



Universidade de Brasília
Departamento de Estatística

**Determinantes da duração do desemprego no biênio 2021-2022:
Acesso ao emprego formal, informal e inatividade no Brasil segundo um
modelo de riscos competitivos**

Laura Cristina Melo Teixeira

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2023**

Laura Cristina Melo Teixeira

**Determinantes da duração do desemprego no biênio 2021-2022:
Acesso ao emprego formal, informal e inatividade no Brasil segundo um
modelo de riscos competitivos**

Orientador: Prof. Dr. Leandro Tavares Correia

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2023**

Agradecimentos

Agradeço à minha família, em especial à minha mãe, Maria Vilma, por todo o trabalho empenhado ao longo de toda a minha vida. Por toda sua dedicação e amor, sou eternamente grata. Sem você nada disso seria possível.

Aos professores do Departamento de Estatística, quero agradecer a todos na figura do meu orientador, Leandro Correia, que fez apontamentos essenciais para o desenvolvimento deste trabalho. À banca, agradeço na figura do professor Eduardo Yoshio Nakano, que se mostrou muito solícito, dando sugestões que enriqueceram o conteúdo do trabalho.

Não posso deixar de agradecer aos amigos que fiz durante esses anos de graduação. O apoio de vocês tem um valor imensurável para mim. Quero deixar aqui o meu agradecimento a todos, na figura de algumas pessoas muito queridas: Davi Guerra Alves, Jéssica Vasconcelos e Tamara Rodrigues.

Ao meu companheiro, Marcelo Fleury, agradeço por todos os anos de graduação que estive ao seu lado. Sou grata não apenas pelas contribuições que você deu a este trabalho, mas por todo o incentivo para que eu sempre buscasse fazer o meu melhor.

À ESTAT Consultoria, agradeço por ter sido o ambiente em que conheci pessoas maravilhosas, algumas das quais futuramente se tornariam fundadoras, junto a mim, da Tukey Data. E, posteriormente, fundadoras da Associação Democracia Beta. Por todos esses anos de trabalho em conjunto, pelo apoio e compreensão, agradeço a todos na figura de Gabriel Reis Nogueira.

Resumo

O presente trabalho buscou realizar uma análise da recente duração do desemprego (até a transição para o emprego formal, informal ou inatividade) no Brasil, no biênio 2021-2022. Foram utilizados dados da PNAD Contínua como base para este estudo. Investigou-se a relação entre a duração do desemprego e um conjunto de características do indivíduo. Assim, foi ajustado o modelo de regressão log-normal em análise de sobrevivência, avaliado com ou sem a presença de riscos competitivos.

O tempo mediano até o ingresso no emprego formal ou informal para as mulheres foi mais que o dobro do tempo dos homens. Pessoas não brancas apresentaram maior exposição ao emprego informal, com tempo até ingresso na ocupação 8% menor que de pessoas brancas. Os grupos com maior propensão a ingressar na inatividade foram: indivíduos de 14 a 20 anos, com poucos anos de estudo e residentes das regiões Nordeste e Sudeste.

Palavras-chaves: análise de sobrevivência, duração do desemprego, riscos competitivos, log-normal, PNAD Contínua.

Lista de Tabelas

1	Frequência das covariáveis	38
2	Frequência de saída do desemprego por covariáveis	39
3	Estimativas dos coeficientes do modelo de regressão log-normal	48
4	Estimativas dos coeficientes do modelo de regressão log-normal em riscos competitivos	50
5	Frequência das categorias de posição no domicílio agrupadas como “Outros”	61
6	Intervalos de confiança (95%) dos coeficientes estimados no modelo de regressão log-normal em riscos competitivos	62

Lista de Quadros

1	Descrição dos indicadores que compõem a dimensão IVS infraestrutura urbana e seus respectivos pesos	27
2	Descrição das variáveis utilizadas na criação de chaves dos indivíduos na PNADC	29
3	<i>Variance inflation factor</i> (VIF) entre as covariáveis	33
4	Teste Qui-Quadrado de independência entre as covariáveis e a causa de saída do desemprego	40

Lista de Figuras

1	Gráfico ilustrativo de possíveis funções de risco acumuladas	21
2	Fluxograma das possíveis transições do desemprego.	28
3	Fluxograma da construção do banco de dados	30
4	Curvas de sobrevivência por saída do desemprego estimadas segundo ano .	32
5	Curvas de sobrevivência por saída do desemprego estimadas segundo IVS infraestrutura urbana	33
6	Fluxograma do processo de modelagem dos dados	35
7	Curva de sobrevivência por saída do desemprego estimada	41
8	Curvas de sobrevivência por saída do desemprego estimadas conforme co- variáveis (sexo, raça/cor e posição no domicílio)	43
9	Curvas de sobrevivência por saída do desemprego estimadas conforme es- colaridade	44
10	Função de risco acumulada estimada - $\hat{H}(t)$	44
11	Curvas de sobrevivência estimadas pelos modelos paramétricos <i>versus</i> es- timação por Kaplan-Meier	45
12	Resíduos de Cox-Snell por modelo sem covariáveis	45
13	Resíduos de Cox-Snell e resíduos padronizados do modelo de regressão log- normal	49
14	Resíduos de Cox-Snell e resíduos padronizados do modelo de regressão log- normal em riscos competitivos	52
15	Probabilidades de ingresso no emprego formal, informal ou na inatividade por perfil demográfico com intervalos de confiança 95%	54

Sumário

1 Introdução	9
2 Referencial Teórico	11
2.1 Análise de Sobrevivência	11
2.1.1 Dados censurados	12
2.1.2 Estimaco no-paramtrica	13
2.1.3 Estimaco paramtrica	13
2.1.4 Modelos paramtricos usuais	14
2.2 Modelos de regresso em sobrevivncia	16
2.2.1 Modelo de riscos competitivos	17
2.3 Escolha do modelo paramtrico	20
2.4 Anlise de resduos	22
3 Metodologia	25
3.1 Conjunto de dados - PNAD Contnua	25
3.2 Definindo a sada do desemprego	28
3.3 Construo do banco de dados	29
3.4 Seleo de covariveis	31
3.5 Modelagem dos dados	33
4 Resultados	37
4.1 Anlise descritiva	37
4.1.1 Curvas de sobrevivncia no-paramtricas	41
4.2 Escolha do modelo paramtrico	44
4.3 Modelo paramtrico com covariveis sem riscos competitivos	46
4.4 Modelo paramtrico com covariveis em riscos competitivos	49
4.4.1 Probabilidades de sada do desemprego por perfil sociodemogrfico	52
4.5 Limitaes	54
5 Concluso	57

Referências	59
Apêndice	61
A Categoria “Outros” da variável posição no domicílio	61
B Intervalos de confiança dos coeficientes estimados no modelo de regressão log-normal em riscos competitivos	62

1 Introdução

O desemprego de longa duração, que se refere àqueles que estão buscando colocação profissional há mais de dois anos, tem tomado altas proporções no Brasil. Segundo dados da PNAD¹, em 2021, durante o período de recuperação do mercado de trabalho diante da pandemia de covid-19, cerca de 3,6 milhões de pessoas se encontravam nessa situação. Como resultado, a questão do desemprego vem conquistando cada vez mais um papel notável no debate econômico nacional.

Além da taxa de desemprego, amplamente divulgada pelos meios de comunicação e definida como a relação entre o número de pessoas desempregadas e a População Economicamente Ativa (PEA), existem outros conceitos fundamentais para entender a estrutura do desemprego (BARROS; CAMARGO; MENDONÇA, 1997). Entre esses conceitos, como a magnitude do desemprego e a probabilidade de entrada ou saída do desemprego, encontra-se a duração média do desemprego. Enquanto a taxa de desemprego em algum grupo social exprime a extensão do fenômeno de desemprego nesse grupo, a duração média dessa busca por trabalho fornece uma medida da intensidade do desemprego.

Conforme destacado por Barros, Camargo e Mendonça (1997), uma baixa taxa de desemprego em determinado subgrupo da economia pode, no entanto, estar associada a um longo tempo de busca por emprego. Esse período prolongado de procura tem um impacto na qualidade de vida do indivíduo e, conseqüentemente, agrava a situação de pobreza.

Portanto, compreender como esse passivo social está distribuído na sociedade é, em última instância, compreender em que grupos a inclusão social encontra barreiras, favoráveis ao aumento das desigualdades e da pobreza. Assim, esse tipo de estudo toma a função de alicerce para a formulação e implementação de políticas públicas voltadas à inserção no mercado de trabalho.

Júnior e Oliveira (2009) identificaram, por meio da Pesquisa Mensal de Emprego (PME - IBGE) de 2006, que o tempo médio de duração do desemprego era maior em alguns subgrupos. Em especial para as mulheres, que apresentaram um risco de saída do desemprego 47% menor em comparação aos homens.

Outros estudos também investigaram os determinantes do tempo de desemprego no Brasil, como Menezes-Filho e Picchetti (2000), que utilizaram os dados da PME para a região metropolitana de São Paulo em 1997. Algumas das características que se mostraram

¹Pesquisa Nacional por Amostra de Domicílios Contínua - IBGE.

relevantes no aumento da duração média do desemprego foram um nível educacional mais elevado e uma idade mais avançada do indivíduo.

Além disso, Reis e Aguas (2014) analisaram a duração do desemprego até a transição para o emprego formal, informal ou para inatividade, no período entre 2006 e 2013, também por meio da PME. Dentre os resultados, destaca-se que as mulheres apresentaram menor probabilidade de transição para empregos formais e informais, assim como maior probabilidade de ingressar na inatividade.

Mais recentemente, Wroblevski e Cunha (2022) também analisaram os determinantes do tempo de desemprego até o emprego formal, informal ou inatividade. Por meio dos dados da PNAD, entre 2012 e 2019, identificou-se que mulheres, não brancos e menos escolarizados formavam um grupo em desvantagem.

Tendo isso em vista, o presente trabalho busca realizar uma análise da recente duração do desemprego (até a transição para o emprego formal, informal ou inatividade) no Brasil nos anos de 2021 e 2022. Por meio de modelagem estatística em análise de sobrevivência, utilizando regressão log-normal com e sem riscos competitivos, serão diagnosticados os determinantes do tempo de desemprego e comparadas ambas as metodologias. Para alcançar esses objetivos, foram utilizados dados da Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD), uma vez que a PME foi encerrada em 2016.

A estrutura deste relatório consiste em cinco seções principais. O Referencial Teórico, Seção 2, apresenta os conceitos fundamentais de análise de sobrevivência e modelos de regressão em sobrevivência. A Seção 3, Metodologia, descreve em detalhes o conjunto de dados utilizado, bem como as técnicas de modelagem aplicadas. Na Seção 4, os resultados obtidos são apresentados, além de uma breve exposição das limitações do trabalho. Por fim, na Seção 5 são apresentadas as principais conclusões e discussões.

2 Referencial Teórico

2.1 Análise de Sobrevivência

Como descrito em Cox e Oakes (1984), análise de sobrevivência é um conjunto de técnicas centradas em compreender uma variável de estudo (T), definida como o tempo até a ocorrência de um evento de interesse. Utilizada em diversos campos, desde pesquisas médicas até problemas de caráter industrial, a análise de sobrevivência ganha destaque na área econométrica a partir de estudos como Lancaster (1979), um dos primeiros a estimar a função de risco para a duração do desemprego. Posteriormente, Kiefer (1988) faz uma ampla avaliação desses modelos de sobrevivência para dados econômicos, ressaltando a utilidade e aplicabilidade da técnica.

Assim, alguns conceitos são centrais para descrever o tempo T , o qual pode se caracterizar como variável contínua ou discreta, sempre não-negativa. O primeiro desses conceitos é a função de sobrevivência, sendo expressa por

$$S(t) = P(T > t) = \int_t^{\infty} f(t)dt, \quad t \geq 0. \quad (2.1.1)$$

Ou, no caso de tempo discreto,

$$S(t) = \sum_{j=t+1}^{\infty} p(j), \quad t = 0, 1, 2, \dots, \quad (2.1.2)$$

em que $p(t) = P(T = t)$ é a função de distribuição de probabilidades. Ou seja, a função de sobrevivência dá a probabilidade do tempo T ser maior que um determinado valor t , analogamente, representa a cauda superior da distribuição de probabilidades a partir do ponto t .

Na análise da duração do desemprego, $S(t)$ representa a probabilidade de ainda estar na situação de desemprego no tempo t . Ou, em outras palavras, $S(t)$ representa a probabilidade de que a saída do desemprego ocorra após o tempo t .

Uma propriedade importante da função de sobrevivência é que a área sob a curva representa o valor da esperança de T . Ou seja, para tempos contínuos:

$$E(T) = \int_0^{\infty} S(t)dt. \quad (2.1.3)$$

Outro conceito de suma importância é a função de risco, definida como

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}, \quad t \geq 0. \quad (2.1.4)$$

Ou, para tempos discretos,

$$h(t) = P(T = t | T \geq t), \quad t = 0, 1, 2, \dots \quad (2.1.5)$$

Assim, $h(t)$ é o risco instantâneo (ou taxa de falha) de ocorrer o evento de interesse no tempo t , dado que o evento não ocorreu antes de t . O comportamento da função de risco tem a seguinte interpretação: seja T a duração do desemprego, uma $h(t)$ crescente (decrecente) significa que com o passar do tempo o risco do indivíduo sair da situação de desemprego aumenta (diminui).

2.1.1 Dados censurados

Uma questão comum na análise de sobrevivência é a presença de dados censurados. Os dados censurados podem ser entendidos como observações incompletas, onde não foi possível mensurar exatamente o tempo t de sobrevivência. Por exemplo, em um estudo com duração pré-determinada de 6 meses, os indivíduos que não apresentarem o evento de interesse até o fim do estudo são observações censuradas. Isso é, sabe-se que $t > 6$ para cada indivíduo censurado, cada um deles certamente “sobrevive” mais que o tempo do estudo, porém não se tem uma medida exata dos tempos.

Além disso, a censura também pode ocorrer de maneira aleatória, como quando um indivíduo decide sair antes do fim de uma pesquisa, ou outro fato externo inviabiliza a mensuração do tempo até o evento de interesse. De todo modo, essas observações não devem ser descartadas, o que levaria a um viés na análise. A inclusão desses, e muitos outros, casos de censura no processo de estimação é possível através da análise de sobrevivência.

No sentido de caracterizar as censuras, temos os tipos de censura à direita, à esquerda e intervalar.

Censura à direita

É a censura em que a falha (evento de interesse) ocorre após o período de observação. Isto é, o tempo é $t > x$, sendo x o tempo final em que o indivíduo ou material foi observado ainda sem falha. Desse modo, os exemplos apresentados no início da Seção 2.1.1 são censuras à direita. Ademais, esse é o tipo de censura presente nos dados deste trabalho.

Censura à esquerda

De forma análoga, a censura à esquerda é aquela em que o evento de interesse ocorre antes do período de observação. Ou seja, sabemos que $t < x$, sendo x o tempo inicial em que o indivíduo ou material foi observado.

Censura intervalar

Já a censura intervalar pode ser vista como uma mescla das duas anteriores. Neste caso, sabe-se que o tempo de falha está entre dois momentos (x, y) , de tal sorte que $y > t > x$.

2.1.2 Estimação não-paramétrica

A estimação não-paramétrica da curva de sobrevivência cumpre dois objetivos principais: a) servir de análise preliminar, indicando distribuições candidatas para uma futura estimação paramétrica; b) ser o único meio de estimação quando nenhuma distribuição conhecida parece se adequar aos dados.

Dessa forma, o estimador Kaplan-Meier para a função de sobrevivência, também chamado de produto-limite, é um estimador não-paramétrico frequentemente utilizado de tal sorte que (KAPLAN; MEIER, 1958)

$$\hat{S}(t) = \prod_{j:t_j \leq t} \left[\frac{n_j - \lambda_j}{n_j} \right] = \prod_{j:t_j \leq t} \left[1 - \frac{\lambda_j}{n_j} \right], \quad (2.1.6)$$

em que n_j é a quantidade de indivíduos sob risco no tempo t_j e λ_j representa a quantidade de indivíduos que experimentou o evento de interesse nesse tempo. É possível provar que (2.1.6) é um estimador de máxima verossimilhança, conforme detalhado em Johansen (1978).

2.1.3 Estimação paramétrica

Já no caso paramétrico, para uma amostra aleatória t_1, t_2, \dots, t_n de tempos de sobrevivência, considere δ a variável indicadora de censura,

$$\delta_i = \begin{cases} 0, & \text{se } t_i \text{ foi censurado} \\ 1, & \text{se } t_i \text{ foi observado} \end{cases} \quad i = 1, 2, \dots, n. \quad (2.1.7)$$

Assumindo que os tempos são independentes e provenientes da mesma distribuição de probabilidades indexada pelo parâmetro θ , e que as censuras são à direita, a função de verossimilhança pode ser escrita na forma (COLOSIMO; GIOLO, 2006):

$$L(\theta) \propto \prod_{i=1}^n [f(t_i; \theta)]^{\delta_i} [S(t_i; \theta)]^{1-\delta_i}. \quad (2.1.8)$$

É possível ver de imediato que cada observação não censurada contribui para a função de verossimilhança (2.1.8) na forma de $f(t_i; \theta)$, ou $p(t_i; \theta)$ no caso discreto. Por outro lado, as observações censuradas contribuem na medida de $S(t_i; \theta)$.

