 ***
Blood.Sugar  0.035046   0.003454  10.147  < 2e-16 ***
BMI          0.103988   0.016817   6.183 6.27e-10 ***
Age          0.031455   0.007909   3.977 6.98e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 993.48  on 767  degrees of freedom
Residual deviance: 741.67  on 764  degrees of freedom
AIC: 749.67

Number of Fisher Scoring iterations: 5
```

Figure 4. 12: Reduced Model

The simplified model is found to have a satisfactory fit with the data, as evidenced by the null and residual deviance values along with the AIC, which are comparable to those of the full model. Additionally, consistency in the number of Fisher Scoring iterations further suggested that the iterative process used to optimize the model's parameters remained unchanged. Overall, the reduced model demonstrates strong statistical significance for the included predictors and maintains a good fit to the data, as evidenced by the deviance and AIC values.

Moreover, the explanatory variables are a good fit for the binary logistic regression model, as all the coefficients in the model are statistically significant at a 5% level of significance. This allowed for the construction of a diabetes prediction model using the meaningful binomial logit coefficients derived from the significant β values.

Logit (P) = -9.461537 +0.035046 Blood sugar + 0.103988 BMI + 0.031455 Age   Equation 4.1

Logit (P) = $-9.461537 + 0.035046\ X_1 + 0.103988\ X_3 + 0.031455\ X_5$ Equation 4.2

Using the equations above as a basis, the logistic regression model forecasts the probability of diabetes. The log-odds of the result in the equations when all predictor variables are zero are represented by the intercept of -9.461537. The influence of age, blood sugar, and body mass index on the probability of having diabetes is shown by the respective coefficients. Taking all factors equal, for example, an increase of one unit in blood sugar is linked to an increase of 0.035046 in the log-odds of having diabetes. Assuming all other variables remain constant, an increase of one unit in BMI is linked to a 0.103988 rise in the log-odds, while an increase of one unit in age is linked to a 0.031455 increase in the log-odds of having diabetes.

4.3.2 Comparison for the predictors of diabetes

When considering the risk factors for diabetes, it's important to understand the various indicators that are positively associated with the condition. Age, BMI, and blood sugar levels are all significant factors that can contribute to the likelihood of developing diabetes. By comparing the ranges associated with the highest risk for each of these factors, valuable insights can be gain into the interplay between these indicators and the increased susceptibility to diabetes. By examining the ranges associated with the highest risk for age, BMI, and blood sugar levels, a comprehensive understanding of the complex relationship between these factors and the development of diabetes can be gained. This knowledge can empower individuals to take proactive steps towards maintaining their health and well-being.

Comparison for age

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.152060	0.1244273	-9.2589017	2.065457e-20
Age_cat30-39	1.088344	0.2024493	5.3758853	7.620730e-08
Age_cat40-49	1.419123	0.2269699	6.2524730	4.040035e-10
Age_cat50-59	1.450553	0.3020240	4.8027748	1.564818e-06
Age_cat60-65	0.102238	0.4564416	0.2239892	8.227657e-01

Figure 4. 13: Comparison for Age

According to the estimates and statistical values, it was clear from the results that the risks of diabetes differed depending on the age group. The estimated risk of diabetes at a reference age (ages 21 to 29) is represented by the intercept value. Compared to the reference age group, the estimates for the age groups 30-39, 40-49, 50-59, and 60-65 are notably positive, indicating an increased risk of diabetes. The age range of 50–59 has the highest estimate (1.450553) among these groups, indicating a significant increase in the risk of diabetes. Additionally, this estimate's reasonably high z-value and low p-value suggest its statistical significance. As a result, of the age categories mentioned, the 50–59 age group is found to have the highest risk of diabetes.

Comparison for blood sugar

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.307834	0.2405046	-9.595800	8.326879e-22
Blood.Sugar_cat100-143	1.556580	0.2653709	5.865674	4.473113e-09
Blood.Sugar_cat144-199	3.096292	0.2865046	10.807127	3.184880e-27

Figure 4. 14: Comparison for Blood Sugar

The estimates for blood sugar categories 100-143 and 144-199 are positive, suggesting an increase in the risk of diabetes compared to the reference blood sugar level (blood sugar category of 44-99). Among these categories, the blood sugar category of 144-199 have the highest estimate (3.096292), supported by a very low p-value and a notably high z-value. Therefore, based on these findings, the blood sugar category of 144-199 presents the highest risk of diabetes among the specified blood sugar ranges.

Comparison for BMI

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.628801	0.3913637	-6.717028	1.854682e-11
BMI_cat25-29.9	1.396657	0.4288584	3.256686	1.127209e-03
BMI_cat30-44.6	2.472695	0.4020423	6.150336	7.731891e-10

Figure 4. 15: Comparison for BMI

According to the findings, estimates for BMI categories 24-29.9 and 30-44.6 are positive, suggesting an increase in the risk of diabetes compared to the reference BMI category (BMI category of 18.2-24.9). This showed a clear positive correlation between BMI and the risk of diabetes, with those in this category having a far higher risk of the disease than those in the reference group. The BMI category of 30–44.6 has the highest estimated risk of diabetes (2.472695), along with a high z-value of 6.150336 and a highly significant p-value of 7.731e-10, among these categories. The data indicate that BMI is a strong predictor of diabetes risk overall, with the highest risk group having a BMI between 30 and 44.6.

4.4 Model validation

In the creation and evaluation of logistic regression models, model validation is an essential stage. It involves evaluating the performance and generalizability of the model to ensure its reliability in making predictions on new data. The researcher employed various validation techniques to gain confidence in the model's predictive capabilities. A logistic regression model validation procedure includes several methods, which are described below, such as the Wald test, the Hosmer-Lemeshow goodness of fit test, the analysis of deviance tables (Type III tests), and others.

4.4.1 Wald Test

coeffient	Estimate	Wald_Statistic	P_value	Significaant
(Intercept)	-9.46153688	-12.431832	0.000000e+00	TRUE
Blood.Sugar	0.03504576	10.147266	0.000000e+00	TRUE
BMI	0.10398848	6.183420	6.272742e-10	TRUE
Age	0.03145455	3.976817	6.984385e-05	TRUE

Figure 4. 16: Wald Test Results

Based on the results of the Wald test, it is evident that the intercept coefficient yielded a Wald statistic of -12.4318 with an extremely low p-value of 0.000000e+00, indicating its significant impact. Similarly, the coefficient for blood sugar demonstrated a Wald statistic of 10.147 with a p-value of 0.000000e+00, highlighting its statistical significance. Additionally, the coefficient for BMI produced a Wald statistic of 6.183, accompanied by a p-value of 6.2727e-10, signifying its significant impact on the outcome. Lastly, a p-value of 6.984e-05 and a Wald statistic of 3.976 for the age coefficient further supported its statistical significance. Based on the remarkably low p-values for each coefficient, it may be inferred that all the variables have a statistically significant effect on diabetes.

4.4.2 Analysis of deviance table (Type III test)

