UNIVERSITY OF NAIROBI

Comparing Accelerated Failure Time and Cox PH Models:A Case Study on Employee Attrition

BY

ELIZABETH GITAU

I56/37232/2020

A Thesis Submitted to the Department of Mathematics for Examination in Partial Fulfillment of this Requirements for the Award of Degree of Master of Science in Mathematical Statistics of the University of Nairobi

November 2022

Abstract

The goal of this project, is to compare the AFT model and Cox PH model using the employee attrition data set. Survival analysis examines the desired outcome until the occurrence of the event. Although Cox PH together with AFT models have been widely utilized in survival time predictions, AFT models are least used in employee attrition. Therefore, the goal of this research is to conduct survival analysis on the employee attrition data set to narrow down on the specific factors that will benefit the employer using both models and the best method to use. Using R, the Accelerated Failure Time model gave favourable outcome compared to the Cox PH model. The main factors that have a significant impact on the survival attrition include, the job role(Research Scientist, Sales Executive), home to job distance, work life balance, level of satisfaction in job and nature of travel. Furthermore, the Generalized Gamma AFT model offers the most outstanding fit for the observed data. The research will serve as a focal point for surviving analysis models in predicting employee attrition, enlightenment in the analysis and deepen the context of survival analysis.

Master Thesis in Mathematics at the University of Nairobi, Kenya.
ISSN 2410-1397: Research Report in Mathematics

DISTRIBUTOR: Department of Mathematics, University of Nairobi, Kenya

Declaration and Approval

I the undersigned declare that this dissertation is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.

16/11/2022

Signature Date

ELIZABETH GITAU
Reg No. I56/37232/2020

In my capacity as a supervisor of the candidate's dissertation, I certify that this dissertation has my approval for submission.

16th November 2022

Signature Date

Dr Idah Orowe
Department of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: orowe@uonbi.ac.ke

Dedication

This project is dedicated to me.

Contents

Bibliography 32

Acknowledgments

First, I thank the Almighty God for giving me the faith and perseverance to complete the project. My respect to my parents for been my pillar in my academic journey and thus far. My sincere gratitude and thanks to my supervisor Dr. Idah for her guidance in the whole process. Lastly, my classmates for the encouragement in completing the project.

Elizabeth Gitau

Nairobi, 2022.

1 Introduction

1.1 Background Information

The field of statistics known as survival analysis examines the desired outcome until the occurrence of the event. As a result, it's referred to be "time to event analysis," employed in a variety of fields including medicine, manufacturing, transportation, e-commerce, human resources, and engineering. When a patient is diagnosed with a terminal condition such as cancer, medical science understands how long they will live. In the manufacturing sector, time to events such as when a car battery dies permanently is of interest. When it comes to predicting when an employee will quit a job and measuring employee retention and satisfaction, human resources can help. Given that the focus is on machine modeling or components linked with electronics, engineering has participated in the research of survival analysis known as "failure time analysis." As a result, different advancements in the field of survival analysis have been incorporated into various fields. However, there are minor differences in the techniques utilized, such as duration analysis in economics.

Complications stemming from censored observations infiltrated statistical methodology developed primarily in the second part of the twentieth century. We will concentrate on frequentist methods in our application, despite the fact that Bayesian Survival approaches [14] have been substantially established and are growing in popularity for survival data. A number of textbooks have been developed on the same: Lawless [18], as well as Oakes [6], Fleming and Harrington [10], and Klein and Moeschberger [15] are only a few examples.

Kaplan and Meier [16] made a significant contribution to the non-parametric analytic method. They function well with similar class samples, but don't assess any particular variables are linked to survival duration. But since survivorship durations are very seldom distributed normally, and redacted data results in lacking survivorship times, this loophole leads to regression method application, but it still falls short in survival data. The Cox PH model is frequently included in survival analytical data processes when variables are included because of its ease. The underlying hazard rate, according to the model, is a function of the uncorrelated variables and not of the hazard function's shape. The original has been extended through modifications.

The semi-parametric method has no assumptions about how the event's underlying risk evolves over period of time therefore the Cox PH model is more largely used compared to parametric methods for analyzing time-to-event data. Exponential, lognormal, Weibull, log-logistic and other hazard distributions are examples of hazard distributions. The relative hazard is calculated using both semi parametric and parametric approaches. Modeling of actual failure times is possible with some distributions. During the nth percentile occurrence of individuals is achieved, The result of an intrigue could be explicitly quantified via the accelerated failure models' as a measure of connection. It is anticipated that the fundamental risk will match a Weibull. using time-to-event data.

1.2 Problem Statement

Although Cox PH together with AFT models have been widely utilized in survival time prediction, a dilemma limits the accuracy of these prediction methods. Limited collection sizes and censored input continue to be a barrier to training reliable and precise models Cox classification models. Despite all these, AFT models are least used in employee attrition. As a result, the goal of this research is to conduct survival analysis on the employee attrition data set to narrow down on the specific factors that will benefit the employer.

1.3 Objectives

Main Objective:

To compare AFT to Cox PH models in the employee attrition to determine "survival" of employees.

Specific Objectives:

1) To identify factors affecting employee attrition using survival analysis.

2) To compare survival probabilities obtained using AFT and Cox PH.

3) To obtain the model appropriate model for the data.

1.4 Significance of the Study

The study findings would reveal the risk variables or the most important covariates that have a substantial impact on employee attrition. A number of characteristics will be examined in this study, including gender, overtime, business travel, and status of marriage, to mention a few. The research will assist in identifying the risk of employee attrition in the involvement of major circumstances. The findings will also provide a better understanding of how to apply the concepts of standard measure of variability and AIC to the data set.

As a result, offering solutions that can predict staff turnover could be extremely useful to businesses. Furthermore, by combining Survival Analysis with the temporal dimension, it is possible to anticipate when an employee will leave.

1.5 Scope of the Study

While survival analysis has been well documented in various fields, the employee attrition analysis is quite limited. Using a fictional data set of 1470 employees with different attributes that will assist to focus the research within achievable parameters. This is a fictional data set with 9 attributes: 1) Business travel involved 2) Role of the employee in the organization 3) Levels of Work Involvement: Low, Medium, High, and Very High 4) Marital status of the employee 5. Distance between working place and home 6) Yes or No to overtime 7) Poor, Good, Better, Better Working Harmony 8) Gender of the employee 9) Level of satisfactory in the job

2 Literature Review

2.1 Staff Attrition

Staff attrition is defined as the natural methodology by which staff quit their jobs without being immediately replaced, for example, through individual withdrawal or old age. Employee attrition may be quite costly for businesses: according to statistics [24], hiring a substitute for a departing staff expenses employers 33 out of a hundred of that employee's yearly compensation. Furthermore, it may risk production, result in knowledge loss, and lower employee morale. In any firm, attrition is unavoidable.

Turnover is characterized as "the voluntary complete withdrawal of participation in an institution by a participant who acquires financial compensation for partaking in that institution" by Mobley (1982)[23]. Turnover is defined by Denver as well as McMahon (1992) as "the movement of individuals from and to workforce inside an institution [5]." According to Forbes (1971), labor turnover refers to employees leaving an organization, and includes promotions, transfers, and any other internal mobility inside the firm. Controlling attrition is emphasized by Meaghan et al. [21], who claims that the worth of people to a business is a critical factor in its success (2002).

He goes on to say that because this value is intangible and difficult to duplicate, managers should keep an eye on attrition. Mobley [22] suggests a measure for forecasting employee turnover, asserting that term of office constitutes one of the strongest predictors of turnover (1977). In accordance to Firth et al. [9], a number of elements contribute to task stress, an absence of dedication to the company, and work discontentment, each of which induce employees to quit (2007).