O estimador de máxima verossimilhança de θ é aquele que maximiza o valor de $L(\theta)$. Dentre as distribuições de probabilidade comumente utilizadas para modelar o tempo de sobrevivência, destacam-se a distribuição Weibull (para funções de risco monótonas) e as distribuições log-normal e log-logística (as quais permitem descrever funções de risco unimodais).

2.1.4 Modelos paramétricos usuais

Distribuição Weibull

A distribuição Weibull é uma das mais notórias em análise de sobrevivência. Com aplicação em diversas áreas, sua forma é determinada pela função de densidade de probabilidade (COLOSIMO; GIOLO, 2006):

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp \left\{ - \left(\frac{t}{\alpha} \right)^\gamma \right\}, \quad t \geq 0, \quad (2.1.9)$$

em que $\gamma, \alpha > 0$ são, respectivamente, os parâmetros de forma e escala. Assim, as funções de sobrevivência e risco são dadas, respectivamente, por:

$$S(t) = \exp \left\{ - \left(\frac{t}{\alpha} \right)^\gamma \right\} \quad \text{e} \quad h(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1}. \quad (2.1.10)$$

Uma característica fundamental da distribuição Weibull é que seu risco $h(t)$ é uma função monótona, podendo ser crescente, decrescente ou constante. Quando $\gamma > 1$ o risco é estritamente crescente, para $\gamma < 1$ o risco assume forma estritamente decrescente. Por fim, com $\gamma = 1$ o risco é constante e a distribuição é uma exponencial. Logo, a distribuição exponencial é um caso particular da Weibull.

Distribuição Log-logística

Uma variável aleatória T com distribuição log-logística possui a seguinte densidade de probabilidade (COLOSIMO; GIOLO, 2006):

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} (1 + (t/\alpha)^\gamma)^{-2}, \quad t > 0. \quad (2.1.11)$$

Na expressão, α é parâmetro de forma, enquanto γ é de escala, sendo ambos estritamente positivos. Já as funções de sobrevivência e risco são definidas como:

$$S(t) = \frac{1}{1 + (t/\alpha)^\gamma} \quad \text{e} \quad h(t) = \frac{\gamma(t/\alpha)^{\gamma-1}}{\alpha[1 + (t/\alpha)^\gamma]}. \quad (2.1.12)$$

Diferente da distribuição Weibull, a log-logística abriga funções de risco unimodais. Além disso, como o próprio nome sugere, ao tomar o logaritmo de uma variável T com distribuição log-logística, $Y = \log(T)$ segue uma distribuição logística, cuja densidade é dada por (COLOSIMO; GIOLO, 2006):

$$f(y) = \frac{1}{\sigma} \exp\left\{\frac{y - \mu}{\sigma}\right\} \left(1 + \exp\left\{\frac{y - \mu}{\sigma}\right\}\right)^{-2}. \quad (2.1.13)$$

Os parâmetros da log-logística se relacionam com a distribuição logística de tal sorte que $\gamma = 1/\sigma$ e $\alpha = \exp\{\mu\}$.

Distribuição Log-normal

A distribuição log-normal, assim como a Weibull e log-logística, é amplamente utilizada para descrever a sobrevivência de indivíduos ou materiais em diversas situações. Sua densidade é dada por (COLOSIMO; GIOLO, 2006):

$$f(t) = \frac{1}{\sqrt{2\pi t\sigma}} \exp\left\{-\frac{1}{2} \left(\frac{\log(t) - \mu}{\sigma}\right)^2\right\}, \quad t > 0, \quad (2.1.14)$$

em que $\mu \in \mathbb{R}$ é a média do logaritmo do tempo de sobrevivência, e $\sigma > 0$ é seu desvio-padrão.

Note que ao tomar o logaritmo de uma variável T com distribuição log-normal, $Y = \log(T)$ segue uma distribuição normal com média μ e desvio-padrão σ . Assim, as funções de sobrevivência e risco não possuem forma explícita, mas são calculadas respectivamente por:

$$S(t) = \Phi\left(\frac{-\log(t) + \mu}{\sigma}\right) \quad \text{e} \quad h(t) = \frac{f(t)}{S(t)}, \quad (2.1.15)$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada da normal padrão. Tal qual ocorre na distribuição log-logística, o risco da log-normal assume forma unimodal, onde normalmente cresce até um certo ponto máximo e, em seguida, decresce, aproximando-se de zero conforme t aumenta.

2.2 Modelos de regressão em sobrevivência

Na presença de um conjunto de covariáveis, surge a necessidade de compreender a relação entre essas variáveis explicativas e o tempo de sobrevivência. Nesse contexto, os modelos de regressão em sobrevivência relacionam $\mathbf{x}' = (1, x_1, \dots, x_p)$, o vetor de covariáveis que caracterizam os indivíduos, com o tempo T (ou indiretamente com $Y = \log(T)$) por meio de um vetor de coeficientes $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$.

Uma das vantagens de modelar o logaritmo do tempo de sobrevivência é a simplificação na expressão do modelo, que passa a assumir forma linear (COLOSIMO; GIOLO, 2006). O modelo de regressão paramétrico para T assumindo distribuição log-normal é discutido a seguir.

Regressão Log-normal

Assumindo T com distribuição log-normal, definida em (2.1.14), seu modelo de regressão aloca o vetor de covariáveis no parâmetro μ , de tal sorte que $\mu(x) = \mathbf{x}'\boldsymbol{\beta}$ (LAWLESS, 2003). Assim, de acordo com Lawless (2003), obtém-se o modelo:

$$Y = \log(T) = \mathbf{x}'\boldsymbol{\beta} + \sigma Z, \quad (2.2.1)$$

em que Z segue a distribuição normal padrão ($Z \sim N(0, 1)$) e $\sigma > 0$ é o parâmetro de escala. Desse modo, a função de sobrevivência do modelo é expressa por:

$$S(t|\mathbf{x}) = \Phi\left(\frac{-\log(t) + \mathbf{x}'\boldsymbol{\beta}}{\sigma}\right). \quad (2.2.2)$$

Outros modelos na forma (2.2.1) podem ser obtidos com diferentes distribuições de T e, conseqüentemente, Z . Para T seguindo distribuição log-logística, por exemplo, Z assume distribuição logística. Esses modelos são chamados de locação-escala ou modelo de tempo de vida acelerado, pois as covariáveis atuam acelerando ou desacelerando o tempo de vida (COLOSIMO; GIOLO, 2006). Sendo uma classe de modelos muito utilizada, essas regressões são úteis principalmente quando o tempo de sobrevivência de diferentes indivíduos pode variar em ordem de magnitude (LAWLESS, 2003).

2.2.1 Modelo de riscos competitivos

Quando o evento de interesse em análise pode ocorrer por m diferentes causas, sendo as m causas eventos mutuamente exclusivos, ou seja, cada indivíduo só pode experimentar uma delas, esses eventos são chamados riscos competitivos (KLEINBAUM; KLEIN, 2012). Os diferentes modelos de riscos competitivos buscam compreender a relação entre um vetor de covariáveis e o tempo até a ocorrência de cada um dos c_1, c_2, \dots, c_m eventos que “competem” pela falha do indivíduo.

Desse modo, os dados podem ser representados pela tríade (y, c_j, δ) , sendo $Y = \min(T_1, \dots, T_m)$ o tempo até ocorrer a primeira falha, tal que T_j é o tempo até a falha pela causa c_j (MAIA, 2018). Já c_j é a causa de falha j , $j = 1, 2, \dots, m$. Por fim, δ é a variável indicadora de censura, definida como:

$$\delta = \begin{cases} 1, & \text{se } Y \text{ foi observado;} \\ 0, & \text{se } Y \text{ foi censurado.} \end{cases} \quad (2.2.3)$$

A seguir são apresentados alguns conceitos básicos em riscos competitivos, por meio das principais funções que descrevem os tempos de sobrevivência.

Descrição do tempo de sobrevivência em riscos competitivos

Seja $T = (T_1, T_2, \dots, T_m)$ o vetor de variáveis aleatórias representando os tempos de falha por cada causa c_j , $j = 1, 2, \dots, m$. No contexto de riscos competitivos, só é possível observar o mínimo dos tempos de falha para cada indivíduo $Y = \min(T_1, \dots, T_m)$ e sua causa de falha c_j , na ausência de censura.

Assim, a função de sobrevivência conjunta é expressa por:

$$S(t_1, t_2, \dots, t_m) = P(T_1 \geq t_1, T_2 \geq t_2, \dots, T_m \geq t_m). \quad (2.2.4)$$

A função de sobrevivência total é definida como:

$$S_Y(t) = P[Y \geq t] = P[\min(T_1, \dots, T_m) \geq t] = P[T_1 \geq t, \dots, T_m \geq t]. \quad (2.2.5)$$

Já S_j é a função marginal de $S(\cdot)$, levando em consideração apenas a causa c_j . Caso os T_j 's sejam independentes, observa-se que $S_Y(t) = S(t, \dots, t) = S_1(t)S_2(t) \dots S_m(t)$. Além disso, para tempos independentes, a sobrevivência conjunta é dada pelo produto

das marginais, ou seja: $S(t_1, t_2, \dots, t_m) = S_1(t_1)S_2(t_2) \dots S_m(t_m)$.

A partir das funções de sobrevivência apresentadas, é possível definir as seguintes funções de risco (NAKANO; RODRIGUES, 2006):

- (i) $h(t) = P[\text{falhar de alguma causa em } (t, t + \Delta t) | \text{sobreviveu de todos } c_j\text{s até } t]$;
- (ii) $h_j(t) = P[\text{falhar devido a causa } c_j \text{ no período } (t, t + \Delta t) \text{ e } c_j \text{ é o único risco atuando no período} | \text{sobreviveu de } c_j \text{ até } t]$;
- (iii) $g_j(t) = P[\text{falhar devido a causa } c_j \text{ no período } (t, t + \Delta t) \text{ e todos os riscos estão atuando no período} | \text{sobreviveu de todos os } c_j\text{s até } t]$.

A função $h_j(t)$ é denominada função de risco específico, ou função de risco “net”, enquanto $g_j(t)$ é a função de risco “crude”. As funções de risco são calculadas por (NAKANO; RODRIGUES, 2006):

$$g_j(t) = \frac{- (\partial S_j(t_1, t_2, \dots, t_m) / \partial t_j)}{S(t, \dots, t)}, \quad t_1 = t_2 = \dots = t_m = t \quad (2.2.6)$$

e

$$h_j(t) = -\frac{d}{dt} \log S_j(t) \quad (2.2.7)$$

e, finalmente,

$$h(t) = -\frac{d}{dt} \log S_Y(t) = \sum_{j=1}^m g_j(t). \quad (2.2.8)$$

A partir da $g_j(t)$, define-se a $G_j(t)$ como uma subfunção de sobrevivência, isto é, a probabilidade de sobreviver à causa c_j na presença de todas as causas:

$$G_j(t) = \exp \left\{ - \int_0^t g_j(u) du \right\}. \quad (2.2.9)$$

Desse modo, a função de sobrevivência total pode ser reescrita como:

$$S_Y(t) = \exp \left[- \int_0^t h(u) du \right] = \exp \left[- \int_0^t \sum_{j=1}^m g_j(u) du \right] = \prod_{j=1}^m \exp \left[- \int_0^t g_j(u) du \right], \quad (2.2.10)$$

logo nota-se a relação

$$S_Y(t) = \prod_{j=1}^m G_j(t). \quad (2.2.11)$$

Quando os tempos T_j 's são independentes, tem-se que os riscos *net* e *crude* são iguais ($g_j(t) = h_j(t)$), logo $G_j(t) = S_j(t)$, e portanto segue que $S_Y(t) = \prod_{j=1}^m S_j(t)$.

Função de Verossimilhança

A função de verossimilhança no contexto de riscos competitivos é dada por (MAIA, 2018):

$$L = \left\{ \prod_{j=1}^m \prod_{i=1}^{D_j} f_j(t_{ij}) \right\} \left\{ \prod_{i=1}^W S_Y(t_{wi}) \right\}, \quad (2.2.12)$$

em que D_j é o número de indivíduos que falharam pela causa c_j , W representa o número de indivíduos que foram censurados. O termo t_{ij} refere-se ao tempo decorrido até a falha ocorrer em relação à causa c_{ij} . Além disso, w_i representa o i -ésimo indivíduo censurado e t_{wi} é o tempo de censura específico desse indivíduo.

Nota-se que cada indivíduo observado contribui na medida da sua função densidade de probabilidade $f_j(t_{ij})$, enquanto os censurados contribuem na medida da sobrevivência total $S_Y(t_{wi})$. A verossimilhança pode ser reescrita como:

$$L = \left\{ \prod_{j=1}^m \prod_{i=1}^{D_j} g_j(t_{ij}) S_Y(t_{ij}) \right\} \left\{ \prod_{i=1}^W S_Y(t_{wi}) \right\}, \quad (2.2.13)$$

e, utilizando a equação (2.2.11),

$$L = \left\{ \prod_{j=1}^m \prod_{i=1}^{D_j} \left[g_j(t_{ij}) \prod_{j=1}^m G_j(t_{ij}) \right] \right\} \left\{ \prod_{i=1}^W \prod_{j=1}^m G_j(t_{ij}) \right\}. \quad (2.2.14)$$

Por fim, os termos podem ser rearranjados de tal sorte que:

$$L = \prod_{j=1}^m \prod_{i=1}^{D_j} g_j(t_{ij}) \prod_{l=1}^n G_j(t_l) \quad (2.2.15)$$

e

$$L = \prod_{j=1}^m \left\{ \prod_{i=1}^{D_j} [g_j(t_{ij})]^{\delta_{ij}} \right\} \prod_{l=1}^n G_j(t_l) = \prod_{j=1}^m \prod_{i=1}^n [g_j(t_{ij})]^{\delta_{ij}} G_j(t_i). \quad (2.2.16)$$

Todo o desenvolvimento apresentado não tem como pressuposto tempos independentes. Entretanto, para T_j 's independentes é possível reescrever (2.2.16) de tal forma que as diferentes $j = 1, \dots, m$ causas podem ter sua verossimilhança maximizada separadamente. Para tanto, dado uma causa de falha j , basta considerar os tempos associados às

demais causas como censuras e prosseguir a estimação dos parâmetros normalmente. A maximização é usualmente feita por meio de procedimentos numéricos e está implementada em diversos pacotes estatísticos.

Regressão Log-normal em riscos competitivos

Para T_1 , T_2 e T_3 , tempos de sobrevivência independentes, cada um seguindo distribuição log-normal $T_j \sim \text{log-normal}(\mu_j, \sigma_j)$, a função de sobrevivência conjunta é expressa por:

$$S(t_1, t_2, t_3) = S(t_1)S(t_2)S(t_3) = \Phi\left(\frac{-\log(t_1) + \mu_1}{\sigma_1}\right) \Phi\left(\frac{-\log(t_2) + \mu_2}{\sigma_2}\right) \Phi\left(\frac{-\log(t_3) + \mu_3}{\sigma_3}\right), \quad (2.2.17)$$

com $t_1, t_2, t_3 > 0$. No modelo de regressão log-normal, na presença de riscos competitivos, o vetor de covariáveis $\mathbf{x}' = (1, x_1, \dots, x_p)$ será indexado por meio do parâmetro μ , de tal sorte que

$$\mu_j = \mathbf{x}'\boldsymbol{\beta}_j, \quad j = 1, 2, 3. \quad (2.2.18)$$

Desse modo, a função de sobrevivência conjunta do modelo e a função de sobrevivência total são, respectivamente:

$$S(t_1, t_2, t_3) = S(t_1)S(t_2)S(t_3) = \Phi\left(\frac{-\log(t_1) + \mathbf{x}'\boldsymbol{\beta}_1}{\sigma_1}\right) \Phi\left(\frac{-\log(t_2) + \mathbf{x}'\boldsymbol{\beta}_2}{\sigma_2}\right) \Phi\left(\frac{-\log(t_3) + \mathbf{x}'\boldsymbol{\beta}_3}{\sigma_3}\right) \quad (2.2.19)$$

e

$$S_Y(t) = \Phi\left(\frac{-\log(t) + \mathbf{x}'\boldsymbol{\beta}_1}{\sigma_1}\right) \Phi\left(\frac{-\log(t) + \mathbf{x}'\boldsymbol{\beta}_2}{\sigma_2}\right) \Phi\left(\frac{-\log(t) + \mathbf{x}'\boldsymbol{\beta}_3}{\sigma_3}\right). \quad (2.2.20)$$

2.3 Escolha do modelo paramétrico

A escolha da distribuição de probabilidades, no contexto da modelagem paramétrica dos tempos de sobrevivência, é uma etapa que pode ser guiada por técnicas gráficas. Dentre as possibilidades, a análise da curva de risco acumulada estimada $\hat{H}(t)$ dá uma ideia da natureza da $H(t)$ populacional, a qual é definida como (NAKANO, 2017)

$$H(t) = \int_0^t h(u)du, \quad (2.3.1)$$

no caso contínuo, ou, no caso discreto, por

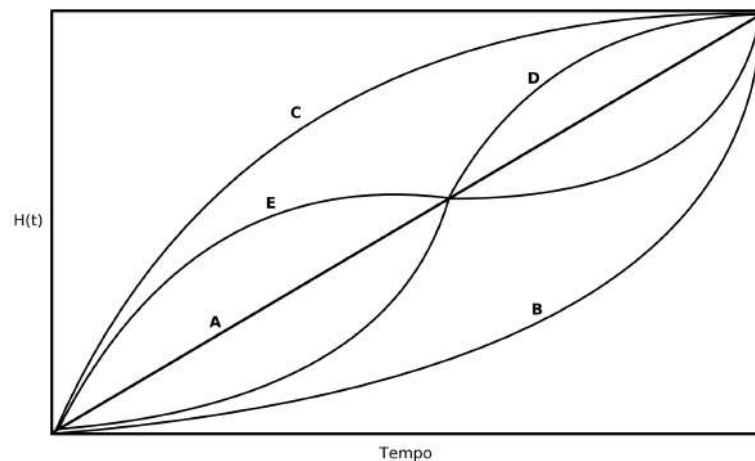
$$H(t) = \sum_{k=0}^t h(k). \quad (2.3.2)$$

A $\hat{H}(t)$ pode ser obtida pelo estimador de Kaplan-Meier (KAPLAN; MEIER, 1958), utilizando a relação $H(t) = -\log S(t)$, válida para tempos contínuos. A forma da $\hat{H}(t)$ permite entender o comportamento da função de risco em estudo, sugerindo distribuições candidatas para cada caso. Algumas possíveis formas que a $\hat{H}(t)$ pode assumir estão exemplificadas na Figura 1.

