



**Universidade de Brasília
Departamento de Estatística**

**Estudo da relação entre comportamentos e bem estar psicológico de
universitários**

Eduardo Felipe Machado Côrtes

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**

Eduardo Felipe Machado Côrtes

**Estudo da relação entre comportamentos e bem estar psicológico de
universitários**

Orientador(a): Prof. 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

Inicialmente, gostaria de demonstrar minha gratidão geral com os amigos que fiz, oportunidades que tive e professores incríveis que tive o prazer de ter aula durante minha caminhada ao longo do curso. Menciono aqui apenas alguns dos meus amigos para que os agradecimentos sejam sucintos: Gabriel Reis, Marcelo Fleury, Laura Teixeira, Jéssica Abreu, Davi Guerra e Silvânia Caribé e muitos outros que, para além de ótimas companhias, também me agregaram muito no meu crescimento acadêmico, pessoal e profissional.

Em especial, agradeço ao João Lucas, meu melhor amigo e quase irmão, por sempre estar me dando suporte quando eu mais preciso e por manter a nossa conexão viva e sintonizada como sempre. Também ao Vida, que para além de meu amor, é uma das companhias mais insuperáveis que eu tenho hoje, mesmo que de longe.

Também agradeço à minha família, que sempre me deu sustento e estrutura para chegar até onde eu cheguei e ainda vou chegar. Que as nossas relações se tornem mais parcimoniosas nessa nova etapa.

E por último mas nem um pouco menos importante, ao meu orientador Leandro e ao professor Luís Vinha, que toparam executar e me ajudar nessa ideia maluca de tema que não facilitou nada nossas vidas, mas que foi muito importante para que eu estudasse mais e mais a fim de conseguir uma solução para todos os problemas. Sem vocês, este resultado nunca teria chegado.

Resumo

A relevância de problemas psicossociais como depressão e ansiedade tem aumentado nas últimas décadas e, principalmente após a pandemia do COVID-19, grupos como dos estudantes universitários tendem a ser mais afetados devido a sua vulnerabilidade na manutenção do bem-estar e do humor. Dessa forma, este trabalho relaciona características demográficas e comportamentos apresentados por universitários de uma universidade do Distrito Federal, coletados através do Questionário Sobre Qualidade de Vida e de Saúde dos Estudantes do ano 2021, com a escala de depressão. Isto é realizado por meio da utilização da análise fatorial para mensurar o fator depressão a partir da escala DASS-21 (Depression, Anxiety and Stress Scale - 21 Items) e do modelo logístico beta inflacionado em zero e um para estudar os efeitos e significâncias das variáveis demográficas e comportamentais. Os resultados demonstraram que o fator depressão é significativamente relacionado com variáveis como, por exemplo, o ato de chorar, de praticar atividades físicas, de fazer uso de álcool, medicamentos ou substâncias e de não se identificar como heterossexual.

Palavras-chaves: Depressão; Humor; Análise Fatorial; Modelo Logístico Beta Inflacionado em Zero e Um; Comportamento; Demografia; Psicometria.

Lista de Tabelas

1	Variáveis do banco tratado relacionadas ao comportamento	17
2	Variáveis do banco tratado relacionadas a características demográficas . . .	17
3	Rótulos de cada variável para os gráficos da análise exploratória	19
4	Relação entre cada parâmetro estimado com sua função de ligação e sua notação de coeficiente	28
5	Coeficientes de regressão estimados para μ do fator depressão	29
6	Coeficientes de regressão estimados para ν do fator depressão	31
7	Coeficientes de regressão estimados para τ do fator depressão	32
8	Medidas resumo dos resíduos do modelo para o fator depressão	35
9	Variáveis que foram significativas para cada esfera do modelo. Em cores, se apresenta em verde e vermelho as significâncias que representam respec- tivamente um impacto positivo e negativo para o indivíduo	36

