



Universidade de Brasília
IE - Departamento de Estatística

**TRAJETÓRIA DE REEMPREGO:
Fatores que influenciam a recolocação do indivíduo no mercado de trabalho
formal**

Fabiana Mariquito de Sousa

Relatório Final

**Brasília
2021**

Fabiana Mariquito de Sousa

**TRAJETÓRIA DE REEMPREGO:
Fatores que influenciam a recolocação do indivíduo no mercado de trabalho
formal**

Orientadora: Prof^ª. Ana Maria Nogales Vasconcelos
Co-orientadora: Valéria Maria Rodrigues Fechine

Relatório Final de Trabalho de Conclusão de Curso apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2021**

Agradecimentos

Antes de tudo, agradeço a Deus todos os dias, por ter trilhado este caminho comigo até aqui, concedendo-me esta inimaginável oportunidade e me sustentando nos obstáculos. Por todas as pessoas e meios pelos quais Ele proveu como bençãos neste seguimento.

Agradeço à minha orientadora, professora Ana Maria, e co-orientadora, Valéria Maria, por me orientarem neste estudo, pelo suporte, pelo conhecimento transmitido e por acreditarem na minha proposta de trabalho.

Aos meus professores, Maria Teresa, Cátia Regina e George agradeço por todo aprendizado, pela paciência e dedicação em ensinar da melhor maneira. Da mesma forma, aos outros professores, não menos importantes, da Universidade de Brasília.

Agradeço à minha mãe, Micheline, minha maior torcida, que sempre esteve ao meu lado, orientando-me desde os meus primeiros passos e dando-me apoio, motivação e amor incondicional.

Ao meu irmão, Kleverton, minha força, meu incentivo e conforto que sempre me traz luz.

Aos meus tios, Amélia Francisca e Geraldo Cristóvão, que me ensinaram valiosas lições de vida, dando-me toda assistência e credibilidade.

À minha família do coração, Maria das Graças, Kristophr, Miguel, Thaís, Kelven e outros, por me acolherem com amor e por me incentivarem a ir em frente sempre.

À minha irmã e amiga, Thuany, por ceder seu ombro nas horas ruins e celebrar todas as minhas conquistas como se fossem suas.

Agradeço aos meus queridos amigos, Thamara e Victor pelo carinho e companheirismo, pela amizade incondicional e por todo apoio que me deram durante esta etapa.

Agradeço aos amigos conquistados neste percurso, Thais, Otávio, Larissa, Leslie, Felipe e Anna, por passarmos juntos por esta jornada com boas risadas, comemorações e orgulho pelas vitórias, e, nos momentos difíceis, com apoio e carinho recíprocos. Que nossa amizade prevaleça por toda vida.

Agradeço aos meus chefes, mentores e colegas de trabalho, que tive a honra de estagiar na Confederação Nacional da Indústria, Anaely, Maria Cecília, Edgar e toda a equipe, no Banco do Brasil, Igor, Rodrigo, Fábio, Winston, Maria Edite, Vinícius, Hayana e equipe que tanto contribuíram pelo meu crescimento profissional e pessoal. Guardo com muito carinho todos os conselhos, orientações e momentos que passei trabalhando e aprendendo ao lado de vocês.

Agradeço a todos o amor, amparo e afeto a mim dispensados em cada dia, sem isso eu não conseguiria.

Resumo

No Brasil, os postos de emprego formais são regidos pela Consolidação das Leis do Trabalho (CLT) a fim de garantir a melhor estrutura de emprego para empregado e empregador. No cenário econômico atual, o mercado de trabalho formal se encontra estagnado na criação de postos desde 2015. Neste trabalho, foi feito um estudo de caso utilizando os dados da Relação Anual de Informações Sociais (RAIS) fornecida pela Secretaria do Trabalho do Ministério da Economia, dos trabalhadores do município de São Paulo com vínculos formais desligados em 2015 e acompanhados ao longo de 2016 a 2018. Tem por objetivo, analisar as probabilidades de retorno ao mercado de trabalho formal com base nas características dos indivíduos e informações prestadas pelos estabelecimentos. A técnica aplicada foi a Regressão Logística Politômica Ordinal que apresentou evidências de maiores probabilidades em direção ao não retorno para indivíduos pretos e pardos, pessoas desligadas de estabelecimento de grande porte, trabalhadores do sexo masculino, pessoas com menos idade, menor remuneração e menos tempo de permanência no vínculo desligado. Deste modo, tornar-se útil a utilização desses resultados para subsidiar discussões que possibilitem entender a dinâmica do mercado trabalho formal e incentivar outras análises sobre o assunto.

Palavras-chave: Desemprego, Mercado de trabalho formal, RAIS, Regressão logística politômica Ordinal.

Lista de Figuras

1	Proporção de indivíduos segundo Ano de Retorno, São Paulo (SP) - 2016 A 2018	18
2	Mosaico da proporção da variável Tipo Salário segundo Ano de Retorno, São Paulo (SP) - 2015 A 2018	21
3	Mosaico da proporção da variável Porte do Estabelecimento segundo Ano de Retorno, São Paulo (SP) - 2015 A 2018	22
4	Mosaico da proporção da variável Grau de Instrução segundo Ano de Retorno, São Paulo (SP) - 2015 A 2018	22
5	Mosaico da proporção da variável Raça-cor segundo Ano de Retorno, São Paulo (SP) - 2015 A 2018	23
6	Mosaico da proporção da variável Motivo de Desligamento segundo Ano de Retorno, São Paulo (SP) - 2015 A 2018	24
7	Mosaico da proporção da variável Sexo segundo Ano de Retorno, São Paulo (SP) - 2015 A 2018	24
8	Mosaico da proporção da variável Setor segundo Ano de Retorno, São Paulo (SP) - 2015 A 2018	25

Lista de Tabelas

1	Descrição das variáveis	9
2	Frequência de indivíduos segundo Ano de Retorno ao mercado de trabalho formal, São Paulo (SP) - 2016 a 2018	18
3	Proporção das características dos vínculos dos trabalhadores Readmitidos, São Paulo (SP) - 2016 a 2018	19
4	Características dos vínculos dos trabalhadores desligados em 2015 segundo Ano de Retorno, São Paulo (SP) - 2015 a 2018	20
5	Medidas descritivas das variáveis Remuneração Média, Idade e Tempo de emprego segundo Ano de Retorno, São Paulo (SP) - 2015 a 2018	26
6	Testes de Ajuste Global	28
7	Análise de significância	29
8	Estimativas das razões de chances e respectivos intervalos de confiança (95%)	30

Sumário

1 Introdução	4
1.1 Objetivos	6
2 Metodologia	7
2.1 Base de Dados	7
2.2 Variáveis.	7
2.3 Manipulação dos Dados	8
2.4 Regressão Logística.	8
2.4.1 Regressão Logística Múltipla	8
2.4.2 Estimação dos parâmetros	10
2.4.3 Teste de Significância do modelo	11
2.4.4 Intervalos de Confiança	12
2.4.5 <i>Odds Ratio</i> (Razão de Chances)	14
2.4.6 Stepwise - Seleção de Variáveis	14
2.4.7 Regressão Logística Politômica	15
2.4.8 Regressão Lógica Política Ordinal	16
2.4.9 Modelo Logito Cumulativo para Variável Resposta Ordinal (4 cate- gorias)	16
2.4.10 Modelo Logito Cumulativo de Chances (Odds) Proporcionais	17
3 Resultados	18
3.1 Análise Descritiva	18
3.2 Análise de Regressão Logística Politômica Ordinal	28
4 Conclusão	32
5 Referências Bibliográficas	34

1 Introdução

O direito ao emprego é um dos fundamentos que se ligam diretamente ao conceito de cidadania, sendo destacado como um dos principais objetivos da administração pública em todo o mundo (SIMÕES; ALVES; SILVA, 2016).

Souza Campos (2016) define trabalho formal como aquele que garante a melhor estruturação da relação de emprego e que favorece o crescimento e proteção, por meio de de direitos e deveres, para o conjunto da sociedade. Na legislação brasileira, o Decreto-Lei nº 5.452, de 1943, regulamenta o emprego formal que se caracteriza pelo contrato de trabalho formalizado nos termos do disposto na Consolidação das Leis do Trabalho (CLT) que garantem uma série de direitos e deveres tanto para o empregador como para o empregado (BRASIL, 1943).

De acordo com a Organização Internacional do Trabalho (OIT, 2015), a análise dos dados do mercado de trabalho é necessária para a formulação e avaliação de políticas públicas. Além de poder analisar os benefícios, custos e os efeitos dessas políticas ao longo do tempo.

Dois dos aspectos mais relevantes na análise do mercado de trabalho formal para o governo e sociedade são o desemprego e a taxa de informalidade. Quando um indivíduo é desligado de um vínculo formal, o mesmo pode ficar desempregado, empreender ou ir para a informalidade. Em relação ao desemprego, um dos indicadores usados para mensurá-lo é a taxa de desocupação. Segundo a Pesquisa Nacional de Amostra de Domicílios (PNAD Contínua), a taxa média anual de desemprego mostra uma tendência de queda entre os anos 2012 (7,35%) e 2014 (6,82%), mas sobe a partir de 2015 (8,50%), e mantém uma tendência de crescimento desde então (IBGE, 2021).

Portanto, torna-se relevante o estudo do tempo que o trabalhador leva, após o desligamento de um vínculo formal de emprego, para ser reempregado em outro vínculo regido pela CLT. Gomes (2016) avaliou que fatores como escolaridade, treinamento, gênero, raça entre outros podem influenciar a recolocação do trabalhador no mercado de trabalho. Aspectos como setor, motivo do desligamento e salário recebido no emprego anterior também podem ser fatores a serem estudados para melhor compreender essa reinserção.

