ANALYSIS OF TIME TAKEN FOR A CASE TO BE DETERMINED IN HIGH COURTS OF KENYA

ALEX KANGOGO KANDIE

A RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN STATISTICS OF THE UNIVERSITY OF EMBU

AUGUST, 2025

DECLARATION

This research proposal is my original work and has not been presented elsewhere for a degree or any other award.

Signature…………………………………… Date……………………………… 28/8/2025

Alex Kangogo Kandie

Department of Mathematics and Statistics

B531/1346/2019

This research project has been submitted for examination with our approval as the University Supervisors.

Signature…………………………………… Date……………………………… 28/8/2025

Dr. Edwin Atitwa

Lecturer, Department of Mathematics and Statistics

University of Embu

Signature…………………………………… Date. ……………………………… 28/8/2025

Dr. Wesley Nyaluke

Lecturer, Department of Mathematics and Statistics

University of Embu

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMN

AIC	Akaike Information Criterion
Cox PH	Cox Proportional Hazard
EDA	Exploratory Data Analysis
FIM	Fisher Information Matrix
ICT	Information Communication Technology
IQRs	Inter-Quartile Ranges
KM	Kaplan Meier
KNBS	Kenya National Bureau of Statistics
MLE	Maximum Likelihood Estimator

LIST OF SYMBOLS

β	beta
θ	theta
ϑ	theta variant
μ	mu
∞	infinity
σ	sigma
Γ	gamma
δ	delta
λ	lambda
α	alpha

ABSTRACT

The Kenyan High Courts face persistent case backlogs, undermining the constitutional mandate for timely justice delivery. This study sought to investigate case resolution time in Kenyan high courts on cases from 2012 to 2022 to identify factors causing delays and propose strategies for improving judicial efficiency. The objectives were to conduct exploratory data analysis to identify factors affecting case determination times, predict resolution times using survival analysis, and compare the precision of Cox Proportional Hazards and parametric models (Weibull, Exponential, Log-normal, Log-logistic, and Gamma). The study analysed 92,405 case records from 40 High Courts in 40 Kenyan counties that had established High Courts as of the study period. The data was sourced from Judiciary Headquarters at Milimani with approval from the Registrar of High Courts. Survival analysis methods, including Kaplan-Meier estimators and Cox and parametric models, assessed case type (Anti-Corruption, Civil, Constitutional, Criminal, and Family) and appeal status, with model performance evaluated using the Akaike Information Criterion (AIC). Results showed Family cases had the longest mean duration (40.19 months, median = 35 months), followed by Civil cases (30.89 months, median = 23 months). Anti-Corruption (14.22 months, median = 8 months) and Criminal cases (15.91 months, median = 8 months) resolved faster, while Constitutional cases averaged 21.54 months (median = 14 months). Appealed cases took longer (mean = 30.08 months, median = 24 months) than non-appealed cases (mean = 26.22 months, median = 15 months, $p < 0.001$). The Weibull model outperformed others (AIC = 713,383.0) compared to the Cox model (AIC = 1,668,250.5), excelling in modelling skewed durations. Recommendations include establishing specialized Family and Civil case divisions with mandatory pre-trial mediation, enforcing strict appellate timelines, adopting the Weibull model for predictive case management, increasing judge recruitment, enhancing training, and expanding and fully implementing the Judiciary Case Management System (JCMS) with a centralized data warehouse to ensure real-time monitoring and data integrity. Promoting alternative dispute resolution can further reduce court congestion, aligning with constitutional goals for timely justice.

CHAPTER ONE

INTRODUCTION

1.1 Background Information

Analysis of time taken during judicial process or procedure is due to its effects on the economy and social welfare. These consequences further extend its impact to the judicial policy significance. Delay in judicial case determination has been branded as low performance or non-performing judicial system. This has not only led to stagnation of the parties involved and loss of significant resources but also eroded public trust in the system. Bielen et al., 2017 argued that overly long duration of court cases increases private and public legal expenditures and perpetuate the uncertainty faced by the disputing parties, redistributes wealth from plaintiffs to defendants, and provides incentives for vexatious litigation which in turn further increases court backlogs.

Tsintzos and Plakandaras (2020) empirical findings, for 22 countries of the EU over the period 2010–2017, suggest that the judicial system and more specifically the judicial budget positively affect productivity. In this regard, governments have strived hard in ensuring that the Judiciary operations run smoothly by making Judiciary an independent entity, allocating enough funds and giving all support required by the Judiciary in delivering its mandate.

In Kenya, this has been a great concern to all levels and arms of government. In the recent past, the Kenyan Judiciary has tried hard in reducing the number of cases at its disposal as evidenced in the establishment of high courts in almost every county and the appointment of Judges. Furthermore, the Judiciary in its efforts to try and reduce the case backlog and correct capturing of data/information, it has integrated ICT in its day-to-day activities. However, despite all these efforts put in place by the government and the Judiciary in particular, access to timely justice has remained a challenge.

According to KNBS (2020), any case that has not been solved within a year is termed as case backlog hence longer time in case determination which has in turn led to huge accumulation of cases in the courts leading to huge case backlogs.

The study on the duration taken to determine the judicial cases is minimal globally due to the complexity associated with gathering the data required. Removing the incomplete cases, the information contained in the unfinished cases is discarded and therefore the fundamental principle of statistics on using all the information is not met and, in this case, survival analysis is considered the best, as it would make sure it utilizes all the information at place as it would be described as ‘censored’ scenarios in order to guarantee data homogeneity. There are different types of cases that are presented in Kenyan high courts. These cases include;

Civil Matters - cases that involve two or more parties that arise from damage, injury, breach of contract, unjust enrichment and other situations.

Criminal Matters - cases brought against individuals accused of engaging in criminal activities such as homicide, assault, fraud and drug possession.

Land Disputes - cases that involve disputes about ownership and/or use of land.

Constitutional Law Matters - cases related to the interpretation of the Constitution.

Employment Disputes - cases involving an employer’s behavior towards an employee such as wrongful termination, discrimination, harassment, and wage violation.

Family Law Matters - cases related to divorce, child custody, adoption, and guardianship.

Anti-corruption and Economic crimes – offenses and misconduct related to corruption and illicit activities that have a significant impact on the economy which may include acts of bribery, embezzlement, fraud, money laundering, abuse of public office, and other illegal activities that undermine the integrity of public and private sectors, distort market forces, and hinder economic development.

This study focused on civil matters, criminal matters, family matters, Anti-corruption and Economical crimes and Constitutional Law Matters, due to the higher number of cases that were filed under these categories.

1.2 Statement of the problem

Delay in Justice has occasioned to citizens losing trust with the Judiciary. There are several factors that influence the time taken by a case in High court to be determined. Globally, different researchers have tried estimating the duration taken in a court of law for a case to be determined. Access to timely justice still remains a mirage for most citizens globally which has been contributed significantly by a number of factors. Davila, (2015) found out that time taken for a case to be determined was contributed by several factors, among them are the sovereignty of the case, type of the case, where criminal cases differed significantly in time with the civil cases.

In Kenya, courts operate at two levels, namely; Superior and Subordinate courts. The decentralization of the Court system has been established with the Supreme Court and the Court of Appeal having their own Presidents and the high court having a Principal Judge as heads of the respective institutions. The high court is established under Article 165 and it consists of a number of judges to be prescribed by an Act of Parliament. The court is organized and administered in the manner prescribed by an Act of Parliament and also has a Principal Judge, who is elected by the judges of the high court from among themselves. The Kenyan high courts' key responsibilities are to hear all civil and criminal cases as well as the appeals from lower courts. Further, this court is premier in interpreting the constitution and supervises the judicial review processes and thus this makes it an ideal court to study its activities in relation to case disposals.

According to Kenya National Bureau (KNBS, 2020), there were 101,588 cases pending in Kenyan high courts in the financial year 2019/2020 and 30,695 new cases filed. This has resulted in huge case backlog in the Judiciary and has led to delay in settling the cases. Gibson et al., (2021) stated that Justice delayed is equivalent to Justice being denied and therefore, a case being determined within the shortest time possible will help the involved parties get their fundamental right to Justice. The increase of case backlog is an indication of weakness in a legal system which erodes public confidence in the "rule of law" (Ngonga, 2019). Despite its efforts to ensure that the time taken for a case to be determined in high courts is minimum, there are several factors that play great role in influencing case determination.

Modelling time taken for a case to be determined in high courts using survival analysis approach was important as it does not only give the approximate time but also ensures that all information is utilized since the incomplete cases are censored and provides information regarding the study. This study therefore, aimed at analyzing time taken for case determination in Kenyan High courts, analyzing predictors to a 'quick' or 'prolonged' case determination and modelling time taken for a case to be determined in a court of law.

1.3 Justification of the Study

By employing the survival analysis techniques, this study sought to provide quantitative insights into the factors influencing case determination time, addressing a crucial gap in knowledge within the Kenyan legal context. Understanding the underlying drivers of court process length can guide policy reform and interventions aiming to accelerate case disposition, reduce backlog, and ensure timely access to justice. Second, using survival analysis models in this study makes available a broader set of methodological approaches to study court systems and demonstrates the value of advanced statistical techniques in legal research. Lastly, the research findings will inform the development of customized strategies and best practices for the Kenyan high courts, leading to a more efficient, transparent, and fair judicial system that is sensitive to the needs of all stakeholders of the legal process.

1.4 Significance of the study

The need to model and analyze the duration of time taken to conclude a court case in Kenyan high courts using a survival analysis approach is multidimensional. This provides crucial information regarding the efficiency and effectiveness of the judicial system, enlightening its citizens on what drives case proceeding time. Understanding these variables can help identify potential inefficiencies or congestion points in the court system and provide clear and actionable recommendations for better case management and resource utilization.

Furthermore, through survival analysis techniques, analysts are able to quantify the impact of various covariates on the determination time of cases. Critical information can be obtained on which elements primarily contribute to lengthening the court case duration, say the type of the case or appeals made. This information will empower policy makers and aid them in setting priorities for reforms or intervention to speed up the completion of court cases.

Moreover, the use of survival analysis models for Kenyan high courts is a contribution to the area of knowledge in the field of legal studies and judicial process analysis. It expands the set of methods usable by researchers of court systems and shows the applicability and usefulness of survival analysis in fields other than those of medicine or engineering.

Overall, this study has significant policy implications for policymakers, legal practitioners, and researchers with an interest in enhancing the efficiency and justice of Kenya's high court system. By identifying the key determinants of case determination time and producing predictive models, it offers useful insights for streamlining court procedure, reducing case backlog, and ultimately improving access to justice for litigants to court proceedings

1.5 Objectives of the Study

1.5.1 General Objective

To analyze time taken for a case to be determined in high courts of Kenya using survival analysis.

1.5.2 Specific Objectives

1. To perform exploratory data analysis of factors determining the survival time of cases in Kenyan high courts.
2. To predict the time for a case to be determined in Kenyan high courts using Survival models.
3. To compare model precision between Cox regression and other parametric survival models.

1.6 Research Hypothesis

1. There is no significant relationship between factors determining survival time in Kenyan high courts.
2. There is no significant difference in factors that predict survival time of cases in Kenyan high courts.
3. There is no significant difference in model precision between Cox regressions to parametric survival models in predicting case survival time in Kenyan high courts.

1.7 Scope of the study

The study was conducted on cases in Kenyan high courts. The data for this study were of cases from 2012 to 2022 from all Kenyan high courts established as at year 2022. The independent variables for this study included in-court factors (type of the case, number of appeals, court's location). The dependent variable was the time at which event of interest was observed and was treated as a continuous variable. The data were secondary data, retrieved from judiciary database. The data were filtered to remain with only the desired variables and treated with utmost confidentiality, for research purposes only and in line with data protection Act 2019.

CHAPTER TWO
LITERATURE REVIEW

2.0 Introduction

This chapter provides a critical overview of theoretical and empirical literature of relevance to the aims of the research: exploratory data analysis (EDA) of determinants of Kenyan High Court case duration, forecasting of case determination times through survival models, and comparison of the accuracy of Cox regression and parametric survival models. The review integrates cross-national and Kenyan research to identify factors that influence judicial delays, the application of survival analysis in legal processes, and model performance comparison. It concludes with a conceptual framework relating case characteristics, judicial processes, and time to case determination, pointing out the research gap addressed in this study.

2.1 Theoretical review

This part examines research in High Courts that has utilized survival analysis to model case determination time and provides a theoretical framework to implement survival models in the context of the Kenyan High Court. Focus is on how survival analysis has been used to model case duration and variable factors influencing delay in the judicial setting.

2.1.1 Survival analysis in high court case duration studies

Survival analysis has also been extensively applied in High Courts globally to measure case resolution durations, establishing proof of judicial efficiency and delay-causing factors. Perez and Figueroa (2006) applied survival analysis to the High Courts in Argentina in order to analyze civil case duration, using the Kaplan-Meier estimator for illustrative purposes of survival probabilities and the Cox Proportional Hazards (PH) model for an estimation of the effect of case complexity and procedural actions. Their findings indicated that complicated civil cases, including multi-party cases, took significantly longer durations to resolve.

2.1.2 Parametric models in judicial contexts

Smyth and Narayan (2016) applied the Cox PH model in the Australian High Court to examine case disposition times for constitutional and appeal cases. The findings in their research identified that appealed cases were 1.5 times more likely to possess a hazard ratio over original jurisdiction cases, representing faster resolution due to procedural streamlining, taking longer for constitutional cases due to their complexity.

Similarly, Dimitropoulos (2018) used the Log-normal model in the European Court of Human Rights (operationally similar to a High Court) to model case settlement times. The study found delays of procedure by evidentiary hearings increased lengths (median = 36 months), with non-monotonic risks most accurately modelled by the Log-normal model, noting its suitability for complex judicial data.

2.2 Empirical Review

This section reviews empirical studies which have examined determinants of court case determination times, focusing on High Courts globally and in Kenya. The review combines evidence-based studies to identify primary determinants of case duration, applications of survival analysis for predicting resolution time, and comparisons of survival model performance. Split by the objectives of the study which is exploratory data analysis of factors that affect case duration; prediction of case determination time using survival models; and comparison of parametric survival models and Cox regression. This review discusses evidence from both Kenyan and international views. It highlights empirical measures of judicial process, appeal status, resource limitation, and case type, and determines gaps in survival analysis application to Kenyan High Courts in a bid to justify the need for further research.

2.2.1 Exploratory Data Analysis of factors determining the survival time of a case

Exploratory Data Analysis (EDA) is a powerful tool for ascertaining the factors that influence the survival time for a case in Kenyan high courts. It provides a scientific method of obtaining relevant information from various sets of court cases. There was a requirement for an examination of such data to gain insight so that the judicial process could be expedited and bottlenecks eliminated. This review of literature will provide an overview of studies that have been conducted on EDA compared to Kenyan high courts.

Amutah’s (2015) study of two of Kenya's juridical systems, contract law and tort law revealed contradictions in the judicial process, i.e., outstanding cases that should have been determined by the courts. Amutah deduced that the delays were mainly caused by the lack of institutional capacity, resources, and strategy to manage cases effectively. This indicates that the justice system in Kenya is in dire need of reforms in order to access justice speedily and efficiently. Moreover, the research enlightens us on how other countries with similar issues of the law would be able to introduce interventions such that the justice system would be enhanced and access resources in a more equitable manner.

Chinyio and Chisawani (2017) research provides important information regarding the use of EDA to determine court effectiveness. Their findings showed that the efficient management and decision-making of court staff are an important determinant of determining court efficiency and success. Moreover, their research reveals that sensitization and training of court staff has a positive impact on court efficiency with consequent timely resolution of cases. Organizations that prioritize training and sensitization of court staff can leverage their studies to optimize court effectiveness and attain maximum timely resolution of cases. Besides, it helps us gain a deeper understanding of how court effectiveness can be quantified using EDA, providing us with a valuable tool for measuring the performance of courts.

The empirical findings of Wakhungu (2018) research supports the view that court systems must place priority on the application of EDA for improved case management and alleviation of the burden. However, an analysis based on the impact of the legal environment on the length of cases is also inevitable. For instance, the existing laws governing court procedure can contribute to case survival time, and reforms by considering the local context may address these. Generally, understanding the legal environment and how it influences case durations is critical in employing effective EDA and court administration.

According to Maseno (2019) study, the findings validated the personalized opportunity technological innovation presents in optimizing court resource management. What its results can do is help inform the development of EDA-driven systems within courts. Comparison of the research to other legal resource management studies has demonstrated that application of such systems can be beneficial towards optimizing the reduction of court case backlog and resolution velocity. While Maseno (2019) concluded by suggesting specific remedies, further research is required to ascertain how the management of court resources could be improved by using an EDA-facilitated system where there are sophisticated processes and numerous interests.