Lista de Figuras

1	Matriz de gráficos cruzando todas as variáveis de comportamento dos universitários.	20
2	Matriz de gráficos cruzando todas as variáveis demográficas dos universitários.	21
3	Matriz de gráficos cruzando as frequências ou distribuições das variáveis demográficas (nas colunas) com as de comportamento (nas linhas).	22
4	Matriz cruzando as variáveis demográficas nas colunas com as de comportamento nas linhas, realizando os respectivos testes de comparação de médias e de independência.	23
5	Gráficos de histograma e densidade empírica (escalonada para acompanhar o eixo de frequência do histograma), à esquerda com todos os pontos e à direita com os pontos de massa em zero e um (em azul) isolados dos outros pontos.	24
6	Primeira linha da grade: distribuição em boxplot dos pontos de fator ansiedade (eixo y) para cada categoria da variável na respectiva coluna. Segunda linha da grade: proporção de zeros e uns em cada categoria de cada variável.	25
7	Primeira linha da grade: distribuição em boxplot e dispersão dos pontos de fator ansiedade (eixo y) para cada categoria da variável na respectiva coluna. Segunda linha da grade: proporção de zeros e uns em cada categoria de cada variável e distribuição de idade em boxplot nos casos zero e um.	26
8	Efeitos fixos das variáveis para o parâmetro μ para o fator depressão	30
9	Efeitos fixos das variáveis para o parâmetro ν para o fator depressão	31
10	Efeitos fixos das variáveis para o parâmetro τ para o fator depressão	33
11	Diagnóstico de estrutura das funções de ligação e de pontos influentes do modelo para o fator depressão	34
12	Diagnóstico de normalidade dos resíduos do modelo para o fator depressão	34
13	Diagnóstico de ajuste e adequabilidade do modelo para o fator depressão com <i>wormplot</i> , <i>DTOP</i> e simulação envelope	35

Sumário

1 Introdução	8
2 Referencial Teórico	10
2.1 Distribuição de probabilidade	10
2.1.1 Distribuição de Bernoulli	10
2.1.2 Distribuição beta	10
2.1.3 Distribuição beta inflacionada em zero e um (BIZU)	11
2.2 Análise fatorial	12
2.3 Modelo	14
2.3.1 Regressão logística beta inflacionada em zero e um (RLBIZU)	14
3 Metodologia	16
3.1 Base de dados.	16
3.2 Software e ferramentas.	18
3.3 Nível de significância.	18
4 Resultados	19
4.1 Análise exploratória	19
4.1.1 Variáveis de comportamento	19
4.1.2 Variáveis demográficas	20
4.1.3 Variáveis comportamentais e demográficas	22
4.1.4 Fator depressão	24
4.2 Variáveis significativas para o fator psicológico	27
4.2.1 Interpretação do modelo	28
5 Conclusão	36
Referências	38

1 Introdução

Nas últimas décadas, a popularidade e relevância dos tópicos acerca de saúde mental têm aumentado, dado que uma a cada 8 pessoas da população mundial em 2019 viviam com transtornos mentais como depressão e ansiedade (WORLD HEALTH ORGANIZATION, 2022b). No Brasil, especificamente a depressão e a ansiedade são a quinta e a sexta maiores causas de anos vividos em incapacidade (MARINHO et al., 2018). Esse cenário tornou-se ainda mais crítico devido à pandemia do COVID-19, em que houve um aumento ainda maior os casos de depressão em 2021 (ROCHA, 2022).

Em especial para estudantes universitários, problemas psicossociais, tal como a depressão, são bastante comuns. Este grupo tende a possuir uma baixa qualidade de vida não somente no domínio psicológico, mas também em outros como o físico, o de ambiente e o nível de independência. Tais condições, quando não diagnosticadas e tratadas adequadamente, podem ocasionar evasões acadêmicas, que são graves para o ensino público, para a sociedade e, principalmente, para o próprio estudante (CERCHIARI et al., 2004).

Esses transtornos mentais afetam o bem-estar e o humor, que é reconhecido como elemento central no comportamento humano. Análises a respeito do monitoramento, autocontrole e autorregulação do humor têm ganhado interesse e espaço nas pesquisas comportamentais, uma vez que comportamentos cotidianos das pessoas incluem como um de seus objetivos a manutenção do bem-estar (THAYER; NEWMAN; MCCLAIN, 1994).

Nesse contexto, pesquisas propostas por Thayer, Newman e McClain (1994) buscaram estudar quais comportamentos e atividades são praticados para lidar com o humor, a fim de categorizar e encontrar estratégias que são utilizadas para sua regulação. Tendo em vista que a depressão é um transtorno mental centrado no humor em si e no impacto causado no interesse e realização de atividades (WORLD HEALTH ORGANIZATION, 2022a), se torna pertinente ter como base o estudo dessas estratégias para o bem-estar como forma de manejar também a depressão em si.

Tendo isso em vista, o presente trabalho visa avaliar quais dos comportamentos levantados pelo estudo de Thayer, Newman e McClain (1994) apresentam evidências estatísticas de impacto no estado psicológico de universitários. Para tal, serão construídas variáveis para indicar a prática dessas atividades pelos estudantes, as suas características demográficas e seus escores numa escala psicométrica para a depressão, esta última me-

didada a partir do uso da análise fatorial sob itens da escala do DASS-21¹ (LOVIBOND, 1983). Utilizando estas variáveis, será ajustado um modelo de regressão logística beta inflacionada em zero e um (RLBIZU) (MARTINEZ, 2008) para então entender seus efeitos no fator depressão e quais são estatisticamente significativas.