```
Analysis of Deviance Table (Type III tests)

Response: Outcome
            LR Chisq Df Pr(>Chisq)
Blood.Sugar  130.014  1  < 2.2e-16 ***
BMI           41.388  1  1.249e-10 ***
Age           15.914  1  6.627e-05 ***
```

Figure 4. 17: Analysis of Deviance Table Results

The logistic regression analysis reveals that all three variables - blood sugar, BMI, and age - are significant predictors of diabetes in Zimbabwe. Furthermore, each variable's intensity of correlation with the outcome is shown by likelihood ratio chi-squared statistics (LR Chisq), with blood sugar (130.014) and BMI (41.388) showing very strong relationships. Age (15.914) also shows a notable association. Degrees of freedom (df) for each variable are equal to 1, indicating that each variable was being treated as a single predictor. The p-values (Pr(>Chisq)) confirmed the significance of each variable, with blood sugar and BMI being highly significant ($p < 2.2e-16$ and $p = 1.249e-10$, respectively) and age being significant ($p = 6.627e-05$). These results suggest that blood sugar, BMI, and age are all important predictors of diabetes in Zimbabwe, with blood sugar and BMI playing particularly critical roles.

4.4.3 Test of Hosmer-Lemeshow Goodness of Fit

```
        Hosmer and Lemeshow goodness of fit (GOF) test

data:  model1$y, model1$fitted
X-squared = 12.722, df = 8, p-value = 0.1218
```

Figure 4. 18: Hosmer-Lemeshow Results

The Hosmer-Lemeshow test was used to assess the goodness of fit of the logistic regression model, and the results showed a chi-squared statistic of 12.722 with 8 degrees of freedom. Given that the observed values do not significantly deviate from the anticipated probabilities, the model appears to be a good match for the data, as indicated by the p-value of 0.1218. This shows that the model's utility for inference and prediction is supported by the fact that it accurately depicts the links between the predictor variables and diabetes in Zimbabwe. The model's conclusions are validated by the acceptable goodness of fit, which offers a solid basis for comprehending the variables linked to diabetes in this population.

4.4.4 Confusion matrix

```
Confusion Matrix and Statistics

          Actual
Predicted   0   1
        0 435 117
        1  65 151

               Accuracy : 0.763
                 95% CI : (0.7313, 0.7927)
    No Information Rate : 0.651
    P-Value [Acc > NIR] : 1.244e-11

                  Kappa : 0.4539

 Mcnemar's Test P-Value : 0.0001566

            Sensitivity : 0.8700
            Specificity : 0.5634
         Pos Pred Value : 0.7880
         Neg Pred Value : 0.6991
             Prevalence : 0.6510
         Detection Rate : 0.5664
   Detection Prevalence : 0.7188
      Balanced Accuracy : 0.7167