2.2.2 Analyzing and predicting time for a case to be determined in high courts

The judiciaries have been experiencing increasingly longer case tendencies, which have resulted in tardy justice. Several statistical approaches such as survival analysis, logistic regression, machine learning, Monte Carlo simulation, System Dynamic model, Agent-Based model, and Discrete Event Simulation model have been applied in order to model the time a case takes to be disposed of by the world's high courts (Ngonga, 2019).

Medvedeva et al. (2020) used machine learning in case duration modeling and prediction in the European Court of human rights. They identified a 75% accuracy in predicting the length of time for a court case to be concluded using machine learning. Thus, further research into the application of survival analysis might come to such similar findings and provide a better and more accurate modeling of the time for a case to be determined by Kenyan high courts.

Davila (2015), found that charge severity, custody days, and counts significantly impacted the length of a case. Pradhan (2018) used an exploratory data analysis in order to study crime prediction in San Francisco and concluded that crime pattern was not stable over time and skewed towards some cities. By looking at the findings of the two studies, it can be argued that a multidimensional set of factors, including charge severity, days in custody, and number of counts, are highly instrumental in determining the time required in processing a criminal case. In addition to this, crime patterns that change depending on geographic location and times can be determined through data analysis.

2.2.3 Predicting time for a case to be determined using Survival model

Survival analysis techniques are generally used to depict the time it requires for an event to happen in the literature (Emmert-streib & Dehmer, 2019; Moore, 2016; Teshnizi et al., 2017). With this, researchers are able to gain greater knowledge regarding the factors that can affect the duration of a particular case (Dirks-Linhorst & Linhorst, 2012). In a study conducted by Perez and Figueroa (2006), the time for civil cases for Argentina was estimated via Kaplan-Meier product limit method and the explored variables using the Cox proportional hazard model. Dirks-Linhorst and Linhorst (2012) also employed Cox regression survival analysis to estimate the time for cases for mentally ill individuals.

Both studies indicate the relevance of survival analysis techniques in explaining the duration of a case. Through the different models of survival analysis, the dependent variables may be examined to make implications of findings inform judicial decisions on case length and/or appropriate treatment of the individuals involved.

2.2.4 Comparison between Cox regression and other Parametric Models

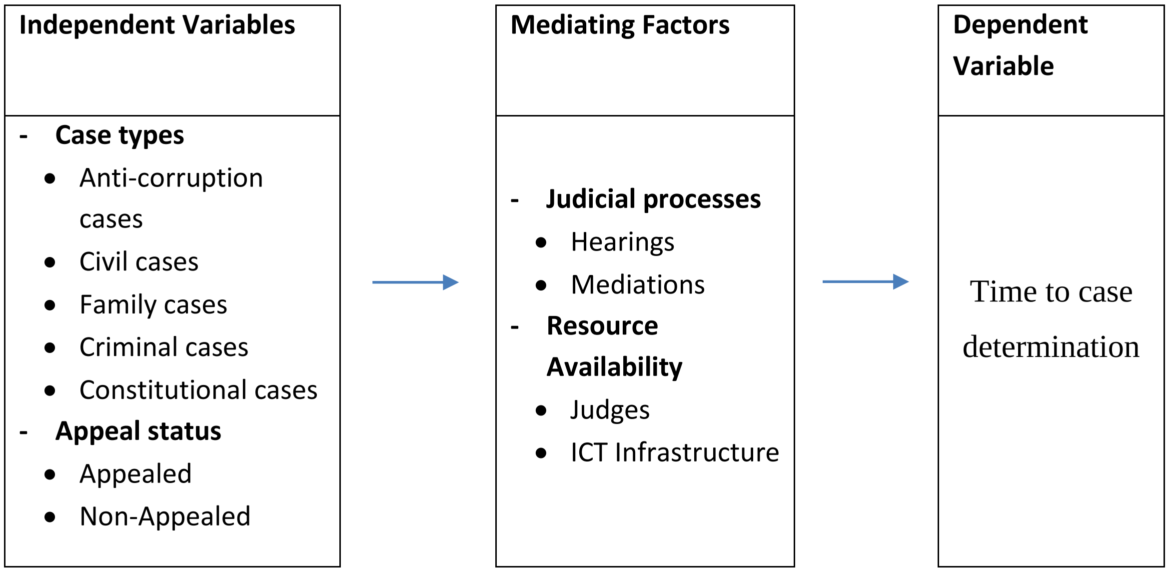
A number of studies have been contrasted in which the accuracy of the Cox regression model, a nonparametric model, is compared to a variety of parametric models for the analysis of cancer and other health datasets. Teshnizi et al. (2017) contrasted the Cox regression model and parametric models to determine which model worked best for survival data. The study found that parametric regression models by Weibull and log-normal provided a better model for the risk factors for acute cancer compared to the Cox regression model. This is consistent with previous research that found that the Akaike information criterion (AIC) advised that the Cox regression model performed the least well among the compared parametric models. This would mean that when trying to model risk factors for acute cancer, the parametric models are the better model to use over the Cox regression model. Further studies are needed, however, to determine whether the performance of parametric and nonparametric models varies according to different types of cancer and severities.

A comparison between the two studies reveals that both the nested case-control study and the Cox proportional hazard model have a lot of potential in modeling time-dependent exposures in different studies. The efficiency and superiority levels of the two models as presented, nonetheless, appear to be dependent on the kind of study being conducted and the nature of the exposure being modeled. For example, Essebag et al. (2005) concluded that the nested case-control model was the better model to investigate time-dependent exposure in the case of their particular research. Nevertheless, the evidence gathered by Zare et al. (2015) indicated that the Cox proportional hazard model was a better model to investigate the survival of gastric cancer patients.

2.3 Conceptual Framework

The conceptual model shows the relationship among independent variables (appeal status, case type), dependent variable (time to case determination), and mediating variables (judicial processes, resource availability). Case type (Anti-Corruption, Civil, Constitutional, Criminal, Family) influences duration due to varying complexity and procedural requirements. Appeal status lengthens duration through additional procedural stages. Judicial processes (hearing, mediation, etc.) and the availability of resources (judges, ICT infrastructure, etc.) mediate these relationships because queueing theory holds that delays are exacerbated by resource constraints (Flinn, 2000).

The model contends that case length is predicted by survival analysis models (Kaplan-Meier, parametric, Cox) through the estimation of the effects of such variables. Well-managed cases and integration of ICT, based on Maseno (2019), can halt delays, in consonance with suggestions to upscale the Judiciary Case Management System (JCMS) (Gakuya, 2022).



2.4 Research Gap

Even though the literature is informative with respect to judicial delays, there are significant gaps in survival analysis for the estimation of case determination times in Kenyan High Courts, particularly in addressing the study objectives. Therefore, for the application of EDA on case duration determinants, studies such as Amutah (2015) and Wakhungu (2018) have listed institutional capacity and procedural complexity as causes of delays in Kenyan courts but focused primarily on lower courts or specific types of cases. These analyses did not employ EDA to examine systematically variables like case type (Anti-Corruption, Civil, Constitutional, Criminal, Family) and appeal status in Kenyan High Courts and thus could not provide a comprehensive snapshot of delay determinants in this context.

Modelling the case determination times using survival models, global studies like Perez and Figueroa (2006) managed to apply Kaplan-Meier and Cox Proportional Hazards models in High Courts in Argentina, which shows their suitability for time-to-event data. This study however did not employ other non-parametric survival models to evaluate its suitability in modelling the court data. In Kenya, however, Ngonga (2019) employed system dynamics rather than survival analysis for Kisumu High Court. The absence of studies using survival models to predict case lengths in Kenyan High Courts is a major gap since such models are best placed to handle time-to-event data and would provide precise predictions for planning in the judiciary.

Further, with respect to contrasting Cox regression and parametric survival models, Bełdowski et al. (2020) and Dimitropoulos (2018) contrasted model performances in Polish and European courts, respectively, and concluded parametric models (e.g., Weibull, Log-normal) to be better for skewed data. Limited research, however, has drawn such comparisons in Kenyan High Courts, where case duration data could have different distributional characteristics due to a range of case types and procedural variations. This gap hinders the establishment of the optimal modeling approach for Kenyan judicial data.

While studies have been done in high courts on judicial efficiency, no study has integrated EDA, survival modeling, and model comparison for Kenyan high courts. This study addresses these gaps by using survival analysis on data from all Kenyan high courts and contributing existing studies on delay determinants, predictive modeling, and optimal model selection to inform judicial reform in Kenya.

CHAPTER THREE

MATERIALS AND METHODS

3.0 Introduction

This chapter outlines the research methodology employed to investigate the time taken for cases to be determined in Kenyan High Courts. The research has three objectives: conducting exploratory data analysis (EDA) to establish factors that influence the duration of cases, making predictions about case determination times using survival models, and comparing the precision of Cox regression and parametric survival models. The chapter presents the study design, data collection, data sources, and analytical procedures, including survival analysis models such as Kaplan-Meier, Cox Proportional Hazards, and parametric models (e.g., Weibull, Log-normal). It also details the steps in data processing, variable specifications, and statistical methods used to offer valid analysis. In presenting a whole framework for the study's methodology, the chapter provides the foundation for addressing the research objectives and formulating workable recommendations for judicial reform in Kenya.

3.1 Data Source and Description

The study used secondary data from 2012 to 2022. The data were retrieved from judiciary databases upon the approval of the registrar of judiciary. The study variables included in-court factors such as type of the case, number of appeals, and testimonial proofs. The time at which the case was filed in a high court of law, the hearings of the case and finally the time when the case is determined or settled. The dependent variable is the time at which the case was disposed. The data were then analyzed using survival models i.e. Kaplan-Meier estimator, Cox regression, Weibull, Exponential, Log-normal and Log-logistic. These survival models were fitted to the data using R statistical software. To compare the performance of the survival models, the Akaike Information Criterion.

3.2 Statistical Diagnostics

Statistical diagnostics play an extremely crucial role in data analysis to determine the validity and reliability of the results. They help identify potential issues, outliers, and patterns within the data, as well while reviewing the assumptions and relevance of the statistical techniques used. This section was utilized check for multicollinearity, homogeneity and normality of the data set.

3.3 Data Analysis

3.3.1 Exploratory Data Analysis

Performing exploratory data analysis (EDA) of the determinants of the survival time of cases in Kenyan high courts to establish any significant relationship between the variables and to test for any violation of Cox regression was carried out by computing and reporting descriptive statistics such as the mean, median, mode, range, and standard deviation. In addition, graphical methods such as histograms, box-and-whisker plots, and scatter plots were used to explore the dataset and identify patterns and trends. During the EDA, the data from past case files and court databases were analyzed, with the aim of determining what variables had a significant role in impacting the survival time of the court cases. Furthermore, the EDA attempted to identify and analyze any pattern, trend or relation in the data, in order to gain further insight into the determinants of the survival time of court cases.

3.3.2 Survival Models

Survival analysis refers to the study of time-to-event phenomena. Survival analysis models covered in this study were, the Kaplan-Meier, Cox Proportional Hazard and the parametric models such as Weibull, Exponential, Log-logistic, Gamma and the Log-normal.

i) Kaplan-Meier Estimator for the Survival Function

The Kaplan-Meier estimator was used to estimate the survival function by plotting a Kaplan-Meier curve, which is a graphical representation of the survival probability over time. The Kaplan–Meier (KM) estimator of a survival function $S_{KM}(t)$ is given by

$$S_{KM}(t) = \prod_{i:t_i<t} \frac{n_i - d_i}{n_i} = \prod_{i:t_i<t} (1 - \frac{d_i}{n_i}) \qquad (3.1)$$

The Kaplan-Meier estimator holds for all $t > 0$ and it depends only on two variables, n_i and d_i which are number in risk at time t_i and number of events at time t_i respectively.

ii) Cox Proportional Hazard

The Cox Proportional Hazard distribution model is a type of survival analysis that can be used to analyze the time taken for a case to be determined in the high courts of Kenya. The model is based on two factors: the baseline hazard function, $h0(t)$, which is unaffected by the predictor variables; and the exponential function, $e^{\beta x}$ which involves the predictor variables. It is assumed that the hazard ratio is constant over time and this is referred to as the proportional hazard’s assumption. The hazard ratio is the ratio of the hazard rate for an individual with the predictor values x to the hazard rate for a reference individual with the same baseline hazard function but different predictor values. The Cox PH model is expressed as

$$h(t,x) = h_0(t)e^{\sum_{j=1}^{p}\beta_j x_j} \tag{3.2}$$

where $h(t,x)$ is the hazard function at time t for an individual with predictor values x_j, $h_0(t)$ is the baseline hazard function that is not affected by the predictor variables, $\beta_{j's}$ are the regression coefficients for the predictor variables $x_{j's}$ and the function $e^{\sum_{j=1}^{p}\beta_j x_j}$ represents the relative hazard ratio for the individual, which is the ratio of the hazard rate for an individual with the predictor values $x_{j's}$ to the hazard rate for a reference individual with the same baseline hazard function but different predictor values. The hazard ratio is assumed to be constant over time, which is the proportional hazards assumption.

ii) Weibull distribution

The Weibull distribution is a two-parameter family of continuous probability distributions that is used to describe a wide range of lifetime phenomena, including the time to failure of mechanical devices. It can be utilized to model time-to-event data in Kenyan high courts, including the time to determination of a case. Making the assumption that the Weibull distribution has parameter $\theta > 0$ and $\beta > 0$ and the survival function is given by,

$$s(t) = e^{-(\theta t)\beta} \tag{3.3}$$

For $t > 0$, the Weibull density function is given by

$$f(t) = \frac{\partial s_{\theta,\beta}(t)}{\partial t} = \beta\theta(\theta t)^{\beta-1}e^{-(\theta t)\beta} \tag{3.4}$$

The hazard function can be obtained by dividing the Weibull density function with the survival function

$$\lambda(t) = \frac{f(t\,;\theta,\beta)}{s_{\theta,\beta}(t)} = \beta\theta(\theta t)^{\beta-1} \tag{3.5}$$

Estimation of both Weibull parameters is done through the maximum likelihood estimation (MLE) technique. The MLE technique optimizes the log-likelihood function with respect to θ and β.

From the high court data, researchers can estimate the Weibull parameters of the time to resolve a case in Kenyan high courts and predict the time to resolve a case in Kenyan high courts based on the estimated parameters.

iv) Gamma Distribution

The Gamma distribution model is a bell-shaped continuous probability distribution. It is commonly used to describe circumstances in which the rate at which events are occurring increases over time. In order for the Gamma distribution model to be achieved from the available data, the density function will be derived from data, and this will be achieved by finding a parametric model from the data.

The gamma distribution with parameter $\beta > 0$ and $\alpha > 0$, the density function is given by

$$f(t) = \frac{\beta^{\alpha}}{\Gamma\alpha} t^{\alpha-1} e^{-\beta t}, t \geq 0 \qquad (3.6)$$

The hazard and survival functions can be estimated from the gamma density function.

v) Log-Logistic Distribution

The Log-logistic distribution is a continuous distribution with two parameters, α, and θ, that describes the survival of a case in a High Court of Kenya. The Log-logistic distribution with the parameters $\theta > 0$ and $-\infty < \alpha < \infty$ is a continuous distribution with the density function given by

$$f(t) = \frac{e^{\theta} \alpha t^{\alpha-1}}{\left(1+e^{\theta} t^{\alpha}\right)^{2}} \qquad (3.7)$$

The survival function is derived from the density function as

$$S(t) = \left[1 + e^{\theta} t^{\alpha}\right]^{-1} \qquad (3.8)$$

The parameters, α and θ, will be estimated from the data set by using the Maximum Likelihood Estimator technique. This will be achieved by finding the log-likelihood function and setting it equal to zero to obtain the ML parameter estimates. The ML estimates for the parameters, α, and θ, then be used to obtain the hazard and survival functions, λ(t) and S(t), respectively. The hazard function λ(t) is obtained by dividing the log-logistic density function by the survival function, S(t), and the survival function, S(t), is obtained from the log-logistic density function.

vi) Log-Normal Distribution

Log-normal distribution is a continuous probability distribution with a bell-shaped probability density function. It is derived from the product of two independent normally distributed random variables. Log-normal distribution is utilized to model the probability of occurrence of an event in random processes with a large number of possible outcomes and a finite chance of occurrence. The Log-normal distribution can be applied in the analysis of the time to decide a case in Kenyan high courts, by transforming the time taken for the case to reach some particular conclusion, to a probability distribution and then examining the resulting variability in the time taken to reach the conclusion.

