A STATISTICAL ANALYSIS ON PREDICTING STUDENTS' DROPOUT RATE

PRESENTED BY

OKAFOR PETER CHUKWUEMEKA

2021554012

A PROJECT SUBMITTED TO THE DEPARTMENT OF STATISTICS

FACULTY OF PHYSICAL SCIENCES, NNAMDI AZIKIWE UNIVERSITY, AWKA

IN PARTIAL FULFILMENT FOR THE AWARD OF BACHELOR OF SCIENCE DEGREE IN STATISTICS

SUPERVISOR: MR CHUKWUENYEM ONYENEKWE

APRIL 2025

CERTIFICATION

This is to certify that this work "**STATISTICAL ANALYSIS OF STUDENTS DROPOUT RATE"** is the original work of Okafor Peter Chukwuemeka and has not been used in full or part for the award of any degree or diploma elsewhere

_____ _____

Okafor Peter Chukwuemeka **Date**

DEDICATION

This work is dedicated to the Almighty God for his Grace and blessings of wisdom and knowledge toward the completion of this project. Special dedication also to my Parents (Mr. and Mrs. Charles Okafor) and my siblings for their support

ACKNOWLEDGEMENT

I would like to express my profound gratitude to those who have been instrumental in the completion of this project.

First and foremost, I extend my sincere appreciation to my parents for their unwavering support, encouragement, and patience throughout my academic journey. Their belief in my capabilities and their constant reassurance during challenging times have been the cornerstone of my perseverance.

I am deeply indebted to my lecturers for their expert guidance, constructive criticism, and generous sharing of knowledge. Their commitment to academic excellence has significantly shaped my understanding of the subject matter and enhanced the quality of this project. Their mentorship has been invaluable in developing my analytical skills and research abilities.

I would also like to acknowledge the academic institution for providing the necessary resources and infrastructure that facilitated the smooth execution of this project.

This accomplishment would not have been possible without the collective support of these individuals. Their contributions have not only helped me complete this project but have also enriched my learning experience.

ABSTRACT

This research project examines the dropout rates among students using descriptive statistical methods. The study analyzes patterns and trends in student withdrawal from educational programs to better understand the scope and nature of this issue. By collecting and analyzing data on student enrollment and departure, this research identifies key factors that may contribute to students leaving their studies before completion. The analysis includes measures such as frequency distributions, central tendency (mean, median, mode), and variability (standard deviation, range) to provide a clear picture of dropout patterns. The findings from this study offer valuable insights for educational institutions seeking to develop effective retention strategies and support systems. This research contributes to the broader understanding of student persistence in educational settings and provides a foundation for more targeted interventions to reduce dropout rates.

TABLE OF CONTENTS

CHAPTER ONE

BACKGROUND OF THE STORY

The Problem of Student Dropout

Many students leave school before finishing their education. This is called "student dropout." When students drop out, it creates problems for the students themselves, for schools, and for society. Students who drop out often earn less money in their lives and have fewer job opportunities. Schools lose funding and may be seen as failing. Communities suffer when fewer people complete their education.

Why We Need to Predict Dropout

In the past, schools would only know students had dropped out after they were already gone. By then, it was too late to help them. Today, schools are trying to find ways to predict which students might drop out before it happens. This way, they can help these students early and prevent them from leaving.

Predicting student dropout means using information about students to guess who might leave school in the future. This is different from just counting how many students have already dropped out.

Using Descriptive Statistics as a Tool

Descriptive statistics is a tool that helps us organize and make sense of information about students. It includes simple ways to summarize data, like:

- Averages (mean, median, mode)
- How spread out the data is (range, standard deviation)
- How often certain things happen (frequency)
- How different factors might be connected (correlation)

These simple statistics help us see patterns in student data that might tell us who is at risk of dropping out.

STATEMENT OF THE PROBLEM

Educational institutions face significant challenges with students leaving before completing their programs. Despite efforts to improve retention, dropout rates remain a persistent issue across various educational levels. When students leave school prematurely, this has negative consequences for the students themselves, educational institutions, and society at large.

The primary problem is that most educational institutions rely on reactive approaches to address student dropout. They typically become aware of dropout issues only after students have already left the institution, at which point intervention is no longer possible. This reactive approach fails to utilize the potential benefit of early identification and timely support that could prevent students from leaving.