According to Griffeth et al [12]. pay and remuneration factors have such a significant impact on employee attrition (2000). [13] Prior studies has shown that commitment to the organization and fulfillment are important factors which influence turnover intention, as per Hom as well as Griffeth (1995).Wanous (1992) concentrates on fresh staff turnover, stating that new staff often leave since one's expectations aren't met, culminating in an infringement of one's psychological empowerment and retention. As per Abassi[1], numerous different factors that drive staff to quit the firm include inefficient as well as poor on-boarding processes, style of management, absence of appreciation, work environment, and absence of an attractive compensation framework (2000). [20]Turnover occurs when entry level personnel are hired, according to Louis (1980). At a certain juncture, a worker will decide to quit the organization, be it for individual or business details. Once turnover exceeds the limit set,warning signs are considered.

2.2 Approaches to Employee Attrition

In 2017, a study on employee attrition with reference to Lanson Toyota was done using chi-square, Anova and correlation to assess the causes of attrition and found out the expectation of employees' retainment.[25] The result showed that employees with experience are not promoted leading to dissatisfaction. Employees looked for better opportunities on receiving low income, no promotion or career growth. In return, the effects of attrition can be drastic ranging from cost increase, training cost to low productivity.

Research on Staff attrition forecast using survival analysis was done in 2019 by Zhu, Jiaxing and Xinjun. [31] They did a study on algorithm for predicting attrition from an occurrence approach that combined outcomes of a survival analysis using ensemble techniques. The outcome demonstrated effective turnover prediction using CoxRF but failed to take into consideration the AFT model. From the findings; gender played a crucial in attrition behavior with female being higher; external factors such as GDP growth contributed greatly, turnover differs greatly across different industries and; Staff turnover is greater for people with better educational qualifications than for those with more common credentials. This research took into consideration key factors associated with employees using Cox. However, a deeper study should be done using AFT for comparisons and determining various predictors having a significant impact on employee attrition.

Kumar [17] with a focus on time of payout, proposed the use of survival analysis since factors vary from one organization to another. The Kaplan-Meier method was used which gave detailed information on its ability to verify effect with three variables; time in organization, status at the end of their time and the study group they are in. For instance, a change in the turnover rate would be due to various effects but a change in shape of the survival curves would indicate with a high probability the effect of incentives and allow for tweaking (2016).

2.3 Non-Parametric and Parametric Models Approaches

Previously, typical machine learning approaches were employed to forecast for certain if or not a worker would then resign their employment. For instance, [27]Rombaut and Guerry (2018) used Logistic Regression to model employee attrition. They discovered sex, hierarchical position, and relationship status had been found among the most vital elements that predicted labor attrition rate by evaluating the outcome of the fitted Regression analysis. Liu et al. (2018) evaluated the performance of a Logistic analysis model using various regulated educational techniques Rf, AdaBoost, and Meaningful Effect Perceptron are examples of in forecasting employee turnover. According to their findings, the best predictive performance was attained by Random Forest and AdaBoost, and the predictors affecting staff turnover were professional level as well as strong employment abilities. On balanced together with unbalanced datasets, Alduayj with Rajpoot(2018) tested a variety of prototypes, that involve RF as well as K-N Neighbours, to determine how well they predicted worker attrition. [2] Through balanced data, they discovered that extra working hours and years of experience were the most important indicators of employee attrition.

Researchers also used techniques other than typical machine learning to predict if a worker would depart an organization. For example, [28] Emadi with Staats used econometrics to model worker turnover, finding that supervisors had a key part in forecasting a staff's choice to quit a position (2020). To forecast labor attrition over time, Fang(2018) used a semi-Markov contingent framework [8].The participants determined that the worker's tenure inside the present role was an important element in forecasting quitting the firm by analyzing the probabilities output by their model. Cai et al. [4] also used graph embedding techniques to predict employee attrition and compared them to machine learning approaches (2020). Employee job level and educational background were shown to be the most important factors driving worker turnover.

A number of research have recently begun to investigate the possibility of survival analysis in the field of employee attrition. E. Lee [19] used the KM estimate with Cox PH methods for calculating nurses survival functions in South Korea (2019). They found that gender and work satisfaction were two of the most important determinants in employee turnover. Assefa, Mariam, Mekonnen, and Derbew [3] used Cox PH to look into why medical academics in Ethiopia abandon their jobs (2017). [30]Attrition was found to be influenced by academic level and age. W. Wang (2019) investigated employee quitting through assessing the output of Cox PH. [26] Among the variables, job level with sex were two of the top drivers of turnover, according to the researchers. Madariaga et al. (2018) used employee data to fit a Cox PH with Logistic model in order to investigate common characteristics that lead to worker turnover. They determined that Cox PH or the Logistic model is in accordance to suggesting that staff pay, sex, age, with relationship status seem to be primary factors of staff turnover after interpreting the output of the fitted models. Survival analysis approaches, on the other hand, were shown to be more suited in assessing employee churn than logistic regression since they can estimate survival probability throughout time rather than only at a single moment in time (Madariaga et al., 2018). Silva, Vieira, Pimenta, and Teixeira [7] used a Cox PH to model employee attrition, with a focus on forecasting low-income employee churn. Sex, age, knowledge level, and years in the organization seem to be found to be within the important characteristics determining turnover. They also pointed out that survival analysis was better suited to forecasting staff turnover than other methods since it could deal with censored data, which meant the results were not skewed compared to other tools could be (2018).

Despite advantage of employing survival techniques in staff churn subject, present research on employment retention has just looked classical analysis techniques such as Kaplan with Cox to date. Furthermore, past survival analysis-based articles on employee attrition have mostly focused on exploring the causes of churn without comparing the prediction capacities of survival models. This is in contrast to previous machine-learning-based studies on employee turnover, in which different algorithms are compared and the primary factors impacting churn are investigated.

Lately, unique longevity deep learning methodologies have been suggested and initially used in health profession. The performance of Cox was evaluated using various approaches in predicting atherosclerosis, such RSF and DeepSurv. The above approaches beat classic Cox model, according to researchers. They also used permutation importance to interpret the most important factors impacting atherosclerosis according to RSF and DeepSurv. Cox, DeepHit with Weibull network were compared to predict individuals suffering from Alzheimer by use of removed elements scanned in the brain by Nakagawa et al. (2020) [29].

They came to the conclusion DeepHit or the Weibull neural performed way more compared to the Cox model when forecasting individuals with a potential to suffer from Alzheimer. [11]To predict liver transplantation, Kantidakis(2020) contrasted an RSF with a Partial Logistic Artificial Neural Network(PLANN) technique to Cox model. They discovered that innovative survival mechanisms outperformed Cox PH. They also used the RSF to investigate the most important indicators in estimating the importance of liver transplantation retrieving aspects.

This study attempts to add to the field by comparing the predicted performance of survival analysis algorithms using Cox PH and AFT models while also looking at the most common reasons of employee attrition.

Consequently, the capacity to get understanding into the most relevant features for forecasting employee attrition will not be harmed by the performance comparison provided by this study. Indeed, earlier research in the health profession investigated the most important predictors used by new techniques for producing forecasts, as stated in the previous chapter above.

On the capacity of employee attrition, previous studies all had the same objective of analyzing fitted models and determining the significance of the variables they utilized. Employee churn is caused by a variety of factors, including career length, gender, job level, overtime, and age, according to studies. We will use unique ways to retrieve the primary turnover predictors because recovering them is prevalent in the employee attrition literature and will not hinder our capacity to do so.

