

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA**

Antônio Fernando Costa Pella

**Identificando fatores associados à ocorrência de
demissão em empresas europeias durante a
pandemia de COVID-19**

Belo Horizonte

2023

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA**

Antônio Fernando Costa Pella

**Identificando fatores associados à ocorrência de demissão em
empresas europeias durante a pandemia de COVID-19**

Monografia apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Especialista em Estatística.

Orientador: Prof. Dr. Guilherme Lopes de Oliveira

Belo Horizonte

2023

2023, Antônio Fernando Costa Pella.
Todos os direitos reservados

Pella, Antônio Fernando Costa

P385i Identificando fatores associados à ocorrência de demissão em empresas europeias durante a pandemia de COVID-19 [manuscrito] / Antônio Fernando Costa Pella — 2023.
30.f. il.

Orientador: Guilherme Lopes de Oliveira.
Monografia (especialização) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística.
Referências: 29-30.

1. Estatística. 2. Covid-19. 3. Colaboradores - Empresa – Desligamento. I. Oliveira, Guilherme Lopes de. II. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística. III. Título.

CDU 519.2 (043)

Ficha catalográfica elaborada pela bibliotecária Belkiz Inez Rezende Costa CRB 6/1510 Universidade Federal de Minas Gerais – ICEX



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br
Tel: 3409-5923 – FAX: 3409-5924

ATA DO 309ª. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE ANTÔNIO FERNANDO COSTA PELLA.

Aos quatro dias do mês de outubro de 2023, às 09:00 horas, com utilização de recursos de videoconferência a distância, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística, para julgar a apresentação do trabalho de fim de curso do aluno **Antônio Fernando Costa Pella**, intitulado: “Identificando fatores associados à ocorrência de demissão em empresas europeias durante a pandemia de COVID-19”, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, o Presidente da Comissão, Professor Guilherme Lopes de Oliveira – Orientador, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Após a defesa, os membros da banca examinadora reuniram-se sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: o candidato foi considerado Aprovado condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 04 de outubro de 2023.

Handwritten signature of Guilherme Lopes de Oliveira in black ink.

Prof. Guilherme Lopes de Oliveira (Orientador)
DECOM / CEFET-MG

Handwritten signature of Guilherme Augusto Veloso in black ink.

Prof. Guilherme Augusto Veloso
GET / UFF

Handwritten signature of Livia Maria Dutra in black ink.

Prof. Livia Maria Dutra
DECOM / CEFET-MG

Resumo

Este trabalho tem como objetivo identificar os fatores associados às firmas que demitiram durante a pandemia de COVID-19 na Europa. Durante este período, as operações de diversas atividades econômicas tiveram que ser reduzidas ou paralisadas, o que acabou por gerar queda no faturamento das firmas e, conseqüentemente, demissões. Todavia, o efeito pode ter sido diferente, a depender das características individuais da firma observada. Para tanto, foram explorados dois modelos de classificação: Regressão Logística e Árvore de Decisão. Os dados foram coletados da *Enterprise Surveys*, uma base de dados do Banco Mundial com informações no nível da firma. O modelo que melhor se ajustou aos dados foi a Regressão Logística, tendo apresentado uma capacidade preditiva moderada. Os resultados mostraram que o percentual de produtos que é exportado, o recebimento de ajuda governamental e o tamanho da firma em termos do total de funcionários são características associadas a uma maior chance de demissão. Estimou-se que a chance de demissão nas empresas que receberam ajuda governamental foi mais de duas vezes maior que nas empresas que não receberam auxílio. Embora possa ser um resultado contraintuitivo, isto pode estar relacionado ao porte ou outras características específicas das firmas analisadas. Uma análise mais minuciosa se faz necessária a fim de entender a relação entre eventos num contexto de alta complexidade.

Palavras-chave: COVID-19. Firms. Demissões.

Abstract

The aim of this study is to identify the factors associated with firms that have made redundancies during the COVID-19 pandemic in Europe. During this period, the operations of various economic activities had to be reduced or paralyzed, which ended up firms' turnover and, consequently, layoffs. However, the effect may have been different, depending on the individual characteristics of the firm observed. Two classification models were explored: Logistic Regression and Decision Tree. The data was collected from Enterprise Surveys, a World Bank database with information at the firm level. The model that best fitted the data was Logistic Regression, with a moderate predictive capacity. The results showed that the percentage of products exported, receipt of government aid and the size of the firm in terms of total employees are characteristics associated with a greater chance of dismissal. It was estimated that the chance of dismissal in companies that received government aid was more than twice as high as in companies that did not receive aid. Although this may be a counter intuitive result, it may be related to the size or other specific characteristics of the firms analyzed. A more detailed analysis is needed in order to understand the relationship between events in a highly complex context.

Keywords: COVID-19. Firms. Layoff.

Lista de ilustrações

Figura 1 – Representação de uma Árvore de Decisão	18
Figura 2 – Distribuição dos valores observados (azul) <i>versus</i> imputados (vermelho) . . .	20
Figura 3 – Relação de demissão com as variáveis contínuas	22
Figura 4 – Proporção de firmas que demitiram por país	23
Figura 5 – Análise das variáveis contínuas por país	24
Figura 6 – Curva ROC da Regressão Logística estimada	27
Figura 7 – Árvore de decisão não calibrada	28

Lista de quadros

Quadro 1 – Descrição das variáveis	15
--	----

Lista de tabelas

Tabela 1 – Observações faltantes por variável	19
Tabela 2 – Estatísticas descritivas	21
Tabela 3 – Firmas que demitiram por país	23
Tabela 4 – Recebeu ajuda <i>versus</i> Demissão	25
Tabela 5 – Indústria <i>versus</i> Demissão	25
Tabela 6 – Inovação <i>versus</i> Demissão	25
Tabela 7 – Resultado da Regressão Logística	26
Tabela 8 – Matriz de confusão da Regressão Logística	27

Sumário

1	INTRODUÇÃO	11
2	REVISÃO DE LITERATURA	12
3	METODOLOGIA	14
3.1	Base de Dados	14
3.2	Modelagem empírica	15
3.2.1	Regressão Logística	15
3.2.2	Árvore de Decisão	17
3.3	Software	18
4	RESULTADOS E DISCUSSÕES	19
4.1	Imputação de Dados Faltantes	19
4.2	Análise Exploratória de Dados	20
4.3	Resultados da modelagem estatística	25
5	CONSIDERAÇÕES FINAIS	29
	REFERÊNCIAS	31

1 Introdução

Com a necessidade de restringir as aglomerações para conter a disseminação do vírus SARS-CoV-2, as atividades econômicas, com exceção das essenciais, foram parcialmente ou totalmente paralisadas em todo o mundo. Muitas firmas com restrição de operação tiveram que reduzir seus custos para sobreviver à nova realidade. Consequentemente, de acordo com [Banco Mundial \(2023\)](#), o Produto Interno Bruto (PIB) global de 2020 caiu 4,4%. Conforme [Chen et al. \(2020\)](#), a Europa, particularmente, foi o primeiro continente depois da Ásia a sofrer com o avanço da pandemia e a sentir os efeitos econômicos.

Para tentar mitigar esses efeitos negativos, a União Europeia, entre outras iniciativas, elaborou o *Support to mitigate Unemployment Risks in an Emergency* (SURE), um apoio financeiro temporário para ajudar os Estados-Membros a combater o desemprego. Tal suporte contribuiu para a taxa de desemprego não subir para além dos 8,1% verificados em 2020, ante os 7,5% em 2019, ano anterior à pandemia. Além disso, o bloco econômico utilizou o Banco de Investimento Europeu para amparar cerca de 100 mil pequenas e médias empresas europeias, garantindo que elas mantivessem a sua liquidez ([D'ALFONSO, 2020](#)).