       'Positive' Class : 0
```

Figure 4. 19: Confusion Matrix

The performance of the logistic regression model in predicting diabetes in Zimbabwe was evaluated using a confusion matrix and various performance metrics. A confusion matrix provided 453 true negatives, 117 false positives, 65 false negatives, and 151 true positives, which translated to an accuracy of 0.763, indicating that 76.3% of predictions were correct. A 95% confidence range for accuracy of 0.7313 to 0.7927 and a no-information rate of 0.651 are two additional metrics used to assess the model's performance and demonstrate the improvement in predictive ability that the model provides. Additionally, the kappa statistic of 0.4539 indicated moderate agreement between predicted and actual classes, adjusted for chance. Finally, Mcnemar's test p-value of 0.0001566 indicated that the confusion matrix was significantly different from a random distribution, further supporting the model's effectiveness in predicting diabetes in Zimbabwe.

Sensitivity analysis reveals that the logistic regression model for predicting diabetes in Zimbabwe achieved a sensitivity of 0.87, indicating that 87% of true positives (individuals with diabetes) are correctly identified. However, the specificity is moderate at 0.5634, suggesting that 56.34% of true negatives (individuals with no diabetes) are correctly identified. The positive predicted value of 0.788 indicates that 78.8% of predicted positives are true positives, while the negative predicted value of 0.6991 shows that 69.91% of predicted negatives are true negatives. Furthermore, the population's prevalence of diabetes is 65.1%, and the model's detection prevalence is 71.88%, meaning that it can identify 56.64% of real diabetes cases. The balanced accuracy of 0.7167 provides a comprehensive measure of the model's overall performance, highlighting its effectiveness in predicting diabetes.

4.4.6 AUC-ROC test

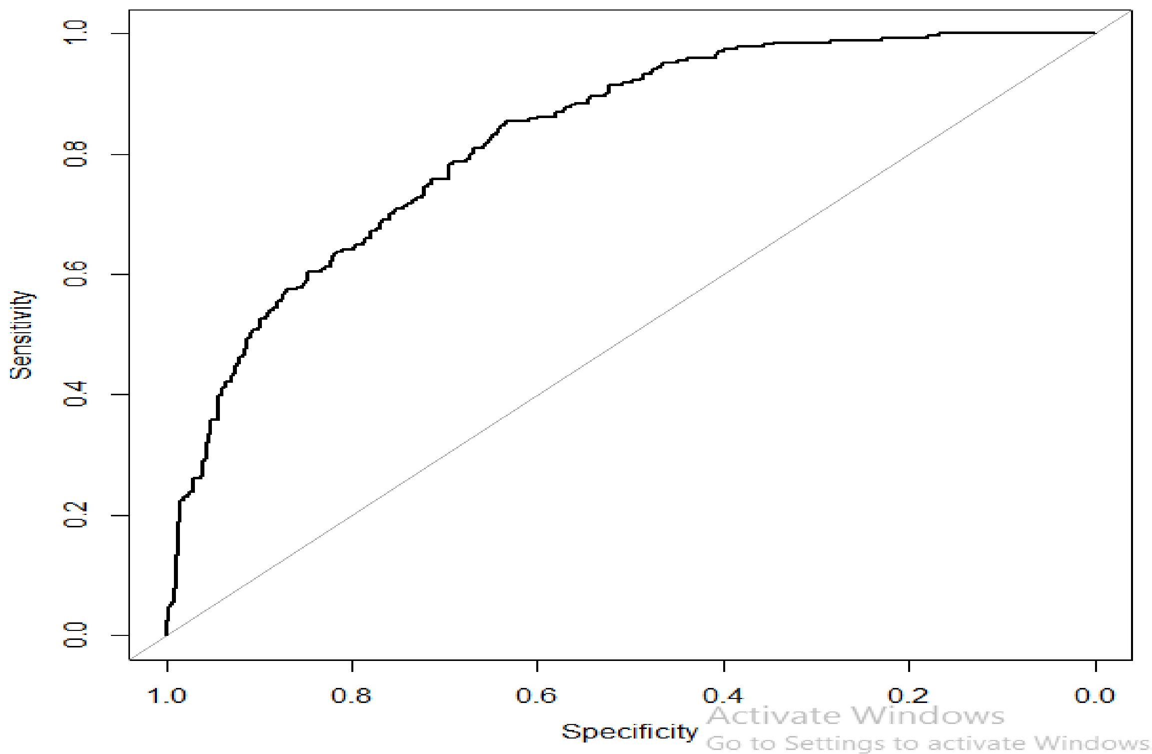

Figure 4. 20: AUC-ROC Curve

An essential assessment tool for evaluating a binary classification model's performance is the AUC-ROC (Area under the Receiver Operating Characteristic) test. By contrasting the true positive rate (sensitivity) with the false positive rate at different classification thresholds, the ROC curve—a graphical plot—illustrates the model's capacity for class discrimination. The area under the ROC curve, or AUC-ROC score, indicates how well the model can discriminate between positive and negative classifications.