A random variable T is said to have a lognormal distribution with parameters $-\infty < \mu < \infty$ and $\sigma > 0$. The density function of T is given by

$$f(t) = \frac{1}{\sigma(2\pi)^{\frac{1}{2}}} t^{-1}\, exp\{-(log\, t - \mu)^2/2\sigma^2\}, t \geq 0 \tag{3.9}$$

Survival function is given below

$$S(t) = 1 - \frac{\varphi(\ln(t) - \mu)}{\sigma} \tag{3.10}$$

where φ denotes the cumulative distribution function of the standard normal distribution.

vii) Exponential Distribution

The exponential distribution model is a type of parametric survival model that is used to analyze the time taken for cases to be determined in Kenyan high courts. The exponential model assumes that the "hazard" (or the probability of a case terminating within an interval of time) is constant over time. The survival function of the exponential model is the probability that a case will not terminate in time "t". Suppose random variable T follows an exponential distribution with parameter θ such that $\theta > 0$. Then, the density function of the exponential distribution with parameter $\theta > 0$ is given by

$$f(t) = \theta e^{-\theta t} \tag{3.11}$$

The survival function is estimated from the exponential function as

$$S(t) = \int_t^\infty f(z, \theta) dz = \int_t^\infty \theta e^{-\theta z} dz = e^{-\theta t} \tag{3.12}$$

Time is treated as a continuous variable in the above models. The Kaplan-Meier estimator and Cox Proportional Hazard Model both use continuous time as a measure, while the Weibull, Gamma Log-Logistic, Log-normal, and Exponential distributions all use continuous time equations to represent the likelihood of a particular event occurring. The parameters that are estimated in these models are also related to the continuous-time function, such as the exponential coefficient (θ) in the Exponential model or the log-normal mean (μ) and variance (σ) in the Log-normal model. Thus, all of the above models treat time as a continuous variable.

3.3.3 Comparison between Cox regression and Parametric Survival Models

To compare the performance of the survival models, model diagnostics and comparison will be performed using the Akaike information criterion (AIC). The better performing model will be applied to estimate the survival time for a case to be determine in Kenyan high court. The Akaike Information Criterion is expressed as follows

$$AIC = 2K - 2\,ln(L) \quad (3.13)$$

Where, K is the amount of degrees of freedom used while L is the likelihood function of the model.

In general, when comparing two or more survival models, the model with the lower AIC is the better fitting model. AIC is also an approximation of the relative quality or goodness of fit of a set of statistical models for a particular set of data. The model with the lowest AIC is the best-fitting model. AIC considers the number of parameters of the model (K) and the likelihood of the model (L). A model with fewer parameters will have a lower AIC in the event that the model fit approximates that of a model with more parameters.

3.4 Parameter Estimation

The Maximum likelihood estimator (MLE) technique will be applied to estimate model parameters. Consider the failure times given by $x_1, x_2, \cdots, x_n$ and also, consider the parametric model $X \sim f\,(x, \vartheta)$(Kleinbaum & Klein, 2012; Wang, 2006). The likelihood function on the basis of the failure times $x_1, x_2, \cdots, x_n$ is given by

$$L(\vartheta) = \prod_{i=1}^{n} f(x_i, \vartheta) \quad (3.14)$$

The MLE of $\hat{\vartheta}$is the value of ϑwhich maximizes the likelihood function $L(\vartheta)$

$$ln\,L\,(\vartheta) = \sum_{i=1}^{n} ln\,f\,(x_i, \vartheta) \quad (3.15)$$

$$v(\vartheta) = \frac{\partial}{\partial\vartheta} ln\,L\,(\vartheta) = \sum_{i=1}^{n} \frac{\partial}{\partial\vartheta} ln\,f\,(x_i, \vartheta) \quad (3.16)$$

The MLE of $\hat{\vartheta}$ satisfies $v(\hat{\vartheta}) = 0$. Then, by Taylor series

$$v(\hat{\vartheta}) = v(\vartheta) + v'(\vartheta) + (\hat{\vartheta} - \vartheta)$$

Thus $\hat{\vartheta} - \vartheta \simeq -\frac{v(\vartheta)}{v'(\vartheta)} = -\frac{1}{v'(\vartheta)}\sum_{i=1}^{n}\frac{\partial}{\partial\vartheta} ln\, f\, (x_i, \vartheta)$ (3.17)

By central limit theorem, ϑ is asymptotically normal expressed $\hat{\vartheta} \sim N\left(\vartheta, I^{-1}(\vartheta)\right)$ where $I(\vartheta)$is the Fisher information matrix (FIM) (Wang, 2006).

$I(\vartheta) = E\left[-\frac{\partial^2}{\partial\vartheta^2} ln\, L\, (\vartheta)\right]$ (3.18)

3.5 Model Diagnostic

The survival model fit will be ascertained through diagnostic methods to check model fit and predictive validity. The Akaike Information Criterion (AIC) will be utilized to contrast overall model fit between the Cox and parametric models, sacrificing goodness of fit in favor of model complexity to identify the model that best fits the distribution of survival times (Moore, 2016). A lower AIC value is an indication of a better-fitting model which balances the number of parameters so that it does not overfit. Predictive accuracy in the Cox Proportional Hazards model will be evaluated through the computation of the concordance index, a measure of how well the model can rank survival times appropriately, with higher values indicating increasing discriminative power (Kleinbaum & Klein, 2012). These diagnostics give a stringent measure of model performance, supplementing the research focus of comparing the precision of Cox regression and parametric survival models for predicting case determination times in Kenyan High Courts.

CHAPTER FOUR

RESULTS AND INTERPRETATION

4.1 Introduction

This section presents the results from the analysis of case determination time in Kenyan high courts through survival analysis methods. The chapter organizes results around the study's three specific objectives using methods described in Chapter 3. The data were processed using Rmarkdown in RStudio software.

4.2 Exploratory data analysis of factors determining the survival time of cases in Kenyan high courts

4.2.1 Summary of cases in Kenyan high courts per case type

The **Table 4.1** below presents the distribution of case types in Kenyan High Courts based on the total number of cases filed between 2012 and 2022.

Table 4.1: *Distribution of case types in Kenyan high courts (2012–2022)*

Case Type	Number of Cases	Proportion in (%)
Anti-corruption cases	418	0.5
Civil cases	31148	33.7
Constitutional cases	7036	7.6
Criminal cases	31707	34.3
Family cases	22096	23.9
Total	92405	100.0

The analysis of case types in Kenyan High Courts using the dataset of 2012-2022 shows that cases that fall under the criminal court are the largest in terms of percentage as it constitutes 34.3 percent, followed by civil cases which racked up 33.7 percent. These two categories combined comprise almost 68 percent of all the cases that are filed meaning that there is a huge need of cases that require the input of the court either in the case of the matter of the population and also in case of the individual. Domestic and familial problems come to the legal system in the forms of family law cases which comprise 23.9 percent of the total caseload. Less common, the constitutional and human rights cases, still form 7.6% of the total caseload demonstrating their relevance in terms of the emergence of the judicial precedence and preservation of the main rights. The lowest number of endorsed cases are in the categories of anti-corruption and economic crimes at 0.5%, although the respective types usually receive significant time and attention by the public and the institutions as they affect governance and accountability.

4.2.2 Normality test

The general case duration distribution in Kenyan high courts in the histogram indicates right-skewedness with a mean of 27.19 months and a median of 18 months (***Figure 4.1***). The fact that SD = 27.28 shows that there is high variance in processing times among the cases. From the density plot on the histogram, most of the cases resolve in relatively shorter timeframes by having high concentration around 0-20 months. By having a long tail beyond 100 months, there is evidence of a subset of smaller cases that take much longer to be resolved. This right-skewedness ensures that most of the cases take relatively faster handling but some take longer delays. This difference between the mean and the median further points to the skewness of the data caused by outliers or longer length of stays and reiterates that the median is a more reliable indicator of what is typical case duration than the mean.

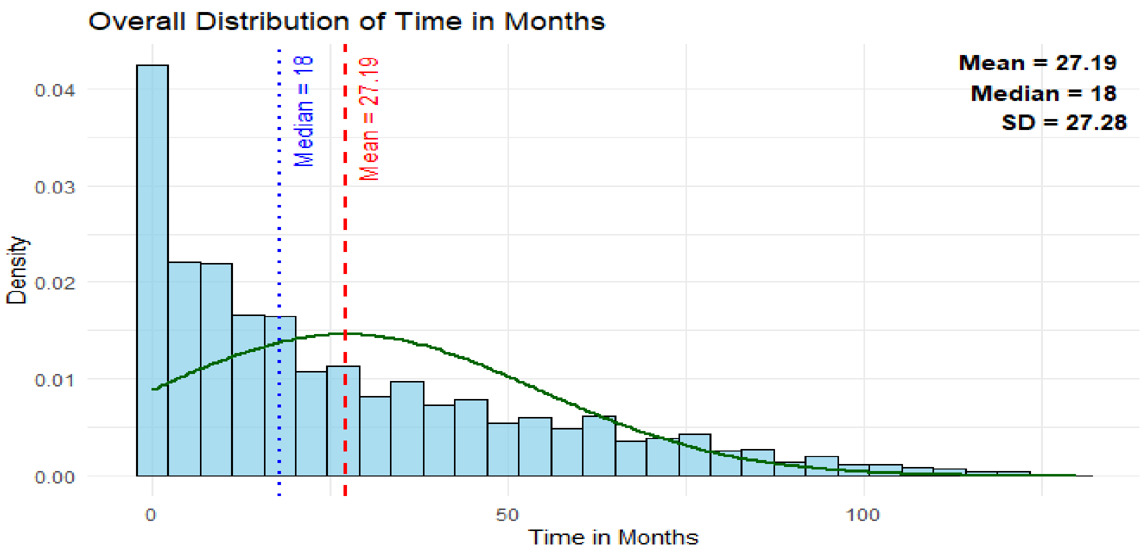


Figure 4.1: ***Overall distribution of time in months***

4.2.3 Case duration by case type

The findings on case durations by case type show significant differences in resolution time across different domains of law in Kenyan high courts shows that on average, cases in the domain of Family law took longest, at a mean of 40.19 months and a median of 35 months (**Table 4.2**), reflecting a commonality of complex emotional, social, or procedural issues involved, thereby extending adjudication durations. By contrast, Civil cases had a mean of 30.89 months and a median of 23 months, reflecting moderate complexity along with relatively longer proceedings. Constitutional law cases, setting aside their fundamental importance to interpreting national laws as well as rights, also had a mean adjudication time of 21.54 months, a median of 14 months, reflecting possible delays owing to their rare nature along with depth of legal reasoning involved. Criminal cases were resolved relatively more swiftly on average at 15.91 months, a median of a mere 8 months, owing perhaps to procedural streamlining of rules governing criminal trial along with emphasis on defendants' right to speedy trial. Lastly, Anti-Corruption cases, though high profile in public interest, recorded a shortest average time of 14.22 months, a median of 8 months, perhaps owing to higher priority by courts, special treatment of such cases, or both, by judges, along with inter-case inconsistencies. The standard deviations also point to variability across categories, specifically among Family as well as Civil cases, where higher dispersion reflects inconsistency of adjudication time perhaps owing to variations in court workload, assignment of judges, or procedural differences in cases along these categories.

Table 4.2: *Summary of time in months grouped by case type*

Case type	Mean time (months)	Median time (months)	Maximum time (months)	Standard deviation
Anti-Corruption and Economic Crimes	14.22	8	72	16.01
Civil cases	30.89	23	130	27.11
Constitutional & Human Rights cases	21.54	14	129	22.24
Criminal cases	15.91	8	126	20.81
Family cases	40.19	35	129	30.14

4.2.4 Distribution by case type

The distributions of types of cases shows positive skewness, as there is a consistent tendency of means to be higher than medians as seen in ***Figure 4.2***. This reveals that there is an extreme set of cases in every type that takes much longer to be resolved than the average (median) case. The longest average durations of Family and Civil cases result in mean resolution periods of around 40.19 and 30.89 months, with high standard deviations of 30.14 and 27.11 months, indicating great variation and frequent outliers.

In contrast, Anti-Corruption and Criminal cases take significantly less time to be resolved, with mean durations of 14.22 and 15.91 months, and tighter spreads, reflecting higher consistency. Constitutional cases come in at mid-level with an average of 21.54 months. Having normal distribution curves helps to visualize spread and pinpoint deviations within symmetric, bell-curve shape as seen with Family and Civil cases where there is evident tail skewness to the right.

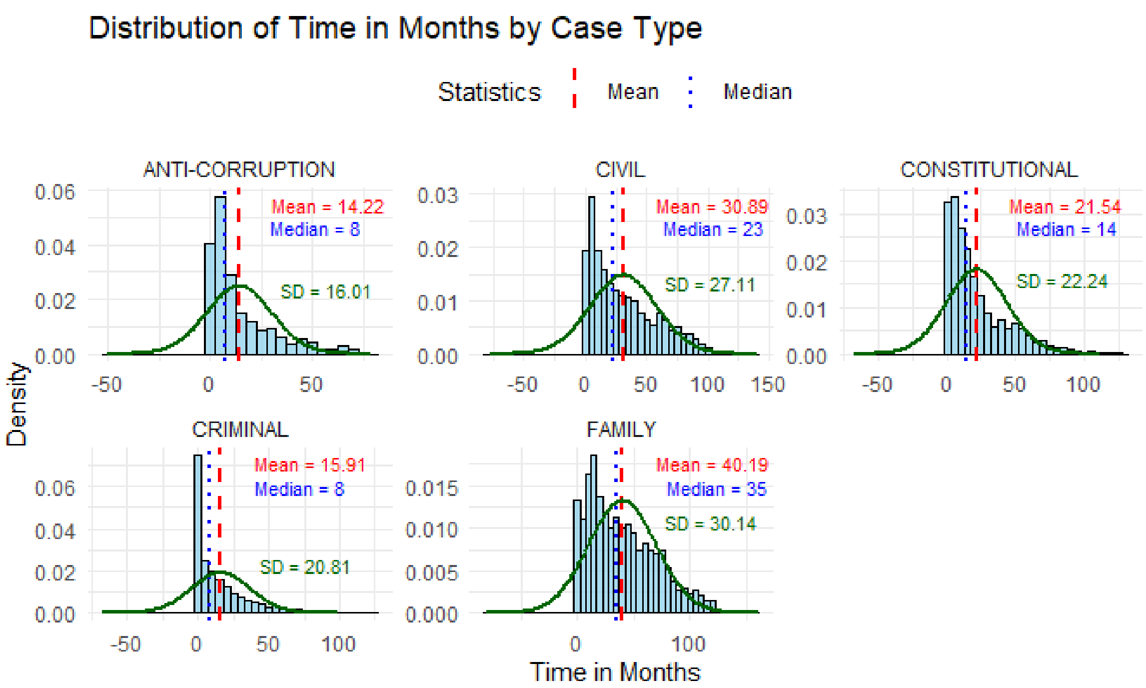


Figure 4.2: Distribution of time in months by the case type

4.2.5 Comparison by appeal status

The findings further showed that based on the number of appeals, cases that underwent one or more appeal stages had a higher average resolution time of 30.08 months, with a median of 24 months (**Table 4.3**). In contrast, cases without any appeal had a mean duration of 26.22 months and a median of 15 months, indicating that the appellate process significantly prolongs case determination. The maximum time recorded for both groups was approximately 130 months, highlighting long-standing delays in some cases. However, the standard deviation was higher for non-appealed cases (28.47 months) compared to appealed cases (23.09 months), suggesting greater variability in processing times among cases that did not go through the appellate system. This observation was statistically confirmed by the Welch Two Sample t-test, which revealed a significant difference in mean durations between the two groups: $t = -20.669$, df = 48,116, $p < 0.001$. The 95% confidence interval for the difference in means ranged from -4.22 to -3.49 months, confirming that cases with appeals take significantly longer to resolve than those without.

Table 4.3: *Summary of case duration in months by appeal status*

Appeal status (0 = No, 1 = Yes)	Mean time (months)	Median time (months)	Maximum time (months)	Standard deviation
0	26.22	15	130	28.47
1	30.08	24	129	23.09

Figure 4.3 below displays a boxplot of resolution time for cases by appeal status. The shape in both categories of time spent in months, No Appeal (0) and Appealed (1), substantiates the previous finding that appealed cases take longer to resolve. The interquartile range or IQR is wider in appealed cases and higher compared to non-appealed ones, a reflection of increased variability. The two categories have multiple high outliers, yet the range of these is larger in appealed cases, an indication that some of these cases in the group take long in processing. The boxplot visually portrays that longer and more varying timelines in resolution are a result of the appeal process.

