Almeida Moraes Neto, Fernando Humberto de; Almeida, Mariana Thais; Brito, Edleide de et al.

Periodical Part

Analysis of factors associated with employee dismissal in non-public companies in the northeast

Revista brasileira de economia de empresas

Provided in Cooperation with:
Universidade Católica de Brasília (UCB), Brasília

Reference: In: Revista brasileira de economia de empresas Analysis of factors associated with employee dismissal in non-public companies in the northeast 24 (2024). https://portalrevistas.ucb.br/index.php/rbee/article/download/15248/12147. doi:10.31501/rbee.v24i2.15248.

Analysis of factors associated with employee dismissal in non-public companies in the northeast.

Resumo: Este estudo compara modelos de Análise de Sobrevivência para avaliar fatores de risco associados ao desligamento de funcionários em empresas não públicas do Nordeste. Os dados utilizados são da Relação Anual de Informações Socioeconômicas (RAIS) de 2015, abrangendo todos os estados da região. Constatou-se que a maioria dos trabalhadores nessas empresas são homens, pardos, brasileiros, com idade entre 18 e 39 anos. Predominam nas capitais e regiões metropolitanas, com salários de até três mínimos e carga horária de 41 a 44 horas semanais. São admitidos por reemprego, permanecem cerca de três meses e são desligados por iniciativa do empregador, sem justa causa. Comparando modelos de sobrevivência com o estimador de Kaplan-Meier e o Critério de Informação de Akaike (AIC), o modelo gama generalizado mostrou-se mais adequado aos dados. As principais covariáveis analisadas incluíram sexo, idade, raça/cor e faixa salarial média em salários mínimos.

Palavras-chave: Análise de Sobrevivência; Demissão de Empregados; Região Nordeste.

Abstract: *This study compares Survival Analysis models to assess risk factors for employee dismissal in non-public companies in Brazil's Northeast. Data were sourced from the 2015 Annual Socioeconomic Information List across all Northeast states. Findings show that most workers working in in these companies are Brazilian men, mixed race, Brazilians and aged between 18 and 39 years old. Workers in capitals and metropolitan regions have a salary range of up to three minimum periods and work from 41 to 44 hours per week. They are hired through reemployment, remain at the companies for approximately three months and are terminated due to unfair termination at the employer's initiative. By comparing survival models with the Kaplan-Meier estimator and using the Akaike Information Criterion, the generalized gamma model emerged as the best fit. In the final model, key covariates included worker sex, age, race/color, and average salary range defined by minimum wage intervals.*

Keywords: *Survival Analysis; Employee Dismissal; Northeast Region.*

Classificação JEL: C41; C44.

Fernando Humberto de Almeida Moraes Neto[1]

Edleide de Brito[2]

Mariana Thais Almeida[2]

Rafael Toledo Costa de Almeida[2]

Lilia Carolina Carneiro da Costa[2]

Adriano Kamimura Suzuki[3]

[1] Department of statistics, Federal University of Bahia.
E-mail: moraesfernando.mat@gmail.com

[2] Department of statistics, Federal University of Bahia.

[3] Department of Statistics, University of São Paulo, São Carlos, São Paulo.

1. Introduction

In Brazil, the Annual List of Socioeconomic Information (RAIS) (RAIS, 2010) deals with meeting the needs of controlling labor activity in the country, providing data for the preparation of labor statistics and making labor market information available to government entities.

According to Oliveira and Simões (2005), the Brazilian Institute of Geography and Statistics (IBGE) has developed and produced a significant set of research to offer information on the various demographic and socioeconomic characteristics of the Brazilian population. In some of these studies, employment times are assessed through descriptive and exploratory data analysis, as in the work of Gomes and Souza (2018). However, these analysis techniques may not be sufficient to identify the risk factors associated with the time of employment until the dismissal of workers in companies.

Other works attempt to identify the risk factors associated with the dismissal of workers in companies, such as: Santos and Nakano (2015) that adjust lognormal regression and Cox regression models to analyze data from formal workers in the Federal District to estimate the distribution of the time spent by workers in a job using the RAIS database from 2002 to 2009, and Arruda et al. (2016) that use data from the National Household Sample Survey (PNAD), from 2003 and 2013, with the application of probit models to investigate the determinants of unemployment in the northeast.

Survival Analysis is an area of Statistics that has seen strong development in the last two decades due to the improvement of statistical techniques and technological advances. It has applications in several areas, such as medicine (Moraes et al., 2021), engineering (Wang et al., 2005) and social sciences (Bolfarine and Bussab, 2005). In the business sector, scholars focus on jobs, unemployment and company survival.