¹21-item version of the Depression Anxiety Stress Scale (a versão de 21 itens da escala de depressão, ansiedade e estresse, LOVIBOND, 1983, tradução nossa)

2 Referencial Teórico

Nesta seção são apresentados os conceitos técnicos específicos utilizados para o estudo.

2.1 Distribuição de probabilidade

Distribuições de probabilidade são modelos teóricos que buscam representar probabilisticamente como se comporta a distribuição de uma variável aleatória, possuindo diversas distribuições propostas para casos específicos ou gerais.

As notações utilizadas são:

- Variáveis aleatórias (v.a.): denotadas por letras maiúsculas, normalmente X e Y ,
- Distribuição de probabilidade de uma variável: denotada por $X \sim \text{Distribuição}(\theta)$, em que θ é o vetor de parâmetros do modelo em questão,
- Probabilidade de um evento: denotado por $P(X = x) = p_X(x)$, a função de probabilidade para variáveis discretas, ou $f_X(x)$, a função densidade de probabilidade para variáveis contínuas, sendo o evento um possível valor (x) que uma variável aleatória (X) pode assumir.

2.1.1 Distribuição de Bernoulli

Esta distribuição visa modelar variáveis aleatórias que podem assumir valores dicotômicos para sucesso ou fracasso de um evento, como por exemplo o evento de cair cara (sucesso) numa moeda ao invés de coroa (fracasso).

Sua distribuição é denotada por $X \sim \text{Bernoulli}(p)$, em que $p \in [0, 1]$ é a probabilidade de sucesso do evento ($X = 1$) e sua função de probabilidade é

$$P(X = x) = \begin{cases} p^x(1-p)^{1-x}, & x \in \{0, 1\} \\ 0, & \text{caso contrário} \end{cases} \quad (2.1.1)$$

2.1.2 Distribuição beta

Dada uma v.a. contínua X , se diz que ela assume uma distribuição beta com a notação $X \sim \text{Beta}(p, q)$, $p, q > 0$ sendo parâmetros de forma, com densidade de probabi-

lidade

$$f_X(x) = \begin{cases} \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}, & y \in (0, 1) \\ 0, & \text{caso contrário} \end{cases} \quad (2.1.2)$$

onde $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ é a função gama.

Para este trabalho, será utilizada uma forma reparametrizada da distribuição beta proposta por Martinez (2008). Definindo $\mu = p/(p+q)$ e $\phi = p+q$, se implica que $p = \mu\phi$, $q = (1-\mu)\phi$, $\mu \in (0, 1)$ e $\phi > 0$.

Assim, a nova notação se torna $X \sim \text{Beta}(\mu, \phi)$, com as novas interpretações dos parâmetros sendo média para μ e precisão para ϕ , também com nova densidade de probabilidade definida como

$$f_X(x) = \begin{cases} \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} x^{\mu\phi-1}(1-x)^{(1-\mu)\phi-1}, & x \in (0, 1) \\ 0, & \text{caso contrário} \end{cases} \quad (2.1.3)$$

2.1.3 Distribuição beta inflacionada em zero e um (BIZU)

Martinez (2008) propôs a distribuição beta inflacionada em zero e um (BIZU) como uma forma de contornar a limitação de domínio da distribuição beta ao permitir que ela inclua os limites do intervalo $[0, 1]$.

Isso é feito a partir da mistura de distribuições entre a beta e a bernoulli. Misturas são basicamente uma forma de unir diversas distribuições diferentes em uma única função ou densidade de probabilidade f de forma ponderada, que é expressa como

$$f_X(x) = \sum_{i=1}^k \pi_i f_i(x), \quad (2.1.4)$$

onde:

- k : quantidade de distribuições sendo misturadas,
- π_i , $i = 1, \dots, k$: peso da i -ésima distribuição da mistura, tendo que $\pi_i \in (0, 1)$ e $\sum_{i=1}^k \pi_i = 1$
- $f_i(\cdot)$, $i = 1, \dots, k$: função ou densidade de probabilidade da i -ésima distribuição.

Dessa forma, no caso da BIZU, temos uma mistura de apenas duas distribuições,

que pode ser definida na forma

$$bizu(x; \alpha, p, \mu, \phi) = \alpha Bernoulli(x; p) + (1 - \alpha)Beta(x; \mu, \phi), \quad (2.1.5)$$

onde

- α : peso da distribuição bernoulli na mistura
- $Bernoulli(.; p)$: função de probabilidade de uma distribuição bernoulli com parâmetro p
- $Beta(.; \mu, \phi)$: função de densidade de uma distribuição beta com parâmetros μ e ϕ .