Este trabalho tem como objetivo verificar se as probabilidades de retorno ao mercado de trabalho formal segundo tempo de desligamento estão associadas às características individuais do trabalhador e de seu vínculo formal do qual foi desligado em 2015. Para o estudo foram consideradas apenas os trabalhadores desligados em 2015 de empresas localizadas no município de São Paulo. Foram utilizados dados da Relação Anual de Informações Sociais (RAIS), fornecidos pela Secretaria do Trabalho do Ministério da Economia e disponibilizados pela Unidade de Estudos e Prospectiva da CNI de forma

não identificada. Para a análise estatística, considerou-se a técnica de Regressão Logística Politômica para os anos 2015 a 2018.

É importante destacar que o momento a que se referem os dados em análise é de recessão econômica, sendo que a criação de empregos formais está estagnada desde 2015 (BARBOSA, 2019). Os resultados apresentados aqui condizem a um momento de recessão econômica, contração dos postos de trabalho formal e demais fatores que prejudicam a situação econômica e social. Nesse sentido, pode-se dizer que o trabalho visa estudar o reemprego em um contexto de baixa geração de empregos formais.

O município de São Paulo foi escolhido como estudo de caso devido a sua maior representatividade no PIB total do país, correspondendo a 10,2% em 2018, a grande movimentação do mercado de trabalho formal, equivalente a 37,01% do total de vínculos formais ativos, e uma das maiores participações econômicas no Brasil (IBGE, 2020). A população desta cidade representa 17,07% da população brasileira, de acordo com a estimativa da população residente de 2018 feita pela PNAD (Ibidem). Além disso, nesse município a taxa de desemprego, após 2015, apresenta tendência de crescimento acima da taxa brasileira para o ano de 2018.

O trabalho está organizado em quatro seções. A primeira traz uma introdução ao trabalho desenvolvido. A seção de metodologia, apresenta a explicação da base de dados, variáveis utilizadas e técnica de análise. Na seção de resultados, são mostrados os dados descritivos das variáveis e as soluções apresentadas pelo modelo. A quarta seção, expõe as conclusões sobre o estudo.

1.1 Objetivos

O objetivo geral é avaliar os fatores que influenciam a reinserção do trabalhador desligado no mercado de trabalho formal entre 2016 a 2018 no município de São Paulo.

Os objetivos específicos são:

- Estudar a aplicação da técnica de Regressão Logística Polinomial ao conjunto de informações de interesse utilizando o *software* estatístico R;
- Avaliar o tempo de recolocação no mercado de trabalho formal no período estudado após o desligamento;
- Analisar a influência das características individuais dos trabalhadores e do vínculo empregatício anterior para a recolocação no mercado de trabalho formal.

2 Metodologia

2.1 Base de Dados

O dados do estudo provém da Relação Anual de Informações Sociais (RAIS) para os anos de 2015 a 2018. A RAIS é uma base da Secretaria do Trabalho do Ministério da Economia.¹

A RAIS é um registro administrativo obrigatório, anual e de âmbito nacional solicitado pela Secretaria do Trabalho aos empregadores onde são declarados os vínculos de trabalho regidos pela CLT (BRASIL. Ministério da Economia. 2021).

Esse sistema de informações tem por objetivo monitorar a atividade trabalhista do país, fornecer dados para elaboração de informações e estudos sobre o mercado de trabalho que possam auxiliar na implementação e monitoramento de políticas públicas de trabalho, emprego e renda, dentre outros (Ibidem).

O registro é preenchido com as informações relativas a todas relações de emprego vigentes no ano-base e são respondidos pelas empresas entre janeiro e abril de cada ano. As informações usados neste trabalho procedem da base de vínculo empregatício.

O banco de dados analisado é constituído por todos os trabalhadores com vínculo CLT que foram desligados no ano de 2015, no município de São Paulo, e acompanhados por um período de três anos.

2.2 Variáveis

Na Tabela 1, são descritas as variáveis consideradas neste estudos e suas respectivas modificações.

Para a análise foram selecionados somente indivíduos com idades entre 18 a 60 anos, com remuneração a partir de um salário mínimo e com pelo menos um mês de tempo de serviço no vínculo desligado em 2015. Não foram considerados vínculos de Jovem Aprendiz pois configuram em vínculos temporários e foram eliminados pessoas com idades acima de 60 anos pois representam um percentual de baixa representatividade.

Algumas categorias das variáveis por terem percentuais irrelevantes e pouco representativos foram filtradas e excluídas. Um desses casos foi a retirada da categoria 'Indígena' da variável 'Raça-Cor' e o outro caso foi a retirada de categorias da variável

¹Os dados utilizados neste trabalho usaram informações de trajetória ocupacional dos trabalhadores formais. As informações foram tratados pela Unidade de Estudos e Prospectiva da CNI e utilizados neste estudo de forma não identificada.

‘Motivo de Desligamento’, só foram considerados os desligamentos com e sem justa causa por parte do empregador e término de contrato de trabalho (visto que representam mais de 80% dos casos), as outras causas como aposentadoria e óbito foram desconsideradas. A categoria de ‘Analfabetos’ da variável ‘Grau de Instrução’ também não foi considerada devido ao seu baixo percentual de representatividade, e da mesma forma a categoria ‘Agropecuário’ da variável ‘Setor’ também não está presente na análise.

As variáveis qualitativas consideradas, exceto ‘Sexo’, tiveram suas categorias reagrupadas para melhorar a distribuição dos dados, pois algumas apresentavam percentuais muito pequenos. Já para os casos das variáveis ‘Setor’ e ‘Porte do Estabelecimento’, ‘Setor’ foi reagrupado com base na classificação feita pela Secretaria do Trabalho que é originária da agregação das Seções da Classificação Nacional de Atividades Econômicas (CNAE 2.0) e está integrada com a quarta revisão da International Standard Industrial Classification of all Economic Activities (ISIC), desenvolvida pela Organização das Nações Unidas (ONU) (BRASIL. Ministério da Economia. 2020) e as categorias do ‘Porte do Estabelecimento’ foi feita com base na classificação do SEBRAE e Dieese (2013) de acordo com o número de empregados do estabelecimento.

Todas as variáveis referentes ao vínculo dos trabalhadores são do momento de desligamento em 2015, as únicas informações dos demais anos são de quando o indivíduo retorna.

2.3 Manipulação dos Dados

Para leitura, limpeza, análise dos dados foi e aplicação da técnica foi utilizado o *software* estatístico livre R.

2.4 Regressão Logística

2.4.1 Regressão Logística Múltipla

A regressão logística é um modelo utilizado para descrever as relações entre uma variável resposta categórica e uma ou mais variáveis explicativas. Esse tipo de modelagem se diferencia do modelo de regressão linear pois a variável resposta é qualitativa categórica e não quantitativa.

Nesse caso Y é uma variável de resposta binária que representa sucesso, $P(Y = 1)$, ou fracasso, $P(Y = 0)$, ou seja, Y tem distribuição Bernoulli e denota-se por $P(Y = 1) = \pi(x)$ e $P(Y = 0) = 1 - \pi(x)$.

Tabela 1: Descrição das variáveis

Nome Técnico	Nome da Variável	Descrição e modificações	Medição
Vínculo Ativo 31/12 e Ano	Ano de Retorno	Variável do ano de retorno ao primeiro vínculo após desligamento em 2015. Leva em consideração o ano de registro da base e se o indivíduo tem ou não vínculo ativo através da variável 'Vínculo Ativo 31/12' codificada em 0 para desligado em 31/12 e 1 para ativo em 31/12.	Categorias: i) 2016. ii) 2017. iii) 2018. iv) Não Retornou.
Tipo Salário	Tipo Salário	Tipo de salário recebido pelo trabalhador. As categorias foram separadas em: i) Mensal. ii) Outros - junção de Quinzenal, Semanal, Diário, Horário, Tarefa e Outros.	Categorias: i) Mensal. ii) Outros.
Tamanho Estabelecimento	Porte do Estabelecimento	Porte do estabelecimento com base no número de empregados. Suas categorias são separadas por: i) Micro: 0 a 9 funcionários. ii) Pequeno: 10 a 49 funcionários.	Categorias: i) Micro. ii) Pequeno. iii) Média. iv) Grande.
Escolaridade após 2005	Grau de Instrução	Grau de Instrução do trabalhador. Categorizada em: i) Nível Fundamental: até o 5ª ano incompleto do ensino fundamental, 5ª ano completo do ensino fundamental, o 6ª ao 9ª ano incompleto do ensino fundamental e ensino fundamental Completo. ii) Nível Médio: ensino médio incompleto e ensino médio completo. iii) Nível Superior: educação superior incompleta, educação superior completa, mestrado completo e doutorado completo.	Categorias: i) Nível Fundamental. ii) Nível Médio. iii) Nível Superior.
Raça Cor	Raça-Cor	Raça-cor do trabalhador. Essa variável foi categorizada em duas classificações que são: i) Preta e Parda: Preta e Parda. ii) Branca e Amarela: Branca e Amarela.	Categorias: i) Preta e Parda. ii) Branca e Amarela.
Motivo Desligamento	Motivo Desligamento	Motivo de desligamento do vínculo do trabalhador. São considerados na análise apenas os motivos de desligamentos associados ao empregador e término de contrato de trabalho. Os demais motivos como transferência, falecimento, aposentaria, mudança de regime, óbito e desligamentos com e sem justa causa por parte do empregado foram desconsiderados.	Categorias: i) Com justa causa - empregador. ii) Sem justa causa - empregador. iii) Término do contrato.
Sexo Trabalhador	Sexo	Sexo do trabalhador.	Categorias: i) Feminino. ii) Masculino.
CNAE Classe	Setor	Setor de atividade econômica do vínculo do trabalhador. Nesta classificação foi considerada os dois primeiros dígitos da variável 'CNAE Classe' e através desses dígitos foram classificados: i) Indústria: dígitos de 5 a 39. ii) Construção: dígitos de 40 a 43. iii) Comércio: dígitos de 45 a 47. iv) Serviços: dígitos de 49 a 99.	Categorias: i) Comércio. ii) Construção. iii) Indústria. iv) Serviços.
Vl Remun (SM)	Remuneração Média (S.M.)	Remuneração média em salário mínimos do trabalhador.	Quantidade de salário mínimos.
Idade	Idade do trabalhador	Idade do trabalhador no ano do desligamento em 2015.	Quantidade de anos.
Tempo Emprego	Tempo de Emprego (em meses)	Tempo de permanência no emprego em Meses.	Quantidade de meses.