In this type of study, there is an interest in estimating the time until the occurrence of an event of interest, which is called failure time. In cases where the failure is not observed, there is a partial observation in the response, called censoring. Censoring can occur for various reasons, such as the withdrawal of an individual from the study or the death of a patient from a cause other than that studied, and is the main characteristic of survival data. If neglected, it can lead to errors in estimates (Colosimo and Giolo, 2006).

In Menezes and Cunha (2014), survival analysis is employed to evaluate the duration of unemployment in Brazil, considering the social characteristics of individuals in the study and economic factors. However, only the Weibull model is utilized as a parametric model. It is necessary to adjust and compare various models to assess the adequacy of the adjustments and the consistency of the results obtained.

The present work aims to evaluate the risk factors associated with dismissals of employees working in non-public companies in the Northeast Region of Brazil using the Survival Analysis approach. Four parametric models will be adjusted and compared to assess their suitability for modeling unemployment. In this way, it is possible to support decision-making that reduces employee dismissal rates.

The remainder of this paper is structured as follows. Section 2 presents the description of the database, sampling and the methodology used in the study. Section 3 discusses the results, whilst Section 4 presents the conclusion and future work.

2. Materials and methods

2.1. Dataset

This work used RAIS microdata from 2015 for the nine northeastern states of Brazil:

Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte and Sergipe. The data were taken from the website of the Labor Statistics Dissemination Program (PDET) of the Ministry of Labor in September 2020, and consist of workers dismissed from non-public companies for reasons of termination at the employer's initiative, indirect termination, dismissal upon request, and termination of contract. Individuals dismissed due to retirement, transfers, or death are beyond the scope of this work and were not considered.

In the initial processing of the database, some inconsistent observations were disregarded, such as time of employment or employee age equal to zero, and categories with less than 5% frequency were grouped, such as for the variable race/color of employees. In addition, the size of the establishment was categorized using the Brazilian Micro and Small Business Support Service (SEBRAE) report in 2013.

The target population comprises all Northeastern workers registered with RAIS in 2015, comprising approximately 12 million and 600 thousand employees. After checking the database, it was decided to use the following variables: employee's time of employment, reason for dismissal, age, weekly contractual working hours, average remuneration range in minimum wages, average remuneration of the worker in nominal value, month date of admission, month of dismissal, nationality, race/color of the worker, sex, type of admission, type of employment relationship, whether the employee was hired, whether the employee was dismissed, level of education and municipality where the establishment is located. The following variables: the size of the establishment, the state where the individual works and the classification of the establishment's municipality as belonging to the metropolitan region or the interior were added to the data.

It was found that 15,410 workers had been employed for more than 420 months or 35 years, which, according to the 2015 retirement rules presented on the Federal Government's website, is the minimum period of social security contribution. These workers are practically all 50 years old or over, work more than 30 hours a week and have a median salary of six minimum wages. Furthermore, 73.8% have completed secondary or higher education, 75.6% have employment in medium and large companies and are mostly in the capitals of the Northeast and 90.0% were not laid off in 2015. This group represents approximately 0.02% of the total number of employees and the characteristics of these workers are similar to those observed in the analysis of the complete data. As workers are expected to remain employed for up to 35 years and discrepant employment lengths can affect the survival and residual analysis results, it was decided to remove these individuals.

In the initial stage of this work, only northeastern workers employed by public companies were considered in the analysis. This is because, in Brazil, government employees have job stability and can only be dismissed if they act illegally (Law No. 8.112 of December 11, 1990).

In addition, sampling techniques were used to reduce the database of employees with non-public employment proportionally to the northeastern states with the aim of making inferences about the population. All analyzes were performed in software R (R Core Team, 2020) (version 4.0.0) using interface RStudio (version 1.4.1717).

2.2. Sampling

Stratified sampling involves dividing a population into homogeneous groups, called strata, according to some known characteristics under study. For each of the strata, samples with corresponding proportions are selected, so that the data is more homogeneous within the strata and heterogeneous between the strata. According to Bolfarine and Bussab (2005), stratification is used to improve the precision of estimates.

We can cite several works that select samples using the stratification approach: Kamienski et al. (2005); Adami et al. (2010); Freitas et al. (2020); Souto et al. (2005).