```
'AUC 0.824686567164179
```

Figure 4. 21: AUC-ROC Test Results

The logistic regression model for predicting diabetes in Zimbabwe obtained an AUC of 0.82469, showing a high degree of accuracy in differentiating between people with diabetes and those without it, according to the AUC-ROC (Area under the Receiver Operating Characteristic Curve) study. The model's performance in classifying true positives (people with diabetes) and true negatives (those without diabetes) is measured by the AUC, where a value of 1 indicates perfect accuracy and a value of 0.5 indicates random chance. The likelihood of a p-value less than 0.05 for the AUC suggests that the model outperforms chance and has statistically significant predictive power. Furthermore, this result indicated that the model performs well in predicting who in Zimbabwe has diabetes, with an AUC of 82.5% indicating good predictive performance.

4.5 Summary

This chapter included an analysis, presentation, and discussion of the diabetes data that was gathered in Zimbabwe. Significant patterns, trends, and connections were found in the data, which shed light on the prevalence, causes, and effects of diabetes in Zimbabwe. The results demonstrate the complexity of diabetes in the nation by supporting certain parts of the literature review and research aims while contradicting others. The findings of this study advance knowledge about diabetes in Zimbabwe by laying the groundwork for further investigation and guiding the development of evidence-based therapies and policy. Additionally, the next chapter discuss the final conclusions and recommendations.

CHAPTER 5

SUMMARY, CONCLUSIONS AMD RECOMMENDATIONS

5.0 Introduction

In this chapter, the results of the statistical analysis of prevalence of diabetes in Zimbabwe are summarized, and recommendations for further study, legislation, and practice are made. The main findings of the study are succinctly summarized in this chapter, emphasizing their importance and implications. The primary findings are summarized in the conclusions section, and actionable lessons from the research are translated into practical strategies for stakeholders to alleviate the diabetes burden in Zimbabwe.

5.1 Summary of findings

Using secondary data from Chitungwiza Central Hospital, which included patients with diabetes between the ages of 21 and 65, the study performed statistical analysis and modeling of diabetes in Zimbabwe. In adult Zimbabweans, a researcher found a statistical correlation between diabetes and a number of risk factors. This required looking at the relationships between the incidence of diabetes and BMI, blood sugar, age, blood pressure, cholesterol, and gender, among other variables. According to the study, age, blood sugar, and BMI were the main indicators of diabetes in adult Zimbabweans. A high-risk profile for diabetes is characterized by a BMI of 30-44.6, blood sugar levels between 144-199, and ages ranging from 50-59.

These factors demonstrated a notable association with incidence of diabetes within the studied population. Conversely, blood pressure, cholesterol, and gender were identified as non-significant predictors, indicating a weaker correlation with the incidence of diabetes. Moreover, the researcher developed a logistic regression model to predict diabetes diagnosis based on the significant predictors identified. Through the use of a confusion matrix, the accuracy of the model was determined to be 76.3%. Additionally, the model's performance was evaluated using an AUC curve, which indicated an accuracy of 82%.

The findings provided valuable insights into the statistical landscape of prevalence of diabetes in Zimbabwe. Additionally, the identification of significant predictors such as BMI, blood sugar levels, and age offers a foundation for targeted interventions and preventive strategies. The performance of the logistic regression model underscores its potential as a tool for predicting diabetes diagnosis based on key risk factors, thereby contributing to more effective disease management and healthcare planning in Zimbabwe.

5.2 Conclusions

In the context of diabetes in Zimbabwe, the statistical analysis has revealed the critical significance of BMI, blood sugar, and age as determinants of this chronic condition, while gender, cholesterol, and blood pressure do not exhibit significant associations. Blood sugar was the most significant factor identified in the study with the lowest p-value and it needs effective blood sugar management through regular monitoring as well as personalized treatment regimens. Moreover, the findings shed light on the key factors that needs attention in the development of targeted interventions and management strategies for diabetes within the Zimbabwean population. With the understanding that age, blood sugar, and BMI are important factors when it comes to diabetes in Zimbabwe, policymakers and healthcare providers can better target their efforts to lessen the disease's effects on the populace and enhance the quality of life for those who suffer from it.

Additionally, the efficiency of logistic regression analysis in predicting diabetes reflects its potential to inform public health initiatives and resource allocation, thereby fostering more targeted and effective interventions within communities. Moreover, the utilization of logistic regression analysis stands as a pivotal asset in advancing proactive healthcare practices and enhancing the overall well-being of populations at risk of diabetes.

5.3 Recommendations