3 Methodology

3.1 Introduction

The method proposed in this thesis leverages the use of AFT and Cox models integrated on Employee Attrition sample points to decide which is the best model. There is also a review of the study site, methodology used, techniques utilized, and moral quandaries associated with the research.

3.2 Data set Description

This study used data set obtained from Human Resources that had common details on the workers (age, compensation, years spent in the firm to name just a few) as well as if one quits or leaves the organization. With 1470 rows (values) and 35 columns, the data set is rather comprehensive.

3.3 Variables of the Predictor

The response The outcome is the survival time measured (attrition on whether the employee stays or leaves) the employment.

The predictor

The factors that are estimated to affect the surviving rate of the employees in employment and are given below. They are a total of 9 attributes speculated to have an impact on the employee attrition.

3.4 Study Design

A study on 1470 subjects was done using the Akaike Information Criterion (AIC) to assess the effectiveness of Cox model and the AFT parametric techniques such as Exponential,

3.5 Method

3.5.1 Survival Analysis

Survival time, often known as failure time, is the most important concept when analysing survivor rate. The term ”survival time” refers to the period of time between the beginning of time and the occurrence of the event of interest. Three conditions must be met in order to accurately predict survival time: A time origin must be clearly specified, a scale for measuring time must be agreed upon, and the definition of an event (sometimes referred to as failure) must be completely obvious. The challenges in survival analysis stem mostly from the fact that only a few people have experienced the event, while others have not undergone the event by the end of the study, leaving their real survival times uncertain. The concept of censoring emerges as a result of this. When we have some information about an individual's survival time but not the exact moment, censoring occurs.

Right censoring, left censoring, and interval censoring are the three types of censoring. If an event occurs after the observed survival time, it is said to be subject to right censoring. Let C stand for the censoring time, which is the amount of time after which the study subject can no longer be watched. Follow up time is another term for the observed survival period. It begins from time 0 and lasts until either the event X or the censoring time C, whichever happens first. The true survival time is such that it is equal to or greater than the observed survival time. Right censoring can occur for a variety of causes, including no event before the study ends, loss to follow-up throughout the study period, or withdrawal from the study for various reasons. Competing hazards may be to blame for the last factor. The actual surviving time is then shorter than the appropriate censored survival time

Censoring might also happen when we notice the presence of a situation but don't know how it started. Left censoring is what we term it in this example, and the actual survival time is shorter than the observed censoring time. Interval filtering occurs when an individual is known to have encountered an event within a specific time interval but the precise survival duration is unknown. Within a certain time interval, the actual occurrence time of an event is known. In survival time data, right censoring is relatively prevalent, but left censoring is uncommon. Interval censoring occurs when the time unobserved survival time falls within a known time interval. For instance, a subject is in employment at time 1 but not in employment at time 2. therefore the subject is interval censored in time interval (t_1, t_2)

3.5.2 Survival Distribution

Let T denote the survival time as a random variable. The survival function, the probability density function, or the hazard function can all be used to describe the distribution of survival times. The probability density and hazard functions are easily given for discrete and continuous T, and the survival function is stated for both discrete and continuous T.

The probability of surviving beyond time is denoted as

$$S(t) = Pr(T > t) = 1 - Pr(T \leq t) = 1 - F(t)$$

The Hazard function is the condition probability of an event occurring instantaneously after the survival to time t.

$$H(t) = lim_{\Delta t \to 0} \frac{Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

1. For discrete case The survival function is

$$S(t) = Pr(T > t) = \sum_{t_j > t} Pr(t_j)$$

   The Hazard rate function is

$$h(t_j) = Pr(T = t_j | T \geq t_j)$$

$$\theta(t_j) = \frac{Pr(t_j)}{S(t_{j-1})}$$

   where j = 1,2,3 ...

2. For continuous case

$$S(t) = \int_t^{\infty} f(u)\, du$$

   The Hazard rate function is

$$h(t) = \frac{f(t)}{S(t)}$$

$$h(t) = \frac{-dS(t)}{dtS(t)}$$

3.5.3 Non-Parametric Methods

The Kaplan-Meier product limit method

Probability of surviving is calculated by use of Kaplan-Meier product limit approach.

$$S = \prod_{j=1}^{k} \left(\frac{n_j - d_j}{n_j} \right), k \leq n, t_j \leq t < t_{j+1}$$

d_j represents the number of failures in t_j, n_j represents the number of at-risk incident cases in t_j, k represents the number of successive observations, and n represents the overall number of incident cases.

The log rank test

The log rank test is a hypothesis test that compares two samples' survival distributions. When the data is properly skewed and filtered, it is appropriate to utilize. Hypothesis: H0: There is no change in survival curves. H1: The survival curves differ from one another.For two groups, the log rank statistic is

$$\frac{(O_2-\Xi_2)^2}{var(O_2-\Xi_2)}\ \chi^2_{G-1}$$

Ξ_{ij} = two groups failure proportion to risk is used to compute the predicted frequency.

$$\Xi_{ij}=\left[\frac{n_{1j}}{n_{1j}+n_{2j}}\right]*(m_{1j}+m_{2j})$$

3.5.4 Cox-Regression Model

This is the regression models that investigate the association between survival time of the subject and the predictor variables. The given model is:

$$h(t/X)=h_0(t)\exp(\beta_1x_1+\beta_2x_2+...+\beta_px_p)$$

$$h(t)=h_0(t)\exp(\beta'x)$$

given

$$X=(x_1,x_2,...,x_p)\prime$$

are prognostic factors for a particular individual, $h_0(t)$ is the standard hazard rate and $\beta'=(\beta_1.\beta_2,...,\beta_p)$ are parameters of vector regression.

The Cox regression model assumes parametric form for the effect of the predictors on the hazard but makes no assumptions regarding the form of $h(t)$ (non-parametric half of model) (parametric part of model). As a result, the model is known as a semi-parametric model. The benefit of the Cox technique is that its ambiguity doesn't make estimation difficult. We can still acquire a fair estimate for regression coefficients, hazard ratio, and modified hazard curves even if the baseline hazard is not supplied. The hazard ratio is a measure of effect. The hazard ratio of two people with distinct variables xx and x* is given by:

$$\frac{h(t|x)}{h(t|x^*)}=\frac{h_0(t)\exp(\beta x)}{h_0(t)\exp(\beta x^*)}=\exp(\sum\beta'(x-x^*))$$

The rate of the hazard is impartial to period.

3.5.5 Partial Likelihood Function for Survival Times Without Tied Survival Times

Calculating $h_0(t)$ using the Cox proportional hazards model. One strategy is trying to maximize the actual figures for probabilities function concurrently in terms of h0 (t) and. Cox, D. R., and Oakes, D. (1984) developed a more popular approach in which a partial probability function for is obtained that is independent of $h_0(t)$. In using Cox PH model, partial likelihood is a strategy for inferring regression coefficients in the existence of problematic coefficients $h_0(t)$. We built the part of likelihood function on the basis of ratio risks model in this section. Let $(t_1 \leq t_2 \leq ... \leq t_r)$ be ordered failure times of r individuals with corresponding covariates $x_1, x_2, ..., x_r$. Let $R(t_i)$ be the group of vulnerable participants prior to t_i. Then the conditional probability

$$L(\beta) = \prod_{j=1}^{r} \frac{\exp \beta'^{x_i}(t_j)}{\sum_{keRt_j} \beta'^{x_i}(t_j)}$$

given $R(t_i)$ is the vulnerability at period t_i When there aren't any links in the set of data, the part likelihood is appropriate. Thus, no two topics will have experienced the same occurrence at the exact moment.