The sample used in this work used stratified sampling proportional to the size of each of the nine northeastern states. The sample size calculation, presented in Bolfarine and Bussab (2005), is given by:

$$n = \frac{\sum_{h=1}^{L} W_h[(1-p_h)p_h]}{\varepsilon^2}, \text{ with } W_h = \frac{N_h}{N}$$

where, e is the margin of error, L is the number of strata, p_h is the proportion of individuals in stratum h with the characteristic of interest, N_h is the total number of elements of stratum h, N is the total number of elements in the population and W_h is the proportion of observations from stratum h.

Table 1 presents some of the values used in calculating the sample size, as well as the sample size per stratum.

Table 1 - Size of strata of employees of companies with non-public employment in the Northeast Region in 2015.

Strata	Region Total	Proportion	Sample Size
Alagoas	500720	0.3142	115
Bahia	2495585	0.3286	572
Ceará	1649567	0.3230	379
Maranhão	669703	0.3286	154
Paraíba	558304	0.3024	128
Pernambuco	1833927	0.3287	421
Piauí	414658	0.3055	96
Rio Grande do Norte	614759	0.3128	141
Sergipe	403655	0.2949	93

Source: Own preparation.

The total sample is made up of 2099 workers working in non-public companies in the Northeast in 2015.

2.3. Survival analysis

In Survival Analysis, the interest is to estimate the time until the occurrence of a certain event of interest, therefore, the response variable is the time to failure. In cases where the event of interest is not observed, we have the presence of censorship (Colosimo and Giolo, 2006). As the aim of this work is to evaluate the factors that impact employee dismissal, the event of interest was defined as employee dismissal. Therefore, survival time is the employment time in months and individuals who were not laid off in 2015 represent the censoring term.

The survival function is one of the main functions for describing survival studies. It is defined as the probability of an observation not failing until a time t, that is, the probability of an observation surviving time t, for $t > 0$. The survival function is given by:

$S(t) = P(T \geq t)$

The presence of censorship in survival data, or the absence of the event of interest, is the main characteristic in this type of study and indicates that the time to failure (the time until the employee leaves the job) is greater than what was recorded. Therefore, specialized statistical techniques are needed to incorporate the information contained in these observations.

The Kaplan-Meier (K-M) estimator, or product limit, is a non-parametric statistical technique for estimating the survival function in the presence of censoring. The K-M estimator for $S(t)$ is given by:

$$\hat{S}(t) = \prod_{(j:t_j<t)} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{(j:t_j<t)} \left(1 - \frac{d_j}{n_j} \right)$$

where $t_1<t_2<\ldots<t_k$ are the k distinct and ordered times in failures, d_j the number of failures in t_j, $j=1,\ldots,k$ and n_j the number of individuals at risk in t_j.

The probabilistic models most used in Survival Analysis to describe failure time are the exponential (E), Weibull (W) and log-normal (LN) models. Below are their respective survival functions presented for the random variable T that represents the failure time,

$$S_E(t) = e^{\frac{t}{\alpha}}, S_W(t) = e^{-\left(\frac{t}{\alpha}\right)^{\gamma}} \text{ e } S_{LN}(t) = 1 - \Phi\left(\frac{[-\log(t) + \mu]}{\sigma} \right)$$

where α is the mean lifetime in the exponential distribution and the scale parameter in the Weibull distribution, μ is the average of the logarithm of failure time, γ is the shape parameter of the Weibull distribution, σ is the standard deviation of the log-normal distribution and Φ is the cumulative distribution function of the standard normal distribution. More information about these distributions can be found in Kundu and Manglick (2004) and Raqab et al. (2018).

Another probabilistic model widely used in survival analysis is the generalized range model (GG). This distribution was proposed in Stacy (1962) and is more flexible than those mentioned previously, as it has three parameters, two for shape and one for scale. Furthermore, the exponential, Weibull and log-normal models are particular cases of this distribution. The density function and survival function follow:

$$f_{GG}(t) = \frac{\gamma}{\Gamma(\kappa)\alpha^{\gamma\kappa}} t^{\gamma\kappa-1} e^{\frac{t^{\gamma}}{\alpha}} \text{ e } S_{GG}(t) = \int_t^{\infty} \frac{\gamma}{\Gamma(\kappa)\alpha^{\gamma\kappa}} u^{\gamma\kappa-1} e^{\frac{u^{\gamma}}{\alpha}} du$$

where α is the scale parameter, κ and γ the shape parameters and $\Gamma(\kappa) = \int_0^{\infty} x^{\kappa-1} e^{-x} dx$.