Additionally, many institutions lack systematic methods to identify students who are at risk of dropping out. While academic records and attendance data are often available, educational administrators frequently lack the tools and methodologies to effectively analyze this information to predict future dropout behavior.

Furthermore, existing prediction models for student dropout are often overly complex, requiring specialized technical knowledge and sophisticated software that many educational institutions cannot access or afford. These complex models, while potentially accurate, present barriers to implementation in real educational settings where simplicity and usability are essential.

The absence of accessible predictive tools means that limited resources for retention efforts are not optimally allocated. Without the ability to identify which students are most at risk, institutions must either implement broad interventions that may not target those who need them most or rely on teacher intuition, which may miss less obvious cases.

OBJECTIVES OF THE STUDY

To develop a predictive model for student dropout rates using descriptive statistical methods that educational institutions can implement as an early warning system.

Our Specific Objectives

1. To identify the most significant statistical indicators that correlate with student dropout behavior across various educational contexts.
2. To analyze the relationship between academic performance metrics (including grades, attendance records, and course participation) and the likelihood of student dropout.
3. To examine how demographic factors contribute to dropout risk and determine which factors have the strongest predictive capability.
4. To design a practical framework that combines relevant descriptive statistical measures into a cohesive prediction model accessible to educational practitioners.
5. To establish threshold values for key indicators that can effectively distinguish between students at high, moderate, and low risk of dropping out.
6. To evaluate the predictive accuracy of the developed statistical model across different educational settings.
7. To formulate implementation guidelines for educational institutions to translate the statistical findings into actionable intervention strategies.
8. To recommend specific early intervention approaches based on the patterns identified through the descriptive statistical analysis.

Scope and Limitations

Scope of the Study

This research focuses specifically on developing a predictive model for student dropout rates using descriptive statistical methods. The scope encompasses the identification, analysis, and application of key statistical indicators that can effectively predict students at risk of dropping out before they leave their educational programs.

The study will examine academic performance metrics including course grades, attendance records, assignment completion rates, and participation levels as potential predictors of dropout behavior. Additionally, demographic factors such as age, gender, socioeconomic status, and family background will be analyzed to determine their predictive value.

The research will be conducted within Nnamdi Azikiwe University, focusing on students enrolled in the university during the academic years 2021/2022. The sample will include approximately 10,000 students from diverse backgrounds to ensure the developed model has broad applicability.

In terms of statistical methodology, the study will employ fundamental descriptive statistical techniques including measures of central tendency, dispersion, correlation analysis, and frequency distributions. The research deliberately constrains its analytical approach to these accessible methods to ensure the resulting predictive model can be implemented by educational practitioners without specialized statistical training.

The scope extends to the development of an early warning system framework based on the statistical findings, including the establishment of threshold values for key indicators that signal varying levels of dropout risk. The study will also provide implementation guidelines for educational institutions to translate the statistical model into practical intervention strategies.

Limitations

While comprehensive within its defined parameters, this study acknowledges several limitations. The research does not employ advanced predictive algorithms or machine learning techniques, as these fall outside the intended

focus on accessible descriptive statistics. The predictive power of the model may therefore be somewhat limited compared to more complex approaches.

The study is confined to quantifiable metrics available through institutional records and standardized assessments. Qualitative factors such as student motivation, educational aspirations, and psychological well-being, while potentially significant in dropout decisions, are beyond the scope of this research except where they can be captured through quantitative proxies.

Geographic and institutional specificities may limit the generalizability of findings to dissimilar educational contexts. The model's predictive accuracy may vary when applied to dramatically different educational systems or cultural settings.

Lastly, while the study will propose intervention strategies based on the predictive findings, the implementation and evaluation of these interventions fall outside the current research scope and would require subsequent investigation.

SIGNIFICANCE OF THE STUDY

This research on predicting student dropout rates using descriptive statistics carries significant implications for various stakeholders in the educational ecosystem. The study's contributions span practical applications, policy development, and academic advancement.

Educational Institutions

For educational institutions, this research provides a practical, accessible methodology to identify students at risk of dropping out before they leave. By employing descriptive statistical techniques that do not require specialized technical expertise, schools and colleges can implement early warning systems regardless of their resource constraints. This democratization of predictive capabilities enables institutions to allocate retention resources more efficiently, focusing interventions on students with the highest dropout risk. The resulting improvement in retention rates can enhance institutional performance metrics, reputation, and financial sustainability.