A forma de parametrização que será utilizada neste trabalho, também proposta por Martinez (2008), é a denotada por $X \sim bizu(\delta_0, \delta_1, \mu, \phi)$. Nesta forma, $\delta_0 = \alpha(1 - p)$ e $\delta_1 = \alpha p$ representam respectivamente os parâmetros de probabilidade dos eventos $X = 0$ e $X = 1$ ocorrerem, visto que consideram as probabilidades de p da bernoulli e α , a probabilidade da bernoulli ser a selecionada para a variável. Sua função de densidade é:

$$f_X(x) = \begin{cases} \delta_0, & x = 0 \\ \delta_1, & x = 1 \\ (1 - \delta_0 - \delta_1)Beta(x; \mu, \phi), & x \in (0, 1) \\ 0, & \text{caso contrário.} \end{cases} \quad (2.1.6)$$

2.2 Análise fatorial

A Análise Fatorial é uma técnica que visa construir fatores que mensuram uma variável latente (não observável, como por exemplo grau de depressão). Para tal, ela se utiliza de variáveis observadas e sua estrutura de covariância, reduzindo suas dimensões para gerar esses fatores (JOHNSON; WICHERN, 2007).

O modelo de análise fatorial pode ser expresso pela fórmula

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{L} \mathbf{F} + \boldsymbol{\epsilon}, \quad (2.2.1)$$

$\begin{matrix} (p \times 1) & & (p \times m)(m \times 1) & + & (p \times 1) \end{matrix}$

onde:

- p é a quantidade de variáveis

- m é a quantidade de fatores a ser criada
- \mathbf{X} é um vetor representando uma observação com p variáveis
- $\boldsymbol{\mu}$ é um vetor representando a média de cada uma das p variáveis
- \mathbf{L} é a matriz de cargas fatoriais, onde l_{ij} é a carga da variável i no fator j
- $\boldsymbol{\epsilon}$ é o vetor de erro (ou fator específico)

Note que o fator específico ϵ_i está associado apenas à i -ésima resposta X_i . Os p desvios $X_1 - \mu_1, X_2 - \mu_2, \dots, X_p - \mu_p$ são expressos em termos de $p + m$ variáveis aleatórias $F_1, F_2, \dots, F_m, \epsilon_1, \epsilon_2, \dots, \epsilon_p$ que são não observáveis. Os vetores aleatórios \mathbf{F} e $\boldsymbol{\epsilon}$ são independentes e satisfazem as seguintes condições:

$$E(\mathbf{F}) = \mathbf{0}, Cov(\mathbf{F}) = \mathbf{I} \quad (2.2.2)$$

$$E(\boldsymbol{\epsilon}) = \mathbf{0}, Cov(\boldsymbol{\epsilon}) = \boldsymbol{\Psi} \quad (2.2.3)$$

em que $\boldsymbol{\Psi}$ é a matriz diagonal.

O método de estimação utilizado neste estudo é o de componentes principais. A matriz de covariância da amostra \mathbf{S} é especificada em termos de seus pares de autovalores e autovetores.

$$\boldsymbol{\Sigma} = \sum_{i=1}^p \lambda_i \mathbf{e}_i \mathbf{e}_i' \cong \sum_{i=1}^m \lambda_i \mathbf{e}_i \mathbf{e}_i' = \begin{pmatrix} \sqrt{\lambda_1} \mathbf{e}_1 & \sqrt{\lambda_2} \mathbf{e}_2 & \dots & \sqrt{\lambda_m} \mathbf{e}_m \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} \mathbf{e}_1' \\ \sqrt{\lambda_2} \mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_m} \mathbf{e}_m' \end{pmatrix} = \mathbf{L}\mathbf{L}' \quad (2.2.4)$$

Em vez de somar de 1 a p (todas as variáveis), agora soma-se de 1 a m (todos os fatores), ignorando os últimos $p - m$ termos na soma. Assim, o objetivo da expressão é aproximar $\sum_{i=1}^p \lambda_i \mathbf{e}_i \mathbf{e}_i'$ de $\sum_{i=1}^m \lambda_i \mathbf{e}_i \mathbf{e}_i'$. A representação em 2.2.4 assume que $\boldsymbol{\epsilon}$ não tem muita importância e são ignorados em $\boldsymbol{\Sigma}$.