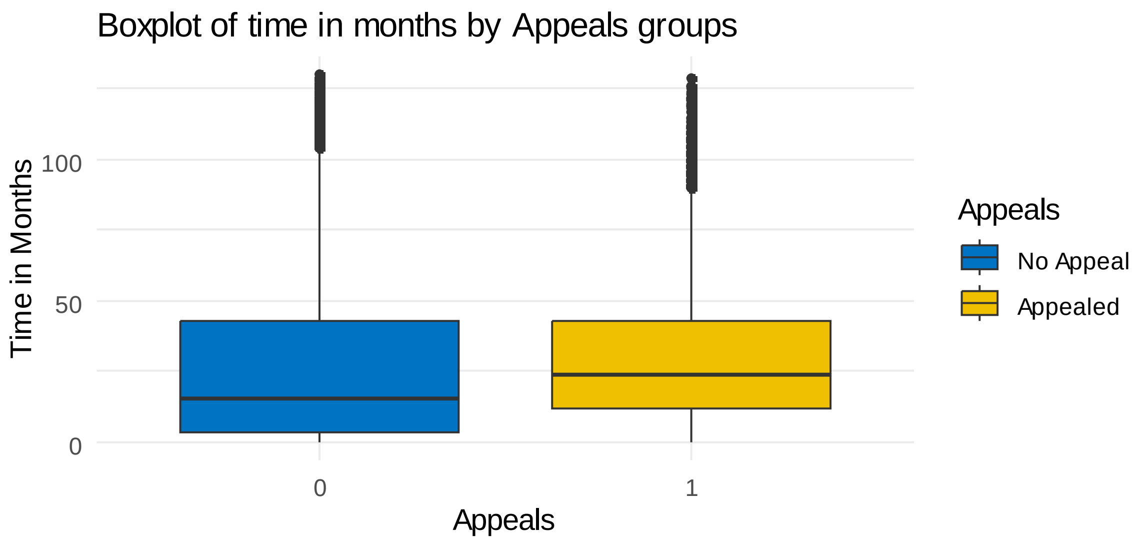


Figure 4.3: Boxplot of case duration (in months) by appeal status

4.2.6 Duration by appeal status and case type

The findings also showed that the duration categorized by case type and appeal status exhibit significant variances in the time it takes to resolve Kenyan high court cases. On average, Family law non-appealed cases had the highest mean duration of 40.36 months and high standard deviation of 30.26 to reflect large processing variability in the group of non-appealed family cases as shown in **Table 4.4**. Family appeals had a lower mean duration of 34.19 months, though with great variability (SD = 24.62). Civil cases also had lengthy duration with a mean of 27.82 months in the absence of appeals and 35.26 months with appeals. Anti-Corruption and Economic crimes had relatively shorter case duration with means of 12.38 (no appeal) and 20.91 (with appeal). Constitutional & Human Rights non-appealed had a mean duration of 21.53 months with those having appeals showing a higher mean of 21.70 months. In essence, Criminal non-appeals took a mean of 12.90 months and appeals took a mean of 22.94 months, showing that appeals increase resolution time in most criminal case categories. All these statistics give general insights into the impact of case type and appeal on the length of time taken in court and which areas of the justice system suffer most from delays.

Table 4.4: *Summary of case duration in months by appeal status and case type*

Case Type	**Appeals** (0 = No, 1 = Yes)	**Mean (months)**	**Median (months)**	**Standard Deviation**
Anti-Corruption and Economic Crimes	0	12.38	6.0	14.97
Anti-Corruption and Economic Crimes	1	20.91	14.0	17.90
Civil cases	0	27.82	16.0	28.41
Civil cases	1	35.26	31.0	24.47
Constitutional & Human Rights cases	0	21.54	14.0	22.23
Constitutional & Human Rights cases	1	21.70	5.0	23.89
Criminal cases	0	12.90	2.0	20.91
Criminal cases	1	22.94	17.0	18.77
Family cases	0	40.36	35.0	30.26
Family cases	1	34.19	31.0	24.62

The results show a comparative boxplot of the distribution of resolution times in months for different case categories, additionally stratified based on appeal status as in figure 4.4 below. Overall across categories, appealing cases tended to take longer than non-appealed ones given higher medians and wider IQRs. The effect is particularly marked in Civil, Constitutional, and Family cases, in which differences between appealed and non-appealed cases are large. Also evident are large numbers of high outliers for all categories but most noticeably for Family and Civil categories, demonstrating that some cases face undue delays. The tighter and shorter distribution in Criminal and Anti-Corruption categories demonstrates more uniform and prompt adjudication processes even in those that do involve an appeal. The figure supports previous statistical results that both case category and having an appeal are major factors in differences in determination time for cases.

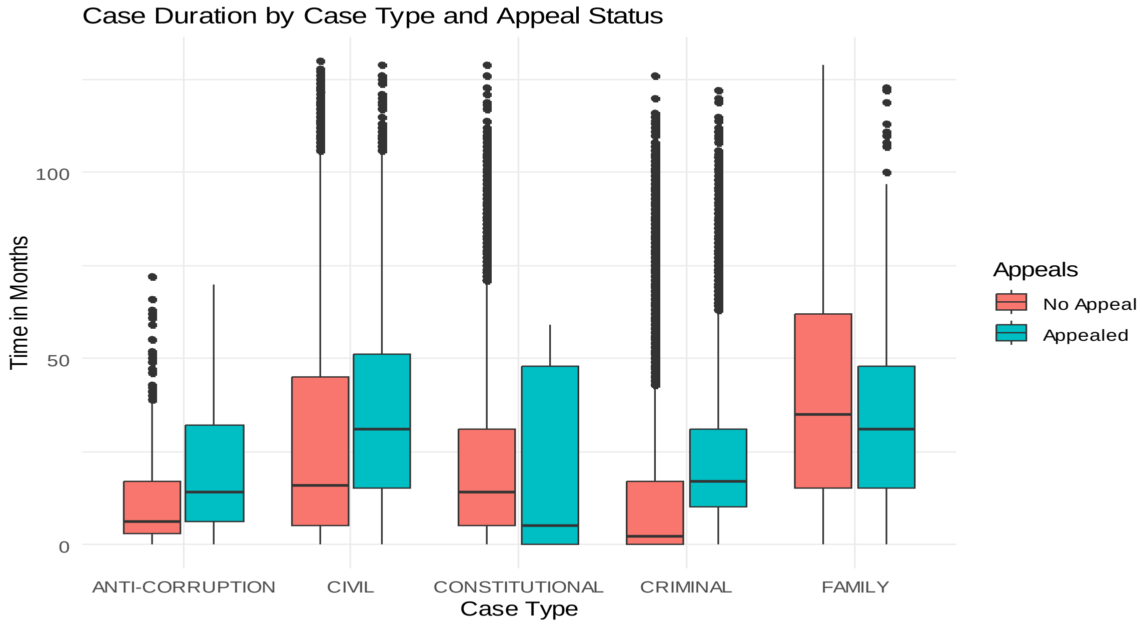


Figure 4.4: Boxplot for case type and appeal status

The exploratory data analysis identified significant imbalances in case determination times between case types and appeal statuses of cases in Kenyan high courts. The persistent right-skewness of case distribution within Family and Civil cases attests to the widespread occurrence of delays in some legal categories. The marked rise in case duration with the appeal process indicates its contribution to inefficiency in court processing. The implication of these results lies in providing essential benchmarking information about case duration structure and behavior, lending rationale to using survival analysis models in subsequent sections to improve prediction and interpretation of factors determining time to case determination.

4.3 Modeling case determination time in Kenyan high courts using survival analysis techniques

4.3.1 Kaplan-Meier survival probabilities

Kaplan-Meier survival analysis provides results showing a declining probability of case determination over time in Kenyan high courts from 0.674 (95% CI = [0.671, 0.677]) at 12 months to 0.002 (95% CI = [0.002, 0.002]) at 120 months as shown in **Table 4.5**. The implications are that a large percentage of cases remain undecided for long durations, and the survival probability decreases dramatically in the first part of the study period and then levels off at decreasing levels as time goes on. At 12 months, 67.4% of cases remain undecided, while 0.2% of cases are still undecided by 120 months. The associated standard errors are small relative to the size of these estimates for all time points. The confidence limits also get smaller as time goes on because of declining variation in survival estimates since progressively fewer remain in the analysis as time increases. These findings show long durations of case resolution in Kenyan high courts and point out the need for reducing delays through interventions for improving judicial efficiency.

Table 4.5: *Kaplan-Meier survival probabilities at yearly time points*

TIME (MONTHS)	SURVIVAL PROBABILITY	STANDARD ERROR	LOWER CI	UPPER CI
12	0.674	0.002	0.671	0.677
24	0.472	0.002	0.469	0.476
36	0.343	0.002	0.339	0.346
48	0.238	0.001	0.235	0.241
60	0.163	0.001	0.160	0.165
72	0.098	0.001	0.096	0.100
84	0.051	0.001	0.049	0.052
96	0.025	0.001	0.024	0.026
108	0.010	0.000	0.010	0.011
120	0.002	0.000	0.002	0.002

The Kaplan-Meier survival curve provides graphical representation of the likelihood that a case is still unresolved over time in Kenyan high courts. The curve begins at 1.00 because all cases are originally unresolved. The survival probability decreases monotonically over time as a cumulative effect of resolved or censored cases. At 12 months, the survival probability decreases to around 0.674, implying that approximately 32.6% of cases are still unresolved as shown in figure 4.5 below. The survival continues, decreasing to 0.002 by 120 months, which implies that a very small fraction of unresolved outstanding cases exists after 10 years.

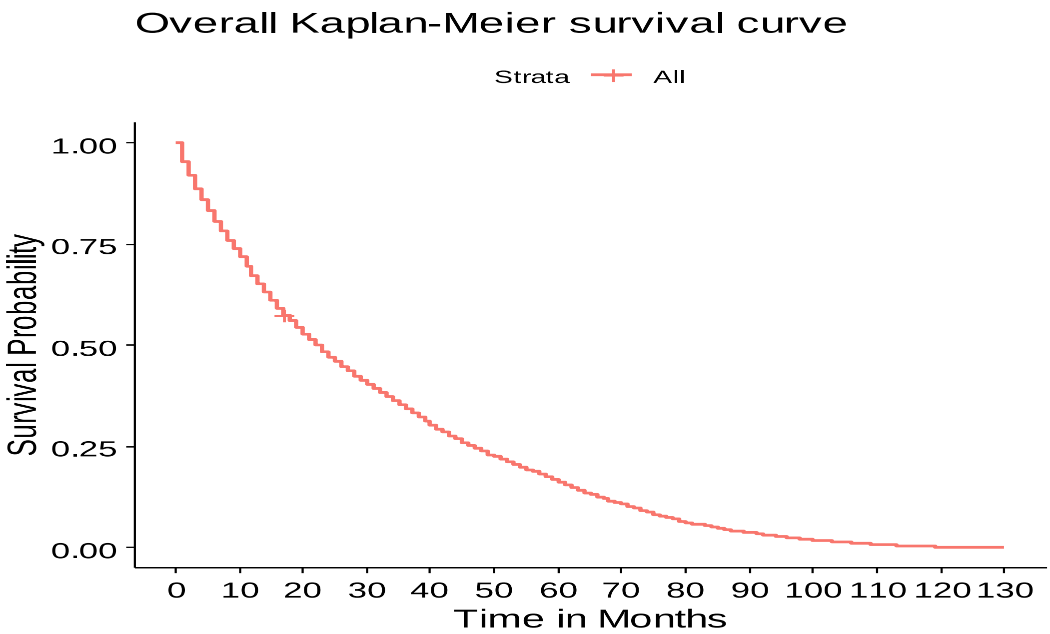


Figure 4.5: Kaplan-Meier survival curve

4.3.2 Kaplan-Meier comparison of cases with and without appeals

The results are summarized in **Table 4.6** below shows survival probabilities for different time intervals for both non-appealed and appealed cases. For non-appealed cases (Appeals = 0), by 12 months, the survival probability is 0.644, illustrating that about 64.4% of non-appealed cases have not been resolved. By 24 months, this drops off to 0.462, illustrating about 46.2% of cases still pending. The survival probability continues its decrease in an overall linear pattern all the way through 120 months, where it equals 0.002 and indicates 0.2% of non-appealed cases remain pending after 10 years. For appealed cases (Appeals = 1), by 12 months, the survival probability is higher at 0.750, illustrating a longer time period before resolution relative to non-appealed cases. By 24 months, the survival probability is 0.499, illustrating a similar pattern yet beginning from a point higher than non-appealed cases. The survival probability continues declining in an overall linear manner through 120 months, where it reaches 0.001 and indicates nearly all cases that were appealed have been resolved or censored. The provided confidence intervals (CIs) for these values reflect an estimation of these values and narrow as time increases because an increasing number of circumstances would lead to a narrowing estimation as time passes due to a reduced number of remaining cases at subsequent time points. These are striking given the long time it takes for appeals in that all time points for appealed cases have higher survival probabilities relative to non-appealed ones throughout. This is reflective of appeals having extensive impacts on durations as these are illustrated as being consistently higher throughout time relative to non-appealed ones.

Table 4.6: *Kaplan-Meier Survival Probabilities at Various Time Points by Appeal Status*

Appeal Status	**Time (Months)**	**Survival Probability**	**Standard Error**	**CI Lower**	**CI Upper**
Appeals = 0	12	0.644	0.002	0.640	0.648
Appeals = 0	24	0.462	0.002	0.458	0.466
Appeals = 0	36	0.348	0.002	0.344	0.352
Appeals = 0	48	0.251	0.002	0.247	0.254
Appeals = 0	60	0.178	0.002	0.175	0.181
Appeals = 0	72	0.112	0.001	0.109	0.114
Appeals = 0	84	0.059	0.001	0.057	0.061
Appeals = 0	96	0.031	0.001	0.030	0.033
Appeals = 0	108	0.013	0.000	0.013	0.014

Appeal Status	Time (Months)	Survival Probability	Standard Error	CI Lower	CI Upper
Appeals = 0	120	0.002	0.000	0.002	0.003
Appeals = 1	12	0.750	0.003	0.745	0.756
Appeals = 1	24	0.499	0.003	0.492	0.505
Appeals = 1	36	0.330	0.003	0.324	0.336
Appeals = 1	48	0.206	0.003	0.201	0.211
Appeals = 1	60	0.124	0.002	0.119	0.128
Appeals = 1	72	0.064	0.002	0.061	0.067
Appeals = 1	84	0.030	0.001	0.027	0.032
Appeals = 1	96	0.010	0.001	0.009	0.012
Appeals = 1	108	0.003	0.000	0.002	0.004
Appeals = 1	120	0.001	0.000	0.000	0.001

The Kaplan-Meier survival curves (**figure 4.6**) for the probability of case determination by time since filing, stratified by Appeal Status (Appealed vs. No Appeal). The curves both have a declining pattern for survival probability over time, yet those that have had an appeal (yellow) have a reduced rate of decline compared to those without an appeal (blue). At 12 months from filing, survival probability for non-appealed cases is about 0.64 and for appealed cases around 0.75. Survival probabilities for both populations begin to approach zero after 120 months but are consistently higher for appealed cases. The log-rank test ($p < 0.0001$) reveals a highly significant difference in these two populations, implying that appealing cases significantly extends file durations. These results suggest that interventions are necessary to remedy delay in the appellate process.