O modelo de regressão logística é dado pela forma:

$$\text{logito}[P(Y = 1)] = \text{logito}[\pi(x)] = \log \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \alpha + \beta x$$

em que o *logito* é a função de ligação g que é aplicada a cada componente $E(y)$ relacionando esta resposta média ao preditor linear $g[E(y)] = \mathbf{X}\beta$, ou seja, conecta a componente aleatória com o preditor linear que neste caso é o logaritmo da chance para a distribuição Binomial.

A partir da equação acima, aplica-se o exponencial dos termos para se obter a $P(Y = 1) = \pi(x)$ da seguinte forma:

$$\pi(x) = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}$$

onde α é o intercepto do modelo e β é o efeito do incremento em x no valor do logaritmo da chance (odds) de $P(Y = 1)$.

Para o caso onde a variável resposta Y pode ser explicada por mais de uma variável explicativa, ou seja, p variáveis preditoras o modelo seguirá a forma:

$$\text{logito}[P(Y = 1)] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p$$

no qual parâmetro β_j representa o efeito do aumento de uma unidade em x_j no logito de $Y = 1$, mantido constantes as outras variáveis.

E dessa forma:

$$g(x) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p$$

na qual sua $P(Y = 1)$ será dada por:

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

2.4.2 Estimação dos parâmetros

Em regressão logística, os parâmetros do modelo são estimados pelo método de máxima verossimilhança. Para o modelo simples temos que a função de probabilidade será:

$$f(y_i, \pi(x_i)) = \pi^{y_i}(x_i)(1 - \pi(x_i))^{1 - y_i}$$

para $y_i = 0, 1$

Dessa forma a função de verossimilhança será:

$$L(\beta) = \prod_{i=1}^n \pi^{y_i}(x_i)(1 - \pi(x_i))^{1-y_i}$$

Aplica-se então, o logaritmo natural em ambos os lados da equação e após isso deriva-se em relação a β_0 e em relação a β_1 igualando a 0. Desse modo, obtêm-se as seguintes equações:

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0$$

e

$$\sum_{i=1}^n x_i [y_i - \pi(x_i)] = 0$$

E pelo mesmo método utilizado no caso univariado, serão obtidas as estimativas para o caso multivariado e a função de máxima verossimilhança será quase idêntica com a ressalva de que $\pi(x)$ agora será definido por:

$$\pi(x) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}}$$

Dessa forma, as equações de verossimilhanças podem ser expressas por:

$$\sum_{i=1}^n [y_i - \pi(\mathbf{x}_i)] = 0$$

e

$$\sum_{i=1}^n x_{ij} [y_i - \pi(\mathbf{x}_i)] = 0$$

para $j=1,2,\dots,p$.

2.4.3 Teste de Significância do modelo

Teste Razão de Máxima Verossimilhança

As comparações entre valores observados e valores preditos que faz uso da função de verossimilhança, utilizam a seguinte expressão:

$$D = -2 \ln \left[\frac{\text{verossimilhança do modelo ajustado}}{\text{verossimilhança do modelo saturado}} \right].$$

Essa estatística D é chamada de deviance (desvio) e sua reescrita é dada por:

$$D = -2\ln[\text{verossimilhança do modelo ajustado}]$$

e para testar a significância das variáveis, usa-se a estatística D da seguinte forma:

$$G = D(\text{verossimilhança sem as variáveis}) - D(\text{verossimilhança com as variáveis})$$

que se tornará

$$G = -2\ln \left[\frac{\text{verossimilhança sem variáveis}}{\text{verossimilhança com variáveis}} \right].$$

Sob a hipótese nula de que os p coeficientes das variáveis no modelo se igualam a zero, a estatística G segue uma distribuição χ^2 com $p = k(j - 2)$ graus de liberdade, onde k é o número de variáveis preditoras e j , o número de categorias da variável resposta Y .

Teste de Wald

Esse teste é usado para testar a significância, separadamente, de cada coeficiente das variáveis presentes no modelo. Sob a hipótese nula de que algum coeficiente β_j para $j = 1, 2, \dots, p$ seja igual a 0, temos a estatística de teste:

$$W_j = \frac{\hat{\beta}_j}{\hat{SE}(\hat{\beta}_j)}$$

seguindo uma distribuição normal. $\hat{\beta}_j$ é uma estimativa do parâmetro e $\hat{SE}(\hat{\beta}_j)$ é a estimativa do erro padrão do coeficiente.

2.4.4 Intervalos de Confiança

A expressão geral do logito estimado para um modelo com p variáveis explicativas será:

$$\hat{g}(\mathbf{x}) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p$$

ou seja,

$$\hat{g}(\mathbf{x}) = \mathbf{x}'\hat{\beta}$$

onde têm-se que o vetor

$\hat{\beta}' = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$ é estimador dos $p + 1$ coeficientes do modelo e o vetor

$$\mathbf{x}' = (x_0, x_1, \dots, x_p)$$

são as p variáveis explicativas do modelo.

Portanto, temos em forma matricial que:

$$\widehat{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}$$

e

$$\widehat{Var}[\hat{g}(x)] = \mathbf{x}'\widehat{Var}(\hat{\beta})\mathbf{x} = \mathbf{x}'(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{x}$$

onde

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

e

$$\mathbf{V} = \begin{pmatrix} \hat{\pi}(x_1)(1 - \hat{\pi}(x_1)) & 0 & \dots & 0 \\ 0 & \hat{\pi}(x_2)(1 - \hat{\pi}(x_2)) & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \dots & 0 & \hat{\pi}(x_n)(1 - \hat{\pi}(x_n)) \end{pmatrix}.$$

Dessa forma, temos que os intervalos de confiança para os coeficientes e logito são dados, respectivamente, por:

$$\hat{\beta}_j \pm z_{1-\frac{\alpha}{2}}\widehat{SE}(\hat{\beta}_j)$$

e

$$\hat{g}(\mathbf{x}) \pm z_{1-\frac{\alpha}{2}}\widehat{SE}(\hat{g}(\mathbf{x})).$$

2.4.5 Odds Ratio (Razão de Chances)

Hosmer e Lemeshow (2002), define o *odds ratio* por uma estatística que calcula a razão entre a chance de ocorrência de $x = 1$ e $x = 0$, ou seja,

$$OR = \frac{\pi(1)/(1 - \pi(1))}{\pi(0)/(1 - \pi(0))}$$

onde

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

Substituindo os termos em OR para $x = 1$ e $x = 0$, temos que

$$OR = e^{\beta_1}$$

ou seja, para cada uma unidade acrescida em x , a chance de "sucesso" aumenta em e^{β} ou o logito (log-chance de "sucesso") aumenta em β unidades.

O intervalo de confiança dessa medida é obtido usando os limites do intervalo de confiança de $\hat{\beta}_j$

$$\hat{\beta}_j \pm z_{1-\frac{\alpha}{2}} SE(\hat{\beta}_j)$$

e assim, o intervalo para odds ratio será dado por

$$[e^{\beta_l}; e^{\beta_s}].$$

2.4.6 Stepwise - Seleção de Variáveis

O método a ser abordado trata da verificação da significância de cada variável ao ser incluída no modelo. Para a modelagem logística, a significância no modelo é garantida pelo teste da Razão de Verossimilhança (TRV). Portanto, a variável mais significativa é aquela que resultar na maior diferença para o logaritmo da verossimilhança.

A verificação é feita seguindo os seguintes passos:

Passo 1: Inicia-se testando um modelo somente com o intercepto e tendo L_0 como o log da verossimilhança desse modelo. Então, são testados modelos individuais para cada variável explicativa, x_j , e comparados suas log verossimilhanças, L_j .

O primeiro passo se encerra quando é obtido a variável com o menor p-valor que seja menor do que o nível de significância estipulado no teste TRV quando comparados ao modelo de intercepto.

Passo 2: Então continuando com o modelo contendo o intercepto e a variável selecionada no passo anterior, continua a busca através de $p - 1$ modelos com as variáveis restante e encontrando suas respectivas log-verossimilhanças. Feito isso, efetua-se novamente o TRV para todos os modelos ajustados versus o modelo obtido no primeiro passo e o critério de escolha da próxima variável permanecerá o mesmo.

Passo 3: Aqui o processo se mantém o mesmo até o último passo, ou seja, incluindo mais uma variável significativa mas aqui torna-se importante a checagem da significância do modelo, pois o acréscimo de mais uma variável pode causar a perda de significância de outras variáveis já incluídas.

Passo n: Nesse passo final, o processo se encerra quando é obtido um modelo onde todas as variáveis detêm p-valor menor que o nível de significância estipulado.

2.4.7 Regressão Logística Politômica

Essa modelagem é usada para o caso em que a variável resposta assume 3 ou mais categorias.