A pandemia de COVID-19 trouxe consequências marcantes para a economia global, afetando empregos e empresas de diversas formas. Estudos mostraram que a condição financeira pré-pandemia das empresas influenciou suas decisões de demissão no início da crise. Além disso, as empresas com predominantemente trabalhadoras mulheres foram mais afetadas em comparação às dominadas por homens. Países europeus enfrentaram desigualdades internas e externas no impacto do emprego devido às características do mercado de trabalho e das medidas de confinamento ([NICOLA et al., 2020](#)).

Dado esse contexto, o objetivo deste trabalho é investigar quais tipos de firmas europeias se viram compelidas a demitir funcionários em resposta à crise econômica causada pela pandemia de COVID-19, identificando fatores associados à ocorrência dessas demissões. Para isso, serão utilizados os microdados da *Enterprise Surveys*, uma pesquisa do Banco Mundial com dados no nível da firma, que incluiu uma rodada especial sobre os impactos da pandemia nos negócios em países da Europa. Os métodos de modelagem estatística baseados em regressão logística e árvores de decisão serão empregados.

O trabalho está dividido da seguinte forma: o Capítulo 2 apresenta uma revisão de literatura de trabalhos que estudaram a relação entre a pandemia e o desemprego. O Capítulo 3 apresenta a base de dados e os modelos utilizados no trabalho. O Capítulo 4 apresenta a análise exploratória dos dados e os resultados das estimações. Por fim, o Capítulo 5 apresenta as considerações finais.

2 Revisão de Literatura

Os efeitos da pandemia de COVID-19 foram intensamente estudados em diversos aspectos. Do ponto de vista econômico, o impacto das medidas restritivas afetou diretamente as receitas das empresas e, conseqüentemente, a força de trabalho. Diversos estudos, desde então, têm investigado os efeitos da pandemia na demissão das firmas, considerando diferentes fatores socioeconômicos e estruturais. Nesta revisão de literatura, analisa-se resultados de estudos que abordaram o impacto da pandemia no desemprego.

Inicialmente, [Couch, Fairlie e Xu \(2020\)](#) investigaram os impactos iniciais da pandemia no desemprego das minorias nos Estados Unidos (EUA). Os resultados revelaram um aumento significativo do desemprego, com impactos desproporcionais nas comunidades afro-americanas e latinas. A análise demonstrou que a taxa de desemprego entre os afro-americanos foi menor do que o esperado com base em recessões anteriores, enquanto a taxa de desemprego entre os latinos alcançou níveis historicamente altos. Esses achados levantam preocupações importantes sobre perdas de renda, desigualdade racial e conseqüências de longo prazo da pandemia nos Estados Unidos.

O estudo de [Nicola et al. \(2020\)](#) abordou os efeitos socioeconômicos da pandemia em diferentes setores da economia global. As medidas de distanciamento social, autoisolamento e restrições de viagens resultaram em uma redução da força de trabalho em todos os setores econômicos, causando uma perda significativa de empregos. O fechamento de escolas e a diminuição da demanda por produtos manufaturados tiveram um impacto negativo adicional no mercado de trabalho. No entanto, setores como suprimentos médicos e alimentos enfrentaram um aumento na demanda. Esse estudo destacou a complexidade dos efeitos da pandemia em diferentes aspectos da economia global.

[Khan \(2023\)](#) utilizou dados do Banco Mundial para investigar o impacto da condição financeira das empresas pré-pandemia no emprego durante a crise. Os resultados indicaram que a condição financeira das empresas desempenhou um papel importante apenas no início da crise sanitária. Empresas com arranjos de teletrabalho, trabalhadores em licença remunerada e aquelas que relataram uma diminuição na demanda por produtos/serviços tinham maior probabilidade de reduzir sua força de trabalho. Além disso, o estudo identificou uma clara influência do gênero no emprego, com empresas que possuíam uma maioria de trabalhadoras mulheres sendo mais propensas a reduzir sua força de trabalho em comparação com empresas onde a maioria dos trabalhadores era homem.

[Dingel e Neiman \(2020\)](#) avaliaram a viabilidade do trabalho remoto durante a pandemia. Através da classificação das ocupações em relação à possibilidade de serem realizadas em casa, eles descobriram que 37% dos empregos nos Estados Unidos poderiam ser feitos remotamente.

No entanto, essa viabilidade variava significativamente entre cidades e setores industriais. Os empregos que podiam ser realizados em casa geralmente tinham salários mais altos e representavam uma parcela considerável da remuneração total nos Estados Unidos. Ao expandir sua análise para outros 85 países, o estudo também constatou que economias de baixa renda tinham uma menor proporção de empregos que poderiam ser realizados em casa.

Fana, Pérez e Fernández-Macías (2020) examinaram os efeitos das medidas de confinamento em diferentes países europeus, incluindo Alemanha, Espanha, Itália, Reino Unido, Polônia e Suécia. Eles observaram que o impacto no emprego foi assimétrico, tanto dentro de cada país quanto entre eles. Os países mais afetados pela pandemia foram também os mais suscetíveis a consequências negativas no mercado de trabalho, devido à sua especialização produtiva e às instituições do mercado de trabalho. Essas descobertas destacam a vulnerabilidade desses mercados de trabalho, caracterizados por altas taxas de desemprego e trabalho precário.

Por fim, o estudo de Kong e Prinz (2020) utilizou dados de busca do Google combinados com informações sobre as datas de implementação de intervenções não farmacêuticas (INFs) nos estados dos EUA. Eles quantificaram o impacto direto de diferentes INFs estaduais no aumento do desemprego durante a pandemia. Os resultados mostraram que as limitações em restaurantes e bares, assim como o fechamento de negócios não essenciais, foram responsáveis por uma parcela significativa dos pedidos de seguro-desemprego. No entanto, outras INFs não tiveram um impacto direto no aumento dos pedidos de seguro-desemprego, indicando a influência de outros fatores, como a queda na demanda do consumidor e as políticas implementadas por empresas privadas e instituições.

Em resumo, a revisão de literatura destaca a complexidade e a amplitude dos efeitos da pandemia na demissão das firmas. A viabilidade do trabalho remoto, a condição financeira das empresas, as medidas de contenção adotadas pelos governos e a estrutura do mercado de trabalho foram fatores importantes que influenciaram o grau de demissão durante a crise. Além disso, questões de gênero e desigualdades raciais surgiram como aspectos relevantes a serem considerados.

3 Metodologia

3.1 Base de Dados

Os dados utilizados neste trabalho são provenientes do estudo do Banco Mundial chamado de *Enterprise Surveys*¹. Esta pesquisa é realizada por meio de questionários aplicados aos responsáveis por um grupo diversificado de empresas, predominantemente localizadas em países em desenvolvimento, incluindo regiões da América do Sul, América Central, México, Leste Europeu, África e Ásia. A base de dados abrange informações de mais de 130.000 empresas em mais de 100 países, detalhando aspectos críticos do ambiente de negócios, como acesso a financiamento, corrupção, infraestrutura e desempenho organizacional, informações estas coletadas através de entrevistas com executivos. Desde o ano 2000 até o início da pandemia, a pesquisa foi conduzida uma vez em 46 países diferentes e pelo menos duas vezes em mais de 50 países, assegurando uma amostra representativa de firmas industriais em cada um deles. O questionário, desenhado para extrair informações profundas sobre o ambiente de negócios, operações, estrutura física, interações com o governo, fornecedores e concorrentes, é respondido por um indivíduo designado ou com responsabilidade de liderança na empresa.

Como forma de conhecer o impacto da pandemia nas firmas, o Banco Mundial implementou uma rodada extra do estudo entre 2020 e 2021, e incluiu questões relativas à pandemia. Para fins desta pesquisa, foram coletados dados dos seguintes países: Alemanha, Áustria, Bélgica, Dinamarca, Espanha, Finlândia, França, Países Baixos, Luxemburgo e Suécia.

As empresas selecionadas em cada país foram as que responderam com informações relativas ao ano fiscal de 2019, de modo que pode-se obter informações antes do início da pandemia. O número total de empresas nessa condição foi de 3.779, no somatório total dos dez países mencionados, que juntos representaram cerca de 90 % do PIB europeu em 2020.