To estimate the model parameters, the maximum likelihood method was used, which consists of finding the parameter estimates that maximize the likelihood function $L(\theta)$:

$$L(\theta) = \prod_{i=1}^{r} f(t_i;\theta) \prod_{i=r+1}^{n} S(t_i;\theta)$$

where the first r observations represent the failures and the remaining observations represent the censures observed in the experiment, t_i is the time until the event of interest in each element of the sample, θ is the parameter or vector of parameters to be estimated, $f(t_i;\theta)$ is the probability function associated with t_i and $S(t_i;\theta)$ is the survival

function associated with t_i.

The likelihood function, $L(\theta)$, shows that the contribution of uncensored observations is its density function and the contribution of censored information is its survival function (Colosimo and Giolo, 2006).

In this work we used stratified sampling so the likelihood used was the weighted likelihood, which is given by:

$$L(\theta) = \prod_{h=1}^{L} \left[\prod_{i=1}^{n_h} f(y_{hi} \mid \theta) \right]^{w_h}$$

More details of this weighted likelihood can be seen in Wang (2001); Romero et al. (2023).

2.4. Residuals analysis

To assess the adequacy of the models, it is essential to carry out residual analysis. Through residual analysis it is possible to evaluate the distribution of errors, in addition to examining other different aspects of the model such as the detection of atypical observations. It can also be used to compare models, to analyze which one best fits the data.

As these are parametric models, the residuals martingal and deviance were used. The plots of residuals versus time may indicate possible violations of the model (Colosimo and Giolo (2006). If the residual points versus time present random behavior and symmetry around zero, the model is considered to be adequate to the data.

The residues martingal are defined by:

$$\hat{m}_i = \delta_i - \hat{e}_i$$

where δ_i is the failure indicator variable (0 if the individual does not present the failure and 1 if he does) and $\hat{e}_i$ the Cox-Snell residuals (Cox and Snell, 1968). However, the residuals martingal have an upper limit equal to 1 and are not symmetric around zero. A suitable solution is to use waste *deviance*.

The deviance residuals make the martingal residuals more symmetric around zero and make it easier to check for outliers. They are defined by:

$$\widehat{d}_i = \text{sinal}(\hat{m}_i)[-2\hat{m}_i + \delta_i \log(\delta_i - \hat{m}_i)]^{1/2}$$

Another type of residual that can be used is the quantile residual (Dunn and Smyth, 1996), which provides additional insights by examining the distribution of residuals across specific quantiles of the data. This type of residual transforms the observed values to quantiles of the model's predicted distribution, allowing for the assessment of model fit in a way that accommodates non-normality in the data.

Furthermore, the comparison between the models can be made through Worm Plots that evaluate the empirical distribution through the theoretical quantiles. The Worm Plots were proposed by Buuren and Fredriks (Buuren and Fredriks, 2001) to verify the regions of the explanatory variables in which the model presented violations. The vertical axis of the graph depicts the difference between the theoretical and empirical distributions for each observation, and to be considered adequate to the data, these points must be within the confidence bands.

2.5. Akaike Information Criterion

The models were also compared using the Akaike Information Criterion - AIC (Akaike, 1974) in which the best model is the one with the lowest value in this selection criterion. The AIC of a model is given by:

$$AIC = 2k - 2\ln(\hat{L})$$

where k is the number of model parameters and $\hat{L}$ the maximum estimated value for the likelihood function.

3. Results and Discussion

Initially, the database for all states in the Northeast Region, represented by employees working in non-public companies, totaled more than nine million observations and 21 variables. After structuring the database, a descriptive and exploratory analysis of the data was carried out with the aim of understanding the behavior and patterns of the data. Some of the results of these analyzes will be presented below.

Table 2 - Characteristics of employees of companies with non-public employment in the Northeast Region in 2015.