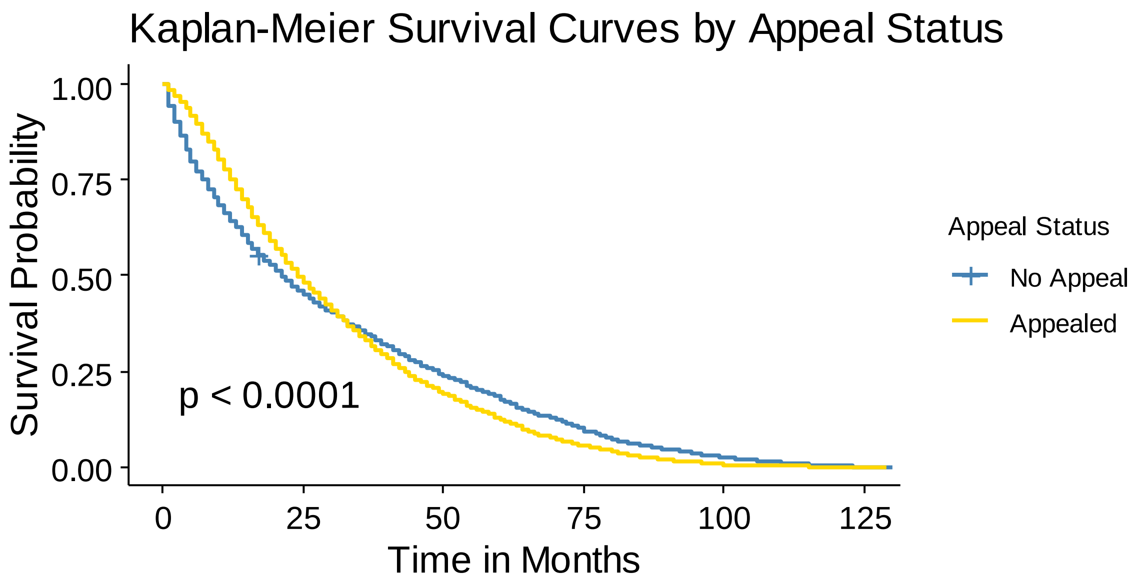


Figure 4.6: Kaplan-Meier survival curves stratified by appeal status (No Appeal and Appealed)

Further, the Kaplan-Meier survival curves provided rich insights related to probabilities of case determination over time by case type and by appeal status as shown in **figure 4.7**. The analysis demonstrates differences in survival probabilities by different legal categories for both non-appealed and appealed cases. In non-appealed cases, survival probabilities decrease linearly over time, although the rate of decrease differs by type of case. Anti-corruption and economic crimes have the quickest resolution, while family law has the longest resolution. Civil, constitutional, and criminal cases are in between, and civil cases have moderate long durations. A log-rank test ($p < 0.0001$) confirms marked differences in survival times between these categories. When including the effect of appeals, survival probabilities decrease more slowly in each type of case because of delays created by the appealing process. Even while declining overall, the survival probabilities in appealed cases have consistently higher values in each time duration compared to non-appealed cases. Family law and civil and constitutional matters are the slowest in resolution, although anti-corruption continues to have faster resolution even after including the effect of appeals. The log-rank test ($p < 0.0001$) once more confirms statistical significance in these differences.

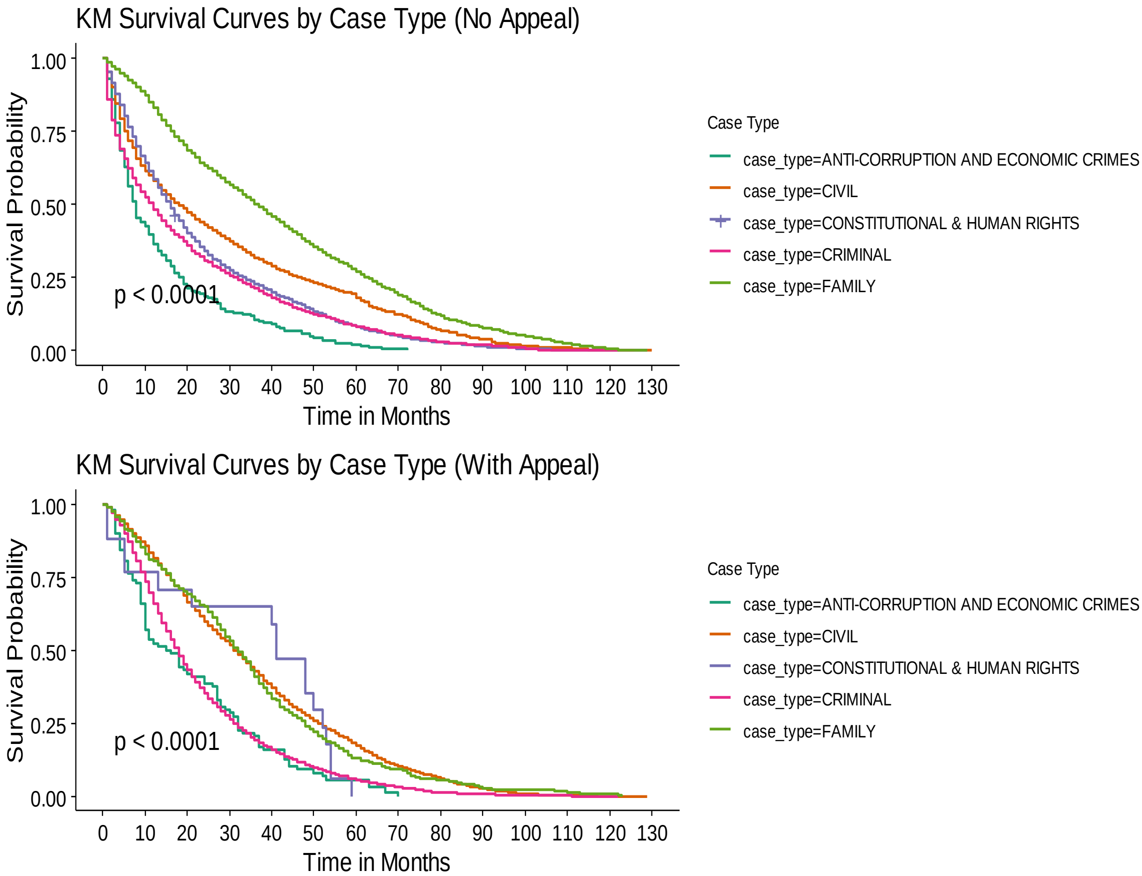


Figure 4.7: Kaplan-Meier survival curves by case type and appeal status

4.3.3 Kaplan-Meier comparison of type of cases

The curves in **figure 4.8** show probabilities over time for different legal categories. The analysis provides noticeable differences in survival probabilities for different legal categories and indicates varying time needed for different categories of cases for resolution. All curves demonstrate a monotonically declining pattern, as the odds of having a still unresolved case decrease as time progresses for all categories of cases. The rate of decline changes significantly for different categories of cases based on differences in complexity, procedure requirements, and resource appropriation. The rate of decline starts off steep for the anti-corruption category of cases (red curve), implying that these are speedily resolved compared to other categories. The survival probability for anti-corruption cases is about 0.75 after 12 months, thus about 75% of these are unresolved by then. With time elapsing, survival probability drops yet further down towards near-zero after 120 months, implying that nearly all these anti-corruption cases are resolved within 10 years. The civil category of cases (yellow curve) displays moderate decline in survival probabilities and a slower rate of resolution compared to anti-corruption but faster compared to constitutional and family issues. The survival probability for civil cases is about 0.80 after 12 months, implying a larger percentage of civil cases are unresolved compared to the anti-corruption category. Just as for other categories of cases, survival probability for civil cases tends towards zero after 120 months, implying that most of these are resolved within the same 10 years. The constitutional category of cases follows a similar pattern as for civil cases but has a linearly progressive fall in survival throughout time. After 12 months, survival probability for constitutional issues is around 0.85, implying longer duration compared to civil issues. The curve progressively drops and reaches near-zero level after 120 months just as for other categories of cases. Criminal category of cases (blue curve) shows a rapid drop in survival probabilities even compared to anti-corruption issues. The survival probability for criminal issues is about 0.70 after 12 months, implying about 70% of these are unresolved. Over time, criminal case survival probability also converges toward zero after 120 months, implying that the majority of cases are settled within 10 years. Family law cases (purple curve) have the lowest decline in survival probability, reflecting longest durations compared to all categories. The survival probability for family cases is around 0.90 within 12 months, implying that 90% of family cases are not settled. The survival probability for family cases continues decreasing but it is consistently higher compared to other categories during the study period even after 120 months. The log-rank test finding of $p < 0.0001$ for highly significant differences in survival probabilities between categories implies that case type is a good determinant of case duration since family law cases take significantly longer compared to other categories. The survival curves indicate that some categories of cases naturally take longer time for settlement, for example family law cases.

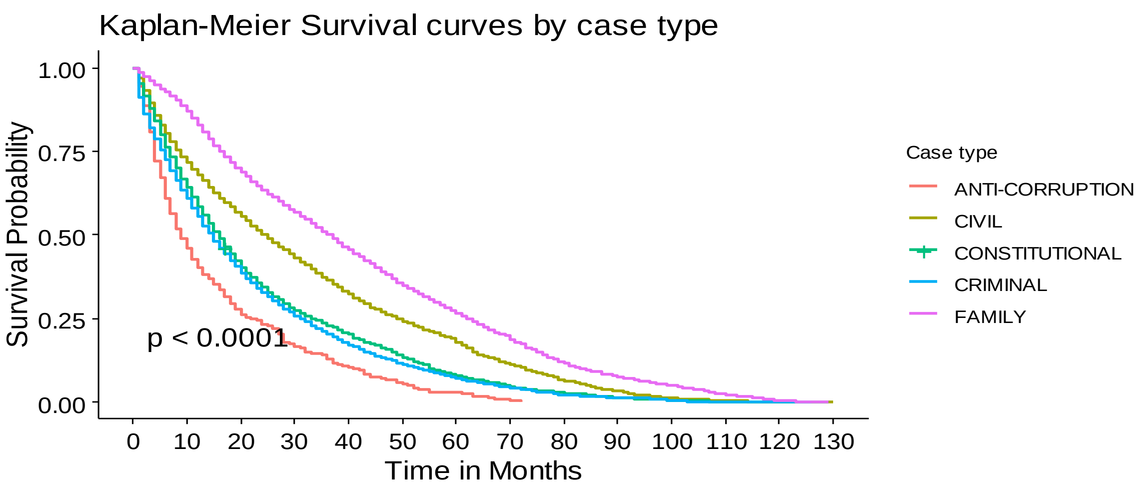


Figure 4.8: Kaplan-Meier survival curves by case type and appeal status

4.3.4 Cox Proportional Hazards Model

The Cox proportional hazards model including both appeals and case type (**Table 4.7**) identifies significant relationships between the predictors and case determination duration. All case types have significantly lower hazards in comparison to the reference category (Anti-Corruption). In particular, civil cases (HR = 0.502, 95% CI [0.456, 0.553], $p < .001$) and family law cases (HR = 0.331, 95% CI [0.301, 0.365], $p < .001$) have the longest durations, as suggested by their considerably lower hazard ratios. Also having significantly lower hazard ratios are criminal (HR = 0.930, 95% CI [0.844, 1.024], $p < .001$) and constitutional cases (HR = 0.625, 95% CI [0.566, 0.690], $p < .001$). In addition, having an appeal significantly decreases the case resolution hazard (HR = 0.732, 95% CI [0.720, 0.744], $p < .001$), as it illustrates again the appeal's association with longer case durations. The concordance index of this model at 0.647 indicates its the ability to distinguish between quickly and slowly resolved cases at a moderate level. These findings suggest case type and appeals as significant predictors of case determination time.

Table 4.7: ***Cox Proportional Hazards Model including case type and appeals***

Predictor	**β (SE)**	**HR**	**95% CI**	***z***	***p***
Civil Case	-0.690 (0.049)	0.502	[0.456, 0.553]	-13.988	< .001
Constitutional & Human Rights Case	-0.470 (0.050)	0.625	[0.566, 0.690]	-9.321	< .001
Criminal Case	-0.073 (0.049)	0.930	[0.844, 1.024]	-1.484	0.138
Family Law Case	-1.105 (0.049)	0.331	[0.301, 0.365]	-22.336	< .001
Appeal (Yes vs. No)	-0.312 (0.008)	0.732	[0.720, 0.744]	-37.963	< .001

Note. HR = Hazard Ratio; CI = Confidence Interval

For the exclusive estimation of the appeals effect, a separate Cox model was run with just the appeals variable as the predictor (**Table 4.8**). From this analysis, appeals are found to be associated with significantly higher hazard of longer case durations (HR = 1.062, 95% CI [1.046, 1.079], $p < .001$). Even though the effect is small in size, it is strong statistically, replicating the delaying effect of appeals on case closure. The concordance index for the model is 0.487 and implies low prediction power with appeals in isolation. However, the positive significant association means that appeals cases are likely to take longer to close relative to non-appealed cases.

Table 4.8: ***Cox Proportional Hazards model for appeals only***

Predictor	**β (SE)**	**HR**	**95% CI**	***z***	***p***
Appeal (Yes vs. No)	0.060 (0.008)	1.062	[1.046, 1.079]	7.658	< .001

Note. HR = Hazard Ratio; CI = Confidence Interval

In addition, another Cox model was estimated to consider the impact of case type on survival time in isolation (**Table 4.9**). The findings confirm that in comparison to anti-corruption cases, all other case types have significantly lower hazards of resolution. Civil cases have 0.459 (95% CI [0.414, 0.508], $p < .001$) hazard ratio while family law cases have the lowest hazard ratio at 0.331 (95% CI [0.299, 0.366], $p < .001$), the longest durations. Constitutional and criminal cases also have significantly lower hazards at hazard ratios of 0.641 (95% CI [0.578, 0.712]) and 0.696 (95% CI [0.628, 0.770]), respectively, both $p < .001$. A concordance index of 0.594 for the model indicates a moderate discriminative ability. These findings support the inference that case type has a significant effect on the time it takes for a case to be decided.

Table 4.9: *Cox Proportional Hazards model by case type only*

Predictor	**β (SE)**	**HR**	**95% CI**	***z***	***p***
Civil Case	-0.779 (0.052)	0.459	[0.414, 0.508]	-15.003	< .001
Constitutional & Human Rights Case	-0.444 (0.053)	0.641	[0.578, 0.712]	-8.374	< .001
Criminal Case	-0.363 (0.052)	0.696	[0.628, 0.770]	-6.982	< .001
Family Law Case	-1.107 (0.052)	0.331	[0.299, 0.366]	-21.238	< .001

Note. HR = Hazard Ratio; CI = Confidence Interval.

4.4 Comparing model precision between Cox regression and other parametric survival models

i. Weibull model

The Weibull model requires an assumption of a particular shape for the hazard in time and serves as a flexible substitute for the Cox model. All predictors were statistically significant at $p < .001$ as demonstrated in **Table 4.10**. Relative to the comparison category (most likely Commercial cases), cases in the category of Civil cases ($\beta = 0.674$, $p < .001$), Constitutional and Human Rights cases ($\beta = 0.432$, $p < .001$), Criminal cases ($\beta = 0.310$, $p < .001$), as well as Family Law cases ($\beta = 0.992$, $p < .001$) had all positive coefficients and longer survival time. An appeal also significantly lengthened the time to case closure ($\beta = 0.130$, $p < .001$). The Weibull scale parameter was 0.903, and hence we have modest departure from the exponential distribution. The model's likelihood ratio chi-square was 6221.5 (df = 5, $p < .001$), representing significant model fit.

Table 4.10: *Parametric survival model (Weibull Distribution)*

Predictor	**β (SE)**	**z**	**p-value**
Intercept	2.767 (0.0467)	59.31	< .001
Appeals	0.130 (0.0078)	16.60	< .001
Civil Case	0.674 (0.0469)	14.37	< .001
Constitutional & Human Rights Case	0.432 (0.0480)	9.01	< .001
Criminal Case	0.310 (0.0470)	6.60	< .001
Family Law Case	0.992 (0.0470)	21.08	< .001
Log(scale)	-0.102 (0.0028)	-36.36	< .001

ii. Exponential model

The exponential model makes the assumption of constant hazard across time, which is a strong assumption. All predictors were significant as indicated in **Table 4.11**. Civil ($\beta = 0.679$), Constitutional ($\beta = 0.443$), Criminal ($\beta = 0.307$), and Family Law ($\beta = 1.017$) cases again predicted longer resolution times compared to the reference case type. Appeals added $\beta = 0.153$ ($p < .001$) to predicted time. Although simple in nature, the exponential model fitted the data significantly ($\chi^2 = 5560.28$, df = 5, $p < .001$), although less well than the Weibull model.

Table 4.11: *Parametric survival model (Exponential Distribution)*

Predictor	**β (SE)**	**z**	**p-value**
Intercept	2.714 (0.0516)	52.57	< .001
Appeals	0.153 (0.0086)	17.75	< .001
Civil Case	0.679 (0.0519)	13.07	< .001
Constitutional & Human Rights Case	0.443 (0.0531)	8.35	< .001
Criminal Case	0.307 (0.0520)	5.91	< .001
Family Law Case	1.017 (0.0521)	19.53	< .001

iii. Log-normal model

The log-normal approach predicts a log-normal distribution of survival times and accommodates non-monotonic hazard functions. In **Table 4.12**, all of the predictors were statistically significant. Most notably, appeals had a strong effect ($\beta = 0.506$, $p < .001$), which indicates that an appeal significantly delays cases. The largest coefficient was for Family Law cases ($\beta = 1.260$), signifying the longest resolution times. The scale parameter of the model was 1.12 and the likelihood ratio chi-square was 9606.11 ($p < .001$), which implies good model fit and flexibility.