As variáveis selecionadas para este estudo são apresentadas no [Quadro 1](#). Como o objetivo é identificar quais atributos aumentam a probabilidade de a firma ter demitido em razão da pandemia, a variável resposta, categórica e binária, é *demitiu*, que indica 1 se a firma demitiu algum funcionário no período e 0, caso contrário. Tal variável é construída a partir da seguinte pergunta do questionário: "D.6 *Since the outbreak of COVID-19, how many workers have been laid off due to the COVID-19 outbreak?*". Se o número respondido for maior que 0, então a classificação é 1. Caso contrário, a firma é classificada com 0.

Para contornar o problema de dados faltantes, o que levaria a perda de 475 observações, optou-se por fazer imputação de dados. Para isso, comparou-se o desempenho de três métodos de imputação: o *Random Forest*, a árvore de decisão e o método *sample*. Este último toma como

¹ Disponível para fins de pesquisa em: <https://www.enterprisesurveys.org/en/enterprisesurveys>.

Quadro 1 – Descrição das variáveis

Nome da variável	Tipo de variável	Descrição
demitiu	Categórica	1: a firma demitiu; 0: caso contrário.
country	Categórica	Representa cada um dos 11 países incluídos na amostra.
industria	Categórica	1: manufatura; 2: varejo; 3: outros serviços.
idade	Contínua	Representa a idade da firma no ano de 2020.
exp	Contínua	Representa o percentual de vendas que é exportado.
inovacao	Categórica	1: se a firma lançou um novo produto ou um novo processo; 0: caso contrário.
ajuda	Categórica	1: se a firma obteve alguma ajuda governamental; 0: caso contrário.
funcionarios	Contínua	Representa o número de funcionários da firma.

Fonte: Elaboração própria.

base amostras aleatórias criadas a partir dos dados observados. Em outras palavras, o *sample* imputa dados faltantes selecionando aleatoriamente valores do conjunto de dados observado.

Para decidir o melhor método, removeu-se 10% dos dados observados da variável resposta e imputou-se os dados com cada uma das técnicas. O método com maior percentual de acerto foi o escolhido. O algoritmo utilizado foi o *mi*ce, do pacote R de mesmo nome, que aplica um método para tratar dados faltantes, criando múltiplas imputações para dados multivariados ausentes (BUUREN; GROOTHUIS-OUDSHOORN, 2011).

3.2 Modelagem empírica

Para identificar as características associadas às firmas que demitiram em resposta à queda da atividade econômica durante a pandemia, pode-se recorrer aos modelos de classificação. A seguir, são apresentados os dois modelos utilizados no estudo, a Regressão Logística e a Árvore de Decisão.

3.2.1 Regressão Logística

Para estimar a relação das variáveis independentes $\mathbf{x} = (x_1, x_2, \dots, x_k)$ quando temos uma variável dependente Y qualitativa, nesse caso binária, indicando se a firma demitiu ou não demitiu algum funcionário em função da pandemia de COVID-19, pode-se recorrer à regressão logística. Conforme Greene (2003), essa é uma regressão em que a forma funcional é uma função bijetora com uma curva sigmoide, que garante que o contradomínio esteja restrito ao intervalo entre 0 e 1, e também pode ser entendida como a probabilidade do evento ocorrer, neste caso, a demissão de funcionário(s). Sob esta abordagem, o valor esperado da variável resposta para uma dada configuração das variáveis independentes \mathbf{x} pode ser representado por

$$E(Y_i|\mathbf{x}_i) = \Lambda(\mathbf{x}'_i\beta) = \frac{e^{\mathbf{x}'_i\beta}}{1 + e^{\mathbf{x}'_i\beta}}, \quad (3.1)$$

onde $\mathbf{x}_i(x_{1i}, x_{2i}, \dots, x_{ki})$ representa o vetor de valores observados para as variáveis independentes na firma i , para $i=1, \dots, n$, e $\beta = (\beta_1, \beta_2, \dots, \beta_k)$ é o vetor de parâmetros a serem estimados.

Por se tratar de uma variável binária, a qual se adequa a uma distribuição Bernoulli, a expressão $\Lambda(\mathbf{x}'\beta)$ denota a probabilidade de ocorrência do evento de interesse, ou seja, $P(Y_i = 1)$ para a i -ésima unidade amostral. Como nesta função os coeficientes não são lineares em relação à variável resposta, é comum a aplicação do logaritmo natural (\ln) na razão entre a probabilidade do evento ocorrer sobre a probabilidade de o evento não ocorrer, gerando a chamada chance (*odds*, em inglês), de forma a tornar a interpretação no sentido linear:

$$\ln \left[\frac{\Lambda(\mathbf{x}'\beta)}{1 - \Lambda(\mathbf{x}'\beta)} \right] = \ln \left(\frac{\frac{e^{\mathbf{x}'\beta}}{1+e^{\mathbf{x}'\beta}}}{\frac{1}{1+e^{\mathbf{x}'\beta}}} \right) = \ln e^{\mathbf{x}'\beta} = \mathbf{x}'\beta. \quad (3.2)$$

Dessa forma, os coeficientes β são na verdade os coeficientes da regressão em que a variável dependente é o logaritmo natural da chance de ocorrência do evento.

A interpretação dos resultados do ajuste do modelo também é comumente feitas em termos da chamada razão de chances (OR, do inglês *odds ratio*). A OR é dada, como o nome sugere, pela razão entre as chances de ocorrência do evento em dois contextos representados por \mathbf{x}_0 e \mathbf{x}_1 . Neste caso, $OR(\mathbf{x}_0, \mathbf{x}_1) = \exp \mathbf{x}'_0\beta / \exp \mathbf{x}'_1\beta$ de modo que, se consideramos, sem perda de generalidade, que temos apenas uma variável independente e tomarmos $x_1 = x_0 + 1$, então $OR(x_0, x_0 + 1) = \exp(\beta_1(x_0 + 1 - x_0)) = \exp(\beta_1)$. Portanto, a razão de chances $OR(x_0, x_0 + 1)$ pode ser interpretada como sendo a magnitude com a qual a chance de $Y = 1$ se modifica (aumentará se $\beta_1 > 0$ e diminuirá se $\beta_1 < 0$) pelo acréscimo de 1 unidade na covariável x . No caso de um modelo com múltiplas covariáveis, a interpretação é feita separadamente para cada covariável na forma descrita acima, considerando que as demais sejam mantidas fixadas.

Os parâmetros β podem ser estimados por diferentes métodos. Neste caso, o método de estimação aplicado foi o de máxima verossimilhança. Detalhes teóricos sobre o processo de estimação podem ser encontrados em [McCullagh e Nelder \(1989\)](#).

Depois de estimados os coeficientes β e, conseqüentemente, a relevância das variáveis associadas a cada um destes, a análise de significância pode ser feita por meio de testes de hipóteses. Neste estudo, foi utilizado o resultado associado ao teste de Wald ([MCCULLAGH; NELDER, 1989](#)) e o nível de 10% de significância foi adotado. A análise da significância foi feita posteriormente à aplicação do algoritmo *stepwise* para encontrar a combinação de variáveis independentes que gera o melhor valor para o Critério de Informação de Akaike (AIC, em inglês). Resumidamente, o AIC é uma medida utilizada para avaliar e comparar modelos estatísticos. Quanto menor o valor do AIC, melhor é o ajuste do modelo aos dados ([AKAIKE, 1974](#)).

Além da análise da significância do modelo, uma parte importante no contexto de métodos de classificação é a análise da capacidade preditiva do modelo. Isto envolve uma comparação entre os valores reais observados para a variável resposta Y com os seus valores preditos, denotados por \hat{Y} .