Individuals diagnosed with diabetes mellitus should follow prescribed medication schedules, maintain a nutritious diet, get regular exercise, and periodically check their blood sugar levels. They should also alter their way of living to include things like cutting back on alcohol and tobacco, controlling their stress, getting enough sleep, and keeping a healthy weight. The Zimbabwean government should create a comprehensive national plan for managing and preventing diabetes, as well as implementing population-wide diabetes screening programs, launching national health promotion and education campaigns, investing in diabetes clinics and healthcare facilities, improving training for medical professionals, and creating laws supporting physical activity and a healthy diet. The government should also set up counseling services and support organizations for people with diabetes. Individuals and the government may collaborate to effectively manage and prevent diabetes in Zimbabwe by adhering to these recommendations.

5.4 Areas for further research

More research is required in a number of crucial areas, even if this study offers valuable insights into the analysis of diabetes in Zimbabwe. Initially, it is important for future research to comprehend the part that social and cultural factors play in diabetes care and prevention in Zimbabwe. To further investigate the patterns and prevalence of diabetes complications as well as the influence of socioeconomic factors on diabetes prevalence in Zimbabwe, more study is required. Moreover, it is the hope of the researcher that future research is needed in health disparities related to diabetes, assessment of healthcare utilization patterns among individuals with diabetes, evaluation of the effectiveness of health promotion and prevention programs targeted at diabetes, and analysis of the economic burden of diabetes in Zimbabwe.

The areas for further research enhance the understanding of diabetes in Zimbabwe, inform evidence-based policies and interventions, and contribute to improved prevention, management, and healthcare delivery for individuals with diabetes in the country.

5.5 Summary

The statistical analysis of prevalence of diabetes in Zimbabwe was conducted at Chitungwiza Central Hospital, focusing on patients aged 21 to 65. The goal of this study was to develop a statistical model and analysis of diabetes in Zimbabwe in order to find better methods for managing and preventing the disease. Furthermore, the study's conclusions demonstrated that, in contrast to gender, blood pressure, and cholesterol, blood sugar, age, and BMI have a favorable association with diabetes. This study, which concentrated on the statistical analysis of diabetes in Zimbabwe, offered insightful information that can guide focused actions and public health policies meant to lessen the incidence and consequences of diabetes in that nation. To address this urgent public health issue, more research and coordinated efforts toward prevention and management are necessary.

REFERENCES

American Diabetes Association (2020). Understanding diabetes. Retrieved from https://www.diabetes.org/diabetes/understanding-diabetes

American Diabetes Association (ADA) (2017). Standards of Medical Care in Diabetes. Diabetes Care, 40(Suppl. 1), S1-S149.

American Diabetes Association (ADA). (2023). What is diabetes? Retrieved from https://www.diabetes.org/diabetes/what-is-diabetes

American Diabetes Association. Type 1 and Type 2 Diabetes (2021). Retrieved from https://www.diabetes.org/diabetes/types-of-diabetes

Balfour, D. L., Zaslavsky, A. M., & Stange, K. C. (2018). Effects of socioeconomic position on diabetes: Systematic review and meta-analysis. PLoS Medicine, 15(10).

Bandura, A. (2006). Social cognitive theory of mass communication. In J. Bryant & D. Zillmann (Eds.), Media effects: Advances in theory and research (3rd ed., pp. 121-153). Mahwah, NJ: Lawrence Erlbaum.

Bax, M., van den Bos, G. M., van Dijk, L., van Weel, C., Slijs, I., & Heine, R. J. (2017). Exploring the interplay between self-reported benefits and barriers in medical adherence: Results from a cross-sectional study in patients with type 2 diabetes. Patient Preference and Adherence, 11(6), 1385–1395.

Bernardo, J. M. (2022). 'The Forgotten Art of Statistical Modeling: From Selection of Variables to Model Checking and Effectual Prediction'. International Statistical Review; Volume 90(1), 86-102.

Bowers, D. (2020). Statistical Methods for Psychology, 6th ed. Boston: Cengage Learning.

Brennan, M. (2021). Multivariate Statistics for the Social Sciences, 6th Ed. New York: Routledge.

Brezo, J. A., Royal, C. D., Francis, A. D., & Headings, V. E. (2006). A general model of behavior for medication self-management in chronic illness. Medical Care, 44(9), 868-877.

Brodeur, A., Okhrin, Y., and Todd, R., (2021). Multicollinearity in econometrics: A survey. Foundations and Trends in Econometrics, Volume 6(1): 1-71.

Centers for Disease Control and Prevention (CDC) (2023). Diabetes: Types of Diabetes. Retrieved from https://www.cdc.gov/diabetes/basics/diabetes.html

Centers for Disease Control and Prevention (CDC) (2023). Diabetes: Types of Diabetes. Retrieved from https://www.cdc.gov/diabetes/basics/diabetes.html

Centers for Disease Control and Prevention. (2020). Social determinants of health. Retrieved from https://www.cdc.gov/socialdeterminants/index.htm.

Centers for Disease Control and Prevention. (2021). "Incidence." CDC.gov. Retrieved from https://www.cdc.gov/diabetes/statistics/basics/incidence

Crandall, S. M., & Doljanac, K. M. (2019). Social cognitive theory: application to diabetes and its management. Journal of Health Psychology, 24(7), 675-693. https://doi.org/

Cryer, J. D., and Chang, B. T., (2019). 'Data analysis with R: Learning statistics with R and R Studio'. Chemical Rubber Company Press.

Daniel, M. T., Pyykkö, I., Bakhtiari, A., Groop, L. C., & Zimmet, P., (2011). Diabetes at the dawn of the 21st century: Trends, challenges, and perspectives. Diabetes Research and Clinical Practice, 91(3), 245-256.