A interpretação do gráfico é tal que (NAKANO, 2017): caso a curva encontrada siga um comportamento linear (A), a função de risco segue uma forma constante, portanto o modelo exponencial é adequado. Já a curva convexa, B, ou côncava, C, indicam uma função de risco monotonicamente crescente ou decrescente, respectivamente. Neste caso, é recomendável o uso da distribuição Weibull, pois é capaz de modelar funções de risco monótonas. Curvas similares ao visto em D, ou seja, convexa e em seguida côncava, sugerem uma função de risco unimodal. Dessa forma, pode-se sugerir o uso das distribuições log-normal ou log-logística.

Por fim, a curva E indica uma função de risco em forma de banheira, as distribuições teóricas possíveis são apresentadas na literatura, apesar de serem complexas (COLOSIMO; GIOLO, 2006). Vale salientar que as distribuições log-normal e log-logística também acomodam funções de risco decrescentes, o que as tornam adequadas também para o caso C.

Figura 1: Gráfico ilustrativo de possíveis funções de risco acumuladas



Fonte: Nakano (2017)

O uso do risco acumulado para escolha do modelo possui uma vantagem teórica em relação à curva TTT (tempo total em teste) (BARLOW; CAMPO, 1975), que é a contribuição de todas as observações, inclusive as censuradas, na construção do gráfico. Assim, em dados com alta proporção de censuras, a curva TTT pode levar a interpretações equivocadas sobre a forma da função de risco, já que as censuras não são consideradas (NAKANO, 2017).

2.4 Análise de resíduos

Após o ajuste do modelo de sobrevivência, é de interesse avaliar a qualidade desse ajuste em relação aos dados em estudo. Nesse sentido, as duas técnicas descritas a seguir são úteis na tarefa de investigar o ajuste global do modelo (COLOSIMO; GIOLO, 2006).

Cox-Snell

Os resíduos de Cox-Snell são definidos genericamente como (COLOSIMO; GIOLO, 2006):

$$\hat{e}_i = \hat{\Lambda}(t_i | \mathbf{x}_i), \quad (2.4.1)$$

em que $\hat{\Lambda}(\cdot)$ denota a função de risco acumulada segundo o modelo utilizado. Portanto, para o modelo log-normal, os resíduos são dados por:

$$\hat{e}_i = -\log \left[1 - \Phi \left(\frac{\log(t_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}})}{\hat{\sigma}} \right) \right]. \quad (2.4.2)$$

Caso o modelo ajustado seja adequado, espera-se que \hat{e}_i siga uma distribuição exponencial padrão. Assim, é possível fazer o uso de métodos gráficos para avaliar a adequabilidade do modelo escolhido. Uma das possibilidades é graficar \hat{e}_i versus $\hat{\Lambda}(\hat{e}_i)$, onde, no cenário ideal, obtém-se uma reta de inclinação 1. Isso deve-se ao fato de que $\hat{\Lambda}(\hat{e}_i) = -\log(\hat{S}(\hat{e}_i))$, em que $\hat{S}(\hat{e}_i)$ é a função de sobrevivência dos resíduos estimada via Kaplan-Meier e, para a exponencial padrão, essa estimativa deve se aproximar da curva $S(t) = \exp(-t)$.

Resíduos Padronizados

Ainda no contexto de avaliar se o ajuste global do modelo é satisfatório, os resíduos padronizados são calculados por meio de (COLOSIMO; GIOLO, 2006):

$$\hat{\nu}_i = \frac{\log(t_i) - \mathbf{x}'_i \hat{\beta}}{\hat{\sigma}}. \quad (2.4.3)$$

Os resíduos padronizados seguem distribuições diferentes a depender do modelo paramétrico escolhido. Em especial, se o modelo log-normal ajustado for adequado, os valores $\hat{\nu}_i$ devem ser uma amostra censurada da distribuição normal padrão. Desse modo, tomando $\hat{e}_i^* = \exp(\hat{\nu}_i)$, é possível comparar graficamente as sobrevivências de \hat{e}_i^* obtidas com Kaplan-Meier com as sobrevivências obtidas parametricamente pelo modelo log-normal padrão. Quanto mais o gráfico se aproximar de uma reta de inclinação 1, melhor é o ajuste do modelo aos dados.

3 Metodologia

Com o intuito de identificar os determinantes da duração do desemprego no Brasil, para o biênio 2021-2022, este estudo propõe o uso de dados da Pesquisa Nacional por Amostra de Domicílios Contínua, na sua versão trimestral. A análise dos dados é centrada em técnicas de sobrevivência. Sendo assim, toda a análise foi realizada por meio do *software* livre R (R CORE TEAM, 2013).

3.1 Conjunto de dados - PNAD Contínua

A Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua), implementada desde 2012 em todo território nacional pelo IBGE, é a base de dados utilizada neste trabalho. Atualmente, a PNAD Contínua é a principal fonte de dados e estatísticas públicas em relação a força de trabalho e inserção laboral da população brasileira, além de abordar temas suplementares em determinadas edições, a depender das necessidades conjunturais.

A variável objeto deste estudo, o tempo de duração do desemprego, é coletada em meses por meio do seguinte enunciado: “Até o dia (último dia da semana de referência da pesquisa), fazia quanto tempo que estava sem qualquer trabalho e tentando conseguir trabalho?”, para todos os respondentes não ocupados. Durações de desemprego menores que 1 mês são agrupadas na categoria “Menos de 1 mês”. Já os tempos de busca maiores que 2 anos passam a ser contabilizados apenas anualmente.

Dessa forma, é importante salientar que, segundo IBGE (2022, p. 36): “As pessoas não ocupadas na semana de referência são formadas pelas pessoas que foram classificadas como desocupadas ou fora da força de trabalho nessa semana”. Dessa definição surge a necessidade de aclarar outras igualmente importantes, como a definição de pessoa desocupada:

[...] pessoas sem trabalho em ocupação nessa semana que tomaram alguma providência efetiva para consegui-lo no período de referência de 30 dias, e que estavam disponíveis para assumilo na semana de referência. Consideram-se, também, como desocupadas as pessoas sem trabalho em ocupação na semana de referência que não tomaram providência efetiva para consegui-lo no período de referência de 30 dias porque já o haviam conseguido e iriam começá-lo em menos de quatro meses após o último dia da semana de referência. (IBGE,

2022, p. 36).

Já o trabalho em ocupação é definido, segundo IBGE (2022, p. 34), como:

Trabalho remunerado em dinheiro, produtos, mercadorias ou benefícios (moradia, alimentação, roupas, treinamento etc.) na produção de bens e serviços; ou trabalho sem remuneração direta ao trabalhador, realizado em ajuda à atividade econômica de membro do domicílio ou parente que reside em outro domicílio, que recebe em conjunto a remuneração pelo trabalho.

As pessoas fora da força de trabalho são definidas como (IBGE, 2022, p. 37) “[...] pessoas que não estavam ocupadas nem desocupadas nessa semana”. Casos de pessoas fora da força de trabalho são, por exemplo, estudantes que dedicam seu tempo integralmente aos estudos e não estão em busca de trabalho. Por fim, a procura de trabalho é entendida como:

[...] a tomada de alguma providência efetiva para consegui-lo, ou seja, o contato estabelecido com empregadores; a prestação de concurso; a inscrição em concurso; a consulta a agência de emprego, sindicato ou órgão similar; a resposta a anúncio de emprego; a solicitação de trabalho a parente, amigo, colega ou por meio de anúncio; a tomada de medida para iniciar o próprio negócio mediante a procura de local, equipamento ou outros pré-requisitos; a solicitação de registro ou licença para funcionamento do empreendimento etc. (IBGE, 2022, p. 35).

Desse modo, o tempo de duração do desemprego será analisado segundo as seguintes covariáveis:

1. Ano: ano de referência da pesquisa, mais especificamente o ano em que o indivíduo aparece pela primeira vez na base em procura de emprego. Neste trabalho são analisados os anos de 2021 e 2022²;
2. Grandes Regiões: construído com base na UF do domicílio (Norte, Nordeste, Centro-Oeste, Sudeste e Sul);

²No ano de 2020, em função da pandemia da covid-19, a aplicação da PNAD Contínua enfrentou diversos desafios, acarretando na coleta dos dados via telefone, conforme explicado na nota técnica emitida pelo IBGE em 6 de agosto de 2020. Por conta dessa diferença metodológica, o ano de 2020 não entrou no escopo dessa pesquisa.

3. Sexo: homem, mulher;
4. Idade: são considerados os indivíduos com idade entre 14 e 70 anos;
5. Cor/raça: branca, preta, amarela, parda e indígena. As categorias foram agrupadas em brancos (branca e amarela) e não brancos (preta, parda e indígena);
6. Condição no domicílio: pessoa responsável, cônjuge ou companheiro(a), filho(a) e outros;
7. Escolaridade: variável medida pelos anos de estudo. Os valores possíveis são: sem instrução e menos de 1 ano de estudo, 1 a 4 anos de estudo, 5 a 8 anos de estudo, 9 a 11 anos de estudo, 12 a 15 anos de estudo e 16 anos ou mais de estudo.

Além das variáveis explicativas enumeradas, que estão na base da PNADC, foi incorporada a variável Índice de Vulnerabilidade Social (IVS) dimensão infraestrutura urbana. O IVS é um índice sintético elaborado pelo IPEA, que busca mensurar a vulnerabilidade social em diversos recortes espaciais brasileiros (IPEA – Instituto de Pesquisa Econômica Aplicada, 2018).

Sendo composto por três dimensões, a sua dimensão de infraestrutura urbana busca refletir as condições de mobilidade urbana e acesso ao saneamento básico. Os indicadores que compõem esse subíndice medem a disponibilidade dos serviços de água e esgoto, coleta de lixo e o tempo de deslocamento casa-trabalho (variável *proxy* da mobilidade urbana). Detalhes dos indicadores estão presentes no Quadro 1.

Quadro 1: Descrição dos indicadores que compõem a dimensão IVS infraestrutura urbana e seus respectivos pesos

Dimensão	Indicador	Peso
IVS infraestrutura urbana	Percentual de pessoas em domicílios com abastecimento de água e esgotamento sanitário inadequados	0,3
	Percentual da população que vive em domicílios urbanos sem serviço de coleta de lixo	0,3
	Percentual de pessoas que vivem em domicílios com renda per capita inferior a meio salário mínimo e que gastam mais de uma hora até o trabalho no total de pessoas ocupadas, vulneráveis e que retornam diariamente do trabalho	0,4

Fonte: IPEA – Instituto de Pesquisa Econômica Aplicada (2018)

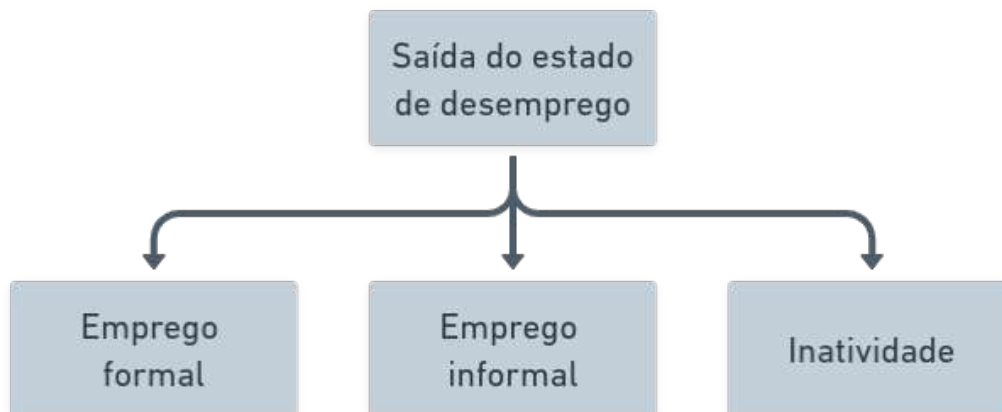
Assim, foi adicionado o IVS infraestrutura urbana do ano de 2021 ao conjunto de covariáveis, obtendo o subíndice para a UF de cada indivíduo presente no conjunto de dados. Com isso, buscou-se acrescentar uma variável explicativa que não fosse relacionada a características do indivíduo (como sexo, idade, raça/cor), e sim ao contexto que ele está inserido.

3.2 Definindo a saída do desemprego

A saída do estado de desemprego pode ocorrer por diferentes razões. Apesar de diversos estudos econométricos tratarem a saída do desemprego como um evento único a ser modelado, essa abordagem apresenta certas fragilidades. Entre elas, o fato de que as variáveis explicativas podem estar relacionadas de diferentes formas com o tempo até cada transição do desemprego. Nesse caso, as estimativas dos coeficientes do modelo se tornam viesadas e, portanto, uma interpretação equivocada do fenômeno é feita.

Tendo isso em vista, este trabalho busca investigar os determinantes da duração do desemprego para cada possível saída do desemprego, ilustradas na Figura 2. Assim, o emprego formal foi caracterizado como aquele onde o indivíduo possui carteira assinada e/ou contribui para instituto de previdência. Caso a ocupação não atenda esses critérios, será considerada como emprego informal. Por outro lado, a inatividade caracteriza a saída do desemprego por conta do fim da procura por trabalho, em outras palavras, o indivíduo sai da força de trabalho.

Figura 2: Fluxograma das possíveis transições do desemprego.



Fonte: Elaboração própria.

Essa escolha metodológica, a mesma adotada em Wroblevski e Cunha (2022), dentro do escopo das técnicas de análise de sobrevivência, acarreta no uso dos modelos de riscos competitivos. Outras escolhas a respeito da modelagem são explicadas em mais detalhes na Seção 3.5.

3.3 Construção do banco de dados

Sabidamente, a PNADC é uma pesquisa que permite o acompanhamento longitudinal dos domicílios amostrados, por meio de algumas variáveis de identificação presentes no banco. Entretanto, o mesmo não vale para os moradores dos domicílios. Um mesmo domicílio pode contar com moradores diferentes ao decorrer das entrevistas, portanto, até o momento, o IBGE não divulga chaves para acompanhamento longitudinal dos indivíduos.

Apesar disso, é possível fazer um pareamento por meio de características demográficas das quais espera-se imutabilidade. Nesse sentido, foram utilizadas a data de nascimento, sexo do respondente e número de ordem (valor fixo que identifica o indivíduo dentro do domicílio) para realizar o pareamento, em conjunto com as variáveis que formam a chave do domicílio.

Desse modo, foi possível criar uma chave identificadora para cada indivíduo e, a partir disso, acompanhar os indivíduos em estado de desemprego longitudinalmente em cada entrevista trimestral da PNADC, entre 2021 e 2022. As variáveis utilizadas na elaboração dessa chave estão descritas em mais detalhes no Quadro 2.