Table 4.12: *Parametric survival model (Log-normal Distribution)*

Predictor	β (SE)	z	p-value
Intercept	2.110 (0.0579)	36.45	< .001
Appeals	0.506 (0.0096)	52.53	< .001
Civil Case	0.627 (0.0582)	10.77	< .001
Constitutional & Human Rights Case	0.537 (0.0596)	9.02	< .001
Criminal Case	0.181 (0.0584)	3.11	.0019
Family Law Case	1.260 (0.0584)	21.57	< .001
Log(scale)	0.115 (0.0025)	46.37	< .001

iv. Log-logistic model

The log-logistic model is similar to the log-normal but is capable of accommodating a more flexible hazard function, especially with heavy tails. **Table 4.13** shows all of the predictors were very significant. Appeals were related to an extended time to resolution (β = 0.484), and cases of Family Law had the greatest coefficient (β = 1.330). The scale factor was 0.643 and the chi-square was 9515.4 ($p < .001$), showing good fit. The log-logistic model, as for the log-normal, has the advantage of being able to model non-monotonic hazards.

Table 4.13: *Parametric survival model (Log-logistic Distribution)*

Predictor	β (SE)	z	p-value
Intercept	2.125 (0.0583)	36.46	< .001
Appeals	0.484 (0.0097)	49.91	< .001
Civil Case	0.704 (0.0586)	12.00	< .001
Constitutional & Human Rights Case	0.588 (0.0599)	9.80	< .001
Criminal Case	0.237 (0.0588)	4.02	< .001
Family Law Case	1.330 (0.0587)	22.64	< .001
Log(scale)	-0.442 (0.0029)	-152.07	< .001

v. Gamma model

The Gamma model has the advantage of flexible shape to capture skewed survival time distributions. Coefficients from the model (**Table 4.14**) point to coherent patterns: all of the

Civil, Constitutional, Criminal, and Family Law cases have positive coefficients suggesting longer survival times compared to the reference category. Appeals once more were found to be associated with slower resolution. Although the model was found to be a good fit overall, the log-likelihood was marginally poorer in comparison to log-normal and log-logistic models.

Table 4.14: *Parametric survival model (Gamma Distribution)*

Predictor	**Coefficient**
Appeals	0.279
Civil Case	0.371
Constitutional & Human Rights Case	0.079
Criminal Case	0.284
Family Law Case	0.261

4.4.1 Model fit comparison using Akaike Information Criterion (AIC)

To identify the best model for time to case determination, the Akaike Information Criterion (AIC) was utilized to contrast overall fit of the five survival models. A lower AIC value signifies a better-performing model, penalizing for complexity to circumvent overfitting. A summary of AIC for all the models is presented in **Table 4.15**.

Table 4.15: *AIC Comparison across models*

Model	**Degrees of Freedom (df)**	**AIC**
Cox Proportional Hazards	5	1,668,250.5
Weibull	7	713,383.0
Exponential	6	714,643.8
Log-normal	7	721,860.8
Log-logistic	7	723,709.4
Gamma	7	713569.9

From the comparison, it is obvious that the Weibull model has the best fit to the data with the lowest AIC (713,383.0), substantially superior to the widely used Cox proportional hazards model with an AIC of more than two times as high (1,668,250.5). This difference implies that although the Cox model is good for semi-parametric interpretation and estimation of hazard ratios (as described under Objective 2 in Table 9), it is not as precise as fully parametric models in describing the survival time distribution for court cases data.

Gamma model as well as AIC of 713,569.9 ranks high in approaching the Weibull model's performance in modeling time-to-event dynamics. The Weibull model's slight edge in AIC does suggest its extra structure being more coincident with the hazard patterns in the data.

The exponential model, although nested in the Weibull model, does not perform as well as Weibull with an AIC of 714,643.8, showing the added flexibility of the Weibull distribution is warranted. Log-normal and log-logistic models also fare less well than Weibull with AICs of 721,860.8 and 723,709.4 respectively. This implies that although these models better capture the effects of skewness and non-monotonic hazard shapes, they do not model the dynamics of the data as well as the Weibull model.

The inferential findings and model fit statistics point to the same conclusion that the Weibull model best describes time until determination of a case in the Kenyan High Courts. It is more informative than the Cox model because it accommodates the shape of the hazard function, and it estimates the case type and appeal status effects more accurately.

CHAPTER FIVE

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This chapter interprets and contextualizes the findings from this study against the backdrop of current literature, including the organization of the judicial system in Kenya. The discourse follows from the specific aims in this study by connecting quantitative findings with general judicial performance themes, specifically those mentioned in Chapter 1, as well as literature examined in Chapter 2. Despite the extensive reforms in Kenya's judiciary, the backlog of undecided cases remains a systemic issue, and knowledge about how long it takes to decide various kinds of cases holds the key to resolving inefficiencies. The use of survival analysis in this analysis made it possible to evaluate thoroughly case duration thus providing information on underlying trends and predictors of judicial delays.

5.2 Exploratory data analysis of factors determining the survival time of cases in Kenyan high courts

The exploratory examination reveals significant case times to resolution variation across different legal areas under the Kenyan high courts. The variations reflect representations of systemic characteristics inherent in Kenya's judicial system, including differential case management approaches, judicial specialization, and work inefficacies in specific divisions. These variations align with observations from Amutah (2015), whose delays he assigned primarily to institutional capacity, resource constrains, as well as to effective case management strategies.

More time in family and civil cases is similar to the complexity of issues typically engaged, e.g., personal relationships, inheritance issues, and property matters. These types of cases entail lengthy procedural steps, e.g., social inquiry reports and strings of hearings, resulting in longer adjudication. The delays are compounded by the scarcity of resources such as fewer judicial officers to deal with these divisions and continuous adjournments, especially in the law of family cases where social and emotional issues are usually rampant. The constitutional guarantees of family and property rights also creates layers of judicial complexity, therefore elongating case lifecycles. This concurs with Chinyio and Chisawani (2017), who noted that court staff training and sound decision-making are crucial to the reduction of case backlogs.

In contrast, criminal and anti-corruption cases are resolved faster. This could be attributed to the special rules of procedure and specialized judicial machinery available for the speedy disposal of these cases, a reflection of the constitutional priority accorded to the right to a fair and speedy trial. Having special divisions such as the Anti-Corruption and Economic Crimes Division enables focused monitoring and case management, which works to limit unnecessary delays.

The right-skewed distribution of case durations, where most of the cases were closed within shorter time frames and others with considerable delays, indicates inherent judicial backlogs. These long-standing cases may result from structural inefficiencies such as judicial transfers, case backlog, adverse file management, and litigant-related adjournments, all causing stresses on resources and irregular timelines. These are corroborated by Maseno (2019), who illustrated how technology can enhance court resource management and reduce backlog.

Appeals have an enormous bearing on case length, and cases undergoing one or more appeals are longer. The multi-level appellate system in Kenya, while essential for reasons of securing justice and legal consistency, provides additional procedural steps that render case times longer. The bearing is strongest for civil and family cases, where appeals contribute more towards case length than for criminal and anti-corruption cases, where statutory limitation of appeals provides checks on delays. This is consistent with Wakhungu (2018), who emphasized the need for EDA-based enhancements in case management improvement and appeal process burden reduction and is consistent with Ngonga (2019), who also highlighted the relevance of statical analysis techniques in case duration modeling in the globe.

In view of the delays experienced especially in family and civil cases, the study suggests that the promotion of more use of Alternative Dispute Resolution (ADR) processes could be a strategic initiative to check case backlog as well as reduce time periods. ADR methods such as mediation and arbitration can resolve disputes faster and in concordance without delving into long court procedures. Encouraging ADR aligns with Kenya's broader efforts at judicial reform aimed at unclogging courts and providing access to justice in a timely fashion, particularly in matters expected to be litigated for many years and appealed for many years after that. This accords with the findings of Amutah (2015) and Wakhungu (2018), which call for expedited institution-level reforms through the expansion of ADR mechanisms.

The interrelation between case type, appeal status, and the potential role of ADR supports differential judicial interventions for various legal categories. For instance, enhanced pre-trial ADR procedures in family and civil cases could filter out disputes, thereby reducing the number of cases that proceed to full trial and appeal stages. Meanwhile, criminal and corruption cases may still rely on their existing expedited judicial processes, complemented where possible by focused case management. These findings confirm Emmert-Streib and Dehmer (2019), who emphasized that survival analysis techniques represent a robust paradigm for analyzing time-dependent events in judicial proceedings.

These findings confirm the relevance of case complexity, procedural structure, appeal dynamics, and settlement possibilities as core factors of judicial efficiency in Kenyan High Courts. The analysis also demonstrates valid survival analysis techniques for explaining the time-dependent nature of case disposal and informing targeted reforms to streamline delays and improve access to justice.

5.3 Predicting case determination time in Kenyan high courts using survival analysis

The study also attempted to determine the length of time for which a case would linger in Kenyan high courts through survival analysis methods. With the use of Kaplan-Meier estimators and proportional hazards models in Cox regression, the analysis created significant points around how interaction with time affects nature of cases and status of appeals to determine judicial effectiveness. The study focuses on the dynamic interaction between systemic institutional facts and case-specific variables governing how quickly, or gradually, justice is administered in the Kenyan context.

Kaplan-Meier estimates of survival, reflecting overall probability of a case not being settled over time, fell precipitously in the initial months after case filing. Most cases are not settled after one year, with survival probability decreasing further successively, suggesting many get weeded out within a few years while many are subject to long delays. The trend conforms with an alleged backlog bedeviling Kenya's judiciary and corroborating national reports of increasing caseload pressures at the country's high courts. The trend reflects a judicial system where early resolution can occur yet long-term clogs within it hinder general efficiency, particularly in more complex or contested cases. These results support Medvedeva et al. (2020), who demonstrated that statistical and machine learning approaches can be used to predict case durations globally.

When we disaggregate by appeal status, we see a stark contrast between appealed and non-appealed ones. At every time point, appealed cases had increased survival probabilities, indicating that they were unresolved longer than non-appealed ones. This aligns with the procedural nature of Kenya's legal system, where appeals create additional levels of litigation, each subject to independent judicial review. Each level of appeals either in High Court, Court of Appeal, or potentially up to the Supreme Court, adds time because records are submitted, hearings are re-listed, and lawyers prepare arguments. The survival curves more clearly showing that the appeals process creates a substantial delaying impact on case resolution even if we compare it with complex yet non-appealed ones. These findings confirm Dirks-Linhorst and Linhorst (2012), who argued that multivariate models are necessary to understand the dynamics of case duration and inform policy decisions.

Kaplan-Meier survival plots also exposed the varying rates of resolution by case type. Anti-corruption and criminal cases exhibited more rapid decreases in survival probabilities, suggesting more rapid resolution. This might be a reflection of special initiatives implemented to expedite resolution in these categories, including the setup of anti-corruption and economic crime divisions and compliance with constitutional mandates for timely trial in criminal cases. Civil, constitutional, and most notably family law cases exhibited more gradual survival decreases, suggesting longer time before case resolution. This can be explained by the inherent complexity of those cases. Family cases are often associated with custody over children, inheritance, and affective disputes requiring social enquiry reports and frequent adjournments. Civil and constitutional cases are often likely to require large amounts of proof, legal argument, and review, leading to longer timelines.

One of the most telling parts of the analysis was comparing survival probabilities both stratified by case type and by appeal status. This interaction brought out compounded effects upon case resolution times. Civil and family cases involving appeals had the most declines in survival probability at a very gradual slope, suggesting considerably longer durations. This further suggests not only are certain categories of cases inherently complex, but appellate processing adds to this complexity. Conversely, even in anti-corruption or criminal ones being appealed, resolution was faster than in unappalled civil or family cases. This indicates a certain legal category rewards procedural efficiency through either prioritization or streamlined processing, not uniformly applied to all case types.

The Cox proportional hazards models reinforced these findings. The model including both case type and appeal status showed each predictor had a significant, independent impact on case resolution time. Hazard ratios for family law cases and civil law cases were considerably lower than those for the reference category (anti-corruption), showing a considerably lower chance of early resolution. The variable, case status was also significant, with cases with appeals far less likely to resolve speedily. These findings confirm survival curve readings and confirm the predictive power of the model for those factors contributing to delays.

Even in models measuring individually the impacts of case type and appeal status, the same trends continued. Appealed cases were shown to be independently associated with longer resolution times. Alone, however, the appeal status had less predictive power than when coupled with case type, implying that together, these variables are essential in portraying a complete picture of judicial delays. The case type alone was a strong predictor, most especially for family and civil cases, whose hazard ratios were lowest and therefore had the slowest case progress.

Collectively, they offer a detailed picture of judicial performance within Kenyan high courts. They demonstrate that despite progress at the systemic level such as investment in specialist courts, integration with information and communication technology, the judicial system remains marred by severe challenges in some dimensions, most notably where appeals are concerned, and in family and civil law cases. These delays undermine the promise of timely justice in the constitution and indicate that targeted reforms are necessary to correct procedural inefficiencies, as well as case-type-specific ones. In conclusion, survival analysis methods have proven to be essential to determine the structural and procedural variables affecting case determination length in Kenyan high courts. The recurring identification of both appellation status and case type as key predictors extending resolution time serves to underscore both the imperative need for streamlining appeals procedures; standardizing case handling procedures across divisions, and redistributing resource allocation to the most heavily weighted legal categories. These would not merely help to advance speed but also build confidence in the judicial process.

5.4 Comparing model precision between Cox regression and parametric survival models

Comparison of Cox Regression and Parametric Models in Terms of Model Precision

Comparing survival models provides key insights into the ability of different statistical models to represent the patterns of case determination time in Kenyan High Courts. While the Cox proportional hazards model has traditionally been popular due to its semi-parametric character and ease of interpretation, results from this investigation demonstrate that parametric models, and especially the Weibull model, outperform them based on model fit and predictive ability. These results align with Teshnizi et al. (2017), who compared parametric and non-parametric survival models and concluded that parametric models like Weibull are more suitable for modeling hazard patterns in cancer data. Akaike Information Criterion (AIC) was employed as an initial indicator of model quality that balances model complexity against goodness of fit. Among all the models compared, the most well-fitting model with the lowest AIC value was the Weibull model. This indicates that it is the one that most accurately describes the underlying hazard structure of time to case determination. Its parametric flexibility allows it to model hazard rates that are increasing or declining with time, which is particularly ideal to model legal case lengths that are not expected to follow a constant hazard pattern.

The Gamma model's performance was close to that of the best models, indicating that it could also accurately capture skewed survival distributions, although not as efficiently as the Weibull. The Exponential model, a subset of the Weibull with a time-invariant hazard, also had a greater AIC, highlighting that its parsimony does not fit the changing risk of resolution.

Log-normal and Log-logistic models, although helpful in modeling non-monotonic hazards and skewness, fared less well, suggesting that while they are able to capture some of the complexities of the data, they could over-fit or distort the temporal dynamics of the Kenyan court processes.

Although the Cox model is still of value because of its intuitive interpretability and capacity to compare the relative covariate effect without assuming a form of the hazard function, the large AIC of the model in the current investigation confirms that it is less accurate in modeling true survival times than are its fully parametric alternatives. The result confirms previous results suggesting that Cox regression is not optimal where the shape of the hazard function is known or can effectively be estimated.

Practically speaking, these results have serious bearing. With survival times varying by case type and appeal status, using the Weibull model enables predictions that are more precise and case delay dynamics that are better understood. These assist the administrators of a court in pinpointing areas of greatest delay and instituting evidence-based reforms in those legal categories or appeal phases. For the Kenyan system of a high court, already criticized for delays and backlogs, that kind of analytical rigor is essential to planning with a proper allocation of resources, judicial policy-making, and transparency in the delivery of justice. The greater fit and flexibility of the Weibull model make it best suited to modelling case determination times within the Kenyan High Courts. It is a statistically robust and practically informative tool that contributes to deeper comprehension of judicial efficiency, and that surpasses conventional Cox regression in prediction and inference.

5.5 Conclusion

Case determination time analysis of Kenyan high courts through survival analysis methods brought into light a number of important observations that are vital to enhance judicial efficiency and minimize delays in the dispensation of justice.

The research findings strongly refute all three hypotheses, demonstrating that there are significant relationships and differences in the factors influencing the survival time of cases in Kenyan High Courts. The initial hypothesis that stated that the relationship between factors that have an impact on survival time was not significant was not accepted in view of the outcomes of the exploratory data analysis and survival modeling. The numbers showed that the type of case and the decision on appeal were important factors in determining the length of time that a case takes and the longest cases were under Family Law (mean = 40.19 months) and the shortest under Anti-Corruption and Economic Crimes (mean = 14.22 months). These results shows that a few types of the law are intrinsically more time-consuming in terms of solving them since they are more complex, involve procedural requirements, and they are less of judicial concern. Also, cases delayed through appeals take very long to be determined as the Kaplan-Meier curves show the increased survival probability of an appeal case to that of non-appealed. This tends to back up the fact that, both type and existence of appeal significantly influence the period during which a case takes.