Educational Practitioners

Teachers, counselors, and administrators stand to benefit significantly from this research as it equips them with clear, interpretable indicators of student risk. The study translates complex student data into actionable insights that can inform day-to-day interactions with vulnerable students. By identifying specific threshold values for key metrics, practitioners gain concrete guidelines for when and how to intervene, allowing for more personalized and timely support strategies.

Students and Families

The ultimate beneficiaries of this research are students who might otherwise drop out of their educational programs. Early identification and intervention can significantly alter educational trajectories, leading to improved academic outcomes, greater career opportunities, and enhanced lifetime earnings. For families, particularly those with limited educational background who may not recognize early warning signs, the structured identification system developed through this research serves as a critical safety net.

Policy Makers

At the policy level, this research contributes valuable insights for developing evidence-based retention strategies. By identifying the most significant predictors of dropout behavior, policy makers can design targeted initiatives that address root causes rather than symptoms. The statistical model provides a framework for assessing the potential impact of various policy interventions and for monitoring their effectiveness over time.

Academic Community

For the academic community, this study bridges an important gap between theoretical dropout prediction models and practical application. By demonstrating how descriptive statistics can be effectively employed for predictive purposes, the research challenges the assumption that only complex algorithms can yield useful predictions. This contribution to the methodology of educational research may inspire similar approaches in other areas where accessibility and practicality are priorities.

Economic and Social Impact

From a broader societal perspective, reducing student dropout rates through improved prediction and intervention has substantial economic and social benefits. Higher educational attainment correlates with reduced unemployment, lower incarceration rates, better health outcomes, and greater civic participation. The economic return on investment from preventing dropouts significantly exceeds the cost of implementing early warning systems, making this research valuable from a cost-benefit perspective.

In summary, this study's significance lies in its potential to transform reactive dropout management into proactive retention strategies through accessible statistical methods, benefiting individuals, institutions, and society at large.

DEFINITION OF TERMS

Student Dropout: The permanent withdrawal of a student from an educational institution before completing their intended program or course of study. In this research, a dropout is operationally defined as a student who has not attended classes or participated in academic activities for a continuous period of 30 days without formal withdrawal or transfer documentation.

Dropout Rate: The percentage of students who leave an educational institution without completing their program within a specified time frame. For this study, dropout rate is calculated as the number of dropouts divided by the total number of enrolled students in a given cohort, multiplied by 100.

Descriptive Statistics: Mathematical methods used to summarize, organize, and present data in a meaningful way. In this study, descriptive statistics include measures of central tendency (mean, median, mode), measures of dispersion (standard deviation, range, variance), frequency distributions, and correlation coefficients.

Predictive Model: A statistical framework that uses historical data to identify patterns and relationships for the purpose of forecasting future outcomes. In this research, the predictive model refers specifically to the structured approach using descriptive statistics to estimate the probability of student dropout.

At-Risk Student: A student who demonstrates characteristics or behaviors that have been statistically associated with an increased likelihood of dropping out. Within this study, at-risk students are those whose statistical indicators exceed the established threshold values for dropout risk.

Early Warning System: A systematic approach to identifying students who show signs of academic or engagement difficulties that may lead to dropout. In this research context, the early warning system refers to the practical implementation of the predictive model that flags students based on statistical indicators.

Academic Performance Metrics: Quantifiable measures of student achievement and engagement, including but not limited to grades, grade point average (GPA), assignment completion rates, test scores, and class participation records.

Retention: The continued enrollment of students in consecutive terms until program completion. For this study, retention is measured as the percentage of students who remain enrolled from one academic period to the next.

Intervention Strategies: Targeted actions or programs designed to address specific risk factors and prevent student dropout. In this research, intervention strategies refer to the evidence-based approaches recommended based on the findings of the predictive model.

Threshold Value: A specific cut-off point in a statistical measure that distinguishes between different levels of dropout risk. In this study, threshold values are established for each predictive indicator to categorize students into low, moderate, and high-risk groups.

Socioeconomic Status (SES): A measure of a student's or family's economic and social position in relation to others. In this research, SES is operationalized through indicators such as family income, parental education level, parental occupation, and access to resources.

Predictive Accuracy: The degree to which the statistical model correctly identifies students who will eventually drop out. For this study, predictive accuracy is measured through the percentage of correctly classified cases when the model is applied to historical data with known outcomes.