Baseado nos valores $\hat{\beta}$ estimados para os parâmetros do modelo, o critério para predição/classificação de uma unidade amostral entre as categorias binárias (demitiu ou não demitiu) é feita considerando-se que

$$\hat{Y} = \begin{cases} 1 & \text{se } \Lambda(\mathbf{x}'\hat{\beta}) \geq k \\ 0 & \text{se } \Lambda(\mathbf{x}'\hat{\beta}) < k \end{cases} \quad (3.3)$$

O parâmetro k é um valor limiar dentro do intervalo $[0, 1]$ utilizado para definir se o indivíduo, que neste caso é a firma, será classificada como 0 ou 1. Para determinar qual será o valor de k , recorre-se à curva *Receiver Operating Characteristic* (ROC). A curva ROC representa graficamente a relação entre a taxa de verdadeiros positivos (Sensibilidade) e a taxa de falsos positivos (1 - Especificidade) para diferentes pontos de corte em um modelo de classificação. Uma análise detalhada a respeito deste critério de avaliação do modelo pode ser encontrado em [Cristiano \(2017\)](#). Quanto maior a área sob a curva ROC (comumente denotada por AUC, do inglês *area under the curve*) melhor a capacidade preditiva do modelo. Neste contexto, o valor k adotado será aquele que maximiza a AUC, sendo o ponto da curva mais próximo às coordenadas (1,1).

3.2.2 Árvore de Decisão

De acordo com [Breiman \(2017\)](#), uma árvore é uma estrutura hierárquica que pode ser definida como um conjunto de nós, em que o primeiro nó é a raiz da árvore, e os demais possuem uma relação de "parentesco" com o nó inicial. Os nós subsequentes ao primeiro são os filhos em relação ao anterior, e irmãos entre si, quando estão no mesmo nível. A cada ramificação da árvore é tomada uma decisão quanto à classificação das unidades amostrais alocadas no nó envolvido, conforme exemplificado na [Figura 1](#). O procedimento segue até que não sejam indicadas mais etapas de decisão.

A escolha do nó inicial pode ocorrer por meio do cálculo do Índice de Gini de cada variável i . Ele mede o grau de impureza das variáveis, e aquela que apresentar o menor valor é escolhida como a raiz dessa árvore. De certa forma, ele mede a probabilidade de uma observação aleatória ser classificada incorretamente quando escolhida aleatoriamente. Formalmente, o índice Gini é definido como:

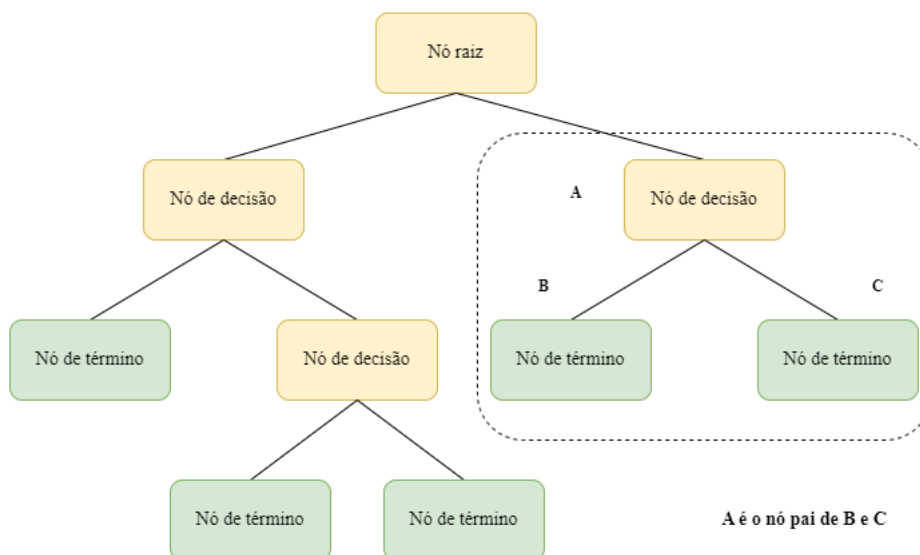
$$Gini = 1 - \sum_{i=0}^n p_i^2, \quad (3.4)$$

em que $0 \leq Gini \leq 1$ e p_i é a proporção de observações em cada variável. Quanto mais próximo o índice estiver de 0, menor o grau de impureza do atributo. Em outras palavras, quanto mais baixo for o Índice de Gini, menor será a probabilidade de erros de classificação.

As observações são então divididas em dois ou mais subgrupos, constituídas por uma fração do total das observações. A cada novo nó, o Índice de Gini é recalculado para as variáveis remanescentes, seguindo o processo de seleção com base na menor impureza. Esse procedimento continua até que se atinja um critério predefinido, como um valor mínimo do Índice de Gini,

indicando que as observações em um nó são praticamente iguais. Desta forma, este método encontra o atributo que produz a menor heterogeneidade dos dados em cada nó. A Figura 1 apresenta uma representação genérica de uma árvore de decisão.

Figura 1 – Representação de uma Árvore de Decisão



Fonte: Elaboração própria.

Os nós finais, onde não ocorrem mais divisões, são conhecidos como folhas ou nós terminais. Eles simbolizam as categorias ou valores que a árvore busca prever. O caminho da raiz até uma folha, através de várias divisões, constitui uma regra de decisão. Assim, a árvore de decisão em si forma um grupo de regras que classifica ou prediz novas observações. Mais detalhes sobre o método podem ser encontrados em [Witten e James \(2013\)](#).

3.3 Software

A análise foi feita usando o *software* R ([R Core Team, 2020](#)). A imputação dos dados foi feita usando o pacote *mi*ce ([BUUREN; GROOTHUIS-OUDSHOORN, 2011](#)). O processo de estimação dos parâmetros do modelo de regressão logística foi feito através da função *glm()* do R usando o seu método *default* de estimação, denominado *iteratively reweighted least squares method* ([MCCULLAGH; NELDER, 1989](#)). A seleção de variáveis por meio do *stepwise* foi feita usando a função *step* do pacote *stats* ([HASTIE; PREGIBON, 1992](#)). O algoritmo para árvore de decisão foi aplicado por meio da função *rpart* do pacote de mesmo nome ([BREIMAN, 2017](#)).

4 Resultados e Discussões

4.1 Imputação de Dados Faltantes

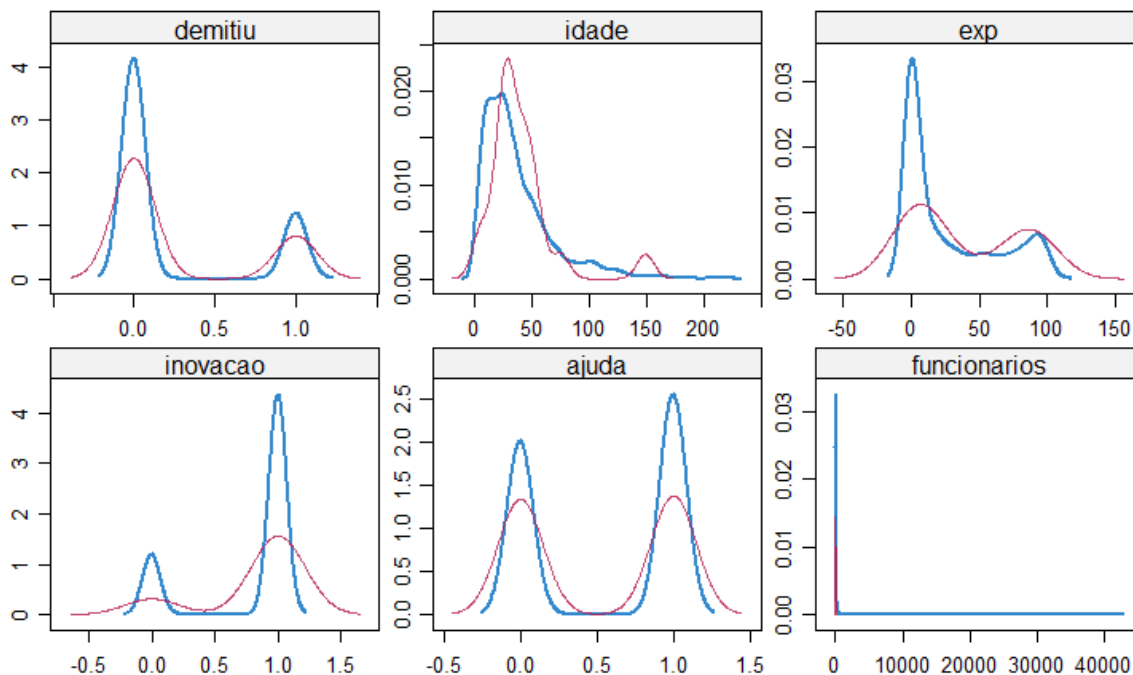
A [Tabela 1](#) apresenta a quantidade de observações faltantes por variável. Após o processo de imputação, as variáveis não possuem nenhuma observação ausente.