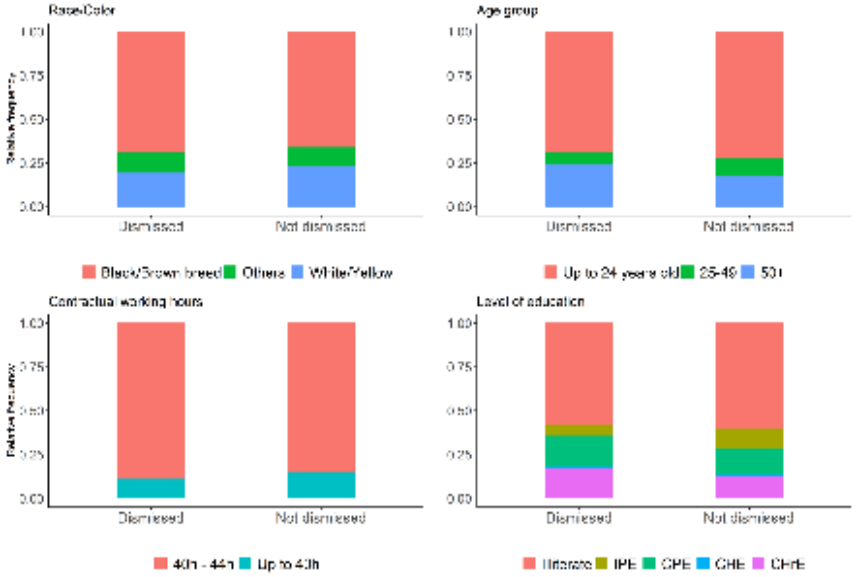
Feature	Employees	%
Situation		
Dismissed	2939611	32.20
Not dismissed	6201267	67.80
Situation		
Admitted	3093230	33.80
Not Admitted	6047648	66.20
Sex		
Female	3224328	35.30
Male	5916550	64.70
Regions		
Capitals	4270901	46.70
Metropolises	1556239	17.00
Interior	3313738	36.30
Total	9140878	100.00

Source: Own preparation.

According to Table 2 it was noted that the percentages of workers dismissed and admitted in 2015 are close, and that there is a predominance of male workers, corresponding to 64.7% of registered workers at RAIS, which have links to non-public companies. Furthermore, most employees are concentrated in the capitals.

Initially, the analyzes were carried out considering the classification of municipalities into capital, metropolis and interior. However, it was noted that workers in capitals and metropolises had similar characteristics and it was decided to combine these two classifications. The analyzes continued considering the interior and metropolitan region of the capital.

Figure 1 - Situation of employees by categories of qualitative variables. In the educational level variable, the categories are Illiterate, IPE = Incomplete elementary education, CPE = Completed elementary education, CHE = Completed secondary education and CHrE = Completed higher education.



Source: Own preparation.

In Figure 1 it is noted that the majority of employees of companies with non-public employment in the Northeast declare themselves black or mixed race. Furthermore, self-declared black or brown individuals have the highest proportion of dismissals and represent approximately 70% of those terminated. It is observed that the majority of employees are between 25 and 49 years old, and that the age group of 50 years or more has lower proportions of dismissals compared to other age groups.

Employees working in non-public companies in the Northeast have, for the most part, a weekly contractual working hour of 40 to 44 hours. It is also noted that there is a higher proportion of individuals dismissed in this category compared to the category of up to 40 contractual hours per week. Approximately 60% of workers in non-public companies in the Northeast have completed high school.

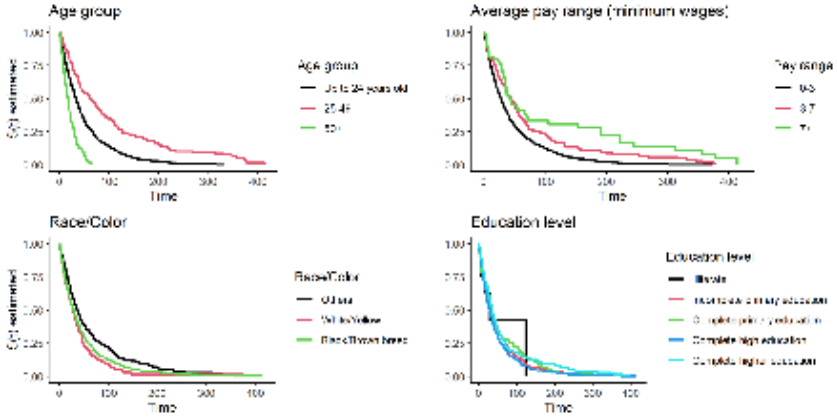
The main reason for dismissal of employees from non-public companies in the northeastern states is termination without just cause at the initiative of the employer, which represents 68% of the total number of employees terminated. This reason is followed by the termination of the employment contract, which represents 17% of the total number of workers dismissed.

To compare the effect of variable categories on the survival time of employees of non-public companies in the Northeast, graphs were created with the survival curves for the covariates.