The second hypothesis, according to which the absence of significant difference in factors that predict the survival time was present was also not supported. Strong evidence was that the different type of cases experiences a dissimilar hazard ratio and that the Family Law cases are the least likely to be closed quickly (HR = 0.331). The hazard ratio in civil cases was moderate (HR = 0.502), and the hazard ratio in the Criminal cases was bigger (HR = 0.696), which implies they were to be closed more frequently and quickly. Appeals also decreased the risk of determining the case (HR = 0.732), which confirms the that they have a great impact on the case longevity. The statistical differences observed across case types and appeal statuses clearly contradict the hypothesis that there is no significant variation in the factors predicting survival time.

Lastly, the third hypothesis that stated that no difference in model accuracy between Cox regression and other parametric survival models also is rejected. When the model performance was compared with the use of Akaike Information Criterion (AIC), parametric models, Weibull model performed better than the Cox proportional hazards model. The AIC value of a Weibull model has been the smallest (713,383.0); the Gamma model almost as small (713,569.9); however, the AIC value of the Cox model is very high (1,668,250.5 Lastly, the third hypothesis that stated that no major distinction in model accuracy existed between Cox regression and other parametric survival models also failed. When the model performance was compared with the use of Akaike Information Criterion (AIC), parametric models, especially the Weibull model, performed better than the Cox proportional hazards model. The AIC value of a Weibull model has been the smallest (713,383.0); the Gamma model almost as small (713,569.9); however, the AIC value of the Cox model is very high (1,668,250.5). It implies that parametric models are more predictive and data-fitting, representing the shape of the hazard function compared with the semi-parametric Cox model. The advantage of the Weibull model in the accuracy, and its ease in including different patterns of hazards also portrays on the significance of choosing the suitable survival models according to the nature of the data collected. All in all, the results verify the fact that model precision differs deeply and the parametric models are a more convincing method of the time of the survival of a case within the Kenyan High Courts.

5.6 Recommendations

On the basis of the foregoing research findings and the current conditions of the Kenyan high courts, the following are recommendations are made to promote judicial efficiency, eliminate backlogs, and enhance access to timely justice

a) To simplify case management of complex legal fields

Family Law cases took the most time to end, with unappalled Family Law cases taking a mean of 40.36 months, whereas Criminal cases took the least at 12.90 months. These differences support the necessity of interventions based on legal domain. The following can be implemented:

i. Create specialist divisions within the High Court to deal with intricate cases like Family Law and Civil Litigation.

ii. Implement compulsory pre-trial conferences and mediation processes to foster settlement at an earlier stage and curtail the number of cases that go to trial.
iii. Establish case tracking systems that identify delayed hearings or procedural delays to allow judicial oversight in advance.

b) Revamp the appellate process to minimize delays

Appeals were also found to add substantially to case determination time. For example, appealed Family Law cases also averaged 34.19 months compared to 40.36 months for unappalled ones with appeals contributing substantially to time even when they did not exactly double the time. In this regard, the following action points can be considered:

i. Impose strict timeframes for appeal filings and hearings, with automatic penalties for unjustified delays or late filings.
ii. Establish fast-track appellate chambers to deal with time-sensitive cases of children custody, injunctions, or business conflict where speed is of the essence.
iii. Encourage electronic filing and remote hearings in the appellate process to minimize logistical delays and accelerate decision-making.

c) Use the Weibull Model in predictive planning of cases

The Weibull model proved superior to other survival models, including the Cox regression model, with the lowest AIC of 713,383.0, signifying better fit and predictability. It also better represented the shape of the hazard function compared to semi-parametric models. Thus, the following action points can be implemented:

i. Implement the Weibull model in court management dashboards to predict anticipated resolution times of incoming cases by type and appeal status.
ii. Use these predictions to manage judicial time more effectively, prioritize intricate cases, and establish realistic expectations of litigants.
iii. Train court administrators and clerks in the application of statistical forecasting tools to aid evidence-based decision-making in managing case flow.

d) Improve judicial capacity and training

Judicial capacity is a persistent bottleneck, particularly where there are large numbers of cases. The evidence points to considerable regional and subject-matter variations in case length that could be compounded by a disproportionate allocation of judges and facilities. The following can be implemented to address the issue:

i. Increase hiring and deployment of new judges, especially to overburdened courts and under-resourced areas.
ii. Offer frequent training programs in contemporary case management methods, digital tools, and best practices on deciding complex and time-critical cases.
iii. Promote peer mentoring and sharing of best practices by judges through judicial forums and specialist training centers.

e) Enhance ICT integration and data infrastructure

Notwithstanding continuous efforts at digitization, irregular data gathering and reporting are still a problem. The judiciary's existing database may also gain from increased integration and access. The following action points can be implemented:

i. Expand deployment of the Judiciary Case Management System (JCMS) to facilitate real-time monitoring of all cases from the filing stage to the determination stage.
ii. Establish a central judicial data warehouse available to policymakers, researchers, and judicial administrators to support public policy making and performance benchmarking.
iii. Adhere to the Data Protection Act, 2019, to protect confidential legal data and foster transparency and accountability.

f) Foster Public Awareness and Access to Justice Delays in case determination undermine public confidence in the judicial process and the rights of litigants. Timely access to justice is a constitutional entitlement that has to be preserved. The following can be considered to improve on timely case determination:

i. Initiate public awareness programs to inform citizens of their rights, court process, and legal aid services available.
ii. Invest in legal aid organizations and paralegal networks to support vulnerable populations in navigating the judicial system more effectively.
iii. Promote the application of Alternative Dispute Resolution mechanisms outside of formal courts, primarily to deal with small civil and family disputes.

Kenya's high courts are severely challenged by case backlogs and delayed delivery of justice. The judiciary can, however, overcome these inefficacies through targeted reforms based on sound statistical models such as the Weibull survival model, combined with institutional reform and technological advancement. Through the above proposals, Kenya can fulfill its constitutional promise of promptly delivering justice, boosting public confidence, and developing a more just legal system.

REFERENCES

Amutah, M. (2015). Exploratory data analysis in Kenyan court systems: An empirical study of the delays and challenges. Journal of Business and Management, 17(3), 75-80.

Bielen, S., Grajzl, P., & Marneffe, W. (2017). Understanding the Time to Court Case Resolution : A Competing Risks Analysis.

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival Analysis Part III: Multivariate data analysis - Choosing a model and assessing its adequacy and fit. British Journal of Cancer, 89(4), 605–611.

Chinyio, E., & Chisawani, G. (2017). Exploratory data analysis of factors influencing court effectiveness in Kenya. Journal of Law, Culture and Humanities, 17(2), 93-95.

Davila, C. (2015). Research and eData: Factors Influencing Criminal Case Processing Time in McLean County Courts, Department of Sociology.

Dimitropoulos, G. (2018). Judicial efficiency and economic growth: Evidence from the European Court of Human Rights. Journal of Law and Courts, 6(2), 341–366.

Dirks-Linhorst, P. A., & Linhorst, D. M. (2012). Recidivism Outcomes for Suburban Mental Health Court Defendants. American Journal of Criminal Justice, 37(1), 76–91.

Emmert-streib, F., & Dehmer, M. (2019). Introduction to Survival Analysis in Practice. Machine Learning and Knowledge Extraction, 1013–1038.

Essebag, V., Platt, R. W., Abrahamowicz, M., & Pilote, L. (2005). Comparison of nested case-control and survival analysis methodologies for analysis of time-dependent exposure. BMC Medical Research Methodology, 5(5), 1–6.

Flinn, C. J. (2000). Modes of interaction between divorced parents. International Economic Review, 41(3), 545–578.

Gakuya, D. (2022). Technology and judicial efficiency in Kenyan courts. African Journal of Law and Technology, 5(1), 23–39.

Gibson, G. M., Lord, T., & Case, N. (2021). Justice Delayed is Justice Denied.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53(282), 457–481

Kleinbaum, D. G., & Klein, M. (2012). Survival Analysis: A Self-Learning Text.

KNBS. (2020). Statistical Abstract 2020.

Maseno, J. O. (2019). Impact of ICT on judicial efficiency in Kenyan courts. Journal of African Law, 63(2), 201–220.

Maseno, P. (2019). Exploratory data analysis of Kenyan High Court cases. Journal of Law and Technology, 32(2), 18-23.

Medvedeva, M., Vols, M., & Wieling, M. (2020). Using machine learning to predict decisions of the European Court of Human Rights. Artificial Intelligence and Law, 28(2), 237–266.

Moore, D. F. (2016). Applied Survival Analysis Using R.

Ngonga, E. O. (2019). A system dynamics model for resolution of legal disputes: a case of kisumu high court. Sustainability (Switzerland), 11(1), 1–14.

Perez, C., & Figueroa, J. (2006). The Duration of Civil Cases. Asurvival Analysis. Perspectives, 1(1).

Pradhan, I. (2018). Exploratory Data Analysis And Crime Prediction In San Francisco. SJSU Scholar Works, 51.

Teshnizi, S. H., Mohammad, S., & Ayatollahi, T. (2017). Comparison of Cox Regression and Parametric Models: Application for Assessment of Survival of Pediatric Cases of Acute Leukemia in Southern Iran. 18, 981–985.

Tsintzos, P., & Plakandaras, V. (2020). The judiciary system as a productivity factor ; the European experience. Economics Letters, 192, 109257.

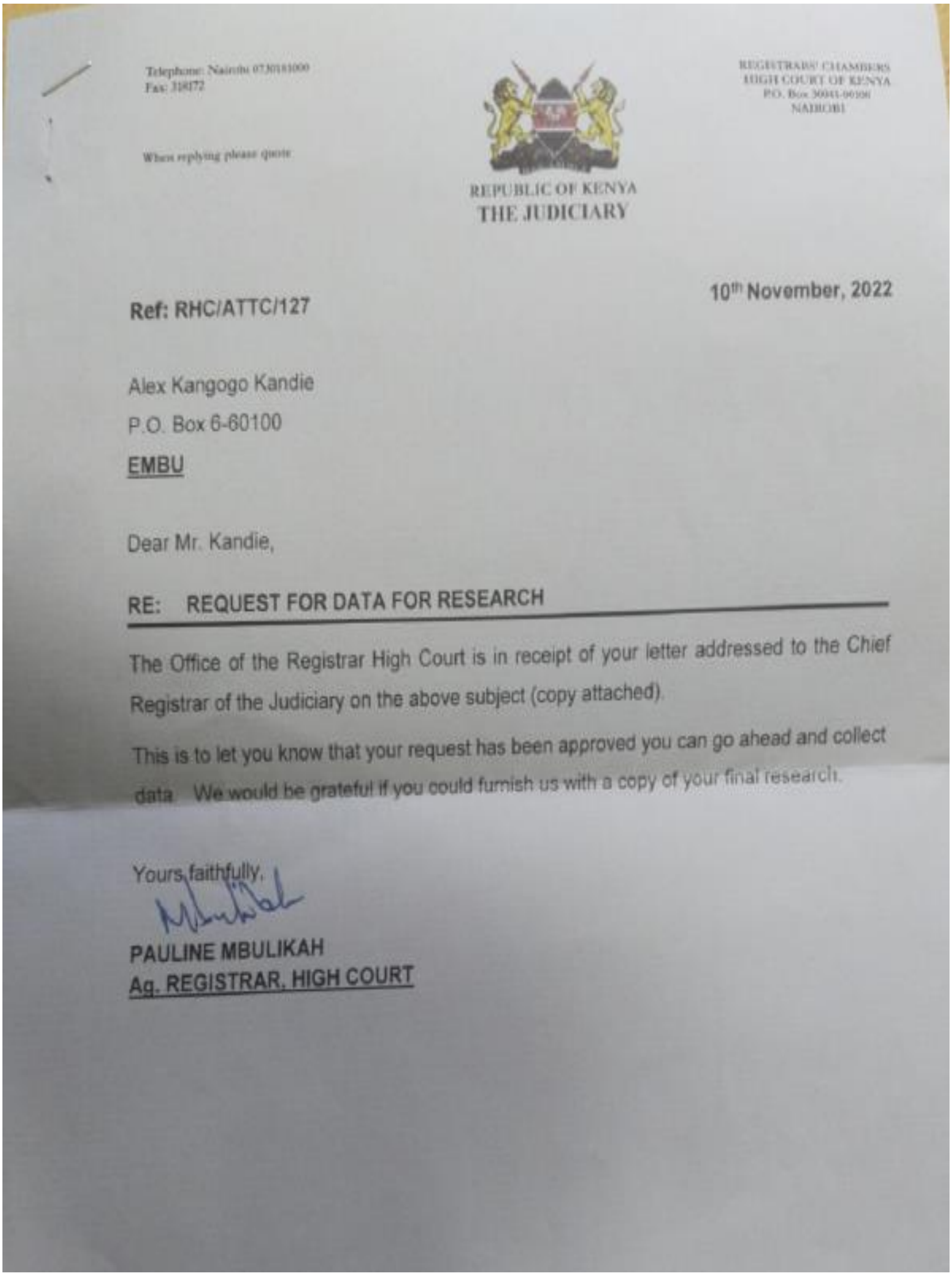
Smyth, R., & Narayan, P. K. (2016). The effect of judicial resources on case disposition times in the Australian High Court. *Journal of Empirical Legal Studies, 13*(4), 611–636.

Wakhungu, B. (2018). Exploratory data analysis of judicial process in Kenya. International Journal of Social Sciences and Management, 6(2), 9-16.

Wang, M. (2006). Summary Notes for Survival Analysis.

APPENDICES

Appendix 1: Approval to collect data from judiciary



Telephone: Nairobi 0730181000
Fax: 318172

When replying please quote

REPUBLIC OF KENYA
THE JUDICIARY

REGISTRARS' CHAMBERS
HIGH COURT OF KENYA
P.O. Box 30041-00100
NAIROBI

10th November, 2022

Ref: RHC/ATTC/127

Alex Kangogo Kandie
P.O. Box 6-60100
EMBU

Dear Mr. Kandie,

RE: REQUEST FOR DATA FOR RESEARCH

The Office of the Registrar High Court is in receipt of your letter addressed to the Chief Registrar of the Judiciary on the above subject (copy attached).

This is to let you know that your request has been approved you can go ahead and collect data. We would be grateful if you could furnish us with a copy of your final research.

Yours faithfully,

PAULINE MBULIKAH
Ag. REGISTRAR, HIGH COURT

Alex Kangogo Kandie,

P. O Box 6-60100,

Embu, Kenya.

Tel: 0724983757/ 0753504522

11/10/2022

Chief Registrar of the Judiciary,

P. O Box 30041 – 00100,

Nairobi, Kenya.

Dear Sir/Madam,

RE: REQUEST FOR DATA FOR MY RESEARCH

Greetings Chief Registrar!

I am Alex Kangogo Kandie a student at the University of Embu, registration number B531/1346/2019 pursuing a Master of Science in Statistics.

Regarding the above-mentioned subject, I am writing to request comprehensive data from (2010-2020 if applicable) on court cases for my research entitled "**Modelling survival time for a case to be determined in High Courts of Kenya**". One of the objectives is to determine significant factors that are associated with different durations in disposing of a court case. Some of my key parameters in the study are **demographic factors, time of case disposal, type of case, testimonial proof presented, number of parties involved, number of appeals** e.t.c. The findings from this research have enormous benefits to the judiciary since they will help in policy review and implementation.

I express my hearty gratitude in examining my request for data. The data will be used purposely for this research and it will be treated with the utmost confidentiality all protocols will be followed, and privacy regulations adhered to in accordance with the data protection Act 2019. If you have any questions or concerns, my contact information is 1346@student.embuni.ac.ke.

Attached to this letter is my recommendation from the School of Pure and Applied Sciences (SPAS).

Thank you.