Tabela 1 – Observações faltantes por variável

Variável	Número de observações faltantes
demitiu	270 (7,1%)
idade	20 (0,5%)
exp	26 (0,6%)
inovacao	12 (0,3%)
ajuda	270 (7,1%)
funcionarios	10 (0,2%)

Fonte: Elaboração própria.

Dentre os métodos de imputação testados, o método *sample* se mostrou mais preciso, acertando cerca de 67% das observações. Dessa forma, optou-se por essa estratégia de imputação. A [Figura 2](#) apresenta, em azul, a distribuição dos dados observados e, em vermelho, a distribuição dos dados imputados para cada variável com dados faltantes. Observa-se que, de modo geral, as distribuições dos dados imputados se assemelham com às distribuições dos dados observados, validando a análise realizada.

Figura 2 – Distribuição dos valores observados (azul) *versus* imputados (vermelho)

Fonte: Elaboração própria.

4.2 Análise Exploratória de Dados

A presente seção faz uma análise exploratória dos dados da pesquisa. A [Tabela 2](#) apresenta as estatísticas descritivas das variáveis, em que `demitiu` é a variável dependente. Como foi utilizado o método de imputação para evitar o problema de dados faltantes na amostra, todas as variáveis têm seus dados completos. Observa-se que 870 firmas, 23% da amostra, demitiram em função da pandemia. Em média, as firmas têm 3,65 anos, exportam cerca de 27% da produção, 78% delas lançaram um novo produto ou algum novo processo, 56% recebeu ajuda governamental por conta da pandemia e possuem 89,69 funcionários.

Por outra medida de tendência central, a mediana, sabe-se que ela separa os 50% menores valores dos outros 50% maiores. Quando tal medida é menor do que a média, existem valores extremos no grupo superior que atuam mais intensamente no sentido de induzir a média para cima. Olhando para as variáveis contínuas, isso ocorre, por exemplo, com a variável `idade`. Enquanto a média é 36,65, a mediana é 28. Ou seja, 50% das firmas tinham mais de 28 anos no momento da realização da pesquisa. Tanto `exp`, quanto `funcionarios`, também têm essa característica. Na variável `exp`, 50% firmas exportaram mais de 8% da sua produção. Para a variável `funcionarios`, verifica-se que metade das firmas tem mais de 21 funcionários.

Adicionalmente, analisando a dispersão dos dados das variáveis quantitativas, observa-se que a variação nos dados é bastante significativa em algumas das variáveis contínuas. O

Tabela 2 – Estatísticas descritivas

Estatísticas	demitiu	idade	exp	inovacao	ajuda	funcionarios
Observações	3779	3779	3779	3779	3779	3779
Mínimo	0	1	0	0	0	1
Máximo	1	220	100	1	1	42703
1º Quartil	0	16	0	1	0	9
3º Quartil	0	48	50	1	1	51,5
Média	0,23	36,55	27,13	0,78	0,56	89,69
Mediana	0	28	8	1	1	21
Soma	870	138010	102510	2959	2114	338952
Variância	0,18	9,30	1186,57	0,17	0,25	725076,32
Desvio-padrão	0,42	3,05	34,45	0,41	0,50	851,51

Fonte: Elaboração própria.

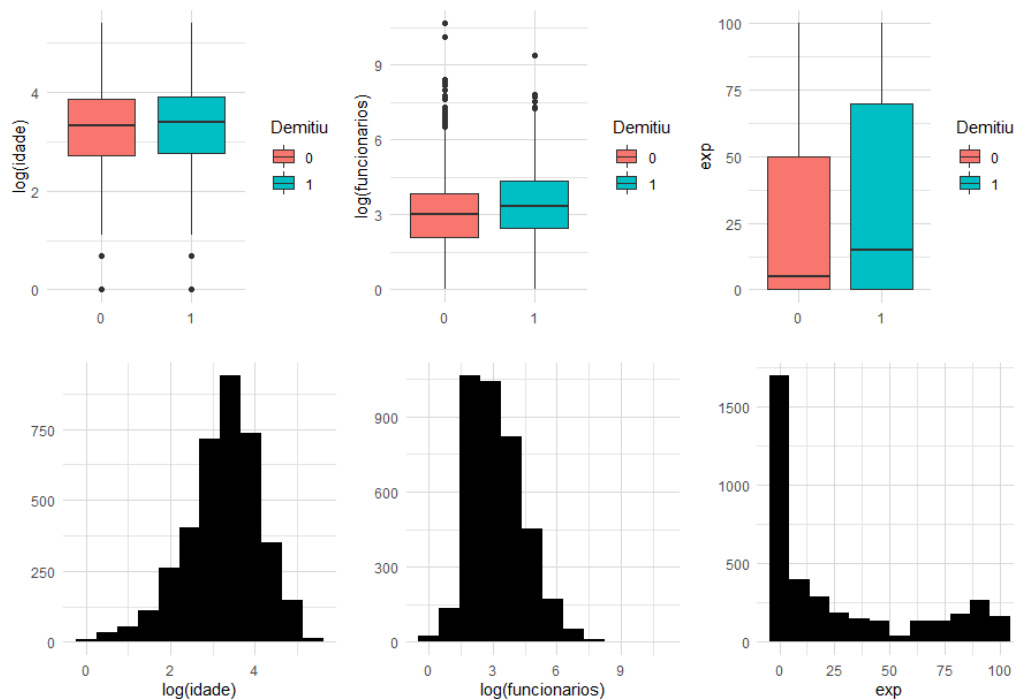
desvio-padrão ajuda a entender quão dispersos são os valores em torno da média. Na variável *idade*, o desvio-padrão é 3,05, indicando que as idades das firmas estão geralmente concentradas em torno da média, com a maioria delas dentro de uma faixa de três anos acima ou abaixo da média. Para a variável *exp*, o desvio-padrão é 34,45, o que sugere uma distribuição mais alargada dos percentuais exportados pelas firmas, indicando uma maior heterogeneidade. A variável *funcionarios* tem o desvio-padrão mais alto de 851,51, refletindo uma ampla gama de tamanhos de firmas dentro da amostra. Esta variabilidade indica a presença de algumas firmas muito grandes ou muito pequenas que distorcem a distribuição.

Aprofundando a investigação das variáveis quantitativas, a [Figura 3](#) apresenta os box-plots das variáveis quantitativas, separando entre firmas que demitiram e firmas que não demitiram, bem como os histogramas desses atributos. Conforme apresentado a seguir, as variáveis utilizadas são na verdade o logaritmo de *idade* e de *funcionarios*, uma transformação que minimiza a influência de *outliers*, evitando distorções no modelo.

Ao considerar o logaritmo da *idade*, existe uma diferença que pode ser definida como sutil, com o box-plot das firmas que demitiram posicionado levemente acima daqueles que não demitiram, indicando uma idade mediana ligeiramente maior para essas empresas. A diferença torna-se mais pronunciada, mesmo que não muito, ao examinar o logaritmo de *funcionarios* e a variável *exp*; os box-plots das firmas que demitiram são consistentemente posicionados um pouco acima das firmas que não demitiram, sugerindo que as tais firmas tendem a ter um número maior de funcionários e uma proporção ligeiramente maior de exportações. Em relação

aos histogramas, destaca-se que o logaritmo de idade e de funcionarios aparentam ter uma distribuição normal, a primeira com a média centrada entre 3 e 4, e a segunda com a média centrada entre 2 e 3. O histograma de exp indica que a maioria das firmas, com mais de 1500, não eram exportadoras, pois a barra do histograma com maior frequência está posicionada em 0.