3.5.6 Accelerated Failure Time Model

The AFT model explains how certain variables and survival probabilities are related. The group's AFT model with variables $(X_1, X_2, ..X_p)$ is expressed below

$$S(t/x) = S_0(t/n(x))$$

given $S_0(t)$ is the standard survival function, n is the enhanced element (rate of survival times to any constant benefit of $S(t)$). The enhanced element is calculated using the formula below.

$$n(x) = \exp(a_1 x_1 + a_2 x_2 + ... + a_p x_p)$$

The coefficient impact are considered on the temporal scale, to be consistent and progressive in an accelerated failure time model, meaning, the coefficients influences surviving on a consistent element. The AFT model's corresponding log-linear form with regard to time is provided by

$$logT_i = \mu + \alpha_1 X_{1i} + \alpha_2 X_{2i} + ... + \alpha_p X_{pi} + \sigma \varepsilon_i$$

Given μ as intercept, σ as scaling factor ε_i as random variable with a presumption of a specific distribution. Examples of Aft model include, exponential, Weibull, log- logistic, log-normal, and gamma. They are labeled after T distribution and not the log of it or ε_1.

Weibull AFT Model

The hazard function is applicable if the survival rate T has distributions with size and shape variables for the ith individual under the AFT model is

$$h_i(t) = \frac{1}{(\theta_i(x))\gamma} \lambda \gamma(t)^{\gamma-1}$$

Given $\theta_i = \exp \alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_p x_p$ for every i having p predictor coefficients, giving the surviving periods an AFT property,

If T_i has a Weibull distribution (Gumbel distribution) then the an extreme value distribution is e_i. The Weibull distribution's surviving function is given as

$$S_{\varepsilon i}(\varepsilon) = \exp[-\exp(\varepsilon)]$$

The Weibull surviving of AFT representation is

$$S_i(t) = \exp\left[-\exp\left(\frac{-\mu - \alpha_i X_{1i} - ... - \alpha_p X_{pi}}{\sigma}\right) t^{\frac{1}{\sigma}}\right]$$

The hazard function of the the model's AFT representation is

$$h_i(t) = \frac{1}{\sigma} t^{\frac{1}{\sigma}-1} \exp\left(\frac{-\mu - \alpha_i X_1 i - ... - \alpha_p X_{pi}}{\sigma}\right)$$

The survival median time is

$$t(50) = \exp\left[\sigma log(log2) + \mu + \alpha^t x_i\right]$$

Log Logistic AFT Model

The survival and hazard function are

$$S(T) = \frac{1}{1 + exp^{\theta} t^k}$$

$$h(t) = \frac{\exp^{\theta} kt^{k-1}}{1 + \exp^{\theta} t^k}$$

Assuming the surviving periods have log-logistic distributions with variable and k, the hazard function for the ith person underneath the AFT model is, wherein theta and k are undetermined variables.

$$h(t) = \frac{\exp(\theta - klog\eta_i)kt^{k-1}}{1+\exp(\theta - klog\eta_i)t^k}$$

Where $h_i = \exp \alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_p x_p$ for individual i with p explanatory variables. As a result, the ith individual's surviving period is a log-logistic distribution with estimates k log, and the k. Hence, log-logistic distribution is said to have an AFT property. The survival and hazard function are given as

$$S_i(t) = [1 + t^{\frac{1}{\sigma}} \exp\left[\frac{-\mu - \alpha_1 X_{1i} - ... - \alpha_p X_{pi}}{\sigma}\right]]^{-1}$$

$$h(t) = \frac{1}{\sigma t}[1 + t^{\frac{1}{\sigma}} \exp\left[\frac{-\mu - \alpha_1 X_{1i} - ... - \alpha_p X_{pi}}{\sigma}\right]]^{-1}$$

The survival median time is

$$t_i(50) = \exp(\mu + \alpha^t x_i)$$

Log Normal AFT Model

The standard surviving function and hazard function are provided when surviving periods get to be considered having a distribution that is log-normal.

$$\S_0(t) = 1 - \phi\left(\frac{logt - \mu}{\sigma}\right)$$

$$h_0(t) = \frac{\phi\left(\frac{logt}{\sigma}\right)}{\left(1 - \phi\left(\frac{logt}{\sigma}\right)\right)\sigma t}$$

Given μ is the intercept, σ is the scale parameter and random variable; $\phi(x)$ is the standard normal distribution's cumulative density function. The ith individual's survival function becomes

$$S_i(t) = S_0(t/\eta_i) = 1 - \phi\left[\frac{logt - \alpha' x_i}{\sigma}\right]$$

Where $\eta_i = \exp \alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_p x_p$. Hence, the log survival period of the individual i having a normal $(\mu + \alpha' x_i, \sigma)$. The log normal has an AFT property.

Generalized Gamma AFT Model

The Generalized Gamma distribution with parameters λ, γ and α is given by

$$f(t) = \frac{\alpha\lambda^{\alpha\gamma}}{A(\gamma)} t^{\alpha}\gamma^{-1} \exp\left[-(\alpha t)^{\alpha}\right]$$

$t > 0$, $\gamma > 0$, $\alpha > 0$ Its distribution is without an enclosed structure for both the surviving as well as hazard functions. Special instances of the generalized gamma model include the exponential, Weibull, and log-normal models. If $\alpha = \gamma = 1$, its distribution gets to be an exponential distribution; if $\gamma = 1$ it becomes the Weibull distribution; and if γ approaches to infinity, the log-normal distribution.

4 Results

4.1 Introduction

In the current chapter, outcome of the analysis are addressed. The analysis was done using R software. By using Cox and AFT, the findings provide details on the study subjects and how various variables affect the outcome of interest, which is employee attrition. There are 1426 observations and nine attributes.

4.2 Survival Data Summary

There is a total of 1426 observations with 44 exclusions after data cleaning and 221 number of failures. Attrition with indicator Yes = 1 denotes the event of occurrence. Failure: Status = 1, time: In Years and total number of Observations: 1426

n	events	median	0.95LCL	0.95UCL
1426	221	40	32	NA

The average length of time until attrition occurs was determined being 40 years; meaning 50% of the workers survive attrition to at most 40 years, while the other percentage survive attrition for over 40 years period. At this point in time, the cumulative survival function equals 0.5. Summary in Figure 1. Using the chi-square statistic and a common testing technique known as the log-rank test, this article explains how the statistical significance of K-M curves for a number of groups is determined. When two K-M curves are statistically comparable, it signifies that there is no facts to suggest that genuine curves for population survival known as probabilities differ when comparing the two curves in a broad sense.