Field, A. P., 2014. An introduction to statistical methods, 5th Ed. Sage. London: Thousand Oaks, CA.

Geboers, H., Ruiz, C., Hasthorpe, M., Tran, D., Dhana, N., & Grant, J. (2017). Applying the health belief model to enhance health literacy among immigrant populations: A randomized controlled trial. Health Education & Behavior, 44(4), 409

Glanz, K., Rimer, B. K., & Viswanath, K. (2008). Health behavior and health education: Theory, research, and practice. San Francisco, CA: Jossey-Bass.

Glanz, K., Rimer, B. K., & Viswanath, K. (2015). Health behavior and health education: Theory, research, and practice

Gupta, S. N., & Gupta, M. M., 2016. 'Logistic Regression and Chi-Square Test for Goodness of Fit in the Presence of Censored Data. Advances in Systems, Control, and Applications', Volume 8, 273-281.

Hales, C., Garcia, A., Harris, S., & Watkins, S. (2018). A socioecological model for the management of type 2 diabetes: What will be most beneficial for patients? Journal of the American Association of Nurse Practitioners, 30(5), 360–367.

Hjelm, K., & Mufunda, E. (2012). Diabetes in Africa: An impending public health disaster? Acta Paediatrica, 101(8), 727-731.

Hosmer, D. W., and Lemeshow, S., (2021). Applied Logistic Regression, 4th Ed. Hoboken, New Jersey: John Wiley and Sons.

Hosmer, D. W., Lemeshow, S., and May, S., (2017). Applied Logistic Regression, 3rd ed. Hoboken, New Jersey: John Wiley and Sons, lnc.

Humphreys, B. and Kausel, E. (2018). A Practical Guide to Basic Modeling Skills in Science and Engineering, Princeton, NJ: Princeton University Press.

Institute for Health Metrics and Evaluation (IHME) (2020). Global Burden of Disease: Diabetics (High-level Summary) (Version 2020). Seattle, WA: Institute for Health Metrics and Evaluation. Retrieved from https://ghdx.healthdata.org/gbd-results-tool

Insulin (2018). American Diabetes Association. Retrieved from https://diabetes.org/diabetes/medications/insulin

International Diabetes Federation, 2013. Global diabetes statistics. Brussels, Belgium: International Diabetes Federation. Available at: https://www.idf.org/e-library/epidemiology-research/global-diabetes-statistics.

Kalra, R., Sethi, D., & Bhutani, G. (2021). The health belief model: A theoretical update. International Journal of Medical Research & Health Sciences, 10(2), 46–51

Kampfumhukutu, M., Muula, A., Dao, L., & Siziya, S. (2017). Analysing diabetes as a time series in Zimbabwe. Journal of Global Health, 7(1), 010601.

King, H. R. E., Bain, G. A., & McMillan, J. (1998). Relationship between macrovascular and microvascular complications and quality of life in diabetes: results from the UK Prospective Diabetes Study (UKPDS 36). Diabetes Care, 21(10), 1624-1629.

Kruk, M. E., Dunn, R. L., Stewart, A. L., Kirtland, K., & Perencevich, E. N. (2018). Theoretical frameworks for improving the uptake of patient safety initiatives. Patient Safety in Surgery, 12(1), 231.

Ladefoged, O., Pedersen, L. F., Hansen, T., Kyvik, K. O., Pluess, M., Trombetta, C., and Hansen, J. T. (2015). The importance of genetic and environmental factors in childhood for risk of type 1 and type 2 diabetes: A Danish population-based study. Diabetologia, 58(4), 803-811.

Lancet Diabetes Endocrinol (2014). 'Global Burden of Metabolic Risk Factors for Chronic Diseases Collaboration. Cardiovascular disease, chronic kidney disease, and diabetes mortality burden of cardio-metabolic risk factors from 1980 to 2010: A comparative risk assessment'. Volume2 (8):634–647.

Li, N., Li, B., Zhou, C., Cui, Q., Zhang, X., Yao, Z., & Xiong, S. (2018). Social support and health outcomes in type 2 diabetes mellitus: a systematic review and meta-analysis. PLOS ONE, 13(2), e0192396. https://doi.org/10.1371/journal.

Li, X., Yin, L., Yang, H., Chen, L., Jia, W., & Li, W. (2019). Socioeconomic status, biopsychosocial factors, and type 2 diabetes incidence: A systematic review and meta-analysis. BMC Public Health, 19(1), 1-15. https://doi.org/

Lochner, F. M., & O'Campo, P. (2019). The social determinants of health: How do they affect diabetes? Current Diabetes Reports, 19(5), 11.

Lohr, S. M., and Shih, H., 2017. 'Data presentation: A practical guide for creating effective charts, tables, and presentations'. Wiley.

Madebwe, T., Chadambuka, C., & Shumba, D. (2017). Prevalence and risk factors of diabetes mellitus among hypertensive patients in Harare, Zimbabwe. International Journal of Diabetes and Endocrinology, 2(3), 1-6.

Mayo Clinic. (2021, December 29). Type 1 diabetes. Retrieved from https://www.mayoclinic.org/diseases-conditions/type-1-diabetes/symptoms-causes/syc-20353011

Mbanya, J. N. C. (2015). Diabetes in Africa: A continent-wide perspective. Diabetes Research and Clinical Practice, 102, 1-4.

McCombe, R., LaPelle, J. C., and Gavurin, D. (2019). The process of preparing a research design. Research Design Review, Volume 4, 1-3.

McCulloch, S. G., Weiner, D. J., & Chew, G. T. (2019). Community health: A population-based nursing approach (4th Ed.). Jones & Bartlett Learning.

McEwen, L. M., Walsh, B. T., Hebert, P. L., & Young, B. A. (2014). Diabetes self-management intervention development using the socioecological model of health behavior: A critical review. Health Promotion Practice, 15(3), 380–388.

McGavock, E. M., Francioli, E., Nourse, J. W., & Zimmet, P. Z. (2021). Worldwide burden and variation in diabetes prevalence from 1980 to 2019. Diabetologia, 64(7), 1618-1627.

Meisner, C., Pershing, L., O'Brien, T., Yeh, S. (2020). Applying the health belief model to improve medication adherence among African American patients with diabetes. Diabetes Spectrum, 33(2), 146-154.

Mudiay, T. K. (1997). Prevalence of diabetes mellitus in African populations: A review. Diabetes Research and Clinical Practice, 38(2), S55-S63.