Alex Kangogo Kandie

Appendix 2: List of courts and counties

Serial Number	County Number	Court Name	County
1.	1	Mombasa High Court	Mombasa County
2.	3	Malindi High Court	Kilifi County
3.	4	Garsen High Court	Tana River County
4.	6	Voi High Court	Taita Taveta County
5.	7	Garissa High Court	Garissa County
6.	10	Marsabit High Court	Marsabit County
7.	12	Meru High Court	Meru County
8.	13	Chuka High Court	Tharaka-Nithi County
9.	14	Embu High Court	Embu County
10.	15	Kitui High Court	Kitui County
11.	16	Machakos High Court	Machakos County
12.	17	Makueni High Court	Makueni County
13.	18	Nyandarua High Court	Nyandarua County
14.	19	Nyeri High Court	Nyeri County
15.	20	Kerugoya High Court	Kirinyaga County
16.	21	Muranga High Court	Murang’a County
17.	22	Kiambu High Court	Kiambu County
18.	23	Lodwar High Court	Turkana County
19.	24	Kapenguria High Court	West Pokot County
20.	26	Kitale High Court	Trans-Nzoia County
21.	27	Eldoret High Court	Uasin Gishu County
22.	30	Kabarnet High Court	Baringo County
23.	31	Nanyuki High Court	Laikipia County
24.	32	Naivasha High Court	Nakuru County
25.	32	Nakuru High Court	Nakuru County
26.	33	Narok High Court	Narok County
27.	34	Kajiado High Court	Kajiado County
28.	35	Kericho High Court	Kericho County
29.	36	Bomet High Court	Bomet County

30.	37	Kakamega High Court	Kakamega County
31.	38	Vihiga High Court	Vihiga County
32.	39	Bungoma High Court	Bungoma County
33.	40	Busia High Court	Busia County
34.	41	Siaya High Court	Siaya County
35.	42	Kisumu High Court	Kisumu County
36.	43	Homabay High Court	Homa Bay County
37.	44	Migori High Court	Migori County
38.	45	Kisii High Court	Kisii County
39.	46	Nyamira High Court	Nyamira County
40.	47	Milimani Anti-Corruption & Economic Crimes Division	Nairobi County
41.	47	Milimani Civil Division	Nairobi County
42.	47	Milimani Commercial & Tax Division	Nairobi County
43.	47	Milimani Constitutional Law & Human Rights Division	Nairobi County
44.	47	Milimani Criminal Division	Nairobi County
45.	47	Milimani Family Division	Nairobi County
46.	47	Milimani Judicial Review Division	Nairobi County

Appendix 3: Rmarkdown source code for the output in HTML format

````
---
title: "ANALYSIS OF TIME TAKEN FOR A CASE TO BE DETERMINED IN HIGH COURTS OF KENYA"
author: "Kandie Alex"
date: "`r Sys.Date()`"
output: html_document
---

```{r,message=FALSE}
# Load required packages
# install.packages("sass")
library(sass)
library(survminer)
library(dplyr)
library(ggplot2)
library(viridis)
library(knitr)
library(kableExtra)
library(flexsurv)
library(kableExtra)
```

```{r}
# Import the data
data <- read.csv("E:/Project/high court cases - Final  dataset.csv")

# Check for any non-positive survival times
summary(data$time_in_months <= 0)

# Removing records with non-positive survival times for non-parametric tests
clean_data <- data[data$time_in_months > 0, ]
```

```{r}
# Rename level names
data$case_type <- factor(data$case_type,
                levels = c("ANTI-CORRUPTION AND ECONOMIC CRIMES",
                       "CIVIL",
                       "CONSTITUTIONAL & HUMAN RIGHTS",
                       "CRIMINAL",
                       "FAMILY"),
                labels = c("ANTI-CORRUPTION", "CIVIL", "CONSTITUTIONAL",
                       "CRIMINAL", "FAMILY"))
```

```{r, warning=FALSE}
# Exploratory Data Analysis (Objective 1)
##################################################################

# 4.2.1
````
````

```
# Correct one-> Group by case_type and summarize time_in_months
summary_stats <- data %>%
  group_by(case_type) %>%
  summarize(
    mean_time = mean(time_in_months, na.rm = TRUE),
    median_time = median(time_in_months, na.rm = TRUE),
    min_time = min(time_in_months, na.rm = TRUE),
    max_time = max(time_in_months, na.rm = TRUE),
    sd_time = sd(time_in_months, na.rm = TRUE)
  )
# Print formatted table
kable(summary_stats, caption = "Summary statistics of time in months grouped by case type")
%>%
  kable_styling()


# 4.2.2
# Histogram of time variables
# Create a histogram of time_in_months
# Compute descriptive statistics
mean_time <- round(mean(data$time_in_months, na.rm = TRUE), 2)
median_time <- round(median(data$time_in_months, na.rm = TRUE), 2)
sd_time <- round(sd(data$time_in_months, na.rm = TRUE), 2)

# Create a data frame for the normal curve
x_vals <- seq(min(data$time_in_months, na.rm = TRUE),
              max(data$time_in_months, na.rm = TRUE),
              length.out = 200)
normal_curve <- data.frame(
    x = x_vals,
    density = dnorm(x_vals, mean = mean_time, sd = sd_time)
)

# Create histogram with normal curve and labeled stats
ggplot(data, aes(x = time_in_months)) +
    geom_histogram(aes(y = after_stat(density)),
                   bins = 30, fill = "skyblue", color = "black", alpha = 0.7) +
    labs(title = "Overall Distribution of Time in Months",
         x = "Time in Months",
         y = "Density") +
    theme_minimal() +

    # Add normal curve
    geom_line(data = normal_curve, aes(x = x, y = density),
              color = "darkgreen", linewidth = 1) +

    # Add vertical line for mean
    geom_vline(xintercept = mean_time, color = "red", linetype = "dashed", linewidth = 1) +
    annotate("text", x = mean_time, y = Inf, label = paste("Mean =", mean_time),
             vjust = 1.5, hjust = 1.1, color = "red", angle = 90, size = 4) +
```

```
  # Add vertical line for median
  geom_vline(xintercept = median_time, color = "blue", linetype = "dotted", linewidth = 1) +
  annotate("text", x = median_time, y = Inf, label = paste("Median =", median_time),
       vjust = 1.5, hjust = 1.1, color = "blue", angle = 90, size = 4) +

  # Annotate summary statistics in top-right corner
  annotate("text",
       x = Inf, y = Inf,
       label = paste("Mean =", mean_time, "\nMedian =", median_time, "\nSD =", sd_time),
       hjust = 1.1, vjust = 1.1, size = 4, color = "black", fontface = "bold")



# 4.2.3
# Distribution by case type

# Compute mean, median, and standard deviation for each case_type
summary_stats <- data %>%
  group_by(case_type) %>%
  summarise(
    Mean = mean(time_in_months, na.rm = TRUE),
    Median = median(time_in_months, na.rm = TRUE),
    SD = sd(time_in_months, na.rm = TRUE)
  ) %>%
  ungroup()

# Generate normal curve data
normal_curves <- summary_stats %>%
  rowwise() %>%
  do({
    case = .$case_type
    mean = .$Mean
    sd = .$SD
    x_vals = seq(mean - 4 * sd, mean + 4 * sd, length.out = 200)
    data.frame(
      case_type = case,
      x = x_vals,
      density = dnorm(x_vals, mean = mean, sd = sd)
    )
  }) %>%
  ungroup()

# Create label positions for SD text
sd_labels <- summary_stats %>%
  mutate(
    x = Mean + SD,  # position label at 1 SD right of mean
    y = dnorm(Mean + SD, mean = Mean, sd = SD),
    label = paste0("SD = ", round(SD, 2))
  )

# Final plot
```

```
ggplot(data, aes(x = time_in_months)) +
  geom_histogram(aes(y = after_stat(density)),
              binwidth = 5, fill = "skyblue", color = "black", alpha = 0.7) +
  labs(
    title = "Distribution of Time in Months by Case Type",
    x = "Time in Months",
    y = "Density"
  ) +
  facet_wrap(~ case_type, scales = "free") +

  # Add vertical lines for mean and median
  geom_vline(data = summary_stats,
         aes(xintercept = Mean, color = "Mean"),
         linetype = "dashed", linewidth = 1) +
  geom_vline(data = summary_stats,
         aes(xintercept = Median, color = "Median"),
         linetype = "dotted", linewidth = 1) +

  # Add text labels for mean and median
  geom_text(data = summary_stats,
        aes(x = Mean, y = Inf, label = paste("Mean =", round(Mean, 2))),
        vjust = 1.5, hjust = -0.3, color = "red", size = 3) +
  geom_text(data = summary_stats,
        aes(x = Median, y = Inf, label = paste("Median =", round(Median, 2))),
        vjust = 3, hjust = -0.5, color = "blue", size = 3) +

  # Add normal curve
  geom_line(data = normal_curves,
        aes(x = x, y = density),
        color = "darkgreen", linewidth = 1) +

  # Add SD labels
  geom_text(data = sd_labels,
        aes(x = x, y = y, label = label),
        color = "darkgreen", vjust = -1, hjust = -0.1, size = 3) +

  # Define colors for mean and median
  scale_color_manual(name = "Statistics",
              values = c("Mean" = "red", "Median" = "blue")) +

  theme_minimal() +
  theme(legend.position = "top")



# 4.2.4
# Group by Appeals and summarize time_in_months
summary_stats <- data %>%
  group_by(Appeals) %>%
  summarize(
```

```
    mean_time = mean(time_in_months, na.rm = TRUE),
    median_time = median(time_in_months, na.rm = TRUE),
    min_time = min(time_in_months, na.rm = TRUE),
    max_time = max(time_in_months, na.rm = TRUE),
    sd_time = sd(time_in_months, na.rm = TRUE)
  )

# Print formatted table
kable(summary_stats, caption = "Summary statistics of time in months grouped by Appeals")
%>%
  kable_styling()

# T-test to compare time by Appeals groups
t_test_results <- t.test(time_in_months ~ Appeals, data = data)

# Print the t-test results
cat("T-test comparing time by Appeals groups:\n")
print(t_test_results)


# Creating the boxplot by appeals
boxplot <- ggplot(data, aes(x = factor(Appeals), y = time_in_months, fill = factor(Appeals))) +
 geom_boxplot() +
 labs(
  title = "Boxplot of time in months by Appeals groups",
  x = "Appeals",
  y = "Time in Months"
 ) +
 scale_fill_manual(
  values = c("0" = "#0073C2", "1" = "#EFC000"),  # Define your custom colors here
  name = "Appeals",
  labels = c("No Appeal", "Appealed")
 ) +
 theme_minimal()

# Print the boxplot
print(boxplot)



# 4.2.5
# Computing summary statistics grouped by case_type and Appeals
# Compute summary statistics
summary_stats <- data %>%
 group_by(case_type, Appeals = factor(Appeals)) %>%
 summarise(
  Mean = mean(time_in_months, na.rm = TRUE),
  Median = median(time_in_months, na.rm = TRUE),
  SD = sd(time_in_months, na.rm = TRUE),
  n = n(),
  .groups = "drop"
```

```
 )

# Print styled table
kable(summary_stats,
     caption = "Summary Statistics of Case Duration by Case Type and Appeal Status") %>%
  kable_styling(full_width = FALSE, bootstrap_options = c("striped", "condensed"))

# Create grouped boxplot
ggplot(data, aes(x = case_type, y = time_in_months, fill = factor(Appeals))) +
  geom_boxplot() +
  labs(
    title = "Case Duration by Case Type and Appeal Status",
    x = "Case Type",
    y = "Time in Months",
    fill = "Appeals"
  ) +
  scale_fill_discrete(labels = c("No Appeal", "Appealed")) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "top") +
  theme_minimal()
```

```{r}
# Create survival object
surv_object1 <- Surv(data$time_in_months, data$Event)
surv_object <- Surv(clean_data$time_in_months, clean_data$Event)

#Objective 2

# Kaplan Meier Survival Curve

# Overall KM model
km_overall <- survfit(Surv(time_in_months, Event) ~ 1, data = clean_data)

# Time points
time_points <- c(12, 24, 36, 48, 60, 72, 84, 96, 108, 120, 132)

# Summary
summary_overall <- summary(km_overall, times = time_points)

# Format
km_overall_df <- data.frame(
  Time = summary_overall$time,
  Survival_Probability = round(summary_overall$surv, 3),
  Standard_Error = round(summary_overall$std.err, 3),
  CI_Lower = round(summary_overall$lower, 3),
  CI_Upper = round(summary_overall$upper, 3)
)

print(km_overall_df)
```

```
ggsurvplot(km_overall,
        data = clean_data,
        xlab = "Time in Months",
        ylab = "Survival Probability",
        title = "Overall Kaplan-Meier survival curve",
        conf.int = TRUE,
        break.x.by = 10,
        xlim = c(0, 130))


# Survival curve by appeal status
km_appeal <- survfit(Surv(time_in_months, Event) ~ Appeals, data = clean_data)

# Summary
summary_appeal <- summary(km_appeal, times = time_points)

# Format
km_appeal_df <- data.frame(
  Appeal_Status = summary_appeal$strata,
  Time = summary_appeal$time,
  Survival_Probability = round(summary_appeal$surv, 3),
  Standard_Error = round(summary_appeal$std.err, 3),
  CI_Lower = round(summary_appeal$lower, 3),
  CI_Upper = round(summary_appeal$upper, 3)
)

print(km_appeal_df)

ggsurvplot(km_appeal,
        xlab = "Time in Months",
        ylab = "Survival Probability",
        title = "Kaplan-Meier Survival Curves by Appeal Status",
        pval = TRUE,
        conf.int = TRUE,
        legend = "right",
        legend.labs = c("No Appeal", "Appealed"),
        legend.title = "Appeal Status",
        palette = c("steelblue", "gold"))

# Subset data
data_no_appeal <- filter(clean_data, Appeals == 0)
data_appeal <- filter(clean_data, Appeals == 1)

# Fit KM models separately
km_no_appeal <- survfit(Surv(time_in_months, Event) ~ case_type, data = data_no_appeal)
km_appeal <- survfit(Surv(time_in_months, Event) ~ case_type, data = data_appeal)

# Create individual plots
plot_no_appeal <- ggsurvplot(km_no_appeal,
```

```
                    data = data_no_appeal,
                    xlab = "Time in Months",
                    ylab = "Survival Probability",
                    title = "KM Survival Curves by Case Type (No Appeal)",
                    pval = TRUE,
                    conf.int = FALSE,
                    legend = "right",
                    legend.title = "Case Type",
                    break.x.by = 10,
                    xlim = c(0, 130),
                    palette = "Dark2")

plot_appeal <- ggsurvplot(km_appeal,
                 data = data_appeal,
                 xlab = "Time in Months",
                 ylab = "Survival Probability",
                 title = "KM Survival Curves by Case Type (With Appeal)",
                 pval = TRUE,
                 conf.int = FALSE,
                 legend = "right",
                 legend.title = "Case Type",
                 break.x.by = 10,
                 xlim = c(0, 130),
                 palette = "Dark2")

# Arrange both plots on one page
arrange_ggsurvplots(list(plot_no_appeal, plot_appeal),
             ncol = 1, nrow = 2)


# Survival plot by case type
km <- survfit(Surv(time_in_months, Event) ~ case_type, data = clean_data)

# Plot survival curves
ggsurvplot(km,
       xlab = "Time in Months",
       ylab = "Survival Probability",
       pval = TRUE,
       conf.int = TRUE,
       break.x.by = 10,
       xlim = c(0, 130),
       data = clean_data,
       legend = "right",
       legend.title = "Case type",
       legend.labs = levels(data$case_type),
       title = "Kaplan-Meier Survival curves by case type")

```

````{r}
#######################################################################
````

```
# Cox Proportional Hazards Model
# Cox PH
overall_coxph <- coxph(surv_object1 ~ case_type + Appeals, data = data)
summary(overall_coxph)

# Cox for appeals
cox_appeals <- coxph(Surv(time_in_months, Event) ~ Appeals, data = clean_data)
summary(cox_appeals)

# Cox by case_type

cox_case_type <- coxph(Surv(time_in_months, Event) ~ case_type, data = clean_data)
summary(cox_case_type)


# Weibull Model, accepts non-zero time and therefore wee use the subste of non-zero data
weibull <- survreg(Surv(time_in_months, Event) ~ Appeals + case_type, dist="weibull", data =
clean_data)
summary(weibull)


# Exponential Model
exponential <- survreg(Surv(time_in_months , Event)~ Appeals + case_type, dist="exponential",
data = clean_data)
summary(exponential)


# Log-normal Model
lognormal <- survreg(Surv(time_in_months , Event) ~ Appeals + case_type, dist="lognormal",
data = clean_data)
summary(lognormal)


# Log-logistic Model
loglogistic <- survreg(Surv(time_in_months , Event) ~ Appeals + case_type, dist="loglogistic", data
= clean_data)
summary(loglogistic)

# Gamma Model
gamma_model <- flexsurvreg(Surv(time_in_months, Event) ~ Appeals + case_type,
                     dist = "gamma", data = clean_data)
summary(gamma_model)

# Objective 3

# Compare models using AIC
AIC(overall_coxph, weibull, exponential, lognormal, loglogistic, gamma_model)
```

```
# End
```