4.3 Log Rank Test

In table 1, overtime, marital status and job role suggest there is statistical difference between survival probabilities since their P-Value are less than 5% level of significance. Implying that an employee not working overtime has a significantly different probability of surviving attrition compared to an employee working overtime. There is a significant difference in the probability of surviving attrition between a divorced employee and a

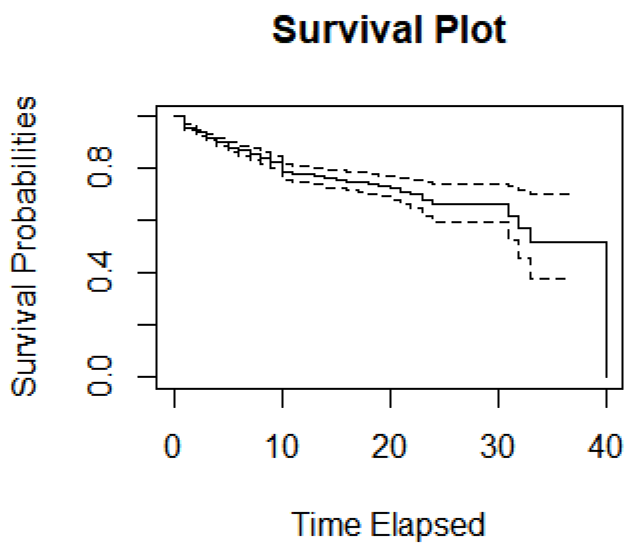
Predictors	Factors	N	Observed	Expected	(O-E)^2/E	P-Value
Gender	Female	588	87	97.0	1.0220	0.2
	Male	882	150	140.0	0.7080	
Overtime	No	1054	110	171.4	22.0000	<2E-16
	Yes	416	127	65.6	57.5000	
Marital Status	Divorced	327	33	53.9	8.1000	3E-12
	Married	673	84	113.1	7.5000	
	Single	470	120	70.0	35.8000	
JobRole	HealthCare	131	9	25.0	10.2583	<2E-16
	HR	52	12	7.1	3.3307	
	Lab Tech	259	62	33.1	25.1629	
	Manager	102	5	27.6	18.5298	
	Director	145	10	25.3	9.2306	
	Research Director	80	2	17.4	13.6719	
	Research Scientist	292	47	38.0	2.1337	
	Sales Executive	326	57	56.1	0.0139	
	Sales Representative	83	33	7.3	91.0995	
WorkLifeBalance	1	80	25	11.9	14.5253	6E-04
	2	344	58	56.7	0.0278	
	3	893	127	143.8	1.9549	
	4	153	27	24.6	0.2295	
Business Travel	None	150	12	24.5	6.4130	3E-05
	Frequently	277	69	45.0	12.8200	
	Rarely	1043	156	167.5	0.7850	
JobSatisfaction	1	289	66	46.0	8.7450	9E-04
	2	280	46	45.5	0.0052	
	3	442	73	70.4	0.0948	
	4	459	52	75.1	7.1129	
JonInvolvement	1	83	28	13.1	17.0800	2E-05
	2	375	71	62.1	1.2800	
	3	868	125	139.0	1.4200	
	4	144	13	22.8	4.2200	

Table 1. Group Survival Function Using Log Rank Test

Predictors	Factors	coef	exp(coef)	s.e(coef)	z	p
Gender	Male	0.1812	1.1987	0.1355	1.338	0.181
	Female(R)					
Overtime	Yes	1.1257	3.0823	0.1307	8.612	2e-16
	No(R)					
Marital Status	Single	1.0445	2.8420	0.1967	5.311	1.09e-07
	Married	0.1919	1.2115	0.2059	0.932	0.351
	Divorced(R)					
JobRole	HR	1.8093	6.1061	0.4606	3.928	8.56e-05
	Lab Tech	1.9013	6.6947	0.3799	5.005	5.59e-07
	Manager	-0.9206	0.3983	0.5791	-1.590	0.111929
	Director	0.2454	1.2781	0.4753	0.516	0.605652
	Research Director	-1.1650	0.3119	0.7923	-1.470	0.141438
	Research Scientist	1.4953	4.4606	0.3873	3.861	0.000113
	Sales Executive	1.2078	3.3460	0.3797	3.181	0.001469
	Sales Representative	2.9118	18.3904	0.4016	7.250	4.16e-13
	HealthCare(R)					
WorkLifeBalance		-0.21416	0.80722	0.08988	-2.383	0.0172
BusinessTravel	Frequently	1.1564	3.1785	0.3130	3.695	0.00022
	Rarely	0.6489	1.9135	0.2997	2.165	0.03039
	None(R)					
JobSatisfaction		-0.21953	0.80290	0.05795	-3.788	0.000152
JobInvolvement		-0.3903	0.6768	0.0865	-4.512	6.41e-06
DistanceFromHome		0.018768	1.018945	0.007423	2.528	0.0115

Table 2. Parameter Estimates of Cox Proportion Hazard on Surviving Attrition

Figure 1. Survival Time to Attrition



single or married employee. An employee job role title differ significantly depending on the role.

4.4 Cox Proportion Hazard

The Wald statistic value is shown in the "z"-designated column.It is shown by the correlation between the standard error and quantity of every coefficient of determination (z = $coef/se(coef)$). The wald statistical examines if the $beta$ parameter of a particular variable deviates significantly from 0. The predictors; overtime, marital status, work life balance, business travel, job satisfaction, job involvement, distance from home and some job roles are statistically significant to surviving attrition.

A positive β coefficient sign indicates that the outcome is worse for people with greater values of that variable since the hazard (risk of attrition) is higher.

From Table 2, Hazard Ratio (HR) 1.1987 of a male employee experiences a higher probability of attrition compared to the female employee. Meaning there is a 0.1812 chance of a female employee surviving attrition compared to a male employee. An employee doing overtime has 3 times possibility of facing attrition contrast to one who does not do overtime. Based on the employee's overtime, an employee with no overtime has 1.1257 chance of surviving attrition compared to an employee doing overtime. A divorced employee has a 1.0445 chance of survival compared to a single employee and 0.1919 compared to a married employee. There is a 1.1564 chance of a non-travel employee to survive compared to a frequent traveller and 0.6489 compared to a rarely travelled employee. Those with higher

Multivariate Cox Analysis Output in R

	coef	exp(coef)	se(coef)	z	p-value	LCL	UCL
GenderMale	0.2490	1.2827	0.1398	1.781	0.074858	0.9753	1.6870
OvertimeYes	1.1994	3.3181	0.1325	9.051	2e-16 *	2.5592	4.3022
,MaritalStatusMarried	0.2220	1.2485	0.2083	1.066	0.286510	0.8301	1.8779
MaritalStatusSingle	0.8569	2.3559	0.1995	4.296	1.74e-05	1.5935	3.4831
JobRoleHR	1.6461	5.1866	0.4622	3.561	0.000369	2.0963	12.8328
JobRoleLabTech	1.6549	5.2325	0.3831	4.320	1.56e-05	2.4695	11.0871
JobRoleManager	-0.6446	0.5249	0.6040	-1.067	0.285866	0.1607	1.7146
JobRoleDirector	0.1887	1.2076	0.4777	0.395	0.692880	0.4735	3.0800
JobRoleResearchDirector	-1.2833	0.2771	0.8274	-1.551	0.120899	0.0548	1.4026
JobRoleResearchScientist	1.0659	2.9035	0.3899	2.734	0.006260	1.3522	6.2347
JobRoleExecutive	1.0900	2.9744	0.3807	2.863	0.004190	1.4105	6.2723
JobRoleRepresentative	2.5230	12.4662	0.4133	6.104	1.03e-09	5.5453	28.0251
WorkLifeBalance	-0.2924	0.7464	0.0938	-3.118	0.001820	0.6211	0.8971
FrequentBusinessTravel	0.9588	2.6087	0.3176	3.019	0.002536	1.3998	4.8615
RareBusinessTravel	0.5305	1.6998	0.3036	1.747	0.0806	0.9375	3.0820
JobSatisfaction	-0.2523	0.7770	0.0579	-4.358	1.32e-05	0.6937	0.8704
JobInvolvement	-0.3381	0.7131	0.0874	-3.871	0.000109	0.6009	0.8463
DistanceFromHome	0.0261	1.0265	0.0078	3.354	0.000796	1.0109	1.0423

Table 3. Multivariate Cox Proportion Hazard on Surviving Attrition

level in job involvement and job satisfaction have lower risk of attrition i.e., 0.3903 and 0.2195. A unit increase in home distance increases the possibility of attrition by 0.0187.