According to Figure 2 it is noted that job survival estimates decline over time. This decline appears to be faster for workers aged between 18 and 24 compared to 25 to 49, with an average salary of up to 3 minimum wages, workers who do not self-declare as white or yellow, and who are illiterate. Other results observed in this study are that the survival curves for the variable sex of the worker were similar for both categories, male and female. In other words, there is an indication that apparently the worker's sex does not influence the length of time they stay in the job. It was also observed that individuals who work in small companies tend to stay in their jobs longer compared

to individuals in micro, medium and large companies. Furthermore, workers who have weekly contractual working hours of up to 40 hours tend to stay employed for longer compared to workers with working hours greater than 40 hours per week.

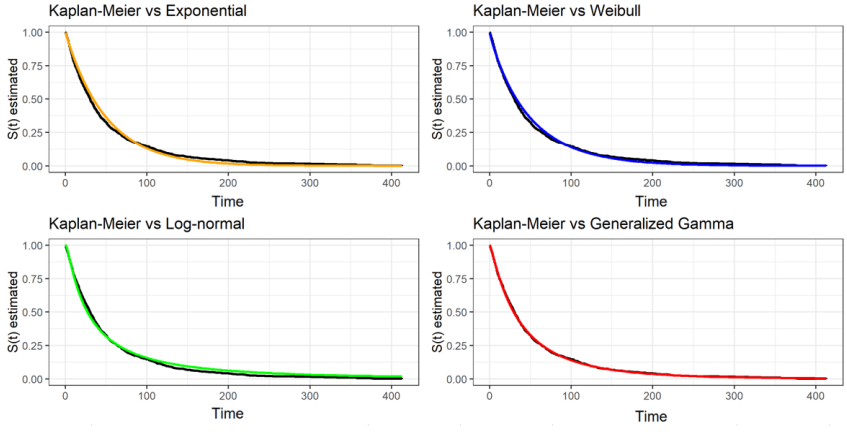
Figure 2 - Survival curves for covariates estimated by Kaplan-Meier.



Source: Own preparation.

The models were then adjusted and through Figure 3, when comparing the survival curves of the proposed models with the survival curve estimated by Kaplan-Meier, it is noted that the best adjusted model is the model with the generalized gamma distribution as it is the closest to the Kaplan-Meier curve.

Figure 3 - Survival curves estimated by the models versus the survival curve estimated by Kaplan-Meier, without the addition of covariates.



Source: Own preparation.

Through Table 3, it is noted that the model with the lowest AIC is the generalized gamma, that is, it presents the best fit among the four adjusted models. With this, the model was adjusted with the generalized gamma distribution considering all variables in the database and the Backward method was used to select covariates. In the result of this method with a significance level of 5%, the best model was observed to be the one that uses the covariates worker sex, race/color, age and average salary.

Table 3 - Comparison of models using the Akaike Information Criterion.

Models	AIC
Exponential	2288.371
Weibull	2288.928
Log-normal	2286.535
Generalized gamma	2279.503

Source: Own preparation.

Therefore, the generalized gamma model was adjusted with the covariate's worker gender, race/color, age and average salary range.

This model is defined by:

$$\mu(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

where β_i represents the estimated coefficients, X_1 corresponds to the worker's sex variable, X_2 represents the employee's self-declared race/color variable, X_3 represents the worker's age in 2015 and X_4 the average remuneration in minimum wages.

The estimates for the shape and scale parameters of the generalized gamma distribution in the model with the addition of the above covariates are given in Tables 4 and 5.

Table 4 - Parameter estimates of the generalized gamma model.

Variables		Estimates
Intercept		2.133
Sex	Male	-
	Female	0.048
Race/color	Others	-
	Black/Brown	0.061
	White/Yellow	0.214
Age (years)		0.035
Average salary (SM)	0-3	-
	3-7	0.292
	7+	0.424

Source: Own preparation.

Table 5 - Estimates of shape and scale parameters.

Estimator	Estimated
$\hat{\alpha}$	4.765
$\hat{\gamma}$	0.758
$\hat{\kappa}$	1.765

Source: Own preparation.