Quadro 2: Descrição das variáveis utilizadas na criação de chaves dos indivíduos na PNADC

Variável	Descrição	Identificador
UPA V1008 V1014	Unidade Primária de Amostragem (UPA) Número de seleção do domicílio Painel	Identifica o domicílio
V2007 V2008/V20081/V20082 V2003	Sexo do respondente Dia, mês e ano de nascimento do respondente Número de ordem do respondente	Identifica o indivíduo

Fonte: Elaboração própria

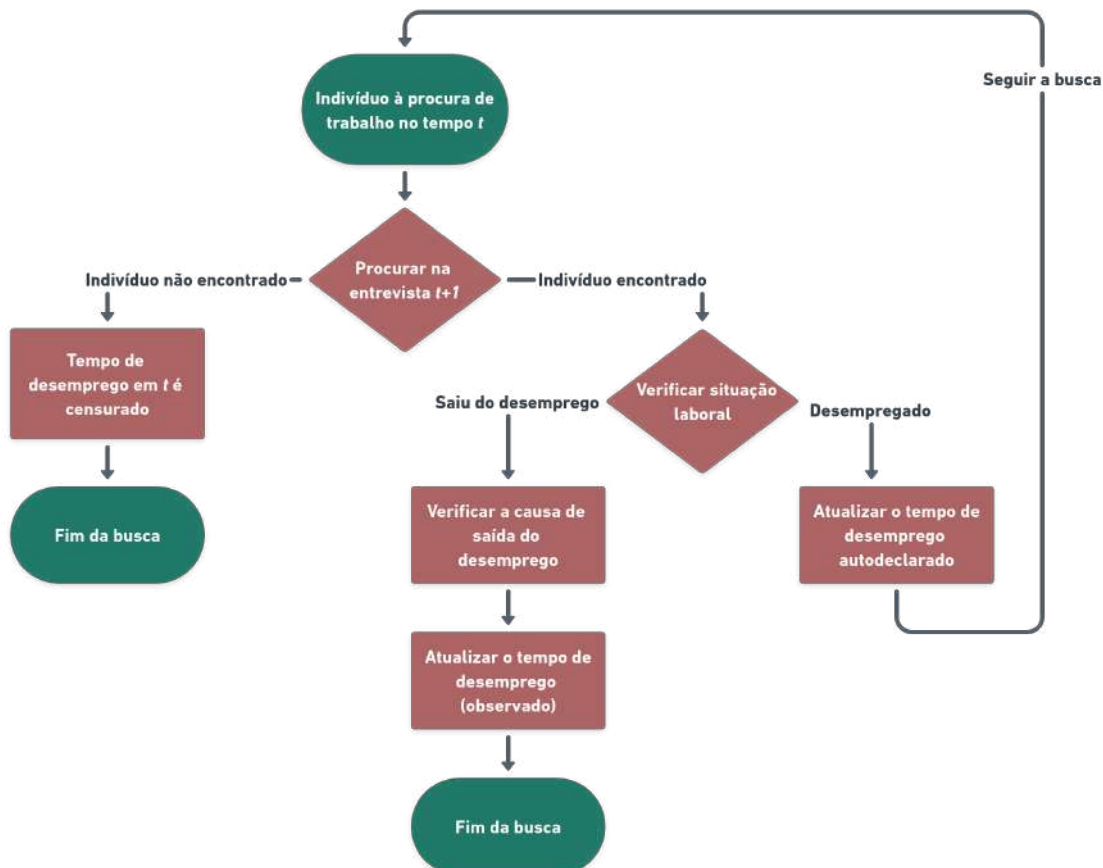
Após a construção das chaves, o banco de dados final foi elaborado por meio de um processo iterativo, que se encontra esquematizado na Figura 3. Para cada indivíduo em busca de trabalho em uma entrevista t , ele será procurado na entrevista seguinte $t + 1$ por meio da sua chave identificadora. Caso o indivíduo não seja encontrado, ele será sinalizado como censurado e seu processo iterativo chega ao fim.

Por outro lado, caso o indivíduo seja encontrado na entrevista $t + 1$, verifica-se qual é sua situação laboral. Constatando-se que o indivíduo saiu do estado de desemprego, busca-se qual foi a causa responsável pela saída (ingresso em emprego formal, informal ou na inatividade) e o seu respectivo tempo de desemprego é atualizado e marcado como observado (não censurado). Nesse ponto o processo iterativo é encerrado.

Alternativamente, é possível que o indivíduo seja encontrado na entrevista $t + 1$

e se constate que ele continua desempregado. Nesse cenário, o tempo de desemprego, em meses, do indivíduo é atualizado conforme sua resposta na entrevista $t + 1$. Assim, o processo iterativo segue acompanhando esse indivíduo nas próximas entrevistas, conforme o algoritmo apresentado, até que seu tempo de desemprego seja censurado ou observado de fato (juntamente com a causa da saída do desemprego).

Figura 3: Fluxograma da construção do banco de dados



Fonte: Elaboração própria

Após repetir o procedimento descrito para todos os indivíduos em busca de emprego, nas 8 entrevistas trimestrais realizadas no biênio 2021-2022, o banco de dados final foi montado. Contando com um total de 86.036 observações, cada registro do banco contém o tempo de desemprego em meses, o indicador de censura e a causa da saída do desemprego, além das covariáveis que caracterizam o indivíduo (sexo, raça/cor, idade, etc.).

Vale ressaltar que, antes de qualquer tipo de manipulação, os dados brutos da PNADC foram submetidos a alguns filtros. Tal limpeza dos dados está descrita a seguir:

- Filtro das idades menores que 14 ou maiores que 70 anos;
- Remoção dos indivíduos com raça/cor “Ignorado”;
- Remoção dos indivíduos com alguma variável não informada.

3.4 Seleção de covariáveis

A seleção de quais variáveis iriam compor o modelo de regressão foi feita levando em consideração alguns critérios. O primeiro deles é que existe um conjunto de variáveis que está presente habitualmente nos estudos econométricos que analisam o tempo de desemprego. Essas variáveis são: sexo, raça/cor, idade, escolaridade, posição no domicílio e região. Tal constatação pode ser verificada em Júnior e Oliveira (2009), Menezes-Filho e Picchetti (2000), Reis e Aguas (2014) e Wroblevski e Cunha (2022).

Tendo isso em vista, o presente trabalho manteve as variáveis supracitadas na análise, a fim de que os resultados encontrados fossem comparáveis com outros estudos da área. De todo modo, a análise descritiva presente na Seção 4.1 deixa claro que essas covariáveis parecem, de fato, ter relação com a duração do desemprego.

Resta decidir sobre a inclusão das variáveis ano e IVS infraestrutura urbana, além de possíveis interações entre variáveis. Nesse sentido, foram testadas algumas estratégias, como a descrita em Colosimo e Giolo (2006), que propõe sucessivas modelagens com diferentes grupos de covariáveis, seguido dos respectivos testes de razão de verossimilhanças (TRV). Entretanto, o alto número de observações ($n = 86.036$) da amostra se mostrou um impeditivo para técnicas baseadas em testes de significância.

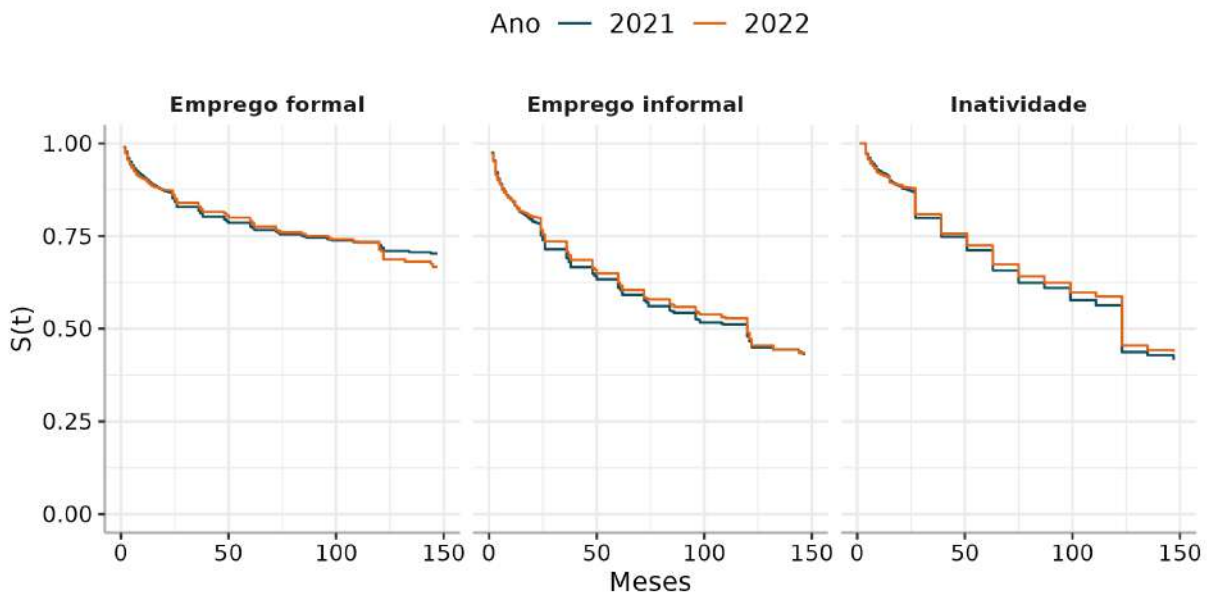
Sabe-se que uma amostra muito grande resulta em uma maior sensibilidade ao realizar qualquer teste de hipóteses, isto é, qualquer pequeno desvio de hipótese nula H_0 pode levar a rejeição de H_0 , mesmo com baixo nível de significância. Desse modo, os TRVs levavam à escolha de todas as covariáveis e termos de interação dupla.

Métodos automáticos *stepwise* também foram testados, utilizando tanto o critério de informação de Akaike (AIC) quanto critério de informação bayesiano (BIC). Em ambos os casos eram escolhidos modelos indesejadamente grandes, com todas as covariáveis e diversos termos de interação.

Assim, decidiu-se que a escolha da inclusão das últimas duas variáveis seria feita de acordo com a análise exploratória, o que permitiu a escolha de um conjunto de covariáveis mais parcimonioso. Em relação ao ano, observa-se por meio da Figura 4 que as

curvas de sobrevivência (estimação não-paramétrica via Kaplan–Meier) são muito similares entre 2021 e 2022, para todas as saídas do desemprego. Portanto, a variável de ano foi descartada para fins de modelagem.

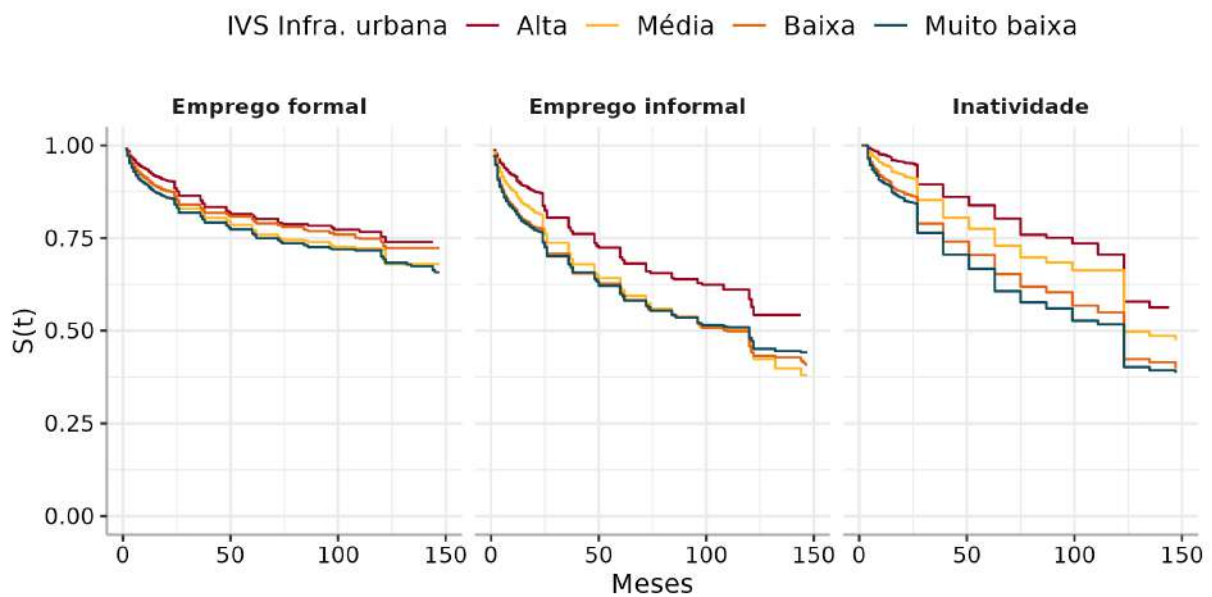
Figura 4: Curvas de sobrevivência por saída do desemprego estimadas segundo ano



Já em relação ao Índice de Vulnerabilidade Social (IVS) infraestrutura urbana, as curvas de sobrevivência para as categorias possíveis da variável - muito baixa, baixa, média e alta, onde muito baixa é o nível de melhor infraestrutura urbana e alta é o pior - revelam diferenças entre os tempos de desemprego. Como nota-se na Figura 5, conforme o índice de vulnerabilidade aumenta, cresce também o tempo médio de desemprego, tanto para os casos de saída para o emprego formal, informal ou inatividade.

Consequentemente, o IVS infraestrutura urbana foi mantido como covariável dos futuros modelos. Apesar disso, surge a questão de uma possível correlação entre a região e o IVS, dado que é natural esperar que algumas regiões estejam associadas com melhores (piores) condições de infraestrutura urbana. Para averiguar tal suspeita, foi calculado o VIF (*Variance inflation factor*) das covariáveis, método comum na detecção de multicolinearidade em modelos de regressão. Os resultados estão apresentados no Quadro 3.

Figura 5: Curvas de sobrevivência por saída do desemprego estimadas segundo IVS infraestrutura urbana



Quadro 3: *Variance inflation factor* (VIF) entre as covariáveis

Sexo	Raça/Cor	Região	Posição no domicílio	Escolaridade	Idade	Infra. urbana
1,061	1,148	2,176	1,629	1,336	1,882	1,929

Segundo Neter et al. (2005), um valor máximo de VIF maior que 10 para alguma variável é um indicativo de multicolinearidade e, conseqüentemente, de que a variância do coeficiente estimado na regressão está “inflado” de forma indevida. Nota-se que o maior valor de VIF observado foi de 2,176 para a região, o que indica que a multicolinearidade não é um problema relevante no conjunto de covariáveis em estudo.

Assim, região e IVS podem ser mantidas na regressão. Em síntese, as covariáveis selecionadas foram: **a)** sexo; **b)** raça/cor; **c)** escolaridade; **d)** posição no domicílio; **e)** região; e **f)** IVS infraestrutura urbana.

3.5 Modelagem dos dados

Na etapa de modelagem dos dados, foram utilizados modelos contínuos, apesar do tempo T de desemprego ser coletado em meses, ou seja, é coletado de maneira discreta. Tal decisão se deve a maior disponibilidade dos modelos contínuos nas rotinas computacionais e a maior facilidade na modelagem de dados contínuos, já que trabalhos com modelagem

em dados discretos em análise de sobrevivência ainda são escassos (BIAZATTI; NAKANO, 2020). Além disso, a grande variabilidade dos valores dos tempos observados (de 1 a 423 meses) é um indicativo de que modelos contínuos podem ser adequados aos dados (NAKANO; CARRASCO, 2006).

Outra escolha metodológica importante foi a de assumir T'_j s independentes, ou seja, os tempos de desemprego para cada causa de saída j foram tratados como variáveis aleatórias independentes. Esse pressuposto, por conta da própria natureza censurada dos dados, não pode ser diretamente testado, se tornando um ponto de alerta.

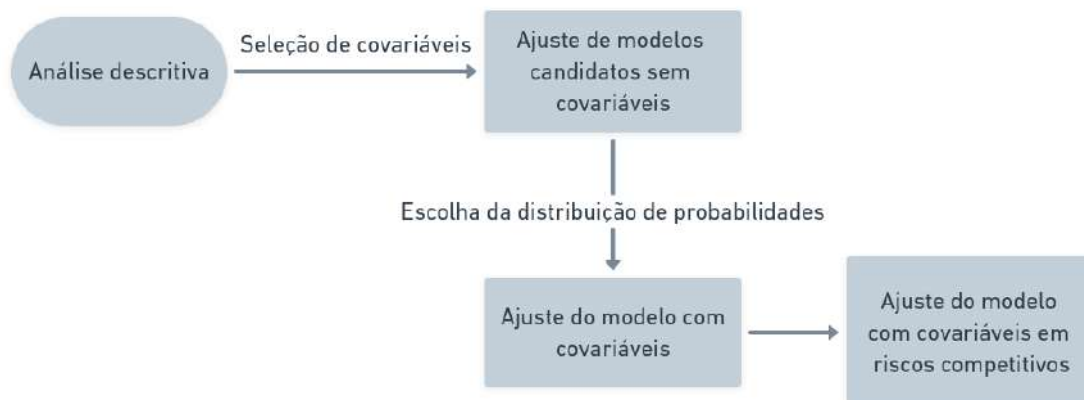
Apesar disso, essa abordagem possibilita a simplificação do processo de modelagem. Como visto na Seção 2.2.1, a estimação dos coeficientes de regressão em riscos competitivos é facilitada no caso de tempos independentes, sem que seja necessário recorrer a modelos mais complexos que abarcam uma estrutura de dependência entre os T'_j s.

Por fim, a esquematização das modelagens utilizadas neste trabalho está presente na Figura 6. Primeiramente, a análise descritiva dos dados possibilitou a seleção das covariáveis para os modelos futuros. Após essa etapa, foi feita uma primeira modelagem do tempo de desemprego sem as covariáveis e sem considerar riscos competitivos, ou seja, considerando apenas a saída do desemprego sem distinguir as causas. Foram ajustados 3 modelos paramétricos candidatos: Weibull, log-normal e log-logística.

A partir disso, os diagnósticos levaram à escolha da distribuição log-normal, que se manteve para as modelagens seguintes, presentes nas seções 4.3 e 4.4, sendo elas:

1. Modelo de regressão log-normal com as covariáveis selecionadas, sem considerar a estrutura de riscos competitivos;
2. Modelo de regressão log-normal com as covariáveis e considerando a estrutura de riscos competitivos.

Figura 6: Fluxograma do processo de modelagem dos dados



Fonte: Elaboração própria

Desse modo, foi possível comparar as técnicas e visualizar de que forma as covariáveis se relacionam com o tempo de desemprego até a transição para o emprego formal, informal ou inatividade. Os resultados e discussões são apresentados na Seção 4.

4 Resultados

4.1 Análise descritiva

A amostra em estudo apresenta um total de 86.036 observações, das quais 55.827 são censuradas (64,89%) e 30.209 são não censuradas (35,11%). Essa alta proporção de censuras é, de certa forma, algo esperado, pois o esquema amostral da PNADC garante que cada domicílio participe da pesquisa por apenas 5 trimestres. Além da saída natural dos domicílios da pesquisa, ainda existe o desafio do acompanhamento longitudinal dos indivíduos, por meio de uma chave criada artificialmente (descrita na Seção 3.3), o que colabora para o aumento das censuras, dada a possível imprecisão desse processo.