Table 3 gives a summary on multivariate cox proportion. From the results, overtime, single marital status, job role for lab technician, scientist, executive and sales representative, work life balance, business travel, job satisfaction, job involvement and distance from home are significant since their p-value is less 0.05.

The p-value for overtime is 2e-16 with hazard ratio of 3.3181 indicating a strong relationship between the employee's overtime and increased risk of attrition. Higher value of overtime is associated with poor survival. A single employee with p value of 1.74e-05 and hazard ratio of 2.3559 indicates a strong relationship with increased risk attrition. Higher value increases chances of attrition. So is the distance from home, frequent business trav-

Model Significance

Test	Value	P_Value
Likelihood ratio test	344.2	C2e-16
Wald test	306.1	$<$2e-16
Score (logrank) test	371.5	$<$2e-16

Table 4. Model Significance

eller and job role for the lab technician, scientist, executive and sales representative. By contrast, the p-value for male gender is 0.0749 with hazard ratio 1.2827 and 95% confidence interval of 0.9753 to 1.6870. Since the interval includes 1, gender makes a smaller impact to the hazard ratio difference adjusting for other predictors. Work life balance has 0.0018 p-value with 0.7464 hazard ratio that is less than 1. This indicates a strong relationship for every unit increase in work life balance there is decreased risk of attrition. In the same light, age, job satisfaction and job involvement. Overall, the model is statistically significant as shown in Table 4 since the p-values 2e-16 for each of the three overall tests (likelihood, Wald, and score) are all significant.

4.5 Accelerated Failure Time

Table 5 gives a summary statement of the Generalized Gamma AFT model. Job satisfaction, involvement, home distance, nature of travel, sales role and overtime are statistically significant with their p-values < 0.05.

The HR of 0.8152 shows that male employee have a lower probability of surviving attrition contrast to female employees. 0.3820 HR indicate a worker doing overtime has a lower chance of surviving attrition contrast to the control group. Under the job role, the control group being health care role, human resources, lab technician, director, scientist,sales executive and sales representative employee have a HR of 0.2140, 0.2383, 0.9511, 03800, 0.7171 and 0.4743 lower chance of surviving employee attrition while manager and research director roles have twice and thrice probability of surviving respectively. Not travelling being the group control, a staff travelling rarely and frequently have a 0.5934 and 0.4479 HR lower chance of surviving attrition. For every unit increase in distance from home, an employee has a HR of 0.9764 lower chance of survival. Work life balance, job satisfaction and involvement have 1.2098, 1.2233 and 1.3987 HR. This implies that for every unit increase, an employee has the respective HR higher chance of surviving. Divorced staff as the control group, a married employee has a HR of 0,8155 lower chance of surviving attrition and a single employee has a HR of 0.4849 lesser opportunity of surviving contrast to the divorced counterpart.

Predictors	coef	exp(coef)	z	p
GenderMale	-0.2042	0.8152	-0.18	0.0536
OvertimeYes	-0.9623	0.3820	-0.47	1.5e-15
MarriedStatus	-0.2039	0.8155	-0.01	0.4142
SingleStatus	-0.7238	0.4849	-0.44	3.8e-05
HR Role	-1.5413	0.2140	-0.36	5.7e-07
LabTech	-1.4339	0.2383	-2.93	1.5e-08
Manager	0.7754	2.1714	9.09	0.0232
Director	-0.0500	0.9511	-0.56	0.6792
Research Director	1.2291	3.4182	2.50	0.0133
Research Scientist	−0.9673	0.3800	-2.56	0.0182
Sales Executive	-0.6099	0.7171	-0.85	0.0018
Sales Representative	-0.7459	0.4743	-4.72	2e-16
Work Life Balance	0.1905	1.2098	0.21	0.0198
Frequently Travel	-0.8030	0.4479	0.60	0.0048
Rarely Travel	-0.5217	0.5934	-0.27	0.0300
Job Satisfaction	0.2015	1.2233	-0.15	0.0015
Job Involvement	0.3356	1.3987	-1.12	0.0201
Distance From Home	-0.0237	0.9764	0.46	0.0006

Table 5. Estimates on Survival Attrition Using Generalized Gamma AFT Model

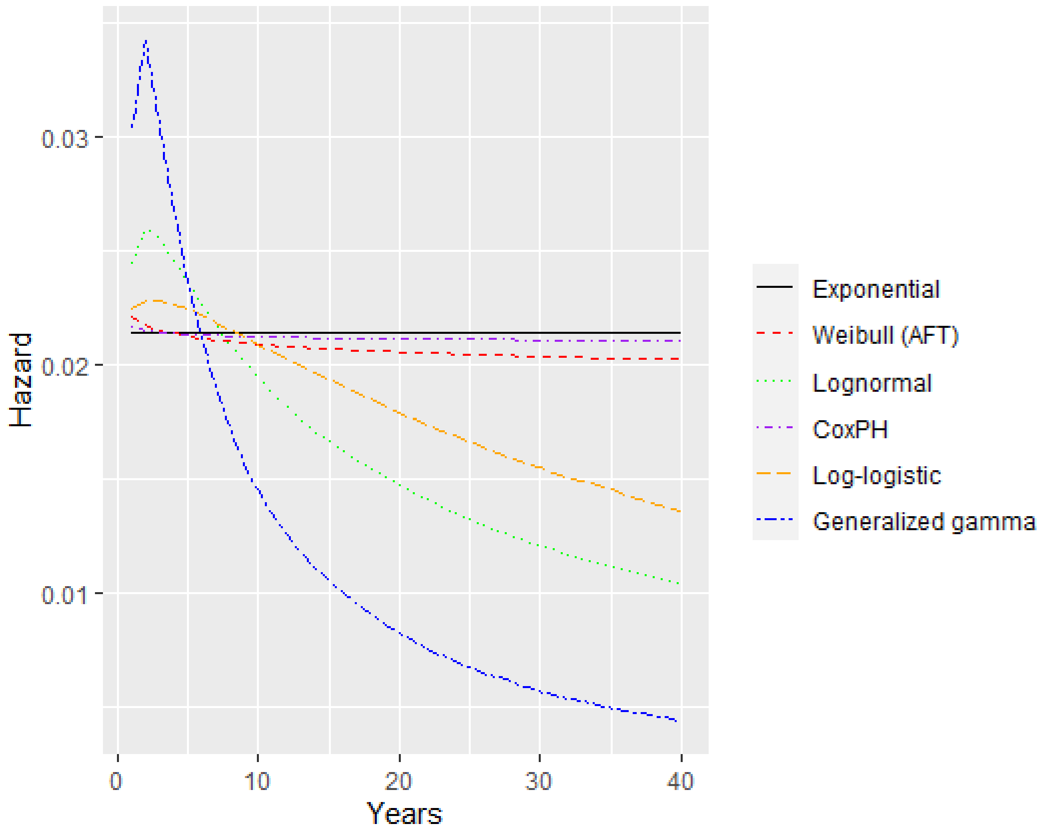
Predictors	CoxPh	Weibull	LogLogistic	Gen-Gamma	Exponential	LogNormal
GenderMale	0.0897	0.0709	0.1021	0.0436	0.0843	0.0986
OvertimeYes	0.0101	1.7e-13	1.5e-14	1.5e-15	8.9e-16	6.8e-14
StatusMarried	0.3510	0.3718	0.2412	0.4142	0.3478	0.2155
StatusSingle	5.21e-05	9.1e-05	3.0e-05	3.8e-05	7.2e-05	3.9e-05
JobRoleHR	4.33e-05	4.6e-05	3.1e-05	5.7e-07	8.5e-05	2.1e-05
LabTech	3.35e-06	3.5e-06	3.6e-07	1.5e-08	8.0e-06	3.1e-07
Manager	0.1679	0.1369	0.0914	0.0232	0.2012	0.0387
Director	0.6562	0.8712	0.8473	0.6792	0.9069	0.8968
ResearchDirector	0.0667	0.0561	0.0512	0.0133	0.0761	0.0252
ResearchScientist	0.0007	0.0009	0.0006	2.7e-05	0.0018	0.0009
Executive	0.0026	0.0055	0.0063	0.0018	0.0063	0.0135
Representative	5.67e-12	7.3e-13	2.3e-13	2e-16	7.8e-12	4.8e-13
WorkLifeBalance	0.0101	0.0098	0.0115	0.0198	0.0085	0.0309
FrequentTravel	0.0023	0.0041	0.0031	0.0048	0.0033	0.0027
RareTravel	0.0414	0.0417	0.0364	0.0300	0.0447	0.0381
JobSatisfaction	0.0001	0.0005	0.0001	0.0015	0.0002	0.0001
JobInvolvement	7.18e-06	8.9e-06	2.9e-05	8.1e-07	8.6e-06	6.2e-05
DistanceFromHome	0.0002	0.0004	0.0009	0.0006	0.0003	0.0006