O modelo consiste em atrelar cada categoria com uma categoria base. Geralmente define-se a última categoria (c) como base, e seu logito é:

$$\log \left[\frac{\pi_j(x)}{\pi_c(x)} \right], \quad j = 1, \dots, c - 1.$$

Para Y , por exemplo, assumindo três categorias, $Y = 0$, $Y = 1$ e $Y = 2$. Os logitos são:

$$g_1(x) = \log \left[\frac{\pi_1(x)}{\pi_0(x)} \right] = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \dots + \beta_{1p}x_p = \mathbf{x}'\beta_1$$

e

$$g_2(x) = \log \left[\frac{\pi_2(x)}{\pi_0(x)} \right] = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \dots + \beta_{2p}x_p = \mathbf{x}'\beta_2$$

e assim, temos

$$\log \left[\frac{\pi_2(x)}{\pi_1(x)} \right] = \log \left[\frac{\pi_2(x)}{\pi_0(x)} \right] - \log \left[\frac{\pi_1(x)}{\pi_0(x)} \right]$$

que para $\alpha = \alpha_2 - \alpha_1$ e $\beta = \beta_2 - \beta_1$ e p variáveis explicativas, o modelo será:

$$\log \left[\frac{\pi_2(x)}{\pi_1(x)} \right] = \alpha_j + \beta_{j1}x_1 + \beta_{j2}x_2 + \dots + \beta_{jp}x_p, \quad j = 1, 2.$$

Para estimar a probabilidade de resposta do modelo acima usamos a seguinte equação:

$$\pi_j(x) = \frac{e^{\alpha_j + \beta_{j1}x_1 + \beta_{j2}x_2 + \dots + \beta_{jp}x_p}}{\sum_{i=1}^2 e^{\alpha_i + \beta_{i1}x_1 + \beta_{i2}x_2 + \dots + \beta_{ip}x_p}} \quad \text{para } j = 1, 2.$$

Generalizando para j categorias, o modelo será:

$$\log \left[\frac{\pi_j(x)}{\pi_c(x)} \right] = \alpha_j + \beta_{j1}x_1 + \beta_{j2}x_2 + \dots + \beta_{jp}x_p, \quad j = 1, \dots, c - 1.$$

2.4.8 Regressão Lógica Poltômica Ordinal

Este caso de modelo logístico multivariado é utilizado quando a variável resposta tem natureza ordinal, ou seja, tem categorias dependentes e ordenadas.

2.4.9 Modelo Logito Cumulativo para Variável Resposta Ordinal (4 categorias)

A probabilidade cumulativa para o resultado da categoria j para a variável Y é:

$$P(Y \leq j) = \pi_1(x) + \pi_2(x) + \pi_3(x) + \pi_4(x), \quad j = 1, 2, 3, 4$$

e tem a ordem

$$P(Y \leq 1) \leq P(Y \leq 2) \leq P(Y \leq 3) \leq P(Y \leq 4) = 1.$$

Os logitos cumulativos serão dados por:

$$\text{logito}[P(Y \leq j)] = \log \left[\frac{P(Y \leq j)}{1 - P(Y \leq j)} \right] = \log \left[\frac{\pi_1(x) + \dots + \pi_j(x)}{\pi_{j+1}(x) + \dots + \pi_c(x)} \right]$$

para $j = 1, \dots, c - 1$. Para este caso de aplicação estão sendo considerados $c = 4$, então os logitos cumulativos são:

$$\text{logito}[P(Y \leq 1)] = \log \left[\frac{\pi_1(x)}{\pi_2(x) + \pi_3(x) + \pi_4(x)} \right],$$

$$\text{logito}[P(Y \leq 2)] = \log \left[\frac{\pi_1(x) + \pi_2(x)}{\pi_3(x) + \pi_4(x)} \right]$$

e

$$\text{logito}[P(Y \leq 3)] = \log \left[\frac{\pi_1(x) + \pi_2(x) + \pi_3(x)}{\pi_4(x)} \right]$$

2.4.10 Modelo Logito Cumulativo de Chances (Odds) Proporcionais

O modelo terá a forma:

$$\text{logito}[P(Y \leq j)] = \alpha_j + \beta' \mathbf{x}, \quad j = 1, \dots, c - 1$$

neste caso, cada logito cumulativo terá seu próprio intercepto.

Para a interpretação do modelo, Agresti (2019) define que o odds ratio (razão de chances) é usado para as probabilidades cumulativas e complementos. Temos para dois valores a e b para x , que

$$OR = \frac{P(Y \leq j \mid x = a) / P(Y > j \mid x = a)}{P(Y \leq j \mid x = b) / P(Y > j \mid x = b)}.$$

Então, o logaritmo dessa razão de chances é a diferença entre os dois logitos para ambos os valores de x . Essa propriedade garante que para $a - b = 1$, as chances de resposta $\leq j$ será multiplicado por e^β para cada acréscimo de uma unidade em x .

3 Resultados

3.1 Análise Descritiva

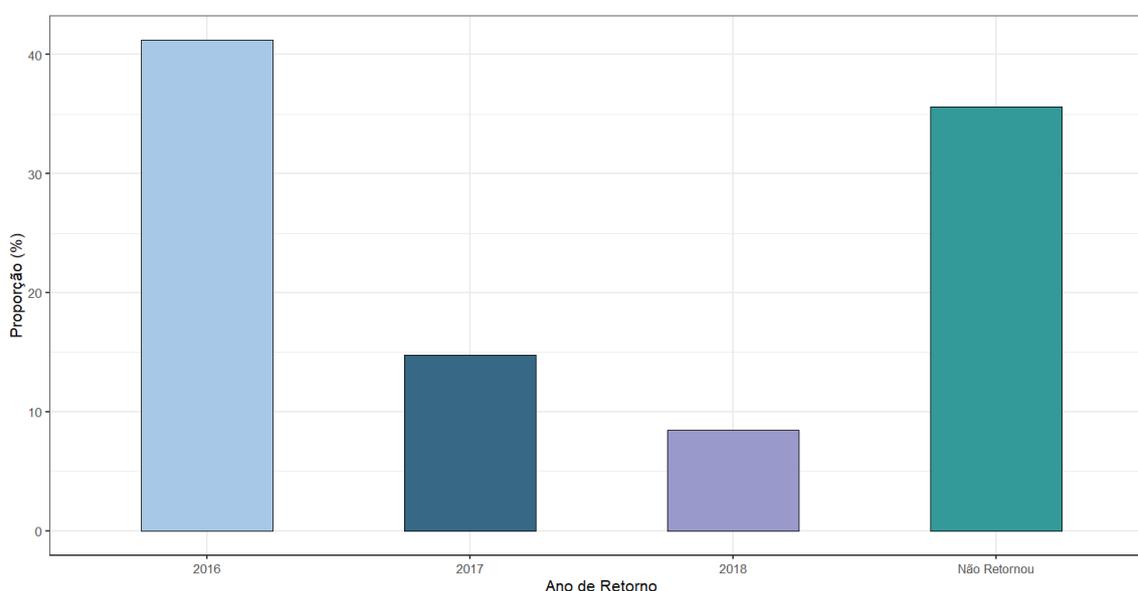
A partir dos dados apresentados, nota-se que 41,2% dos trabalhadores retornaram em 2016, 35,6%, não retornaram ao mercado de trabalho e 23,3% das pessoas retornaram em 2017 e 2018.

Tabela 2: Frequência de indivíduos segundo Ano de Retorno ao mercado de trabalho formal, São Paulo (SP) - 2016 a 2018

Ano de Retorno	Frequência absoluta	Frequência relativa (%)	Frequência acumulada	Frequência relativa acumulada (%)
2016	419.548	41,2	419.548	41,2
2017	150.398	14,8	569.946	56,0
2018	86.003	8,5	655.949	64,4
Não Retornou	362.384	35,6	1.018.333	100,0
Total	1.018.333	100,0	-	-

Fonte: Ministério da Economia, Relação Anual de Informações Sociais (RAIS) - 2015 a 2018.

Figura 1: Proporção de indivíduos segundo Ano de Retorno, São Paulo (SP) - 2016 A 2018



Fonte: Ministério da Economia, Relação Anual de Informações Sociais (RAIS) - 2016 a 2018.

Observa-se que em média 55,3% dessas pessoas foram admitidas em vínculos dos quais recebem remunerações menores que a do vínculo desligado em 2015. Cerca de 74,4% desses indivíduos terminaram o ano ativos em seus anos de reemprego.

Outro fator que foi observado é a permanência desses indivíduos em vínculos no mesmo estado e município. Em torno de 87,56% das pessoas admitidas para os três anos avaliados permanecem no mesmo estado, porém, mais de 25% dos trabalhadores foram admitidos em vínculos que não pertencem ao município de São Paulo.

Tabela 3: Proporção das características dos vínculos dos trabalhadores Readmitidos, São Paulo (SP) - 2016 a 2018

Características	(%)		
	2016	2017	2018
Remuneração			
Igual	1,0	0,9	0,8
Maior	42,1	42,9	46,6
Menor	56,9	56,3	52,6
Vínculo Ativo em 31/12			
Ativo	73,1	74,4	75,8
Desligado	26,9	25,6	24,2
Estado			
Não Permaneceu	11,0	13,0	13,0
Permaneceu	89,0	87,0	87,0
Município			
Não Permaneceu	25,1	27,2	28,4
Permaneceu	74,9	72,8	71,6

Fonte: Ministério da Economia, Relação Anual de Informações Sociais (RAIS) - 2016 a 2018.

Abaixo são apresentadas às análises descritivas bivariadas das variáveis utilizadas no estudo.