Mufunda, J. (2006). Diabetes in Africa: Underdiagnosis and inadequate treatment. Diabetes Research and Clinical Practice, 74(1), 58-66.

Nagoya, Y., and Russell, K. (2018). Application of social determinants of health model in diabetes. The Journal of the American Medical Association, 316(3), 296-298.

Nagoya, Y., and Russell, K. (2018). The biopsychosocial model in public health interventions. Oxford University Press, 2(1), 1-11.

National Institute of Diabetes and Digestive and Kidney Diseases. (2021). "Diabetes statistics." Retrieved from https://www.niddk.nih.gov/health-information/diabetes/overview/statistics

Neal, K., Mellor, D., & Nettleton, C. (2018). The social determinants of health in diabetes. The Lancet, 392(10164), 1749–1762.

Nyabadza, T., Guzha, N., Murphy, M., Demmer, T., Nhachi, K., Nyoni, S., ... & Nhachi, T. (2017). Prevalence and determinants of diabetes and impaired glucose tolerance in Zimbabwe. International Journal of Environmental Research and Public Health, 14(2), 131.

Nyika, C., Shumba, D., Manhambo, N., & Chinangwa, R. (2016). A statistical analysis of diabetes mellitus in Zimbabwe. African Journal of Health Sciences, 16(1), 30-38.

Pan American Health Organization. (2023). "Diabetes." Pan American Health Organization. Retrieved from https://www.paho.org/en/topics/diabetes.

Radhouane, W., Ardenghi, D., Amoussou-Guenou, F., Ataguba, J., Bernabe, F., Ickx, S., ... & Ghassoumi, M. (2020). Prevalence and predictors of type 2 diabetes mellitus in Sub-Saharan Africa: a systematic review and meta-analysis. Diabetes, Obesity and Metabolism, 22(1), 19-28.

Rush, A. J., & Stevens, A. (2013). The biopsychosocial model: A useful framework for understanding diabetes-related distress and its management. Journal of Diabetes and Its Complications, 27(2), 132-139. https://doi.org/

Ryan, C. L., Kim, Y. M., & Kratz, A. S. (2016). A multi-theoretical framework for predicting and intervening in diabetes self-management: The biopsychosocial approach. Annals of Behavioral Medicine, 50(4), 244-260

Salkind, N. J. (2018). "Introduction to Data Analysis." The SAGE Handbook of Data Analysis (2nd ed., pp. 3-22). Sage Publications.

Schoenberg, B. A., Cheng, C. W., Laffel, L. M., & Raskin, P. (2018). The social determinants of diabetes: An important opportunity for patients and healthcare professionals. Diabetes Therapy, 9, 619–633.

Shaw, J. E., Reynolds, K. J., Zimmet, P. Z., Amann, M., Barnett, A. H., Beagley, J., & Dunger, D. B. (2017). The global burden of diabetes: an updated systematic analysis of the incidence, prevalence, and mortality of diabetes by country for 2013. Lancet Global Health, 5(6), e844–e858.

Sperl-Hillen, J., Karabatsos, G., Lawder, D., Linn, C., Tolbert-Burt, T., & Lloyd, L. (2018). Improving community-based diabetes care delivery using the socioecological model. Journal of Community Health, 43(1), 81-92.

Tanyanyiwa, D., Pasipamire, S., Zhou, J., Murambadoro, T., Chadambuka, C., Bagonza, G., & Chimbari, P. (2019). Incidence and mortality trends of diabetes mellitus in Zimbabwe: a modelling study. International Journal of Health Services, 49(3), 315-322.

Vallejo, N., Molina, E., Bosque-Perez, K., Lesko, M., Stein, R., Gorelick, P., Smith, M., Bhatt, S. (2019). Use of the health belief model to improve diabetes self-management among adults with limited literacy skills. Diabetes Research and Clinical Practice, 154, 105873.

Wallenstein, S. (2020). Introductory Econometrics: A Modern Approach, 8th ed. Upper Saddle River, NJ: Pearson.

White, N. (2019). The health belief model: A review of its application and methodological issues. Frontiers in Psychology, 10, 1–13.

Whiting, D. R. (2016). Diabetes in developing countries: Key facts. Journal of Endocrinological Investigation, 39(8), 713-715.

Wild, S., Roglic, G., Green, A., Sicree, R., & King, H. (2004). Global prevalence of diabetes: Estimates for the year 2000 and projections for 2030. Diabetes Care, 27(5), 1047-1053.

World Health Organisation (2016). Global report on diabetes. Geneva, Switzerland: World Health Organisation.

World Health Organization (2005). Diabetes: A challenge for developing countries. In A World Health Report: Make Every Mother and Child Count. Geneva: World Health Organization, pp. 100-108. Retrieved from https://www.who.int/publications/i/item/WHO-RHR-06.01/downloads/diabetes.pdf .

World Health Organization (2021). Global diabetes epidemic: Diabetes factsheet. Retrieved from https://www.who.int/diabetes/publications/diabetes-report-2021/en

World Health Organization (WHO) (2021). Global Diabetes Epidemic: More than 425 million adults are living with diabetes. Retrieved from https://www.who

World Health Organization. (2017). Global report on diabetes. Geneva: World Health Organization.

World Health Organization. (2021). "Diabetes factsheet." World Health Organization. Retrieved from https://www.who.int/news-room/fact-sheets/detail/diabetes

World Health Organization. (2021). "Glossary of terms." World Health Organization. Retrieved from https://www.who.int/health-topics/glossary

World Health Organization. (2022). What is health? Retrieved from https://www.who.int/news-room/fact-sheets/detail/health

World Health Organization. (2023). "Diabetes - WHO." World Health Organization. Retrieved from https://www.who.int/news-room/fact-sheets/detail/diabetes

Zhang, Y., Xu, J., He, X., Guo, X., Wang, X., Shen, Z., & Liu, J. (2017). Prevalence of diabetes mellitus in low- and middle-income countries: a systematic review and meta-analysis. PLoS One, 12(1).

Zimmet P (2000)**:** Globalization, coca-colonization and the chronic disease epidemic: can the Doomsday scenario be averted? Journal of Internal Medicine. pp 247:301-310.

APPENDIX

The whole codes used in the data analysis and model building is found below:

Read Data