Supposing we want to analyze the length of time a white woman, aged 25, has been

employed, with an average salary of 3 minimum wages. We have to:

$$\mu(x) = 2.133 + 0.048 + 0.214 + 0.035*25 + 0.292 = 3.562$$

Therefore, the average time of stay in employment for an individual with these characteristics is 3.56 months. The estimated survival function of the generalized gamma model follows:

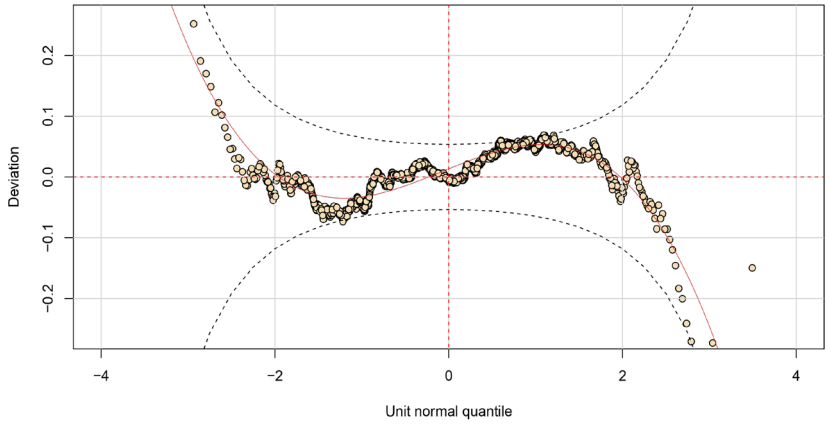
$$\hat{S}_{GG}(t|x)=1-\frac{0.758}{\Gamma(1.765)}\int_0^t z^{0,34}e^{-\left(\frac{z}{4.765}\right)^{0.758}}dz$$

Considering the characteristics mentioned above, the probability of a white woman, aged 25 and with an average salary of 3 minimum wages, remaining in the job for 3.56 months, until leaving, is 71.1%.

To confirm the adequacy of the chosen model, it is necessary to carry out residual analysis, in which it is possible to evaluate the distribution of errors and detect the presence of atypical observations. Therefore, in order to verify whether the generalized range model has violations, the residual analysis was carried out using the *Worm Plot* presented below.

Through Figure 4, it is observed that the differences between the empirical and theoretical distributions for the observations are within the confidence bands. In other words, there is an indication that the generalized gamma distribution model adjusted to the covariates worker sex, race/color, age and average salary has a good fit.

Figure 4 - Warm Plot for the generalized gamma model.



Source: Own preparation.

4. Conclusions

In this work, Survival Analysis models were compared, with the aim of evaluating the risk factors associated with the dismissal of employees working in non-public companies in the Northeast. To this end, data from the Annual Socioeconomic Information List of all states in the Northeast Region in 2015 were used. It was found that workers working in

non-public companies in the Northeast are, for the most part, men, mixed race, Brazilian and aged between 18 and 39 years old. They work in capitals and metropolitan regions, have a salary range of up to 3 minimum wages and work hours of 41 to 44 hours per week. They are hired through reemployment, remain at the companies for approximately 3 months and are terminated due to unfair termination at the initiative of the employer.

By comparing the survival curves of northeastern employees with the Kaplan-Meier estimator and Worm Plot graphs, it was observed that the generalized gamma model presented a better fit to the data in relation to the exponential, Weibull and log-normal models, despite there is censorship of 68% in the database. It was observed that the best average time of stay for workers in Northeastern companies with non-public employment contracts is 34.52 months (approximately 2 years and 9 months), and as time passes these workers begin to be considered more susceptible to shutdowns. Individuals under 24 years old, admitted in 2015, who do not declare themselves white or yellow, with an average salary of up to 3 minimum wages and without some level of education (illiterate) are more susceptible to dismissal.

Subsequently, it is intended to apply more complex models or models stratified by state, to more recent data, with the purpose of more fully evaluating the factors associated with the dismissal of employees with non-public employment in the Northeast.

A significant contribution of this work lies in the analysis of RAIS data, enabling a deeper understanding of the factors that influence employee departures. Understanding the causes behind layoffs allows companies to take more effective measures to improve their internal policies and employees to identify factors that may lead to their layoffs.

It is noteworthy that the results indicate a positive effect of education and experience in extending employment duration. Additionally, younger workers tend to stay in jobs for shorter periods. These findings are essential for the development of public policies that, for instance, encourage young people to complete their studies and receive training aligned with labor market demands.

Acknowledgements

The authors would like to thank the reviewers and editors of the journal for their comments and suggestions that significantly improved this work. Adriano Kamimura Suzuki thanks the support of the National Council for Scientific and Technological Development (CNPq, process no. 302620/2022-2). Mariana Thais Almeida and Rafael Toledo Costa de Almeida thanks the support of the Federal University of Bahia.