Tabela 4: Características dos vínculos dos trabalhadores desligados em 2015 segundo Ano de Retorno, São Paulo (SP) - 2015 a 2018

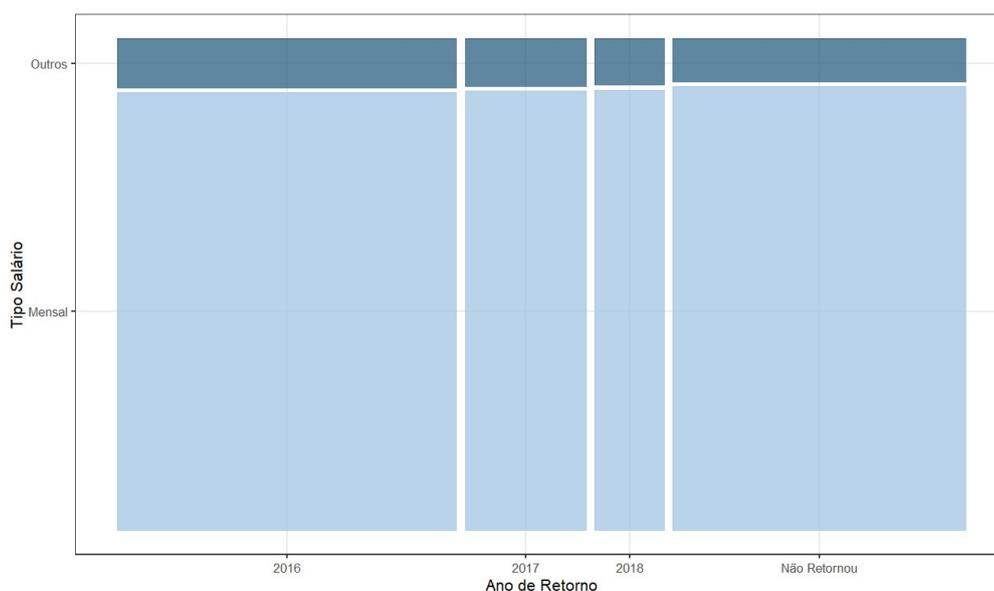
Características	Total		Ano de Retorno							
	n	%	2016		2017		2018		Não Retornou	
			n	%	n	%	n	%	n	%
Total	1.018.333	100	419.548	41,2	150.398	14,8	86.003	8,4	362.384	35,6
Tipo Salário										
Mensal	921.170	90,5	377.403	41,0	135.769	14,7	77.813	8,4	330.185	35,8
Outros	97.163	9,5	42.145	43,4	14.629	15,1	8.190	8,4	32.199	33,1
Porte do Estabelecimento										
Grande	403.671	39,6	171.881	42,6	60.165	14,9	34.154	8,5	137.471	34,1
Média	101.068	9,9	42.700	42,2	14.995	14,8	8.660	8,6	34.713	34,3
Micro	248.142	24,4	94.568	38,1	35.991	14,5	20.686	8,3	96.897	39,0
Pequena	265.452	26,1	110.399	41,6	39.247	14,8	22.503	8,5	93.303	35,1
Grau de Instrução										
N. Fundamental	227.931	22,4	90.407	39,7	31.811	14,0	18.475	8,1	87.238	38,3
N. Médio	605.851	59,5	259.951	42,9	91.999	15,2	52.101	8,6	201.800	33,3
N. Superior	184.551	18,1	69.190	37,5	26.588	14,4	15.427	8,4	73.346	39,7
Raça Cor										
Preta e Parda	419.671	41,2	178.356	42,5	63.182	15,1	36.201	8,6	141.932	33,8
Branca e Amarela	598.662	58,8	241.192	40,3	87.216	14,6	49.802	8,3	220.452	36,8
Motivo Desligamento										
Com justa causa - empregador	41.669	4,1	14.925	35,8	6.603	15,8	4.049	9,7	16.092	38,6
Sem justa causa - empregador	852.639	83,7	353.930	41,5	124.863	14,6	70.756	8,3	303.090	35,5
Término do contrato	124.025	12,2	50.693	40,9	18.932	15,3	11.198	9,0	43.202	34,8
Sexo										
Feminino	402.909	39,6	161.616	40,1	57.614	14,3	33.962	8,4	149.717	37,2
Masculino	615.424	60,4	257.932	41,9	92.784	15,1	52.041	8,5	212.667	34,6
Setor										
Comércio	232.523	22,8	97.703	42,0	34.889	15,0	19.262	8,3	80.669	34,7
Construção	163.091	16,0	69.079	42,4	24.886	15,3	14.018	8,6	55.108	33,8
Indústria	97.591	9,6	39.050	40,0	14.271	14,6	7.700	7,9	36.570	37,5
Serviços	525.128	51,6	213.716	40,7	76.352	14,5	45.023	8,6	190.037	36,2
Quartil Remuneração Média (S.M.)¹										
1	259.291	25,5	100.623	38,8	38.752	14,9	23.092	8,9	96.824	37,3
2	253.996	24,9	107.804	42,4	38.296	15,1	21.845	8,6	86.051	33,9
3	251.541	24,7	111.199	44,2	37.609	15,0	20.908	8,3	81.825	32,5
4	253.505	24,9	99.922	39,4	35.741	14,1	20.158	8,0	97.684	38,5
Quartil Idade²										
1	256.891	25,2	114.153	44,4	43.678	17,0	24.997	9,7	74.063	28,8
2	279.919	27,5	124.130	44,3	42.779	15,3	24.067	8,6	88.943	31,8
3	237.731	23,3	98.468	41,4	34.170	14,4	19.591	8,2	85.502	36,0
4	243.792	23,9	82.797	34,0	29.771	12,2	17.348	7,1	113.876	46,7
Quartil Tempo de emprego³										
1	258.659	25,4	107.921	41,7	39.520	15,3	23.210	9,0	88.008	34,0
2	250.845	24,6	105.666	42,1	38.708	15,4	22.471	9,0	84.000	33,5
3	254.939	25,0	108.881	42,7	38.424	15,1	21.325	8,4	86.309	33,9
4	253.890	24,9	97.080	38,2	33.746	13,3	18.997	7,5	104.067	41,0

Fonte: Ministério da Economia, Relação Anual de Informações Sociais (RAIS) - 2015 a 2018.

¹Quartis de Remuneração Média (S.M.): 1 - [1; 1,38], 2 - (1,38; 1,74], 3 - (1,74; 2,54] e 4 - (2,54; 150].²Quartis de Idade: 1 - [18; 25], 2 - (25; 32], 3 - (32; 40] e 4 - (40; 60].³Quartis de Tempo de emprego (em meses): 1 - [1; 4,9], 2 - (4,9; 12,5], 3 - (12,5; 29,5] e 4 - (29,5; 544].

De acordo com o gráfico (Figura 2) apresentado abaixo e os dados mostrados na Tabela 3, 90,5% dos trabalhadores desligados em 2015 recebiam salário do tipo mensal e somente 9,5% recebiam salário do tipo por hora, diária, entre outros. Em ambas as categorias, a maioria, 41,0% e 43,4%, retornam ao mercado de trabalho em 2016 e já 35,8% dos que recebiam salário mensal e 33,1% dos que recebiam outro tipo de salário, não retornaram.

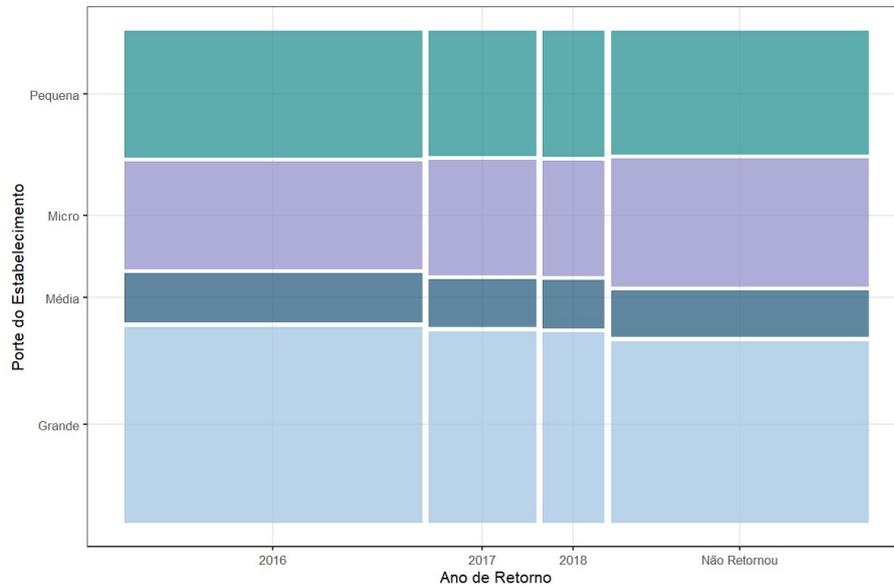
Figura 2: Mosaico da proporção da variável Tipo Salário segundo Ano de Retorno, São Paulo (SP) - 2015 A 2018



Fonte: Ministério da Economia, Relação Anual de Informações Sociais (RAIS) - 2015 a 2018.

Avaliando o porte do estabelecimento em relação ao ano de retorno, percebe-se que das pessoas desligadas, 39,6% eram de empresas de grande porte e dessas, 42,6% retornaram em até um ano após o desligamento. Apenas as pessoas desligadas em micro empresas (24,4%), não retornaram ao mercado de trabalho em sua maioria (39,0%), já para as outras categorias a maior parte retorna em 2016.

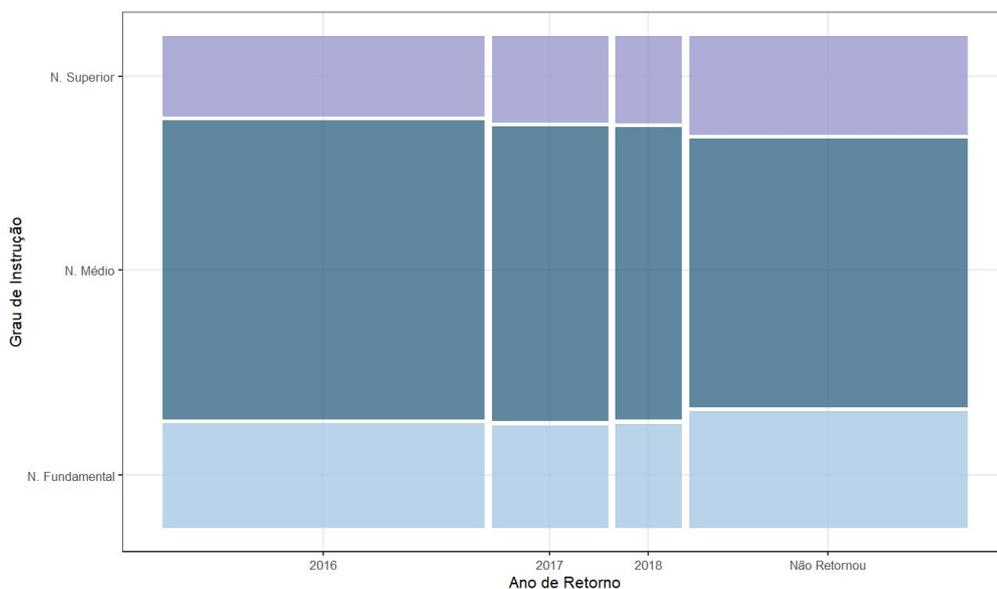
Figura 3: Mosaico da proporção da variável Porte do Estabelecimento segundo Ano de Retorno, São Paulo (SP) - 2015 A 2018



Fonte: Ministério da Economia, Relação Anual de Informações Sociais (RAIS) - 2015 a 2018.