Na Tabela 1 estão apresentadas as frequências de cada variável explicativa na amostra. Observa-se que a amostra é composta por maioria feminina (52,74%) e de raça/cor parda (cerca de 55%). As categorias de raça/cor podem ser agrupadas de tal modo que a categoria branca torna-se o conjunto de brancos e amarelos, totalizando 27.435 (31,89%), e os não brancos se dão pelo total de pretos, pardos e indígenas, totalizando 58.601 (68,11%).

Em relação a idade, existe uma concentração nas faixas etárias mais jovens: 52,95% (45.556) dos indivíduos em busca de emprego tinham entre 14 e 30 anos. Vale ressaltar que as faixas etárias escolhidas para realizar a análise não possuem amplitudes iguais, porém foram escolhidas de forma a impedir um número muito baixo de observações em alguma categoria, o que dificultaria posteriormente o processo de estimação dos coeficientes na regressão.

Sobre a escolaridade, cerca de 45% dos indivíduos em busca de emprego tinham de 12 a 15 anos de estudo. Em contraposição, somente 2,17% da amostra estava na categoria de sem instrução e menos de 1 ano de estudo. No que concerne a posição no domicílio, cerca de 37% são filhos, enquanto aproximadamente 34% são pessoas responsáveis pelo domicílio. Como a variável posição no domicílio permitia muitas categorias de resposta, criou-se a categoria outros, agrupando diversas categorias com pequenas frequências na amostra. Tais frequências podem ser vistas no Apêndice A.

Já em relação ao Índice de Vulnerabilidade Social (IVS) infraestrutura urbana, a maior parte dos indivíduos está em unidades da federação classificadas como muito baixa (45,15%) ou baixa (30,21%) vulnerabilidade social na dimensão de infraestrutura urbana. Ou seja, são maioria as localidades com melhores condições de infraestrutura urbana.

Tabela 1: Frequência das covariáveis

	Frequência	Frequência relativa
Total	86.036	100%
Sexo		
Homem	40.658	47,26%
Mulher	45.378	52,74%
Raça/Cor		
Branca	26.934	31,31%
Preta	10.527	12,24%
Amarela	501	0,58%
Parda	47.631	55,36%
Indígena	443	0,51%
Idade		
14 a 20 anos	18.780	21,83%
21 a 30 anos	26.776	31,12%
31 a 40 anos	17.791	20,68%
41 a 50 anos	13.305	15,46%
51 a 70 anos	9.384	10,91%
Escolaridade		
Sem instrução e menos de 1 ano de estudo	1.865	2,17%
1 a 4 anos de estudo	4.225	4,91%
5 a 8 anos de estudo	14.662	17,04%
9 a 11 anos de estudo	18.333	21,31%
12 a 15 anos de estudo	39.019	45,35%
16 anos ou mais de estudo	7.932	9,22%
Ano		
2021	57.742	67,11%
2022	28.294	32,89%
Posição no domicílio		
Pessoa responsável	28.846	33,53%
Cônjuge ou companheiro(a)	14.835	17,24%
Filho(a)	32.227	37,46%
Outros	10.128	11,77%
Região		
Norte	12.455	14,48%
Nordeste	31.261	36,33%
Sudeste	24.060	27,97%
Sul	9.694	11,27%
Centro-Oeste	8.566	9,96%
IVS infraestrutura urbana		
Muito baixa	38.843	45,15%
Baixa	25.988	30,21%
Média	12.327	14,33%
Alta	8.878	10,32%

Já na Tabela 2, estão apresentadas, apenas para os indivíduos não censurados (30.209), as frequências de saída do desemprego - emprego formal, informal ou inatividade - segundo as variáveis explicativas.

Tabela 2: Frequência de saída do desemprego por covariáveis

	Total	Saída do desemprego		
		Emprego formal	Emprego informal	Inatividade
Total	30.209	7.981 (26,42%)	14.779 (48,92%)	7.449 (24,66%)
Sexo				
Homem	15.264	4.021 (26,34%)	8.250 (54,05%)	2.993 (19,61%)
Mulher	14.945	3.960 (26,50%)	6.529 (43,69%)	4.456 (29,82%)
Raça/Cor				
Branca	9.404	3.133 (33,32%)	4.215 (44,82%)	2.056 (21,86%)
Preta	3.688	966 (26,19%)	1.861 (50,46%)	861 (23,35%)
Amarela	180	40 (22,22%)	83 (46,11%)	57 (31,67%)
Parda	16.776	3.808 (22,70%)	8.539 (50,90%)	4.429 (26,40%)
Indígena	161	34 (21,12%)	81 (50,31%)	46 (28,57%)
Região				
Norte	4.089	815 (19,93%)	2.276 (55,66%)	998 (24,41%)
Nordeste	11.427	1.794 (15,70%)	5.772 (50,51%)	3.861 (33,79%)
Sudeste	8.307	2.770 (33,35%)	3.884 (46,76%)	1.653 (19,90%)
Sul	3.499	1.628 (46,53%)	1.386 (39,61%)	485 (13,86%)
Centro-Oeste	2.887	974 (33,74%)	1.461 (50,61%)	452 (15,66%)
Posição no domicílio				
Pessoa responsável	11.114	2.879 (25,90%)	5.889 (52,99%)	2.346 (21,11%)
Cônjuge ou companheiro(a)	5.636	1.568 (27,82%)	2.787 (49,45%)	1.281 (22,73%)
Filho(a)	10.673	2.804 (26,27%)	4.785 (44,83%)	3.084 (28,90%)
Outros	2.786	730 (26,20%)	1.318 (47,31%)	738 (26,49%)
Escolaridade				
Sem instrução e menos de 1 ano de estudo	733	72 (9,82%)	449 (61,26%)	212 (28,92%)
1 a 4 anos de estudo	1.740	235 (13,51%)	1.065 (61,21%)	440 (25,29%)
5 a 8 anos de estudo	5.278	865 (16,39%)	2.976 (56,38%)	1.437 (27,23%)
9 a 11 anos de estudo	6.076	1.336 (21,99%)	3.028 (49,84%)	1.712 (28,18%)
12 a 15 anos de estudo	13.533	4.225 (31,22%)	6.202 (45,83%)	3.106 (22,95%)
16 anos ou mais de estudo	2.849	1.248 (43,80%)	1.059 (37,17%)	542 (19,02%)
Ano				
2021	20.834	5.473 (26,27%)	10.206 (48,99%)	5.155 (24,74%)
2022	9.375	2.508 (26,75%)	4.573 (48,78%)	2.294 (24,47%)
Faixa etária				
14 a 20 anos	5.613	1.257 (22,39%)	2.442 (43,51%)	1.914 (34,10%)
21 a 30 anos	9.092	2.624 (28,86%)	4.307 (47,37%)	2.161 (23,77%)
31 a 40 anos	6.677	1.871 (28,02%)	3.403 (50,97%)	1.403 (21,01%)
41 a 50 anos	5.206	1.394 (26,78%)	2.766 (53,13%)	1.046 (20,09%)
51 a 70 anos	3.621	835 (23,06%)	1.861 (51,39%)	925 (25,55%)
IVS infraestrutura urbana				
Muito baixa	14.496	3.783 (26,10%)	6.889 (47,52%)	3.824 (26,38%)
Baixa	9.184	2.189 (23,83%)	4.659 (50,73%)	2.336 (25,44%)
Média	4.144	1.233 (29,75%)	2.067 (49,88%)	844 (20,37%)
Alta	2.385	776 (32,54%)	1.164 (48,81%)	445 (18,66%)

De modo geral, entre todas as saídas do desemprego observadas, a maior parcela

dos indivíduos ingressou em empregos informais (48,92%), seguidos dos empregos formais (26,42%). Em menor parcela, mas ainda muito expressiva, outros 24,66% ingressaram na inatividade

Algumas relações chamam atenção na Tabela 2, como o maior percentual de ingresso na inatividade para as mulheres: dentre as mulheres que saíram do desemprego no biênio, cerca de 30% entrou na inatividade, enquanto entre os homens esse percentual foi de aproximadamente 20%. Já o emprego informal é um fenômeno que atingiu mais a parcela masculina, cerca de 54% dos homens ingressou no emprego informal contra cerca de 44% no grupo feminino.

A respeito da raça/cor, aqueles que se autodeclararam brancos formam o grupo com maior proporção de ingresso no emprego formal - cerca de 33%. Em contraponto, dentre os pardos e pretos houve um percentual de entrada no emprego formal de aproximadamente 23% e 26%, respectivamente.

As diferenças entre as regiões também chamam atenção. A região Sul lidera com a maior frequência de ingressos no emprego formal (cerca de 47%), seguida do Centro-Oeste (34%) e Sudeste (33%). Por outro lado, o Nordeste apresenta a menor frequência de entrada no emprego formal (aproximadamente 16%), assim como o maior percentual de indivíduos ingressando na inatividade, cerca de 34%.

Além disso, a escolaridade revela um padrão interessante: quanto mais anos de estudo, maior a proporção de indivíduos ingressando no emprego formal. De forma análoga, nota-se um declínio na frequência de ingressos no emprego informal, conforme o grau de escolaridade aumenta.

A fim de verificar a relação entre as variáveis listadas e as saídas do desemprego, foi feito um teste qui-quadrado de independência para cada subtabela (covariável *versus* saídas do desemprego). Os resultados estão disponíveis no Quadro 4.

Quadro 4: Teste Qui-Quadrado de independência entre as covariáveis e a causa de saída do desemprego

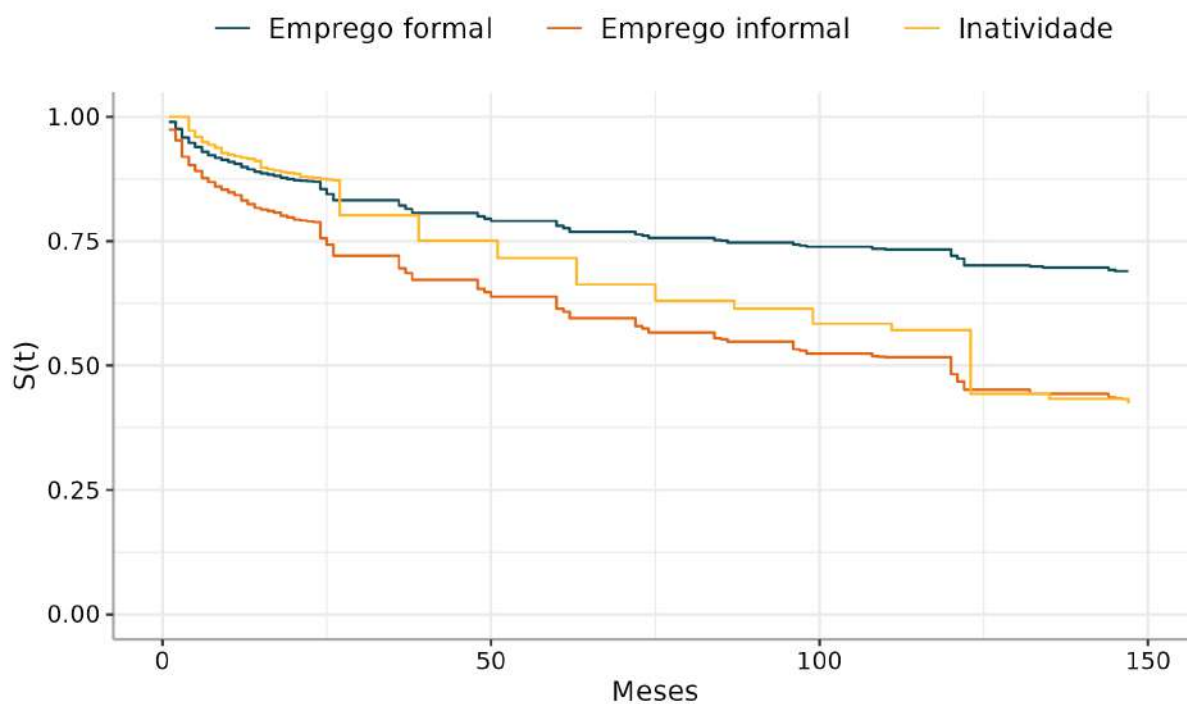
Covariável	Estatística do teste	Graus de liberdade	P-valor
Sexo	484,897	2	< 0,001
Raça/Cor	365,651	8	< 0,001
Região	2.145,563	8	< 0,001
Posição no domicílio	228,073	6	< 0,001
Escolaridade	1.223,849	10	< 0,001
Ano	0,824	2	0,662
Faixa etária	431,323	8	< 0,001
IVS infraestrutura urbana	173,166	6	< 0,001

Para quase todas as variáveis presentes na amostra, rejeita-se a hipótese nula de independência entre tipo de saída do desemprego e a característica em questão (sexo, raça/cor, região, posição no domicílio, escolaridade, faixa etária e IVS). Porém, apenas para a variável de ano não é possível rejeitar a hipótese nula de independência. De fato, o teste confirma aquilo que pode ser visto na Tabela 2: as proporções de ingresso no emprego formal, informal ou na inatividade são todas muito similares entre o ano de 2021 e 2022, não parecendo haver qualquer tipo de relação.

4.1.1 Curvas de sobrevivência não-paramétricas

Na Figura 7, estão apresentadas as curvas de sobrevivência estimadas via Kaplan-Meier para as possíveis saídas do desemprego. É possível notar que, inicialmente a curva da inatividade é a mais elevada, ou seja, é o evento com maior sobrevida. Entretanto, a partir dos 25 meses de desemprego, aproximadamente, o padrão se altera de tal sorte que o emprego formal se torna a causa de saída com maior tempo médio de desemprego, seguido da inatividade e, por fim, do emprego informal.

Figura 7: Curva de sobrevivência por saída do desemprego estimada



A mediana do tempo de desemprego, sem distinguir as causas de saída, é de 27 meses. Já a probabilidade de permanecer desempregado após 1 ano de busca (proba-

bilidade de sobrevivência) é de aproximadamente 68%. É importante observar que, ao calcular a mediana de maneira incorreta, ou seja, sem levar em conta a censura dos dados, obtém-se apenas 6 meses como mediana. Esse fato destaca a incorreção de ignorar a censura em uma análise, o que geralmente leva a interpretações equivocadas.

Na Figura 8 estão ilustradas as curvas de sobrevivência estimadas para as diferentes categorias das covariáveis sexo, raça/cor e posição no domicílio. Em relação ao sexo, as mulheres apresentaram maior tempo de desemprego médio para o ingresso no emprego formal e também no emprego informal, em comparação com os homens. Entretanto, a curva de sobrevivência para a inatividade é similar em ambos os sexos.

Para as categorias de raça/cor, observa-se que os não brancos possuem maior tempo de desemprego médio até o ingresso no emprego formal, quando comparados com os brancos. Por outro lado, pessoas não brancas têm menor tempo médio de desemprego até o ingresso no emprego informal ou na inatividade. Essa observação sugere que os indivíduos não brancos estão mais expostos à entrada no emprego informal e na inatividade.

Já no que concerne a posição no domicílio, é possível notar um tempo de desemprego médio até o emprego informal menor para as pessoas responsáveis pelo domicílio, em comparação com cônjuges ou companheiros(as) e filhos(as). Também é possível destacar que os filhos(as) formam o grupo com menor tempo médio até a inatividade.

Por fim, na Figura 9 estão presentes as curvas de sobrevivência segundo anos de estudo. Constata-se um padrão interessante: conforme aumentam os anos de estudo, decresce de forma expressiva o tempo médio até o ingresso no emprego formal. Nesse sentido, aqueles que possuem 16 anos ou mais de estudo formam o grupo com menor tempo de desemprego até o emprego formal. Apesar disso, observa-se que os indivíduos com poucos anos de estudo têm menor tempo médio até a saída para o emprego informal, especialmente o grupo com 1 a 4 anos de estudo.