Figura 3 – Relação de demissão com as variáveis contínuas



Fonte: Elaboração própria.

É importante também abordar as diferenças existentes entre as firmas dentro do mesmo país. Pode-se, com isso, revelar características que existem entre firmas de um país e outro. A análise subsequente se concentrará em explorar essas diferenças. Inicialmente, tem-se que a Tabela 3 divide, por país, as firmas que demitiram e as que não demitiram. Do total de empresas, 2909 firmas não demitiram, 77%, enquanto que 23% demitiram.

Considerando a quantidade de firmas por país, os Países Baixos possuem a maior quantidade de firmas na amostra, com 755, o equivalente a cerca de 20% do total. Em contrapartida, a Áustria tem o menor número, com 54, o que representa 1,4% da amostra. Levando-se em conta o grupo das empresas que demitiram, as firmas dinamarquesas são responsáveis pelo maior número, 168, representando 19% do total. A Áustria, novamente, tem o menor número de firmas que promoveram demissões em função da pandemia, com 11, ou 1,3% do total. Além disso, quando se olha para firmas de mesmo país, A Suécia tem a maior proporção de firmas que demitiram, com 100, quase 29% do total de firmas suecas. A Espanha, por sua vez, tem a menor proporção, com quase 17% ou, em termos absolutos, igual a 48 firmas.

Tabela 3 – Firmas que demitiram por país

País	Não demitiu	Demitiu
Alemanha	88 (73%)	32 (27%)
Áustria	43 (80%)	11 (20%)
Bélgica	455 (82%)	102 (18%)
Dinamarca	438 (72%)	168 (28%)
Espanha	237 (83%)	48 (17%)
Finlândia	321 (74%)	114 (26%)
França	339 (74%)	118 (26%)
Luxemburgo	123 (75%)	41 (25%)
Países Baixos	619 (82%)	136 (18%)
Suécia	246 (71%)	100 (29%)
Total	2909 (77%)	870 (23%)

Fonte: Elaboração própria.

A Figura 4 apresenta uma análise complementar sobre a separação de firmas por país. A linha pontilhada indica o percentual médio de firmas que demitiram do total da amostra. Observa-se que, dos dez países, seis têm proporção de firmas que demitiram acima da média. São eles: Alemanha, Dinamarca, Finlândia, França, Luxemburgo e Suécia. Como mencionado anteriormente, a Suécia possui a maior proporção. Os percentuais variam 16,8% e 28,9%.

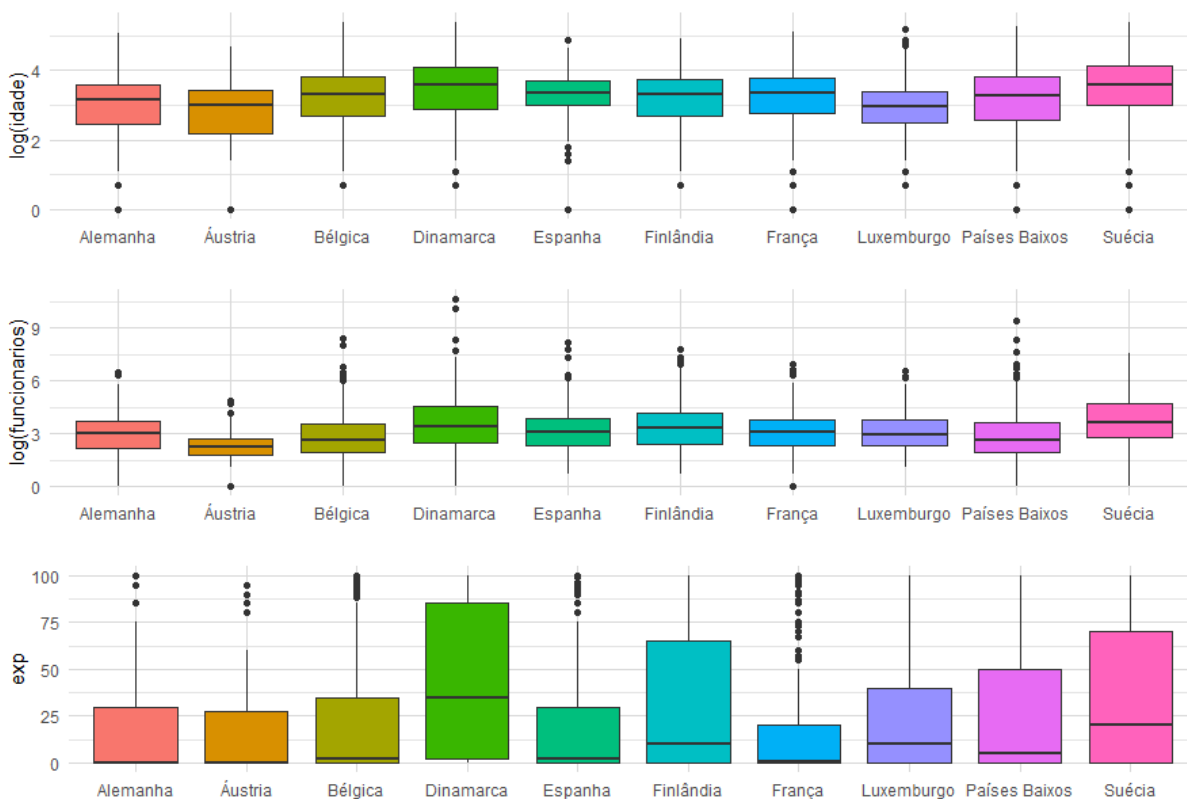
Figura 4 – Proporção de firmas que demitiram por país



Fonte: Elaboração própria.

Relacionando os países com as variáveis quantitativas do estudo, tem-se que, conforme pode ser visualizado na [Figura 5](#), no gráfico superior do logaritmo da variável idade, Dinamarca e Suécia têm a maior mediana dos dez países. Além disso, Espanha e Luxemburgo apresentam uma maior quantidade de valores extremos, indicado pelos pontos acima e abaixo do box-plot. Em relação ao logaritmo da variável funcionários, novamente Dinamarca e Suécia apresentam a maior mediana. Os Países Baixos, por sua vez, têm o maior número de valores extremos, acima do limite superior. Por fim, em relação à variável exp Dinamarca e Suécia também possuem os maiores valores de mediana, enquanto que a França tem a maior quantidade de valores extremos, acima do limite superior.

Figura 5 – Análise das variáveis contínuas por país



Fonte: Elaboração própria.

Avançando na investigação dos dados da pesquisa, recorreu-se às tabelas de contingência para fazer a exploração das variáveis categóricas, relacionando cada uma com a variável dependente demitiu. A [Tabela 4](#) é a relação com a variável categórica ajuda. Essa variável indica se a firma buscou algum tipo de ajuda governamental para poder continuar operando, após ser impactada pela pandemia de COVID-19. Verifica-se que mais de 70% das firmas que tiveram ajuda governamental eram firmas que não demitiram. Adicionalmente, cerca de 68% das empresas que demitiram tiveram algum tipo de ajuda governamental.

A [Tabela 5](#) relaciona a variável dependente com os tipos de indústria incluídos na amostra. A variável industria é dividida em três classes: Manufatura, Varejo e Outros Serviços. As

Tabela 4 – Recebeu ajuda *versus* Demissão

Ajuda	Não demitiu	Demitiu
Não recebeu ajuda	1393 (84%)	272 (16%)
Recebeu ajuda	1516 (74%)	598 (26%)

Fonte: Elaboração própria.

firmas de manufatura compõem a maioria da amostra, com 2128, o equivalente a cerca de 56% do total. As firmas do varejo, por sua vez, têm a menor quantidade, 486, pouco mais de 12% do total da amostra. Considerando a quantidade de firmas que demitiram dentro de cada setor, Manufatura, com 24%, e Outros Serviços, com 23%, têm as maiores proporções.