Os 59,5% dos vínculos desligados tinham grau de instrução de nível médio, os outros 40,5% distribuem-se entre o nível superior e fundamental. Para os níveis fundamental e médio, a maioria dos indivíduos retornou em 2016, 39,7% e 42,9%, mas para o nível superior 39,7% não retornou, representando o maior percentual dos indivíduos dessa categoria.

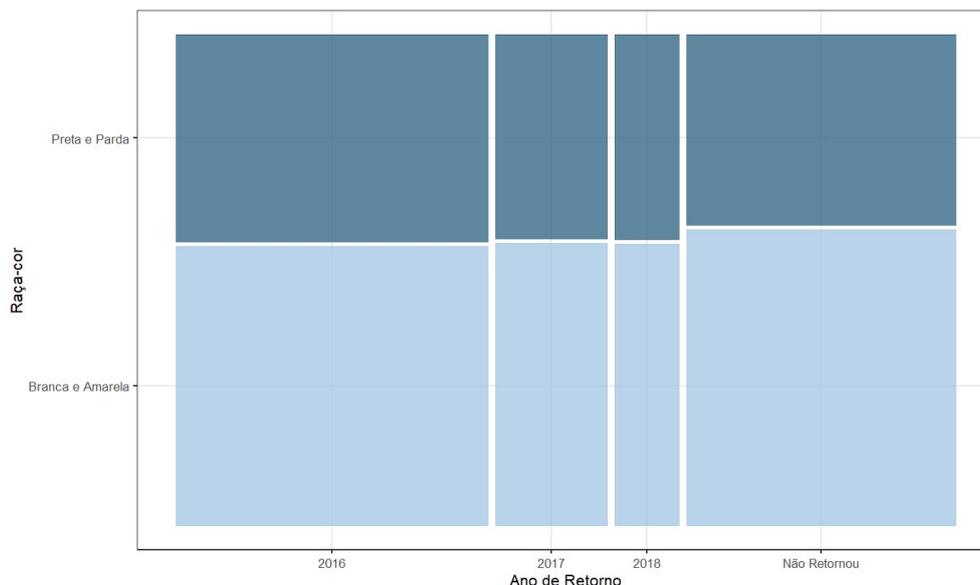
Figura 4: Mosaico da proporção da variável Grau de Instrução segundo Ano de Retorno, São Paulo (SP) - 2015 A 2018



Fonte: Ministério da Economia, Relação Anual de Informações Sociais (RAIS) - 2015 a 2018.

A raça-cor das pessoas analisadas são em maior representatividade, 58,8%, branca e amarela. Para ambas as categorias, a maioria retornou em 2016, 42,5% para preta e parda e 40,3% para branca e amarela e os segundos maiores percentuais de cada categorias, não retornam, 33,8% e 36,8% respectivamente.

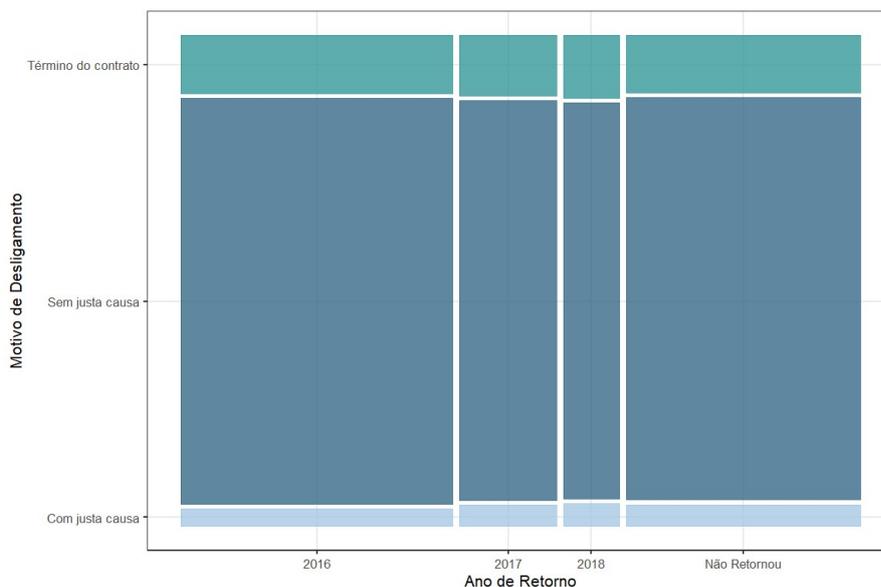
Figura 5: Mosaico da proporção da variável Raça-cor segundo Ano de Retorno, São Paulo (SP) - 2015 A 2018



Fonte: Ministério da Economia, Relação Anual de Informações Sociais (RAIS) - 2015 a 2018.

Dos motivos de desligamento estudados, 83,7% foram vínculos desligados sem justa causa pelo empregador e apenas 4,1% foram desligados com justa causa pelo empregador. O segundo motivo mais registrado, foi o término de contrato de trabalho com 12,2%. Destas categorias, as pessoas desligadas sem justa causa e por término do contrato, a maioria delas 41,2% e 40,9%, retornaram em 2016. Porém, para os indivíduos desligados com justa causa, 38,6%, seu maior percentual não retornaram ao mercado de trabalho.

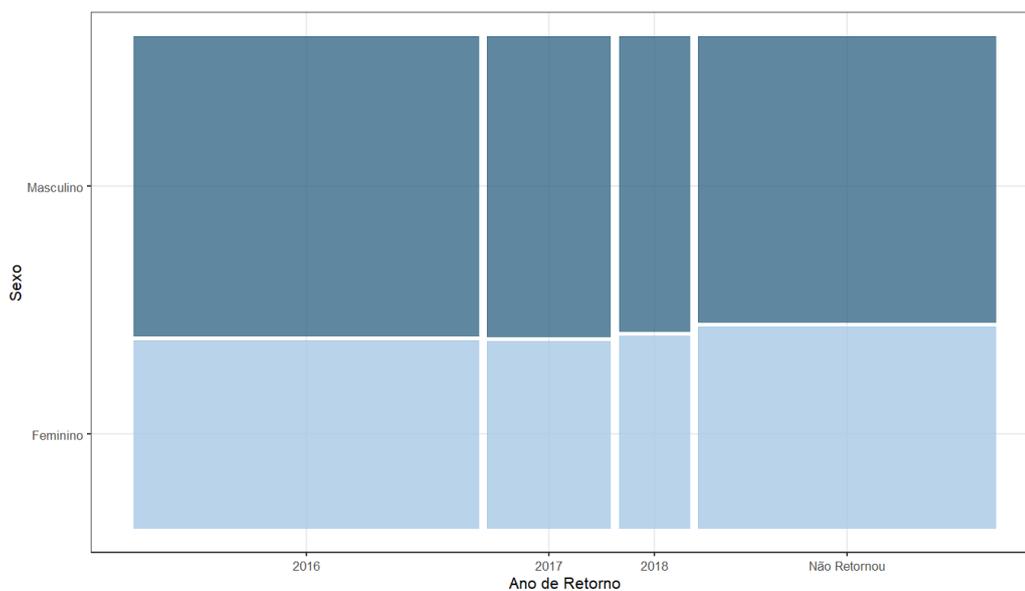
Figura 6: Mosaico da proporção da variável Motivo de Desligamento segundo Ano de Retorno, São Paulo (SP) - 2015 A 2018



Fonte: Ministério da Economia, Relação Anual de Informações Sociais (RAIS) - 2015 a 2018.

Dos dados em análise, 60,4% foram registrados com sexo masculino, ou seja, os vínculos de trabalho são em sua maior parte assumidos por homens. O sexo feminino representam 39,6% dos vínculos estudados. Em ambos os sexos, os maiores percentuais de retorno se encontram no ano de 2016, assumindo 40,1% e 41,9% para os sexos feminino e masculino, respectivamente.

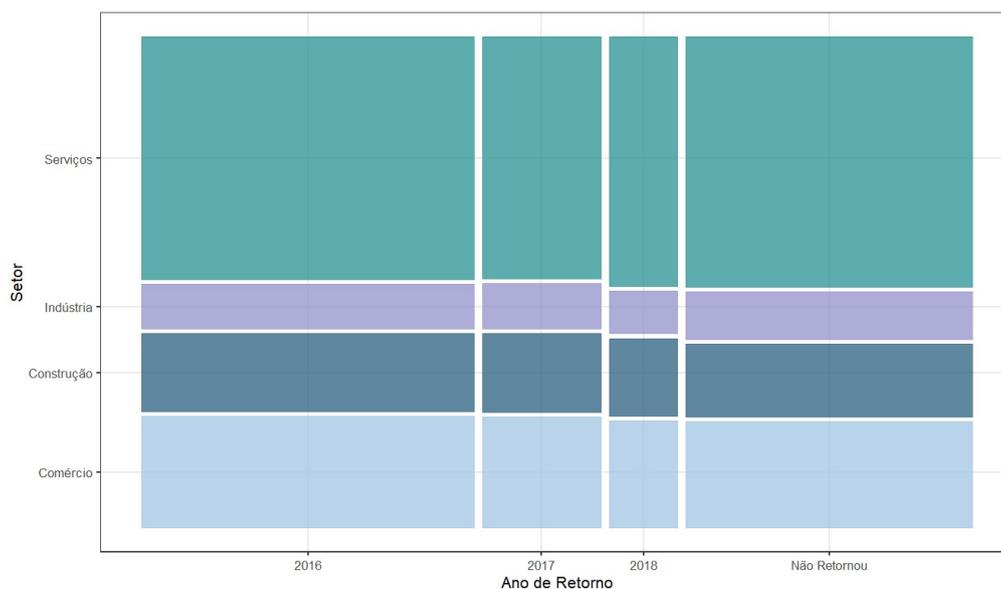
Figura 7: Mosaico da proporção da variável Sexo segundo Ano de Retorno, São Paulo (SP) - 2015 A 2018



Fonte: Ministério da Economia, Relação Anual de Informações Sociais (RAIS) - 2015 a 2018.