Figura 8: Curvas de sobrevivência por saída do desemprego estimadas conforme covariáveis (sexo, raça/cor e posição no domicílio)

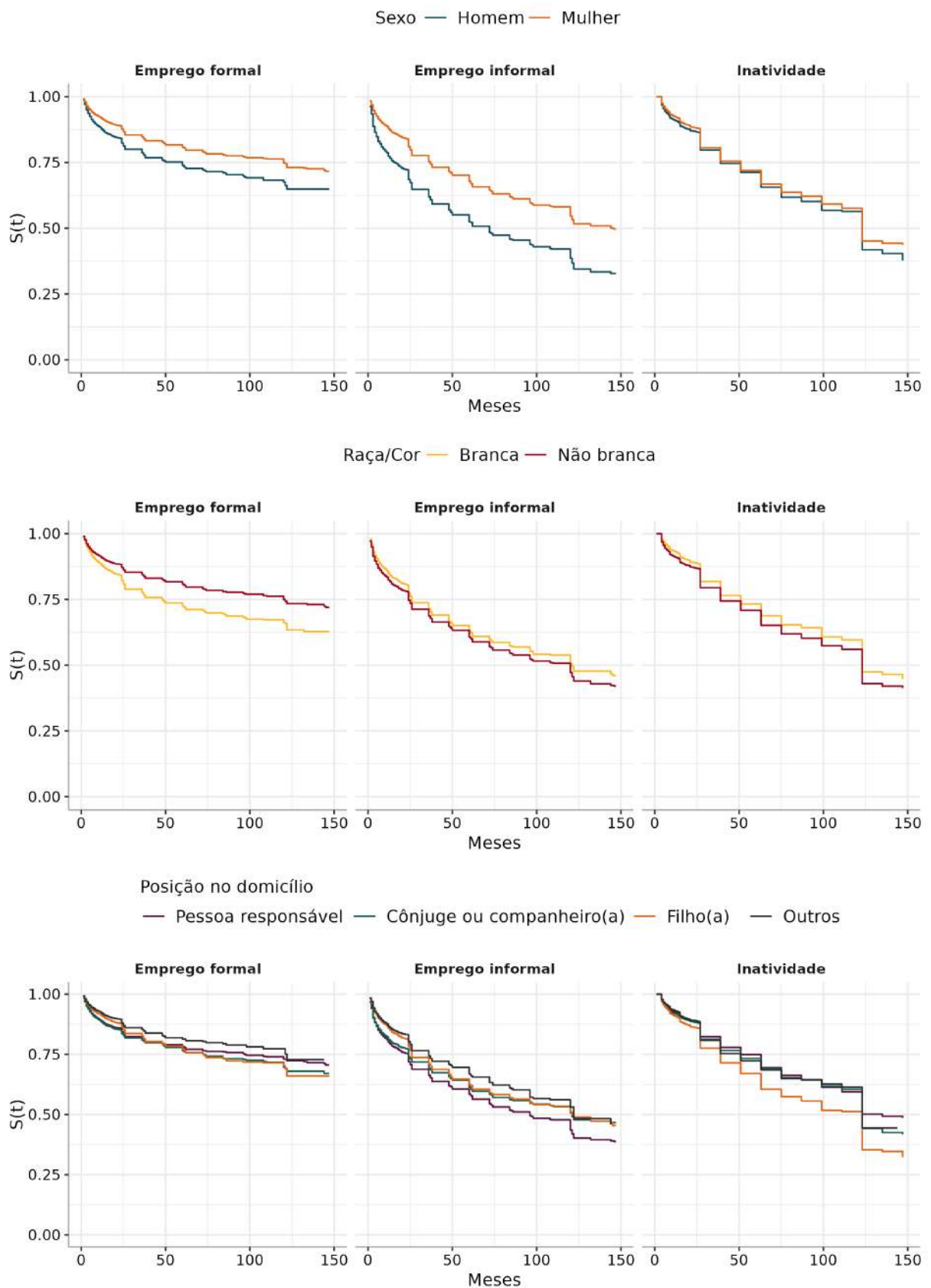
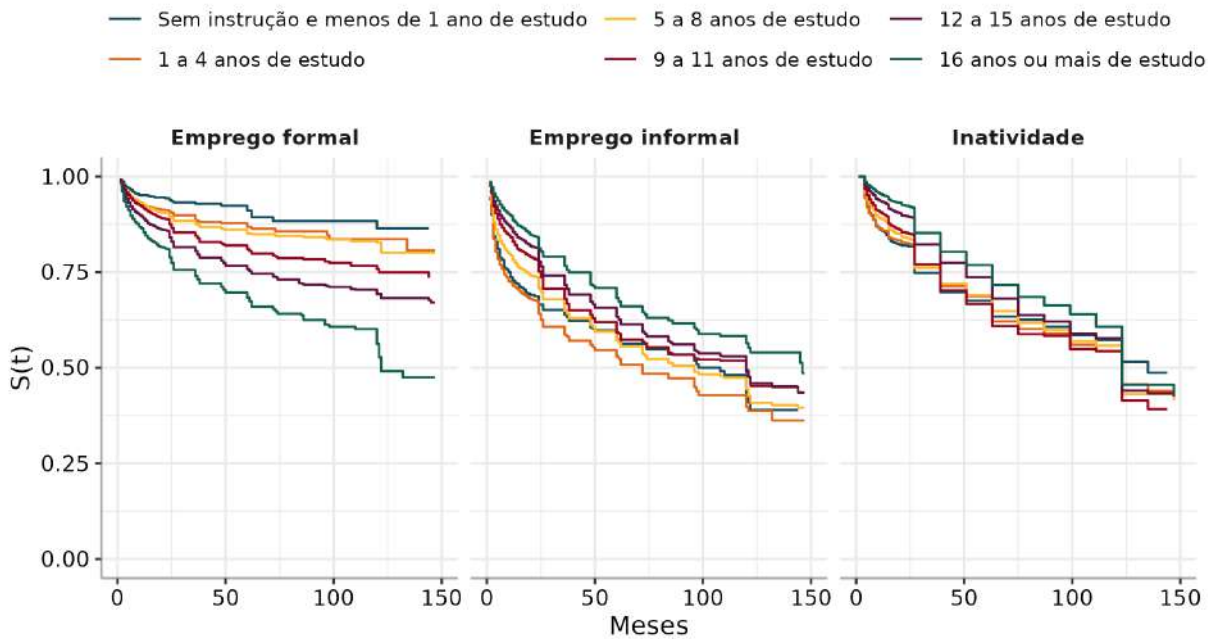


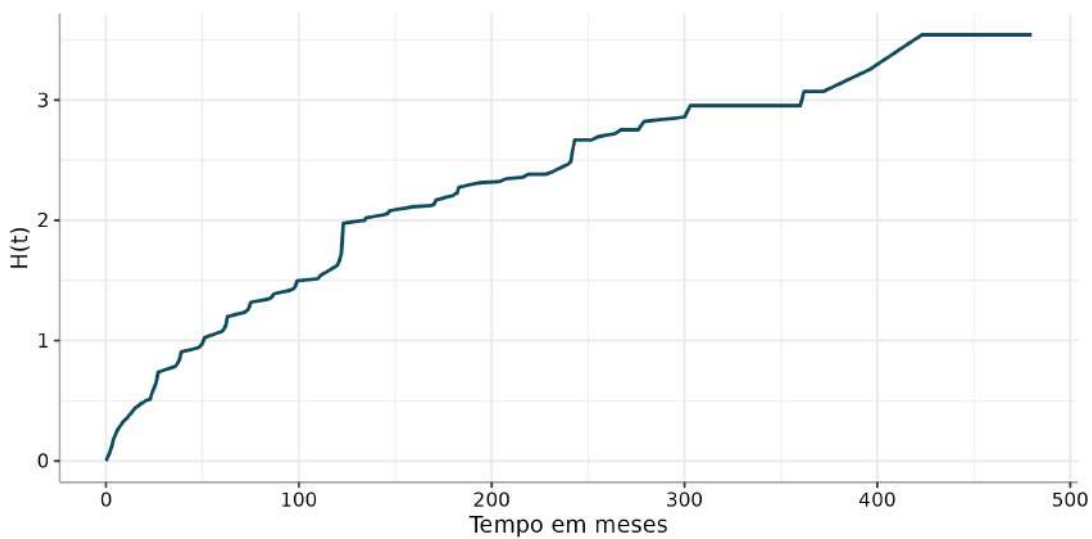
Figura 9: Curvas de sobrevivência por saída do desemprego estimadas conforme escolaridade



4.2 Escolha do modelo paramétrico

No intuito de encontrar distribuições de probabilidade candidatas para modelar a duração do desemprego, foi calculada a curva de risco acumulada estimada $\hat{H}(t)$, presente na Figura 10.

Figura 10: Função de risco acumulada estimada - $\hat{H}(t)$



De modo geral, a curva $\hat{H}(t)$ parece seguir um formato côncavo, o que sugere

uma função de risco monotonicamente decrescente. Assim, as distribuições candidatas são: Weibull, log-normal e log-logística. Para avaliar qual das opções melhor se ajusta aos dados, foram graficadas as curvas de sobrevivência segundo os modelos paramétricos, juntamente com a estimação via Kaplan-Meier, na Figura 11. Além disso, os resíduos de cox-snell para cada distribuição estão presentes na Figura 12.

Figura 11: Curvas de sobrevivência estimadas pelos modelos paramétricos *versus* estimação por Kaplan-Meier

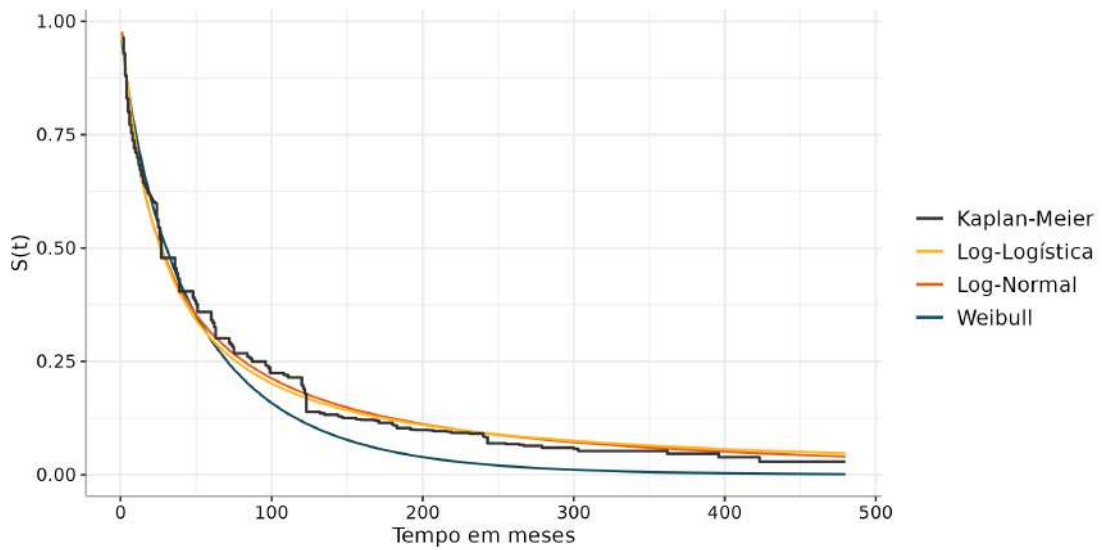
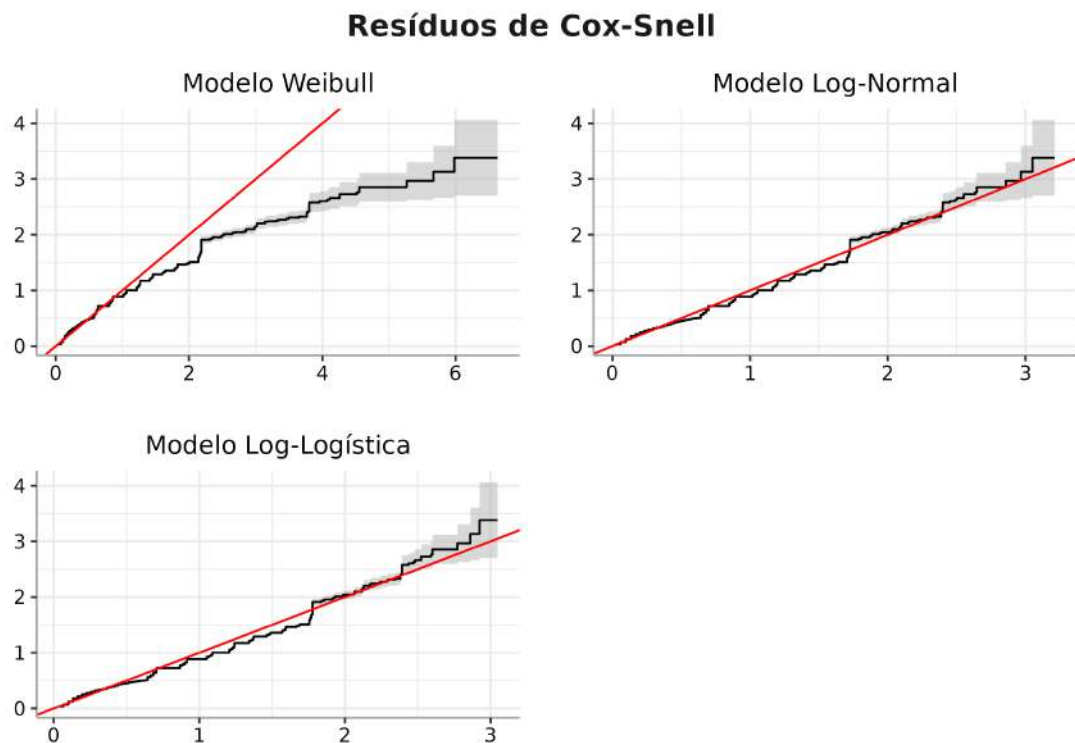


Figura 12: Resíduos de Cox-Snell por modelo sem covariáveis



Por meio das Figuras 11 e 12, fica claro que os modelos log-normal e log-logístico apresentam bons ajustes aos dados, enquanto o modelo Weibull não se mostra satisfatório. Assim, o modelo log-normal é escolhido, pois apresenta um ajuste global ligeiramente melhor que o do modelo log-logístico, como pode ser visto por meio dos resíduos. Dessa forma, os modelos de regressão seguintes são ajustados com base na distribuição log-normal.

4.3 Modelo paramétrico com covariáveis sem riscos competitivos

Na Tabela 3, estão as estimativas dos coeficientes do modelo de regressão log-normal, sem considerar o cenário de riscos competitivos. Ou seja, modelou-se apenas o tempo de desemprego segundo as covariáveis, sem observar as causas de saída. Além das estimativas pontuais, estão presentes os erros padrão e intervalos de confiança planos (com nível de significância $\gamma = 95\%$).

Para cada variável, a primeira categoria presente na Tabela 3 é a categoria de referência. Por isso, essas caselas não estão preenchidas, em verdade, considera-se que os coeficientes nesse caso são 0.

É importante ressaltar que a interpretação dos coeficientes não é feita de maneira direta como em um modelo de regressão linear normal. Em realidade, é necessário exponenciar o coeficiente para chegar no efeito multiplicativo do tempo mediano de desemprego, comparando com o nível de referência em cada variável.

De fato, a mediana da distribuição log-normal (2.1.14) é dada por e^μ . Assim, $e^{\beta p}$ é interpretado como a razão de medianas de dois indivíduos que diferem uma unidade da covariável x_p , mantendo-se fixas as demais covariáveis. Assim, segundo os resultados da Tabela 3, temos que:

- (i) O tempo mediano até a saída do desemprego para as mulheres é cerca de 1,8 ($e^{0,604}$) vezes o tempo mediano dos homens;
- (ii) O tempo mediano de desemprego para pessoas brancas é ligeiramente maior - cerca de 1,05 vezes - que o de pessoas não brancas;
- (iii) Indivíduos com idade entre 51 e 70 anos têm um tempo mediano de desemprego cerca de 1,4 vezes o tempo daqueles com idade entre 14 e 20 anos;
- (iv) Em contraposição ao senso comum, os indivíduos com mais anos de estudo possuem

maior duração do desemprego. Em comparação com o grupo sem instrução, aqueles com 12 a 15 anos de estudo, ou 16 anos ou mais de estudo, possuem tempo mediano do desemprego aproximadamente 26% maior;

- (v) Não existe diferença estatisticamente significativa entre o tempo de desemprego de pessoas responsáveis pelo domicílio e cônjuges ou companheiros(as);
- (vi) Desempregados na região Sudeste apresentam uma duração do desemprego 30% menor do que o tempo daqueles que residem na região Norte;
- (vii) Conforme o aumento do índice de vulnerabilidade social na infraestrutura urbana, cresce também o tempo mediano de desemprego. Indivíduos residentes de locais com IVS infraestrutura urbana classificado como alta têm um tempo mediano de desemprego de cerca de 3 vezes do tempo mediano da categoria IVS muito baixa.