Tabela 5 – Indústria *versus* Demissão

Indústria	Não demitiu	Demitiu
Manufatura	1615 (76%)	513 (24%)
Varejo	398 (82%)	88 (18%)
Outros serviços	896 (77%)	269 (23%)

Fonte: Elaboração própria.

Finalizando a análise das variáveis categóricas, a [Tabela 6](#) relaciona a variável *demitiu* com variável *inovacao*. Observa-se que a maioria das firmas promoveu algum tipo de esforço de inovação, seja lançando um novo produto no mercado, seja implementando um novo processo. Do total de firmas, 2959 inovaram, pouco mais de 78%. Das firmas que demitiram, 707 delas, ou 81% dessa amostra, promoveram algum tipo de processo de inovação.

Tabela 6 – Inovação *versus* Demissão

Inovação	Não demitiu	Demitiu
Não inovou	657 (80%)	163 (20%)
Inovou	2252 (76%)	707 (24%)

Fonte: Elaboração própria.

4.3 Resultados da modelagem estatística

Os resultados das estimações são apresentados a seguir. Para encontrar o melhor modelo de regressão logística, recorreu-se ao algoritmo *step*, que encontra a combinação de variáveis que gera o melhor Critério de Informação de Akaike (AIC, em inglês). Resumidamente, o AIC é uma medida utilizada para avaliar e comparar modelos estatísticos. Quanto menor o valor do AIC, melhor ajustado é o modelo.

A [Tabela 7](#) apresenta o resultado da regressão logística com o melhor valor do AIC que, neste caso, foi de 3915,9. Observa-se que as variáveis independentes utilizadas foram *exp*, *ajuda* e *log(funcionarios)*. Em todas as três, o coeficiente estimado é positivo e estatisticamente significativo a 1%. Dessa forma, obteve-se que as empresas mais exportadoras e de maior porte no que se refere ao número de funcionários apresentaram tendência a demitir. Para os dados analisados, houve evidência de que a ajuda governamental não surtiu efeito em reduzir as chances de demissão no grupo/perfil de empresas consideradas.

Como todos os efeitos foram estimados com sinal positivo [Tabela 7](#), qualquer variação positiva nas variáveis provoca um aumento a probabilidade de uma firma ser classificada como uma firma que demitiu por conta da pandemia de COVID-19. Estimou-se que a chance de demissão nas empresas que receberam ajuda governamental foi mais de 2 vezes maior que nas empresas que não receberam auxílio (Razão de Chances=2,075) e que, para cada incremento unitário no logaritmo do número de funcionários, a chance de demissão aumentou 18,8% (Razão de Chances=1,188).

Tabela 7 – Resultado da Regressão Logística

Variável	Coeficiente	Desvio-padrão	Valor z	Pr(> z)	Razão de Chances
Intercepto	-2,352	0,123	-19,170	0,000	-
exp	0,004	0,001	3,310	0,001	1,004
ajuda	0,730	0,083	8,761	0,000	2,075
log(funcionarios)	0,173	0,031	5,515	0,000	1,188

Fonte: Elaboração própria.

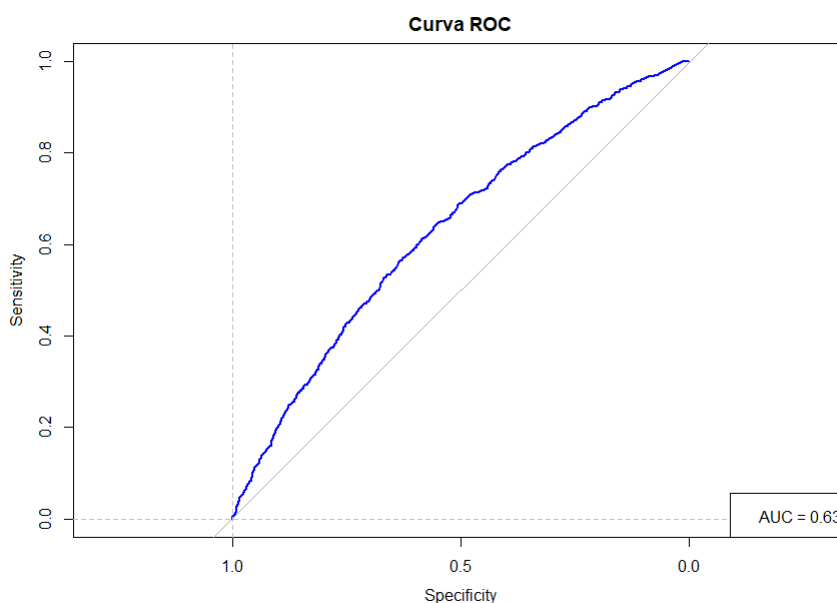
Para definir qual é o valor k que vai classificar a firma como se demitiu ou não, recorreu-se à curva ROC. A [Figura 6](#) apresenta a curva ROC da regressão logística estimada. Verifica-se que a área sob a curva é de 0,63, melhor do que um classificador aleatório. O valor limiar é definido pelo ponto da curva mais próximo às coordenadas (1,1), que seria um classificador perfeito.

Na curva estimada, o ponto mais próximo é o valor 0,23. Assim, considera-se que $k = 0,23$. Para ser classificada como uma firma que demitiu, $\Lambda(\mathbf{x}'\beta) \geq 0,23$. Uma vez definido o valor limiar, pode-se utilizar o modelo para reclassificar as observações, de modo a verificar a acurácia do modelo.

A [Tabela 8](#) apresenta a matriz de confusão, com os valores observados e dos valores preditos da variável dependente *demitiu*. Pelo resultado, pode verificar que o número de observações negativas corretamente classificadas como negativas pelo modelo é de 1855 e o número de observações positivas corretamente classificadas como positivas é de 483. Com isso, a acurácia do modelo é de 61,86%.

A outra metodologia utilizada para identificar os fatores associados às firmas que

Figura 6 – Curva ROC da Regressão Logística estimada



Fonte: Elaboração própria.

Tabela 8 – Matriz de confusão da Regressão Logística

	Predito: 0	Predito: 1
Observado: 0	1855 (63%)	1067 (37%)
Observado: 1	374 (44%)	483 (56%)

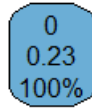
Fonte: Elaboração própria.

demitiram durante a pandemia é a árvore de decisão. A [Figura 7](#) apresenta a árvore gerada. Observa-se que as variáveis utilizadas para calibrar o modelo não foram capazes de serem boas preditoras. Pelo resultado, pode-se verificar que as variáveis independentes não alcançaram um grau relevante no modelo. Isso sugere que outros fatores, não incluídos no modelo atual, podem ter desempenhado um papel crucial nas decisões de demissão das firmas durante a pandemia, segundo o método de árvore de decisão adotado.

Entende-se que esse é um resultado que já poderia ser esperado, visto que a análise descritiva evidenciou o baixo poder discriminatório entre as categorias da variável resposta, com as variáveis independentes tendo distribuições muito similares em ambos os casos. Além disso, pode-se concluir que, com variáveis independentes com baixa correlação com a variável a ser explicada, o melhor método a ser ajustado entre os dois apresentados é o da regressão logística. Como a árvore de decisão não pode ser ajustada, não é possível seguir com a curva ROC e a medição da acurácia do modelo.

Desse modo, das características individuais da firma estudadas neste trabalho, as que são fatores associados a casos de demissão durante a pandemia de COVID-19 são exp, ajuda

Figura 7 – Árvore de decisão não calibrada



Fonte: Elaboração própria.

e $\log(\text{funcionarios})$, com coeficientes associados positivos. Em outras palavras, quanto maior o percentual exportado pela firma do total que é produzido, quanto maior o número de funcionários e se ela obteve ajuda governamental, maior a probabilidade de a firma ser classificada como que tenha demitido. Mesmo assim, com um percentual de acurácia de 61,86%, o modelo estimado pode ser considerado como moderado na sua capacidade de classificar uma firma.