Entre os setores apresentados, o setor de serviços obteve a maior proporção das ocupações entre os desligados em 2015, 51,6%. Os segundo maior setor econômico em desligamento foi o comércio representando 22,8%. Para os quatro setores considerados, todos têm seu maior percentual de retorno no ano de 2016.

Figura 8: Mosaico da proporção da variável Setor segundo Ano de Retorno, São Paulo (SP) - 2015 A 2018



Fonte: Ministério da Economia, Relação Anual de Informações Sociais (RAIS) - 2015 a 2018.

Na Tabela 4 apresentada abaixo, são expostas algumas medidas descritivas para as variáveis quantitativas contínuas consideradas na análise.

Tabela 5: Medidas descritivas das variáveis Remuneração Média, Idade e Tempo de emprego segundo Ano de Retorno, São Paulo (SP) - 2015 a 2018

Ano de Retorno	Medidas Descritivas				
	Média	Mediana	1º Quartil	3º Quartil	Coefficiente de Variação (%)
Remuneração Média (S.M.)					
2016	2,6	1,8	1,4	2,5	134,5
2017	2,6	1,7	1,4	2,5	145,6
2018	2,7	1,7	1,4	2,5	149,1
Não Retornou	3,2	1,7	1,4	2,7	181,5
Idade					
2016	32	31	25	38	28,8
2017	32	30	25	38	29,5
2018	32	31	25	38	29,6
Não Retornou	36	34	27	43	30,4
Tempo Emprego (em meses)					
2016	22,1	12,1	4,8	27,6	131,0
2017	21,9	11,7	4,7	27,0	136,6
2018	21,8	11,4	4,5	26,3	140,9
Não Retornou	29,5	13,9	5,1	34,3	155,0

Fonte: Ministério da Economia, Relação Anual de Informações Sociais (RAIS) - 2015 a 2018.

Nos quartis de remuneração média em salários mínimos vistos na Tabela 3, o primeiro quartil aponta que 25,5% das menores remunerações ficam abaixo de 1,4 salários mínimos. Em todos os quatro quartis, a maioria dos indivíduos retornou ao mercado de trabalho em 2016, e seus segundos maiores percentuais foram para indivíduos que não retornaram.

Na Tabela 4, a maior média e 3º quartil de remuneração são de indivíduos que não retornaram ao mercado, correspondendo a 3,2 salários mínimos e 2,7 salários mínimos, respectivamente. Porém, a maior mediana de 1,8 salários pertence aos que retornaram ao mercado em 2016. Os coeficientes de variação em todos os anos de retorno ficaram acima de 100% o que indica alta variabilidade dos dados desta variável.

Para os quartis de idade, 27,5% dos indivíduos encontram-se entre 25 a 32 anos representados pelo 2º quartil. Para as pessoas na faixa de idade de 40 a 60 anos, representando 23,9% da base total, sua maioria (46,7%) não retornou ao mercado de trabalho, esse comportamento em que a maioria da classe não retornou não é visto para os outros quartis, para as outras categorias seus maiores percentuais de retorno se encontram no ano de 2016. Um fato importante a se destacar é o comportamento de alta do percentual de não retorno ao avançar das idades em cada quartil.

A categoria de não retorno apresentou as maiores medidas de idade em todas as estatística calculadas, 36 anos para média, 34 para mediana, 27 para o 1º quartil, 43 para o 3º quartil e 30,4% de coeficiente de variação, enquanto que, para as outras categorias as

medidas são similares exceto para mediana e coeficiente de variação, mas mesmo assim se mantiveram em valores próximos.

O tempo de emprego, apresenta seu maior percentual de 25,4% em indivíduos que estão no 1º quartil, de 1 a 4 meses e 28 dias. Para todos os três primeiros quartis, ou seja, de 1 a 29 meses e 15 dias, tem seus percentuais de retorno em maior parte no ano de 2016 mas para tempos de permanência no quarto quartil, a maior parte das pessoas, 41,0%, não retornou ao mercado de trabalho.

Assim como na tabela de frequência, na Tabela 4 para a categoria de não retorno, as medidas foram maiores, ou seja, para pessoas que tiveram mais tempo de permanência no vínculo desligado em 2015. Contudo, é importante destacar que as pessoas que retornaram em 2016 tiveram média 22,1, mediana 12,1, 1º quartil 4,8 e 3º quartil 27,6 todas sendo as segundas maiores estatísticas das categorias apresentadas. A variabilidade desses dados também é considerada alta, como pode ser observado pelos coeficientes de variação expostos.

3.2 Análise de Regressão Logística Politômica Ordinal

Como a base possui mais de um milhão de observações, seu tamanho deixaria a modelagem mais sensível a significância. Optou-se por extrair uma amostra aleatória simples de tamanho arbitrário correspondente a aproximadamente a 1% com o intuito de avaliar a significância das variáveis analisadas. A categoria de referência da variável resposta é a categoria "Não retorno", devido a complexidade existente por trás de todos os fatores que desencadeiam esta condição, torna-se relevante comparar as chances do indivíduo pertencer a esta categoria comparado as outras.

O modelo utilizado leva em consideração a ordenação cumulativa da variável resposta, ou seja, 1 para retorno em até um ano após o desligamento em 2015, 2 para retorno em 2017, 3 para retorno em 2018 e 4 para caso de não retorno.

O modelo selecionado pelo método de seleção de variáveis Stepwise contém as variáveis 'Raça-cor', 'Porte Estabelecimento', 'Sexo', 'Idade', 'Remuneração média (em S.M.)' e 'Tempo de emprego (em meses)' este modelo obteve o Critério de Informação de Akaike (AIC) de 24.591,43 e foi o mais parcimonioso dentre todos os testados. Devido a necessidade de reduzir o número de categorias consideradas, as categorias da variável 'Porte Estabelecimento' foram reagrupadas em 'Grande' os médios e grandes estabelecimentos e em 'Pequeno' os micro e pequenos estabelecimentos.

Os testes de ajuste global do modelo, ou seja, testes em que é testado o modelo sem preditores vs modelo como um todo para verificar se o ajuste foi significativamente melhor, resultaram em altamente significativo na análise das quatro categorias da variável resposta.

Tabela 6: Testes de Ajuste Global

Teste	Qui-quadrado	Graus de Liberdade	P-valor
Razão de Verossimilhança	295,2	6	< 2,2e-16
Wald	284,4	6	< 2,2e-16

Como a categoria de referência considerada foi a de não retorno, três modelos cumulativos foram testados em conjunto compondo a regressão logística politômica ordinal. O primeiro trata-se da probabilidade de retorno no ano seguinte após o desligamento em 2015 ($[P(Y \leq 1)]$), o segundo é a probabilidade de retorno até o segundo ano ($[P(Y \leq 2)]$) e o terceiro se trata da probabilidade de retorno até o terceiro ano ($[P(Y \leq 3)]$), ou seja, um cumulativo das probabilidade de retorno ao longo dos três anos avaliados dada a categoria de não retorno.

As estimativas dos coeficientes para os parâmetros são apresentados na Tabela 6.

Foram estimados 9 parâmetros, as expressões dos modelos para as próprias probabilidades cumulativas são:

$$P(Y \leq 1) = \frac{\exp(\alpha_1 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_6 x_6)}{1 + \exp(\alpha_1 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_6 x_6)} \quad (1)$$

$$P(Y \leq 2) = \frac{\exp(\alpha_2 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_6 x_6)}{1 + \exp(\alpha_2 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_6 x_6)} \quad (2)$$

$$P(Y \leq 3) = \frac{\exp(\alpha_3 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_6 x_6)}{1 + \exp(\alpha_3 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_6 x_6)} \quad (3)$$

onde os α_j , com $j = 1, 2, 3$ são os interceptos de cada probabilidade acumulada e os β_p , com $p = 1, 2, \dots, 6$ são os coeficientes das variáveis explicativas.

E as probabilidades das categorias de resposta individuais são dadas pelas diferenças entre elas:

$$P(Y = 1) = P(Y \leq 1) \quad (4)$$

$$P(Y = 2) = P(Y \leq 2) - P(Y \leq 1) \quad (5)$$

$$P(Y = 3) = P(Y \leq 3) - P(Y \leq 2). \quad (6)$$

Abaixo, estão listados os resultados do teste de significância para cada preditor utilizado na modelagem.

Tabela 7: Análise de significância

Variáveis	Coefficiente	Graus de Liberdade	Wald	p-valor
Intercepto 1	-1,58	1	441,8	< 2,2e-16
Intercepto 2	0,27	1	13,6	0,00023
Intercepto 3	1,49	1	365,1	< 2,2e-16
Raça-cor	0,09	1	5,7	0,017
Porte Estabelecimento	-0,13	1	12,5	0,000
Sexo	-0,12	1	9,6	0,002
Idade	-0,02	1	140,9	< 2,2e-16
Remuneração Média (em S.M.)	-0,01	1	8,3	0,004
Tempo de emprego (em meses)	0,00	1	29,9	0,000

De acordo com os resultados expostos, a um nível de significância de 5%, temos evidências estatísticas para rejeitar a hipótese nula de não significância das variáveis explicativas do modelo.