References

Adami, M., Rizzi, R., Moreira, M.A., Rudorff, B.F.T., Ferreira, C.C., 2010. Amostragem probabilística estratificada por pontos para estimar a área cultivada com soja. Pesquisa Agropecuária Brasileira 45, 585–592.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716–723.

Arruda, E.F., Guimarães, D.B., Castelar, I., 2016. Desemprego severo no nordeste brasileiro: uma análise para 2003 e 2013. Revista Econômica do Nordeste 47, 101–116.

Bolfarine, H., Bussab, W. O., 2005. Elementos de amostragem. Editora Blucher.

Buuren, S.v., Fredriks, M., 2001. Worm plot: a simple diagnostic device for modelling growth reference curves. Statistics in Medicine 20, 1259–1277.

Colosimo, E.A., Giolo, S.R., 2006. Análise de sobrevivência aplicada. Editora Blucher.

Cox, D.R., Snell, E.J., 1968. A general definition of residuals. Journal of the Royal Statistical Society: Series B (Methodological) 30, 248–265.

Freitas, R.R., Moura, A.C.D., Sagawe, T.R., Ribeiro, F.G., 2020. Uma proposta de amostragem estratificada para pesquisa de origem e destino (o/d). Revista Latino-Americana de Inovação e Engenharia de Produção 8, 27–38.

Dunn, Peter K., and Gordon K. Smyth. "Randomized quantile residuals." Journal of Computational and graphical statistics 5.3 (1996): 236-244.

Gomes, M. R., Souza, S. de C. I. de., 2018. Assimetrias salariais de gênero e a abordagem regional no Brasil: uma análise segundo a admissão no emprego e setores de atividade. Revista de Economia Contemporânea, 22, 1–31.

Kamienski, C., Souza, T., Fernandes, S., Silvestre, G., Sadok, D., 2005. Caracterizando propriedades essenciais do tráfego de redes através de técnicas de amostragem estratificada, SBRC.

Kundu, D., Manglick, A., 2004. Discriminating between the Weibull and log-normal distributions. Naval Research Logistics (NRL) 51, 893–905.

Menezes, A.I., Cunha, M.S., 2014. Uma análise da duração do desemprego no Brasil 2002-2011. Revista Brasileira de Economia de Empresas 13, 37–58.

Moraes, T.E.N.T.d., Previdelli, I., Silva, G.L.d., 2021. A Bayesian Weibull analysis of breast cancer data with long-term survivors in Paraná state, Brazil. Brazilian Journal of Biometrics 39, 293–310.

Oliveira, L.A.P., Simões, C.C.d.S., 2005. O IBGE e as pesquisas populacionais. Revista Brasileira de Estudos de População 22, 291–302.

Santos, R. O., Nakano, E.Y., 2015. Analysis of job permanence time of formal workers on labor market in Federal District of Brazil via Cox proportional hazard and log-normal models. Brazilian Journal of Biometrics 33, 570–584.

R Core Team, R., 2020. R Foundation for Statistical Computing; Vienna, Austria: 2020. R: A language and environment for statistical computing. URL http://www.R-project.

RAIS, 2010. Ministério do Trabalho e Emprego. Secretaria de Políticas Públicas de Emprego. Manual de Orientação da Relação Anual de Informações Sociais (RAIS) ano-

base 2010. Brasília, DF.

Raqab, M.Z., Al-Awadhi, S.A., Kundu, D., 2018. Discriminating among Weibull, log-normal, and log-logistic distributions. Communications in Statistics-Simulation and Computation 47, 1397–1419.

Romero, A., Govil, S., Yilmaz, G., Song, Y., Scaramuzza, D., 2023. Weighted maximum likelihood for controller tuning. In: 2023 IEEE International Conference on Robotics and Automation (ICRA), IEEE. pp. 1334–1341.

Souto, E., Silva, C., Dourado, G., Gomes, R., Souza, T., Kelner, J., Sadok, D., 2005. Obtenção do consumo de energia em redes de sensores sem fio utilizando amostragem estratificada. 23º Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, Fortaleza 1, 381–391.

Stacy, E.W., 1962. A generalization of the gamma distribution. The Annals of Mathematical Statistics, 1187–1192.

Wang, S.X., 2001. Maximum weighted likelihood estimation. Ph.D. thesis. University of British Columbia.

Wang, Y., Mahboub, K.C., Hancher, D.E., 2005. Survival analysis of fatigue cracking for flexible pavements based on long-term pavement performance data. Journal of Transportation Engineering 131, 608–616.