Table 6. Risk Factor Comparison

Method	LogLikelihood	AIC
Weibull	-908.82	1857.64
LogLogistic	-904.69	1849.38
GeneralizedGamma	-903.40	1848.80
Exponential	-913.84	1865.68
LogNormal	-905.05	1850.10
Cox	-1285.01	2610.01

Table 7. Method Performance Contrast

Figure 2. Comparison Graph



4.6 Comparison

Table 6 shows the risk factor comparison between CoxPH and AFT models. Contrasting the p-values against 0.05 value, research scientist role, executive, work life balance, frequent and rare traveller, job satisfaction and home distance to work place are the most significant. Over time is only significant in all models except Exponential and Log-Normal models.

Table 7 gives a contrast summary of the model performance. The generalized-gamma AFT model has the lowest AIC and the largest log likelihood with 1848.80 and -903.40 respectively. It is merely way better than Log-Logistic model.

Figure 2 shows a simplified plot of all the models involved. The curved-like shape of Log-Normal together Log-Logistic hazards, the unchanging exponential hazard do not fit the data well. The most outstanding models tend to have the hazard decrease monotonically, namely; Weibull, CoxPH, and Generalized Gamma. The best performing model is Generalized Gamma with a flexible curve.

5 Conclusion

5.1 Introduction

In this research paper, the main objective is to contrast the Cox and AFT model. From the first objective, the main factors that have a significant impact on the survival attrition include, the job role, gender, performance of the individual to name a few. Comparing the survival probabilities, the data highly supported the AFT model contrast to the Cox model. The criterion summary indicate that AFT is the most outstanding model of the two. Among the AFT models specifically pin point the best method to use is the Generalized Gamma model.

5.2 Limitations

The data set used for employee attrition is limiting in that it is fictitious data set. The outcome of this data might not stand for using a real data set in the employee set.

5.3 Conclusion

The research paper focus was on

1) Identify the most important factors affecting employee attrition using survival analysis. These include research scientist role, executive role, work life balance, frequent and rare traveller, job satisfaction and home distance to work place across all the models.

2)Compare survival probabilities obtained using AFT and Cox PH. From the results, predictors with higher survival probabilities are female, no overtime, manager and research director roles, divorced and a non-travelling employee. Work life balance, job satisfaction and involvement had a unit increase results to a unit increase chance of surviving. Distance from home had a negative impact i.e., a unit increase in distance decreases the survival chance.

3) Determine the best model using Akaike Information Criteria. AFT model supersedes the Cox PH from the output in table 6. The Akaike Information Criterion is used to determine the method that best analyzes the data set. From the AFT models, Generalized Gamma outshines the other methods with lower AIC value. Cox PH method performed abysmally contrast to the other methods used.

5.4 Recommendations

The definite probabilities of time in survival analysis is not most defined in most cases using AFT. I recommend the use of this process using practical case study and also determining using small sample size and varying use of hazard proportion given that the factors have more than two levels to determine.

6 Appendix

Predictors	coef	exp(coef)	z	p
GenderMale	-0.2171	0.8048	-1.81	0.0710
OvertimeYes	-0.9158	0.4002	-7.37	1.7e-13
MarriedStatus	-0.1567	0.8549	-0.89	0.3718
SingleStatus	-0.6725	0.5104	-3.91	9.1e-05
HR Role	-1.5297	0.2165	-4.08	4.6e-05
LabTech	-1.4288	0.2395	-4.64	3.5e-06
Manager	0.6983	2.0104	1.49	0.1369
Director	-0.0627	0.9392	-0.16	0.8712
Research Director	1.2599	3.5251	1.91	0.0561
Research Scientist	-1.0228	0.3595	-3.31	0.0009
Sales Executive	-0.8466	0.4288	-2.77	0.0055
Sales Representative	-2.3759	0.0929	-7.17	7.3e-13
Work Life Balance	0.2055	1.2281	2.58	0.0098
Frequently Travel	-0.8052	0.4469	-2.87	0.0041
Rarely Travel	-0.5432	0.5808	-2.04	0.0417
Job Satisfaction	0.1792	1.1962	3.48	0.0005
Job Involvement	0.3356	1.3988	4.44	8.9e-06
Distance From Home	-0.0238	0.9764	-3.53	0.0004

Table 8. Estimates on Survival Attrition Using Weibull AFT Model

Predictors	coef	exp(coef)	z	p
GenderMale	-0.2005	0.8182	-1.63	0.1021
OvertimeYes	-0.9664	0.3804	-7.68	1.5e-14
MarriedStatus	-0.2024	0.8167	-1.17	0.2412
SingleStatus	-0.7235	0.4850	-4.17	3.0e-05
HR Role	-1.5455	0.2132	-4.16	3.1e-05
LabTech	-1.4753	0.2287	-5.09	3.6e-07
Manager	0.7335	2.0824	1.69	0.0914
Director	-0.0693	0.9330	-0.19	0.8473
Research Director	1.1973	3.3111	1.95	0.0512
Research Scientist	-0.9979	0.3686	-3.42	0.0006
Sales Executive	-0.7821	0.4574	-2.73	0.0063
Sales Representative	-2.3672	0.0937	-7.33	2.3e-13
Work Life Balance	0.2028	1.2249	2.53	0.0115
Frequently Travel	-0.7973	0.4505	-2.95	0.0031
Rarely Travel	-0.5268	0.5904	-2.09	0.0364
Job Satisfaction	0.2000	1.2214	3.77	0.0001
Job Involvement	0.3365	1.4000	4.18	2.9e-05
Distance From Home	-0.0226	0.9776	-3.32	0.0009