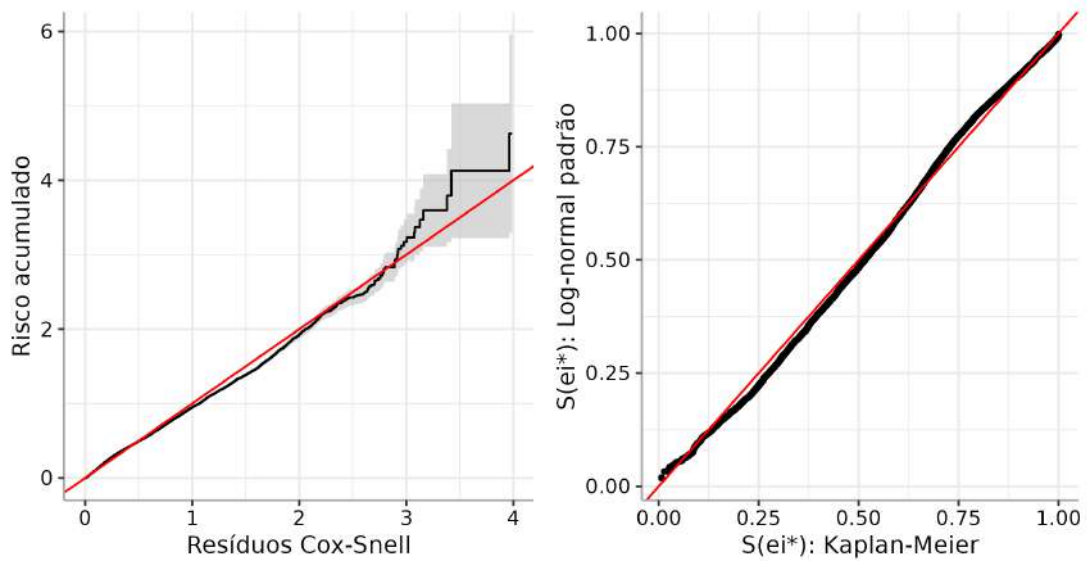
Tabela 3: Estimativas dos coeficientes do modelo de regressão log-normal

		Estimativa	Erro Padrão	IC - 95%
β_0		2,409***	0,059	(2,294 ; 2,524)
$\log(\sigma)$		0,449***	0,004	(0,441 ; 0,457)
Sexo	Homem	-	-	-
	Mulher	0,604***	0,014	(0,576 ; 0,632)
Raça/Cor	Branca	-	-	-
	Não branca	-0,049**	0,016	(-0,080 ; -0,018)
Idade	14 a 20 anos	-	-	-
	21 a 30 anos	0,097***	0,021	(0,056 ; 0,138)
	31 a 40 anos	0,064**	0,025	(0,016 ; 0,112)
	41 a 50 anos	0,093***	0,027	(0,040 ; 0,146)
	51 a 70 anos	0,309***	0,030	(0,249 ; 0,368)
Escolaridade	Sem instrução e menos de 1 ano de estudo	-	-	-
	1 a 4 anos de estudo	-0,084	0,054	(-0,191 ; 0,022)
	5 a 8 anos de estudo	0,112*	0,049	(0,016 ; 0,207)
	9 a 11 anos de estudo	0,238***	0,049	(0,142 ; 0,334)
	12 a 15 anos de estudo	0,229***	0,048	(0,135 ; 0,322)
	16 anos ou mais de estudo	0,228***	0,052	(0,127 ; 0,330)
Posição no domicílio	Pessoa responsável	-	-	-
	Cônjuge ou companheiro(a)	0,020	0,020	(-0,019 ; 0,059)
	Filho(a)	0,325***	0,020	(0,286 ; 0,363)
	Outros	0,457***	0,026	(0,406 ; 0,508)
Região	Norte	-	-	-
	Nordeste	0,131***	0,024	(0,084 ; 0,177)
	Sudeste	-0,368***	0,024	(-0,416 ; -0,320)
	Sul	-0,107***	0,030	(-0,166 ; -0,047)
	Centro-Oeste	-0,412***	0,031	(-0,473 ; -0,352)
IVS infra. urbana	Muito baixa	-	-	-
	Baixa	0,059***	0,017	(0,025 ; 0,092)
	Média	0,482***	0,024	(0,435 ; 0,529)
	Alta	1,077***	0,027	(1,024 ; 1,131)

Nota: Valores significativos ao nível de 0,001***, 0,01** e 0,05*.

Os resíduos de cox-snell e resíduos padronizados do modelo estão presentes na Figura 13. Verifica-se que os resíduos estão bem comportados, o que indica um bom ajuste global do modelo. Dessa forma, as interpretações feitas com base nos coeficientes estimados são, de fato, válidas para o conjunto de dados.

Figura 13: Resíduos de Cox-Snell e resíduos padronizados do modelo de regressão log-normal



4.4 Modelo paramétrico com covariáveis em riscos competitivos

Já os resultados para o modelo de regressão na presença de riscos competitivos estão apresentados na Tabela 4. Nessa modelagem, é possível identificar como as covariáveis estão relacionadas com o tempo até o ingresso para o emprego formal, informal ou para a inatividade. A interpretação dos coeficientes é feita da mesma forma discutida na Seção 4.3.

Na Tabela 4 estão presentes as estimativas pontuais e respectivos erros padrão. Os intervalos de confiança ($\gamma = 95\%$) podem ser vistos no Apêndice B.

Em relação ao sexo, as diferenças entre homens e mulheres são muito evidentes. O tempo mediano até o ingresso no emprego formal para as mulheres é mais que o dobro do tempo dos homens (2,1 vezes). Essa diferença se agrava quando é observado o tempo até o emprego informal: as mulheres levam um tempo mediano de 2,4 vezes do tempo dos homens para ingressar em uma ocupação informal. Já no tempo até o ingresso na inatividade, não há evidência estatística de diferença entre homens e mulheres.

A respeito da raça/cor, não se observa diferença estatisticamente significativa entre os tempos de desemprego até o emprego formal e até a inatividade, comparando brancos e não brancos. Entretanto, pessoas não brancas adquirem um emprego informal em tempo mediano 8% menor que pessoas brancas, aproximadamente.

Tabela 4: Estimativas dos coeficientes do modelo de regressão log-normal em riscos competitivos

	Emprego formal		Emprego informal		Inatividade	
	Estimativa	Erro Padrão	Estimativa	Erro Padrão	Estimativa	Erro Padrão
β_0	6,015***	0,154	3,134***	0,086	3,947***	0,091
$\log(\sigma)$	0,799***	0,008	0,692***	0,006	0,451***	0,008
Sexo						
Homem	-	-	-	-	-	-
Mulher	0,755***	0,029	0,859***	0,022	0,037	0,023
Raça/Cor						
Branca	-	-	-	-	-	-
Não branca	-0,037	0,032	-0,083***	0,024	-0,006	0,025
Idade						
14 a 20 anos	-	-	-	-	-	-
21 a 30 anos	-0,093*	0,045	-0,014	0,033	0,471***	0,031
31 a 40 anos	-0,124*	0,052	-0,075*	0,038	0,551***	0,038
41 a 50 anos	-0,103	0,057	-0,044	0,042	0,591***	0,043
51 a 70 anos	0,306***	0,064	0,245***	0,046	0,458***	0,046
Escolaridade						
Sem instrução e menos de						
1 ano de estudo	-	-	-	-	-	-
1 a 4 anos de estudo	-0,403**	0,149	-0,089	0,077	-0,042	0,083
5 a 8 anos de estudo	-0,322*	0,135	0,152*	0,069	0,107	0,074
9 a 11 anos de estudo	-0,536***	0,134	0,378***	0,070	0,264***	0,074
12 a 15 anos de estudo	-1,017***	0,132	0,502***	0,068	0,453***	0,072
16 anos ou mais de estudo	-1,381***	0,136	0,765***	0,076	0,537***	0,080
Posição no domicílio						
Pessoa responsável	-	-	-	-	-	-
Cônjuge ou companheiro(a)	-0,062	0,041	0,108***	0,030	-0,055	0,033
Filho(a)	0,392***	0,040	0,487***	0,030	-0,038	0,032
Outros	0,451***	0,054	0,607***	0,040	0,165***	0,041
Região						
Norte	-	-	-	-	-	-
Nordeste	0,331***	0,053	0,246***	0,035	-0,130***	0,037
Sudeste	-1,035***	0,051	-0,182***	0,037	-0,139***	0,039
Sul	-1,220***	0,061	0,301***	0,047	0,482***	0,052
Centro-Oeste	-1,114***	0,063	-0,308***	0,046	0,040	0,052
IVS infra. urbana						
Muito baixa	-	-	-	-	-	-
Baixa	0,084*	0,036	0,024	0,026	0,077**	0,026
Média	0,461***	0,048	0,477***	0,037	0,523***	0,039
Alta	1,223***	0,053	1,079***	0,042	0,958***	0,047

Nota: Valores significativos ao nível de 0,001***, 0,01** e 0,05*.

Sobre a faixa etária, os indivíduos de idade mais avançada parecem formar um grupo em desvantagem. Aqueles entre 51 a 70 anos possuem tempo mediano até a

ocupação formal de 1,4 vezes o tempo de indivíduos de 14 a 20 anos. Por outro lado, o grupo mais jovem, de 14 a 20 anos, parece ser o mais exposto ao risco de ingressar na inatividade: todas as outras faixas etárias têm maior tempo até a saída para a inatividade. Em especial, o grupo de 14 a 20 anos possui tempo até a inatividade de cerca de metade (0,55 vezes, em mediana) daqueles entre 41 e 50 anos.

Olhando para a escolaridade, é possível chegar a uma interpretação mais minuciosa que a vista anteriormente na regressão sem riscos competitivos. Considerando o emprego formal, quanto mais anos de estudo maior o impacto na redução do tempo de desemprego. No caso de indivíduos com 16 anos ou mais de estudo, o tempo mediano até a aquisição de um vínculo formal é $1/4$ do tempo daqueles que estão na categoria sem instrução e menos de 1 ano de estudo.

Por outro lado, quanto maior o nível de escolaridade, menor a exposição ao emprego informal e à inatividade. Ou seja, os grupos com mais anos de estudo possuem maiores tempos medianos até a saída para a ocupação informal ou inatividade.

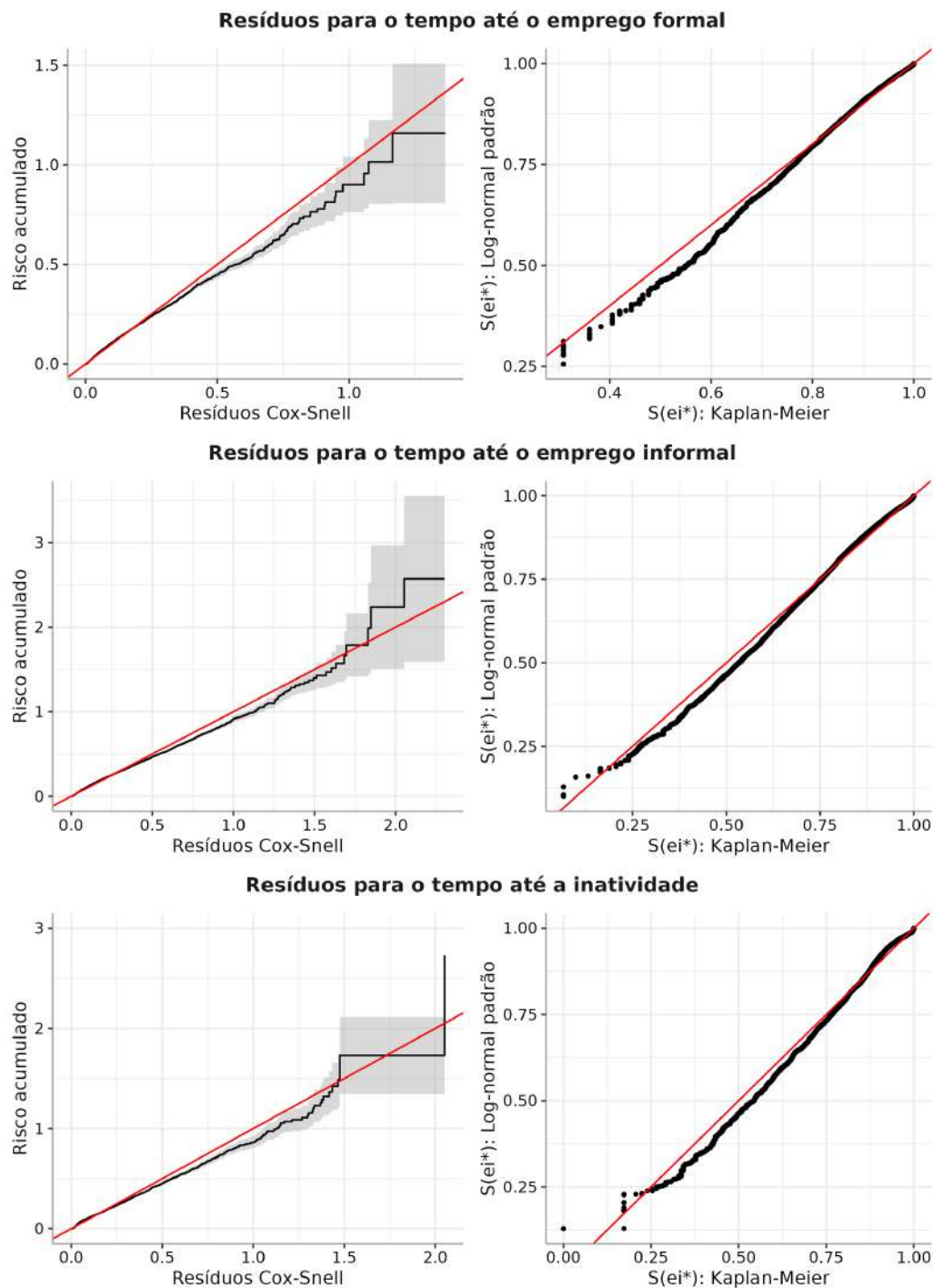
A respeito da posição no domicílio, os cônjuges ou companheiros(as) não apresentam tempos de desemprego significativamente diferentes da pessoa responsável quando se observa a saída para o emprego formal ou inatividade. Porém, para a entrada no emprego informal, os cônjuges possuem tempo mediano cerca de 11% maior.

A região Nordeste chama atenção por ser a única com tempos até o ingresso no emprego formal maiores que o da região Norte. O tempo mediano até o emprego formal é cerca de 39% maior no Nordeste, em comparação com o Norte. Além disso, no Nordeste e Sudeste os indivíduos estão mais expostos à inatividade: o tempo até a inatividade é cerca de 12% e 13% menor, respectivamente, que na região Norte.

Por último, a infraestrutura urbana tem um efeito similar para todas as saídas do desemprego: quanto pior a situação de infraestrutura urbana (mais alto o IVS) maiores são os tempos de desemprego. Isso é válido para o emprego formal, informal ou inatividade. Assim, o tempo mediano até a ocupação formal, na categoria de IVS alta, é cerca de 3,4 vezes o tempo no nível de vulnerabilidade muito baixa.

Na Figura 14, observam-se os resíduos de cox-snell e resíduos padronizados para cada saída do desemprego. Tanto para o ingresso no emprego formal, informal ou inatividade nota-se que os resíduos estão bem comportados. Portanto, conclui-se que o ajuste global do modelo de regressão em presença de riscos competitivos é satisfatório.

Figura 14: Resíduos de Cox-Snell e resíduos padronizados do modelo de regressão log-normal em riscos competitivos



4.4.1 Probabilidades de saída do desemprego por perfil sociodemográfico

Com a regressão em riscos competitivos ajustada, obtém-se 3 curvas paramétricas, uma para cada causa de saída do desemprego. A partir disso, é possível calcular as probabilidades de ingresso em cada causa, dado um conjunto de covariáveis.

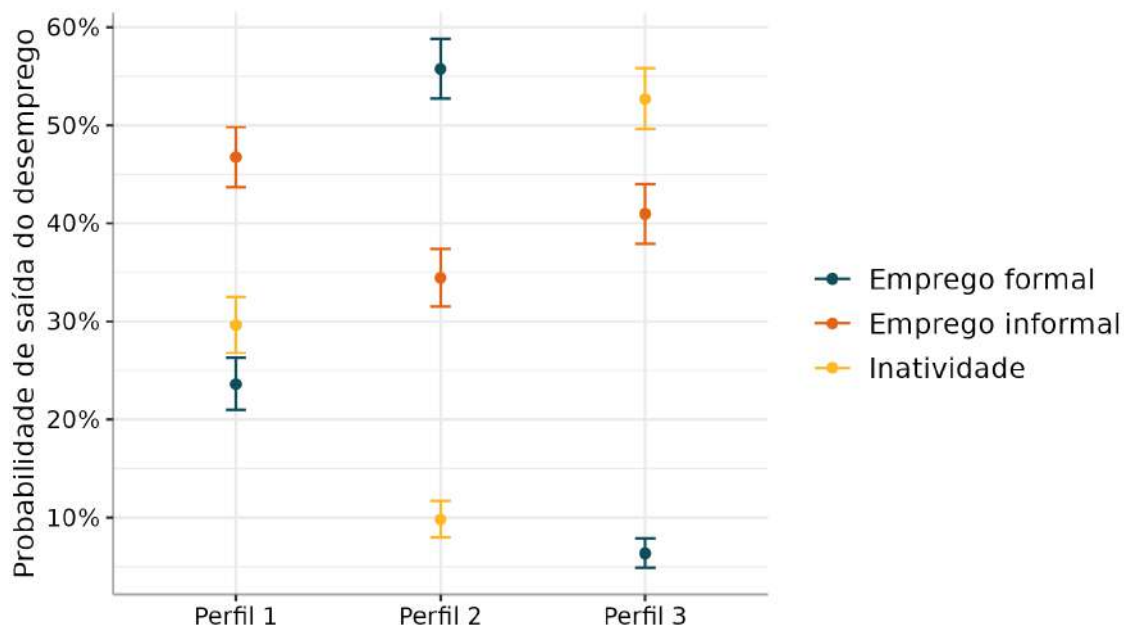
Considere T_1 o tempo até o ingresso no emprego formal, T_2 o tempo até o ingresso no emprego informal e T_3 o tempo até a inatividade. As probabilidades de interesse são: $P[\text{ingresso no emprego formal}] = P(T_1 < \min\{T_2, T_3\})$; $P[\text{ingresso no emprego informal}] = P(T_2 < \min\{T_1, T_3\})$; e $P[\text{ingresso na inatividade}] = P(T_3 < \min\{T_1, T_2\})$.

Dado as covariáveis $\mathbf{x}' = (1, x_1, \dots, x_p)$, o parâmetro de locação da distribuição log-normal é dado por $\mu_j = \mathbf{x}'\boldsymbol{\beta}_j$, $j = 1, 2, 3$. Acrescentando os parâmetros de escala σ'_j s estimados, foi feito um processo de simulação, gerando amostras dos tempos T'_j s por meio da distribuição log-normal. Desse modo, foi possível calcular as probabilidades de saída do desemprego para os 3 perfis descritos a seguir:

- Perfil 1: mulher, branca, 41 a 50 anos, 12 a 15 anos de estudo, região Norte, pessoa responsável no domicílio e IVS infraestrutura urbana muito baixa;
- Perfil 2: homem, branco, 31 a 40 anos, 16 anos ou mais de estudo, região Sul, pessoa responsável no domicílio e IVS infraestrutura urbana muito baixa;
- Perfil 3: mulher, não branca, 14 a 20 anos, sem instrução e menos de 1 ano de estudo, região Nordeste, cônjuge ou companheiro(a) no domicílio e IVS infraestrutura urbana baixa.