Portanto, como o modelo está bem ajustado, as estimativas de razões de chances (*odds ratio*) são pertinentes para expor os efeitos dos preditores sobre a variável resposta vistas na Tabela 7.

Tabela 8: Estimativas das razões de chances e respectivos intervalos de confiança (95%)

Variáveis	Efeito	<i>Odds Ratio</i> (e^{β_j})	Limites de Confiança	Erro Padrão
Raça-cor	Branca e Amarela x Preta e Parda	1,09	[1,016; 1,179]	0,038
Porte Estabelecimento	Grande x Pequeno	0,88	[0,815; 0,943]	0,037
Sexo	Masculino x Feminino	0,89	[0,825; 0,957]	0,038
Idade	-	0,98	[0,973; 0,980]	0,002
Remuneração Média (em S.M.)	-	0,99	[0,977; 0,995]	0,005
Tempo de emprego (em meses)	-	0,99	[0,995; 0,997]	0,001

As chances estimadas de retorno de uma pessoa de raça-cor preta e parda esteja na direção do não retorno em vez de na direção de retorno em 2016, 2017 e 2018 é igual a 1,09 vezes as chances estimadas para branca e amarela, mantendo constante todas as demais variáveis, isto é, chances na direção de não retorno são cerca de 9% superior para pessoas pretas e pardas em relação a brancas e amarelas.

Para porte de estabelecimento, as chances de retorno para um indivíduo desligado de empresa de pequeno porte esteja na direção de não retorno em contraposição ao retorno em um dos 3 anos estudados é 0,87 vezes as chances estimadas para indivíduos desligados de estabelecimentos de grande porte, mantendo constante todas as outras variáveis. As chances em direção ao não retorno são cerca de 12% inferior para indivíduos desligados em empresas de pequeno porte em comparação a pessoas desligadas em estabelecimentos de grande porte.

Já para a variável sexo, o sexo feminino tem chance estimada de ano de retorno na direção de não retornar ao mercado de trabalho ao invés de retornar em algum dos três anos 0,88 vezes as chances estimadas para pessoas do sexo masculino mantendo constante as outras variáveis, isto significa cerca de 11% inferior as chances na direção de não retorno comparadas ao sexo oposto.

Como as variáveis 'Idade', 'Remuneração média (em S.M.)' e 'Tempo de emprego (em meses)' são variáveis quantitativas contínuas, a exponencial de seus parâmetros β terão um efeito multiplicativo na chance da probabilidade da variável resposta para cada uma unidade de aumento em x .

O aumento de um ano de idade resulta em 2,3% de diminuição na probabilidade em direção ao não retorno, mantendo constante todas as demais variáveis.

Em remuneração, o aumento de uma unidade na variável resulta em 12,4% de redução na probabilidade em direção ao não retorno quando mantido constante as outras

variáveis.

O acréscimo de um mês no tempo de emprego, resulta em diminuição de 0,3% de contração na probabilidade na direção do não retorno, considerando constante as demais variáveis.

4 Conclusão

Os resultados deste estudo mostraram que 64,5% dos desligados em 2015, retornaram a um vínculo formal em até três anos, 41,2% no primeiro ano e 22,9% nos anos seguintes. Os outros 35,6% não retornaram ao mercado de trabalho formal no período analisado.

Com relação aos fatores associados à probabilidade de retorno, observou-se que:

- Trabalhadores do sexo feminino têm menores chances de não retornar a um vínculo formal em até três anos do que trabalhadores do sexo masculino;
- Indivíduos pretos e pardos têm maiores chances de não retornar ao mercado de trabalho dentro do período avaliado do que brancos e amarelos;
- Pessoas desligadas em 2015 de estabelecimentos de grande porte têm menor probabilidade de retorno do que aqueles que foram desligados de estabelecimentos de pequeno porte;
- O acréscimo de uma unidade salarial, mantendo as demais variáveis constantes, diminui as probabilidades em direção ao não retorno;
- O aumento de um ano de idade, reduz a probabilidade em direção ao não retorno quando mantida as demais variáveis constantes;
- Quando acrescido um mês no tempo de emprego, isso resulta em uma diminuição na probabilidade em direção ao não retorno mantendo-se constantes as demais variáveis.

Esses resultados divergem da análise descritiva em alguns pontos, como por exemplo, foi mostrado que após uma certa faixa de idade (40 a 60 anos) e de tempo de permanência no antigo emprego (maior que 29 meses), o percentual de pessoas que não retornam ao mercado de trabalho é maior que o percentual de pessoas que retornam. Isso sugere que a queda na chance de retorno a partir dos valores mencionados, não está sendo refletida no modelo.

A controvérsias também nos resultados descritivos do salário das pessoas que retornam, conforme mostrado na Tabela 2, os percentuais de pessoas que foram admitidas em empregos que recebem remuneração igual, maior ou menor que o vínculo anterior, mais de 50% dos indivíduos retornaram para vínculos em que recebem menos do que recebiam em 2015.

O estudo de caso realizado para o município de São Paulo, trouxe informações relevantes sobre a dinâmica do mercado de trabalho formal e a sua associação às carac-

terísticas dos trabalhadores. Será interessante replicar o presente estudo a outras regiões do país afim de verificar diferenças e similaridades.

5 Referências Bibliográficas

AGRESTI, A. **Introduction to Categorical Data Analysis**. 3rd edition. Florida: John Wiley & Sons, Inc, 2019.

BARBOSA, Rogério Jerônimo (2019): **Estagnação Desigual: desemprego, desalento, informalidade e a distribuição da renda do trabalho no período recente (2012-2019)**, Mercado de trabalho: conjuntura e análise - Artigos, No. 67, Instituto de Pesquisa Econômica Aplicada (IPEA), Brasília, 2019. Disponível em: <http://repositorio.ipea.gov.br/handle/11058/9949>. Acesso em: 1 out. 2021.

BRASIL. **Decreto-lei nº 5.452, de 1 de maio de 1943. Aprova a consolidação das leis do trabalho**. Lex: coletânea de legislação: edição federal, São Paulo, v. 7, 1943.

BRASIL. Instituto Brasileiro de Geografia e Estatística. **Oito municípios detinham 25% do PIB do país em 2018**. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases>. Acesso em: 14 de abril de 2021.

BRASIL. Instituto Brasileiro de Geografia e Estatística. **SIDRA – Pesquisa Nacional por Amostra de Domicílios Contínua Trimestral – Tabela 4099 (Taxas de desocupação e de subutilização da força de trabalho, na semana de referência, das pessoas de 14 anos ou mais de idade)**. Disponível em: <https://sidra.ibge.gov.br/tabela/4099>. Acesso em: 14 de abril de 2021.

BRASIL. MINISTÉRIO DA ECONOMIA – SECRETARIA DO TRABALHO. **Grupamento de Atividade Econômicas**. Disponível em: <ftp://ftp.mtpps.gov.br/pdet/microdados/NOVO%20CAGED/Comunicado%20-%20Grupamento%20de%20Atividades%20Econ%F4micas.pdf>. Acesso em: 13 de outubro de 2021.

BRASIL. MINISTÉRIO DA ECONOMIA – SECRETARIA DO TRABALHO. **RAIS**. Disponível em: <http://www.rais.gov.br/sitio/sobre.jsf>. Acesso em: 14 de março de 2021.

BRASIL. MINISTÉRIO DA EDUCAÇÃO. **Informações Gerais sobre a PNAD – Objetivos da RAIS e do CAGED**. Disponível em: . Acesso em: 14 de março de 2021.

GOMES, Magno Rogério. **Discriminação salarial de gênero e alocação por setor econômico dos trabalhadores no Primeiro emprego, Reemprego e Remanescentes: regiões Sul e Nordeste do Brasil**. 131 f. Dissertação de Mestrado em Economia Regional. - Centro de Estudos Sociais Aplicados, Universidade Estadual de Londrina, Londrina, 2016.

HOSMER, D.W.; LEMESHOW, S. **Applied Logistic Regression**. 2nd edition. New York: John Wiley Sons, 2000.

ORGANIZAÇÃO INTERNACIONAL DO TRABALHO - GENEVA. **National employment policies: A guide for workers' organisations**, 2015. Disponível em: <https://www.ilo.org/wcmsp5/groups/public/---ed_emp/---emp_policy/documents/publication/wcms_334917.pdf>. Acesso em: 10 outubro de 2021.

SEBRAE-NA/ Dieese. **Anuário do trabalho na micro e pequena empresa**, 2013, p. 17. Disponível em: <www.sebrae.com.br/Sebrae/Portal%20Sebrae/Anexos/Anuario%20do%20Trabalho%20Na%20Micro%20e%20Pequena%20Empresa.2013.pdf>. Acesso em: 13 outubro de 2021.

SILVA, Fábio Bruno da. MARTINS, Caroline Miriã Fontes. REZENDE, Vânia Aparecida. **A recessão econômica de 2015 e seus impactos nos requerimentos de seguro desemprego no Brasil**. Revista Interdisciplinar Científica Aplicada, Blumenau, v.13, n.1, p. 45-67, TRI I 2019. ISSN 1980-7031.

SIMOES, Pedro Henrique de Castro; ALVES, José Eustáquio Diniz and SILVA, Pedro Luis do Nascimento. **Transformações e tendências do mercado de trabalho no Brasil entre 2001 e 2015: paradoxo do baixo desemprego?**. Rev. bras. estud. popul. [online]. 2016, vol.33, n.3, pp.541-566. ISSN 0102-3098.

SOUZA CAMPOS, Guilherme Caldas. **O emprego formal no Brasil dos anos 2000: um estudo da Relação Anual de Informações Sociais (2003-2013)**. 1 recurso online (118 p.). Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Economia, Campinas, SP, 2016.