```
DiabetesData<-read.csv (file.choose (),header = T)
```

Descriptive Statistics

```
summary (DiabetesData)
```

Outlier Detection

```
boxplot (DiabetesData$Blood.Sugar,main="BOXPLOT OF BLOOD SUGAR")
boxplot(DiabetesData$Blood.Pressure,main="BOXPLOT OF BLOOD PRESSURE")
boxplot(DiabetesData$BMI,main="BOXPLOT OF BMI")
boxplot(DiabetesData$Cholesterol,main="BOXPLOT OF CHOLESTEROL")
boxplot (DiabetesData$Age,main="BOXPLOT OF AGE")
```

Correlation Matrix

```
library (stats)
corr_matrix<-cor(DiabetesData)
print(corr_matrix)
```

Logistic Regression Full Model

```
library(car)
```

```
model<-glm(Outcome~Blood.Sugar+Blood.Pressure+BMI+Cholesterol+Age+Gender,data =
DiabetesData,family = "binomial")

summary(model)
```

Variance Inflation Factor

```
vif(model)
```

Durbin-Waston Test

```
library(lmtest)

dwtest(model)
```

Logistic Regression Reduced Model

```
model1<-glm(Outcome~Blood.Sugar+BMI+Age,data = DiabetesData,family = "binomial")
```

Comparison of predictors of diabetes

```
library(stats)

library(caret)


DiabetesData$Age_cat<-cut(DiabetesData$Age,breaks = c(21,30,40,50,60,65),

                       labels = c("21-29","30-39","40-49","50-59","60-65"))

model2<-glm(Outcome~Age_cat,data = DiabetesData,family = "binomial")

summary(model2)

coef(summary(model2))
```

```
DiabetesData$Blood.Sugar_cat<-cut(DiabetesData$Blood.Sugar,
                        breaks = c(44,100,140,199),
                  labels = c("44-99","100-143","144-199"))
model3<-glm(Outcome~Blood.Sugar_cat,data = DiabetesData,family = "binomial")


summary(model3)
coef(summary(model3))


DiabetesData$BMI_cat<-cut(DiabetesData$BMI,breaks = c(18.2,24.9,29.9,44.6),
                  labels = c("18.2-24.9","25-29.9","30-44.6"))
model4<-glm(Outcome~BMI_cat,data = DiabetesData,family = "binomial")
summary(model4)
coef(summary(model4))
```

Wald Test

```
coeffients<-coef(model1)
vcov_matrix1<-vcov(model1)
wald_statistic<-coeffients/sqrt(diag(vcov_matrix1))
p_value<-2*(1-pnorm(abs(wald_statistic)))
signicant<-p_value<0.05
print(data.frame(coefficient=names(coeffients),Estimate=coeffients,Wald_Statistic=wald_statisti
c,P_value=p_value,Significant=signicant))
```

Analysis of Deviance (Type 111)

```
library(car)

Anova(model1,type = "III")
```

Hosmer-Lemeshow Goodness of Fit

```
library(ResourceSelection)

hosmerlem<-hoslem.test(model1$y,model1$fitted,g=10)

print(hosmerlem)
```

Confusion Matrix

```
library(caret)

predictions<-predict(model1,type = "response")

predictions_class<-ifelse(predictions>0.5,1,0)

true_labels<-ifelse(DiabetesData$Outcome==1,1,0)

confusion_matrix<-confusionMatrix(table(Predicted=predictions_class,Actual=true_labels))
```

AUC-ROC Curve

```
library(pROC)

predictions<-predict(model1,type = "response")

true_labels<-ifelse(DiabetesData$Outcome==1,1,0)

roc_curve<-roc(true_labels,predictions)

plot(roc_curve)

auc<-auc(roc_curve)
```

```
print(paste("AUC",auc))
```