Table 9. Estimates on Survival Attrition Using Log-Logistic AFT Model

Predictors	coef	exp(coef)	z	p
GenderMale	-0.2046	0.8149	-1.65	0.0986
OvertimeYes	-0.9606	0.3826	-7.49	6.8e-14
MarriedStatus	-0.2129	0.8081	-1.24	0.2155
SingleStatus	-0.7215	0.4860	4.11	3.9e-05
HR Role	-1.5314	0.2162	-4.26	2.1e-05
LabTech	-1.4014	0.2462	-5.12	3.1e-07
Manager	0.8447	2.3273	2.07	0.0387
Director	-0.0432	0.9577	-0.13	0.8968
Research Director	1.2489	3.4867	2.24	0.0252
Research Scientist	-0.9172	0.3996	-3.32	0.0009
Sales Executive	-0.6690	0.5122	-2.47	0.0135
Sales Representative	-2.2890	0.1013	-7.23	4.8e-13
Work Life Balance	0.1754	1.1917	2.16	0.0309
Frequently Travel	-0.7903	0.4536	-2.99	0.0027
Rarely Travel	-0.5028	0.6048	-2.07	0.0381
Job Satisfaction	0.2065	1.2294	3.82	0.0001
Job Involvement	0.3281	1.3884	4.01	6.2e-05
Distance From Home	-0.0244	0.9758	-3.43	0.0006

Table 10. Estimates on Survival Attrition Using Log-Normal AFT Model

Predictors	coef	exp(coef)	z	p
GenderMale	-0.2461	1.2790	-1.73	0.0843
OvertimeYes	-1.0937	2.9854	-8.04	8.9e-16
MarriedStatus	-0.1957	1.2162	-0.94	0.3478
SingleStatus	-0.7977	2.2205	-3.97	7.2e-05
HR Role	-1.7381	5.6867	-3.93	8.5e-05
LabTech	-1.6158	5.0320	-4.47	8.0e-06
Manager	0.7157	0.4888	1.28	0.2012
Director	-0.0539	1.0554	-0.12	0.9069
Research Director	1.3922	0.2485	1.77	0.0761
Research Scientist	-1.1400	3.1269	-3.11	0.0018
Sales Executive	-0.9851	2.6780	-2.73	0.0063
Sales Representative	-2.6372	13.9750	-6.84	7.8e-12
Work Life Balance	0.2464	0.7815	2.63	0.0085
Frequently Travel	-0.9688	2.6348	-2.93	0.0033
Rarely Travel	-0.6350	1.8870	-2.01	0.0447
Job Satisfaction	0.2184	0.8037	3.65	0.0002
Job Involvement	0.3919	0.6757	4.45	8.6e-06
Distance From Home	-0.0282	1.0286	-3.59	0.0003

Table 11. Estimates on Survival Attrition Using Exponential AFT Model

Bibliography

[1] Abassi SM, H. K. (2000). *Turnover: the real bottom line.* Public PersonnelManagement, 2 (3):333-342.

[2] Alduayj, S. S. and Rajpoot, K. (2018). "predicting employee attrition using machine learning". *2018 International Conference on Innovations in Information Technology (IIT)*, pages 93–98.

[3] Assefa, T., H. M. D. M. W. e. a. (2017). Medical students' career choices, preference for placement, and attitudes towards the role of medical instruction in ethiopia. *BMC Med Educ*, (96).

[4] Cai, X., Shang, J., Jin, Z., Liu, F., Qiang, B., Xie, W., and Zhao, L. (2020). Dbge: Employee turnover prediction based on dynamic bipartite graph embedding. *IEEE Access*, 8:10390–10402.

[5] Denvir, A. McMahon, F. (1992). *Labour Turnover in London Hotels and the CostEffectiveness of Preventative Measure.* International Journal of HospitalityManagement, Volume 11, No. 2, pp 143 – 154.

[6] D.R. Cox, D. O. (1984). *Analysis of Survival Data.* Chapman Hall.

[7] F. Silva, J. Vieira, A. P. and Teixeira, J. (2018). Duration of low-wage employment: a study based on a survival model. *International Journal of Social Economics*, 45(2):286–299.

[8] Fang, M., Su, J., Liu, J., Long, Y., He, R., and Wang, T. (2018). A model to predict employee turnover rate: Observing a case study of chinese enterprises. *IEEE Systems, Man, and Cybernetics Magazine*, 4(4):38–48.

[9] Firth L, David J Mellor, K. A. M. and ude Loquet, C. (2007). *How canmanagers reduce employee intention to quit?* J. manage.Psychol. 19 (2): 170-187.

[10] Fleming TR, H. D. (1991). *Counting Processes and Survival Analysis.* John Wiley and Sons.

[11] G. Kantidakis, H. Putter, C. L. J. B. A. B. and Fiocco, M. (2020). Duration of low-wage employment: a study based on a survival model. *BMC medical research methodology*, 20(1):277.

[12] Griffeth RW, Hom PW, G. S. (2000). *A Meta-Analysis of Antecedents and Correlates of Employee Turnover: update, moderator tests, and research implications for the next millennium.* J. Manage. 26 (3): 463-88.

[13] Hom P.W., G. R. (1995). *Employee Turnover.* South Western college publishing,Cincinnati, OH pp. 200-340.

[14] J.G. Ibrahim, Chen, M.-H. and Sinha (2001). *Bayesian Survival Analysis.* John Wiley and Sons, first edition.

[15] John P. Klein, M. L. M. (2003). *Survival Analysis: Techniques for Censored and Truncated Data.* Springer, second edition.

[16] Kaplan, E. and Meier, P. (1958). *Nonparametric Estimation from Incomplete Observations, [Journal of the American Statistical Association].* Journal of the American Statistical Association.

[17] Kumar, N. and Yakhlef, A. (2016). Managing business-to-business relationships under conditions of employee attrition: A transparency approach. *Industrial Marketing Management*, 56:143–155.

[18] Lawless, J. (1982). *Statistical Models and Methods for Lifetime Data.* John Wiley and Sons, New York.

[19] Lee, E. (2019). Why newly graduated nurses in south korea leave their first job in a short time? a survival analysis. *Human Resources for Health volume*, (61).

[20] Louis, M. (1980). *Surprise and sense-making: what newcomers experience inentering unfamiliar organizational settings.* Administrative Science Quarterly, Vol. 25 No.2, pp.226-51.

[21] Meaghan Stovel, N. B. (2002). *Voluntary Turnover: knowledge management-friend or foe.* J. intellect. Cap. 3 (3):.

[22] Mobley, W. (1977). *Intermediate Linkages in the Relationship between Job Satisfaction and Employee Turnover.* Journal of Applied Psychology Vol. 62 No.2,.

[23] Mobley, W. (1982). *Employee Turnover: Causes, Consequences and Control.* Addison-Wesley.

[24] Otto, N. (2017). Avoidable turnover costing employers big.

[25] Priya, V. K. and Harasudha, H. H. (2017). *A Study on Employee Attrition with Reference to Lanson Toyoya, Chennai.* Man In India, 97 (22) Part 2 : 115-124.

[26] R. Madariaga, R. O. and Martori, J. (2018). Discrete choice and survival models in employee turnover analysis. *Employee Relations*, 40(2):381–395.

[27] Rombaut, E. and Guerry, M.-A. (2018). Predicting voluntary turnover through human resources database analysis. *Management Research Review*, 41:96 – 112.

[28] Seyed Morteza Emadi, B. R. S. (2020). "a structural estimation approach to study agent attrition". *Management Science, Vol. 66, No, 9)*, pages 3799–4358.

[29] Tomonori Nakagawa, Manabu Ishida, J. N. A. N. S. Y. and Onoda, K. (2020). "prediction of conversion to alzheimer's disease using deep survival analysis of mri images". *Brain Communications, Volume 2, Issue 1.*

[30] Wang, P., Li, Y., and Reddy, C. K. (2019). Machine learning for survival analysis: A survey. *ACM Comput. Surv.*, 51(6):36.

[31] Zhu, Q., Shang, J., Cai, X., Jiang, L., Liu, F., and Qiang, B. (2019). Coxrf: Employee turnover prediction based on survival analysis. pages 1123–1130.