Todos os perfis apresentados estão presentes no banco de dados em estudo. A simulação computacional foi feita gerando $M = 100.000$ replicações de amostras de tamanho $n = 1.000$ da log-normal, para cada perfil. Dessa forma, foi possível gerar além de estimativas pontuais, os intervalos de confiança ($\gamma = 95\%$) a partir dos quantis empíricos. Os resultados estão presentes na Figura 15.

Figura 15: Probabilidades de ingresso no emprego formal, informal ou na inatividade por perfil demográfico com intervalos de confiança 95%



Para o perfil 1, o evento de maior probabilidade é o ingresso no emprego informal (46,74%), seguido da inatividade (29,65%) e do emprego formal (23,61%). Já no perfil 2, ocorre um salto da probabilidade referente à ocupação formal (55,72%), o que se deve especialmente pelas características: homem, 16 anos ou mais de estudo e região Sul.

Finalmente, o perfil 3 chama atenção por ter o ingresso na inatividade como o evento mais provável de ocorrer (52,67%). Esse resultado deve-se, possivelmente, ao impacto gerado pelas características: 14 a 20 anos, sem instrução e menos de 1 ano de estudo e região Nordeste. Nota-se também como a entrada no emprego formal é pouco provável para este perfil, apenas 6,37% de probabilidade.

4.5 Limitações

Uma limitação deste trabalho decorre da própria natureza dos dados. Apesar do tempo de desemprego ser, em essência, uma variável contínua, ela é coletada de maneira discreta (em meses) na PNAD Contínua. De fato, para uma pesquisa amostral é improvável pensar em uma forma de coleta mais minuciosa ou pormenorizada. A limitação central, entretanto, é o fato do tempo de desemprego ser autodeclarado e depender, em grande medida, de um apelo à memória do respondente da pesquisa. Nessa perspectiva, existe uma imprecisão natural das respostas obtidas.

Um outra limitação que se destaca é referente aos pesos amostrais. Para permitir estimativas e extrapolações para a população, a PNAD Contínua disponibiliza os pesos amostrais em cada edição da pesquisa. Tais pesos são transversais, isto é, representam o cenário apenas da semana de referência da pesquisa em questão.

Note que o banco de dados analisado neste trabalho é longitudinal, formado por uma concatenação de várias entrevistas de 2021 a 2022. Nesse sentido, os pesos transversais não foram considerados em nenhuma parte da análise. Em verdade, para incorporar os pesos na análise seria necessário um processo de recalibração. Uma proposta de criação de pesos longitudinais para a PNAD Contínua pode ser vista em Júnior et al. (2019).

5 Conclusão

Analisar a duração do desemprego é buscar entender com que intensidade diferentes grupos são afetados por esse fenômeno. Assim, compreender quais grupos estão em situação de desvantagem é essencial para a elaboração de políticas públicas que promovam a inserção laboral desses indivíduos.

Nesse sentido, este trabalho investigou os determinantes da duração do desemprego no Brasil, para o biênio 2021-2022. Por meio de dados da PNAD Contínua trimestral, analisou-se a relação entre o tempo de desemprego até o emprego formal, informal ou inatividade, segundo um conjunto de covariáveis.

A técnica central foi a análise de sobrevivência, a qual permitiu a elaboração de dois modelos de regressão log-normal: um considerando apenas a saída do desemprego, e outro considerando os tempos até cada tipo de saída, ou seja, um modelo em riscos competitivos.

É possível constatar que a regressão na presença de riscos competitivos trouxe uma análise muito mais detalhada do fenômeno, enquanto a interpretação do modelo mais simples pode levar a conclusões até mesmo equivocadas. Um exemplo é a relação entre escolaridade e tempo de desemprego. No modelo mais simples, observou-se um maior tempo mediano de desemprego para os indivíduos com mais anos de estudo. Contudo, no cenário de riscos competitivos ficou claro que mais anos de estudo geram um impacto positivo na redução do tempo até ingresso no emprego formal, assim como um maior tempo mediano até o emprego informal e inatividade.

As diferenças de sexo também chamam muita atenção, pois revelam uma disparidade na procura por ocupação entre homens e mulheres. O tempo mediano até o ingresso no emprego formal para as mulheres foi mais que o dobro do tempo dos homens (2,1 vezes). Olhando para o emprego informal, a situação se agrava: as mulheres levaram um tempo mediano de 2,4 vezes do tempo dos homens para ingressar em uma ocupação informal. Já em relação à inatividade, não se constatou diferença significativa entre homens e mulheres.

Os efeitos associados a raça/cor também são amplamente discutidos na literatura. Neste trabalho, destaca-se: pessoas não brancas apresentaram maior exposição ao emprego informal, com tempo até aquisição da ocupação informal cerca de 8% menor que o tempo de pessoas brancas. Para o emprego formal e inatividade não se observou diferença significativa entre o tempo de brancos e não brancos.

Sobre a posição no domicílio, o modelo mais simples não identificou diferenças entre a pessoa responsável e o cônjuge ou companheiro(a). Entretanto, com a regressão em riscos competitivos concluiu-se que a entrada no emprego informal, para os cônjuges, leva um tempo mediano aproximadamente 11% maior que o da pessoa responsável pelo domicílio.

Além disso, alguns grupos apresentaram maior propensão ao ingresso na inatividade, sendo eles: indivíduos de 14 a 20 anos, indivíduos com menores níveis de escolaridade (desde sem instrução e menos de 1 ano de estudo até 5 a 8 anos de estudo) e residentes nas regiões Nordeste e Sudeste.

O Índice de Vulnerabilidade Social (IVS) dimensão infraestrutura urbana trouxe um elemento conjuntural à análise, para além da região. A partir disso, ficou claro que os residentes de locais com piores condições de infraestrutura urbana (saneamento básico, coleta de lixo e mobilidade urbana) apresentaram as maiores durações de desemprego, em todas as causas analisadas.

Por fim, o modelo paramétrico em riscos competitivos permitiu o cálculo das probabilidades de ingresso no emprego formal, informal ou inatividade para qualquer conjunto de covariáveis de interesse. Tal processo foi feito por meio de simulação computacional, onde amostras da log-normal são geradas de forma independente.

Tendo em vista os resultados supracitados, depreende-se que o uso do modelo de regressão em riscos competitivos permitiu interpretações muito mais minuciosas que o do modelo mais simples. A complexidade decorrente do uso de riscos competitivos foi mitigada ao se assumir os tempos das causas de falhas T_j 's independentes. Desse modo, a estimação dos coeficientes da regressão se torna similar, do ponto de vista computacional, ao realizado no modelo simples.

Em suma, a regressão em riscos competitivos é uma técnica que se mostrou de grande utilidade na análise do tempo de desemprego, permitindo explorações aprofundadas do fenômeno. A contrapartida que se nota é uma maior complexidade teórica do modelo, mas que não se caracteriza como um impeditivo, nem mesmo parece superar as vantagens listadas.

Sugestões ou propostas futuras incluem a repetição deste estudo para os anos de 2023 e 2024, a fim de verificar se houve mudanças nos determinantes da duração do desemprego no Brasil no período pós-pandemia da covid-19. A análise contínua desses dados pode fornecer percepções valiosas sobre as tendências e os possíveis impactos de longo prazo da crise sanitária nas dinâmicas do mercado de trabalho no país.

Referências

- BARLOW, R. E.; CAMPO, R. A. *Total time on test processes and applications to failure data analysis*. [S.l.], 1975.
- BARROS, R. P. d.; CAMARGO, J. M.; MENDONÇA, R. S. P. d. A estrutura do desemprego no brasil. Instituto de Pesquisa Econômica Aplicada (Ipea), 1997.
- BIAZATTI, E. C.; NAKANO, E. Y. Uma proposta de orientação para o uso de modelos contínuos em dados de sobrevivência discretos. *REMAT: Revista Eletrônica da Matemática*, v. 6, n. 2, p. e4002, jul. 2020.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. 1. ed. [S.l.]: Editora Blucher, 2006. 60–61,124–125,145–146 p.
- COX, D.; OAKES, D. *Analysis of survival data*. [S.l.]: Chapman & Hall/CRC, 1984.
- IBGE. *Pesquisa Nacional por Amostra de Domicílios Contínua: Notas técnicas Versão 1.12*. IBGE Rio de Janeiro, 2022. 34-37 p. Disponível em: https://biblioteca.ibge.gov.br/visualizacao/livros/liv101992_notas_tecnicas.pdf.
- IPEA – Instituto de Pesquisa Econômica Aplicada. *Vulnerabilidade social no Brasil: Conceitos, métodos e primeiros resultados para municípios e regiões metropolitanas brasileiras*. Brasília, 2018.
- JOHANSEN, S. The product limit estimator as maximum likelihood estimator. *Scandinavian Journal of Statistics*, [Board of the Foundation of the Scandinavian Journal of Statistics, Wiley], v. 5, n. 4, p. 195–199, 1978.
- JÚNIOR, A. E. T. et al. Pesos longitudinais para a pesquisa nacional por amostra de domicílios contínua (PNAD Contínua). Instituto de Pesquisa Econômica Aplicada (IPEA), 2019.
- JÚNIOR, J. R. d. A. C.; OLIVEIRA, P. F. A. d. Desigualdade de gênero da duração do desemprego e seus efeitos sobre os salários aceitos no brasil. *Revista Econômica do Nordeste*, 2009.
- KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, Taylor & Francis, v. 53, n. 282, p. 457–481, 1958.
- KIEFER, N. M. Economic duration data and hazard functions. *Journal of Economic Literature*, American Economic Association, v. 26, n. 2, p. 646–679, 1988.
- KLEINBAUM, D. G.; KLEIN, M. *Survival analysis a self-learning text*. 3. ed. [S.l.]: Springer, 2012.
- LANCASTER, T. Econometric methods for the duration of unemployment. *Econometrica*, [Wiley, Econometric Society], v. 47, n. 4, p. 939–956, 1979.
- LAWLESS, J. F. *Statistical models and methods for lifetime data*. 2. ed. [S.l.]: John Wiley & Sons, 2003.

- MAIA, M. A. Modelagem do risco de crédito via modelo de riscos competitivos. Dissertação (Mestrado em Estatística) - Universidade de Brasília, Brasília, 2018.
- MENEZES-FILHO, N. A.; PICCHETTI, P. Os determinantes da duração do desemprego em São Paulo. Instituto de Pesquisa Econômica Aplicada (Ipea), 2000.
- NAKANO, E. Y. Um curso de análise de sobrevivência. *Departamento de Estatística, Universidade de Brasília, Brasília*, 2017.
- NAKANO, E. Y.; CARRASCO, C. G. Uma avaliação do uso de um modelo contínuo na análise de dados discretos de sobrevivência. *TEMA: Tendências em Matemática Aplicada e Computacional*, v. 7, n. 1, p. 91–100, 2006.
- NAKANO, E. Y.; RODRIGUES, J. Uma extensão do modelo weibull bivariado de ryu: Uma aplicação bayesiana em riscos competitivos. *Revista Mat. Est.*, v. 24, n. 4, p. 99–115, 2006.
- NETER, J. et al. *Applied linear statistical models*. 5. ed. Chicago: Irwin, 2005. 408–409 p.
- R CORE TEAM. R: A language and environment for statistical computing. Vienna, Austria, 2013. Disponível em: (<http://www.R-project.org>.)
- REIS, M.; AGUAS, M. Duração do desemprego e transições para o emprego formal, a inatividade e a informalidade. *Economia Aplicada*, SciELO Brasil, v. 18, p. 35–50, 2014.
- WROBLEVSKI, B.; CUNHA, M. Silva da. Duração do desemprego e crise econômica: Uma análise sob riscos competitivos para o brasil no período de 2012 até 2019. *Revista Brasileira de Estudos Regionais e Urbanos*, v. 15, n. 3, p. 489–511, 2022.

Apêndice

A Categoria “Outros” da variável posição no domicílio

Tabela 5: Frequência das categorias de posição no domicílio agrupadas como “Outros”

Posição no domicílio	Frequência	Frequência relativa
Neto(a)	3.082	30,43%
Irmão ou irmã	1.983	19,58%
Outro parente	1.532	15,13%
Genro ou nora	1.331	13,14%
Enteado(a)	1.131	11,17%
Pai, mãe, padrasto ou madrasta	553	5,46%
Convivente	238	2,35%
Agregado(a)	206	2,03%
Bisneto(a)	33	0,33%
Sogro(a)	32	0,32%
Pensionista	5	0,05%
Empregado(a) doméstico(a)	1	0,01%
Parente do(a) empregado(a) doméstico(a)	1	0,01%

B Intervalos de confiança dos coeficientes estimados no modelo de regressão log-normal em riscos competitivos

Tabela 6: Intervalos de confiança (95%) dos coeficientes estimados no modelo de regressão log-normal em riscos competitivos

	Emprego formal		Emprego informal		Inatividade	
	Estimativa	IC - 95%	Estimativa	IC - 95%	Estimativa	IC - 95%
β_0	6,015***	(5,714 ; 6,316)	3,134***	(2,965 ; 3,302)	3,947***	(3,768 ; 4,125)
$\log(\sigma)$	0,799***	(0,783 ; 0,815)	0,692***	(0,680 ; 0,703)	0,451***	(0,436 ; 0,467)
Sexo						
Homem	-	-	-	-	-	-
Mulher	0,755***	(0,698 ; 0,812)	0,859***	(0,816 ; 0,901)	0,037	(-0,007 ; 0,082)
Raça/Cor						
Branca	-	-	-	-	-	-
Não branca	-0,037	(-0,100 ; 0,025)	-0,083***	(-0,131 ; -0,036)	-0,006	(-0,055 ; 0,043)
Idade						
14 a 20 anos	-	-	-	-	-	-
21 a 30 anos	-0,093*	(-0,181 ; -0,006)	-0,014	(-0,079 ; 0,050)	0,471***	(0,410 ; 0,533)
31 a 40 anos	-0,124*	(-0,226 ; -0,022)	-0,075*	(-0,150 ; -0,001)	0,551***	(0,477 ; 0,626)
41 a 50 anos	-0,103	(-0,214 ; 0,009)	-0,044	(-0,126 ; 0,037)	0,591***	(0,507 ; 0,674)
51 a 70 anos	0,306***	(0,180 ; 0,432)	0,245***	(0,154 ; 0,335)	0,458***	(0,367 ; 0,549)
Escolaridade						
Sem instrução e menos de						
1 ano de estudo	-	-	-	-	-	-
1 a 4 anos de estudo	-0,403**	(-0,694 ; -0,112)	-0,089	(-0,240 ; 0,062)	-0,042	(-0,204 ; 0,120)
5 a 8 anos de estudo	-0,322*	(-0,586 ; -0,058)	0,152*	(0,016 ; 0,288)	0,107	(-0,038 ; 0,251)
9 a 11 anos de estudo	-0,536***	(-0,799 ; -0,273)	0,378***	(0,241 ; 0,516)	0,264***	(0,118 ; 0,409)
12 a 15 anos de estudo	-1,017***	(-1,275 ; -0,759)	0,502***	(0,368 ; 0,635)	0,453***	(0,312 ; 0,595)
16 anos ou mais de estudo	-1,381***	(-1,647 ; -1,114)	0,765***	(0,615 ; 0,914)	0,537***	(0,380 ; 0,694)
Posição no domicílio						
Pessoa responsável	-	-	-	-	-	-
Cônjuge ou companheiro(a)	-0,062	(-0,141 ; 0,018)	0,108***	(0,049 ; 0,167)	-0,055	(-0,119 ; 0,009)
Filho(a)	0,392***	(0,313 ; 0,471)	0,487***	(0,427 ; 0,546)	-0,038	(-0,100 ; 0,024)
Outros	0,451***	(0,346 ; 0,557)	0,607***	(0,528 ; 0,685)	0,165***	(0,083 ; 0,246)
Região						
Norte	-	-	-	-	-	-
Nordeste	0,331***	(0,227 ; 0,435)	0,246***	(0,177 ; 0,315)	-0,130***	(-0,202 ; -0,059)
Sudeste	-1,035***	(-1,136 ; -0,934)	-0,182***	(-0,254 ; -0,110)	-0,139***	(-0,216 ; -0,063)
Sul	-1,220***	(-1,339 ; -1,101)	0,301***	(0,208 ; 0,394)	0,482***	(0,379 ; 0,584)
Centro-Oeste	-1,114***	(-1,237 ; -0,991)	-0,308***	(-0,398 ; -0,217)	0,040	(-0,063 ; 0,143)
IVS infra. urbana						
Muito baixa	-	-	-	-	-	-
Baixa	0,084*	(0,013 ; 0,155)	0,024	(-0,026 ; 0,075)	0,077**	(0,026 ; 0,128)
Média	0,461***	(0,367 ; 0,554)	0,477***	(0,405 ; 0,548)	0,523***	(0,446 ; 0,599)
Alta	1,223***	(1,119 ; 1,327)	1,079***	(0,996 ; 1,162)	0,958***	(0,867 ; 1,049)

Nota: Valores significativos ao nível de 0,001***, 0,01** e 0,05*.