5 Considerações Finais

Este trabalho teve como objetivo determinar fatores que levaram as firmas europeias a demitir funcionários devido à crise econômica provocada pela pandemia de COVID-19. Para alcançar tal objetivo, recorreu-se aos microdados da *Enterprise Surveys*, uma pesquisa do Banco Mundial com dados no nível da firma, que incluiu uma rodada especial sobre os impactos da pandemia nos negócios. A análise do trabalho abrangeu informações de 3.779 empresas situadas em dez países europeus, os quais, conjuntamente, representaram aproximadamente 90% do PIB europeu em 2020.

A pandemia de COVID-19 repercutiu profundamente na economia mundial, impactando em diversos aspectos as empresas e os postos de trabalho. Pesquisas indicaram que a saúde financeira das empresas antes da pandemia desempenhou um papel crucial em suas decisões de dispensar funcionários durante os estágios iniciais da crise. Notou-se também que empresas com uma predominância de funcionárias mulheres sofreram mais em comparação às majoritariamente masculinas. Na Europa, os efeitos sobre o emprego variaram devido às especificidades do mercado laboral de cada país e às diferentes estratégias de confinamento adotadas.

Os dados utilizados são da pesquisa *Enterprise Surveys* do Banco Mundial, que coleta informações sobre o ambiente de negócios de firmas industriais em diversos países através de questionários. Estes são respondidos pelos responsáveis das empresas e abordam temas como operação, estrutura, relações governamentais e com concorrentes. Para avaliar o impacto da pandemia, uma rodada extra foi realizada entre 2020 e 2021. Os dados para esta pesquisa são de dez países europeus: Alemanha, Áustria, Bélgica, Dinamarca, Espanha, Finlândia, França, Países Baixos, Luxemburgo e Suécia.

A variável dependente é binária, e indica se uma firma demitiu ou não em função da pandemia. Por isso, utilizou-se dois modelos de classificação: Regressão Logística e Árvore de Decisão. A árvore de decisão não forneceu um resultado do ajuste, impossibilitando avaliações de acurácia através da curva ROC. Com o método da Regressão Logística, em complemento aos estudos já realizados, o presente trabalho identificou o percentual de exportação, a necessidade de ajuda governamental e a quantidade de funcionários como características que favoreceram a decisão de funcionários durante a pandemia. Uma justificativa para o resultado observado com a variável ajuda pode ser que as empresas tenham recorrido à ajuda governamental como um recurso corretivo, após realizarem demissões, ao invés de procurá-la como uma estratégia preventiva para evitar cortes no quadro de funcionários. Além disso, a inclusão de termos de interação entre as variáveis independentes, que não foi explorado no trabalho, pode melhorar a capacidade preditiva do modelo.

O modelo, com acurácia de 61,86%, possui capacidade moderada de classificação correta.

Estimou-se que a chance de demissão nas empresas que receberam ajuda governamental foi 2,075 vezes maior que nas empresas que não receberam auxílio. Isto, de certa forma, pode ser considerado um resultado contraintuitivo, porém pode estar relacionado ao porte ou outras características específicas das empresas analisadas. Uma análise mais minuciosa envolvendo, por exemplo, a taxa de funcionários demitidos em vez de simplesmente a ocorrência ou não de demissão poderia ser útil no entendimento da relação entre eventos num contexto tão complexo. Ressalta-se, ainda, que a incorporação de termos de interação entre as variáveis independentes, aspecto não abordado neste estudo, poderia potencialmente aprimorar a capacidade preditiva do modelo.

O estudo de [Khan \(2023\)](#) também revelou que empresas com restrições de crédito pré-pandemia eram mais propensas a demitir trabalhadores durante a fase inicial da crise. O acesso a financiamento externo, empréstimos bancários e assistência governamental desempenhou um papel crucial na determinação das decisões de emprego a nível da firma.

Como complementação aos estudos realizados neste trabalho, pode-se indicar a aplicação de outros métodos de classificação como, por exemplo, o *Random Forest* e *k*-vizinhos mais próximos (do inglês, *k-nearest neighbours* ou *kNN*), a fim de melhorar a capacidade preditiva do evento em estudo. Além disso, pode-se pensar em realizar a análise estratificada por país, visto que as peculiaridades regionais podem indicar fatores influentes diferentes em cada caso.

Referências

AKAIKE, H. A new look at the statistical model identification. *IEEE transactions on automatic control*, Ieee, v. 19, n. 6, p. 716–723, 1974. Citado na página 16.

Banco Mundial. *GDP growth (annual %)*. 2023. Acesso em: 17 set. 2023. Disponível em: <<https://data.worldbank.org/indicador/NY.GDP.MKTP.KD.ZG>>. Citado na página 11.

BREIMAN, L. *Classification and regression trees*. [S.l.]: Routledge, 2017. Citado 2 vezes nas páginas 17 e 18.

BUUREN, S. van; GROOTHUIS-OUDSHOORN, K. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, v. 45, n. 3, p. 1–67, 2011. Disponível em: <<https://www.jstatsoft.org/index.php/jss/article/view/v045i03>>. Citado 2 vezes nas páginas 15 e 18.

CHEN, S. et al. Tracking the economic impact of covid-19 and mitigation policies in europe and the united states. *IMF Working Papers*, International Monetary Fund, v. 2020, n. 125, 2020. Citado na página 11.

COUCH, K. A.; FAIRLIE, R. W.; XU, H. Early evidence of the impacts of covid-19 on minority unemployment. *Journal of public economics*, Elsevier, v. 192, p. 104287, 2020. Citado na página 12.

CRISTIANO, M. V. d. M. B. *Sensibilidade e especificidade na curva ROC um caso de estudo*. Tese (Doutorado) — Instituto Politecnico de Leiria (Portugal), 2017. Citado na página 17.

D'ALFONSO, A. Temporary support to mitigate unemployment risks in an emergency (sure). EPRS: European Parliamentary Research Service, 2020. Citado na página 11.

DINGEL, J. I.; NEIMAN, B. How many jobs can be done at home? *Journal of public economics*, Elsevier, v. 189, p. 104235, 2020. Citado na página 12.

FANA, M.; PÉREZ, S. T.; FERNÁNDEZ-MACÍAS, E. Employment impact of covid-19 crisis: from short term effects to long terms prospects. *Journal of Industrial and Business Economics*, Springer, v. 47, p. 391–410, 2020. Citado na página 13.

GREENE, W. H. *Econometric analysis*. [S.l.]: Pearson Education India, 2003. Citado na página 15.

HASTIE, T. J.; PREGIBON, D. Generalized linear models. In: _____. *Statistical Models in S*. [S.l.]: Wadsworth & Brooks/Cole, 1992. cap. 6. Citado na página 18.

KHAN, S. U. The firm-level employment impact of covid-19: International evidence from world bank group's enterprise surveys. *Eastern European Economics*, Taylor & Francis, p. 1–34, 2023. Citado 2 vezes nas páginas 12 e 30.

KONG, E.; PRINZ, D. Disentangling policy effects using proxy data: Which shutdown policies affected unemployment during the covid-19 pandemic? *Journal of Public Economics*, Elsevier, v. 189, p. 104257, 2020. Citado na página 13.

MCCULLAGH, P.; NELDER, J. *Generalized linear models.*,(chapman & hall/crc: London.). 1989. Citado 2 vezes nas páginas 16 e 18.

NICOLA, M. et al. The socio-economic implications of the coronavirus pandemic (covid-19): A review. *International journal of surgery*, Elsevier, v. 78, p. 185–193, 2020. Citado 2 vezes nas páginas 11 e 12.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2020. Disponível em: <<https://www.R-project.org/>>. Citado na página 18.

WITTEN, D.; JAMES, G. *An introduction to statistical learning with applications in R*. [S.l.]: springer publication, 2013. Citado na página 18.