Se fatores específicos forem incluídos no modelo, suas variâncias podem ser tomadas como os elementos diagonais de $\boldsymbol{\Sigma} - \mathbf{L}\mathbf{L}'$. Assim, a aproximação pode ser expressa de seguinte maneira:

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi} \quad (2.2.5)$$

em que ψ_i é dada por:

$$\hat{\psi}_i = \sigma_{ii}^2 - \sum_{j=1}^m \lambda_j \hat{e}_{ji}^2 \quad (2.2.6)$$

2.3 Modelo

2.3.1 Regressão logística beta inflacionada em zero e um (RLBIZU)

A técnica de regressão consiste em estabelecer um modelo entre uma variável resposta e suas variáveis explicativas, de forma que os valores assumidos pelas explicativas consigam ser transformados algebricamente para predizer com certo nível de precisão o valor estimado da variável resposta.

Sejam Y_1, \dots, Y_n variáveis aleatórias independentes, cada uma com distribuição $Y_i \sim \text{bizu}(\delta_{0i}, \delta_{1i}, \mu_i, \phi)$, $i = 1, \dots, n$. O modelo de regressão logístico beta inflacionado em zero e um (RLBIZU) é definido pela BIZU (2.1.3) e pelas funções de ligação dos componentes sistemáticos

$$g(\mu_t) = \sum_{i=1}^k X_{ti} \beta_i = \eta_t = \log \left(\frac{\mu_t}{1 - \mu_t} \right),$$

$$H(\delta_{0t}, \delta_{1t}) = (h_0(\delta_{0t}, \delta_{1t}), h_1(\delta_{0t}, \delta_{1t})) = (\zeta_{0t}, \zeta_{1t}) = \left(\log \left(\frac{\delta_{0t}}{1 - \delta_{0t} - \delta_{1t}} \right), \log \left(\frac{\delta_{1t}}{1 - \delta_{0t} - \delta_{1t}} \right) \right), \quad (2.3.1)$$

onde:

- \log : função logaritmo com base e ,
- $\mu_t = E(y_t | y_t \in (0, 1))$,
- $\delta_{0t} = P(Y_t = 0)$,
- $\delta_{1t} = P(Y_t = 1)$,
- $1 - \delta_{0t} - \delta_{1t} = P(Y_t \in (0, 1))$,
- k, k_0 e k_1 : quantidades de variáveis exógenas conhecidas para cada parâmetro,
- $\mathbf{X}_t = (X_{t1}, \dots, X_{tk})^T$, $\mathbf{V}_t = (V_{t1}, \dots, V_{tk_0})^T$, $\mathbf{Z}_t = (Z_{t1}, \dots, Z_{tk_1})^T$: valores observados de cada respectiva quantidade de variáveis exógenas,

- $\beta \in \mathbb{R}^k$, $\rho \in \mathbb{R}^{k_0}$, $\gamma \in \mathbb{R}^{k_1}$: vetores de parâmetros desconhecidos a serem estimados,
- $\eta_t = \mathbf{X}_t^T \beta$: preditor linear de $g(\mu_t)$,
- $\zeta_{0t} = \mathbf{V}_t^T \rho$: preditor linear de $\log \left(\frac{\delta_{0t}}{1 - \delta_{0t} - \delta_{1t}} \right)$,
- $\zeta_{1t} = \mathbf{Z}_t^T \gamma$: preditor linear de $\log \left(\frac{\delta_{1t}}{1 - \delta_{0t} - \delta_{1t}} \right)$.

Assim,

$$\begin{aligned}\delta_{0t} &= P(Y_t = 0) = \frac{e^{\zeta_{0t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}}, \\ \delta_{1t} &= P(Y_t = 1) = \frac{e^{\zeta_{1t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}}, \\ 1 - \delta_{0t} - \delta_{1t} &= P(Y_t \in (0, 1)) = \frac{1}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}}.\end{aligned}$$

3 Metodologia

3.1 Base de dados

A base de dados utilizada é oriunda do Questionário Sobre Qualidade de Vida e de Saúde dos Estudantes, uma pesquisa realizada em 2021 com o intuito de averiguar o estado geral dos integrantes dos campi de uma universidade do Distrito Federal. Este banco possui 1157 respostas de estudantes da graduação e pós-graduação, contendo 161 colunas de variáveis contemplando informações biográficas, demográficas, acadêmicas, de saúde, sociais, comportamentais, psicológicas, entre outras.

Por meio destas informações, foi construído um novo banco de dados tratado e adaptado para este estudo.

É importante ressaltar que o banco de dados original possui um alto viés amostral, visto que muitas das análises exploratórias evidentemente não refletem a realidade dos campi. Ainda, a adaptação realizada para adquirir as informações que seriam úteis para este trabalho foi bastante limitada, mas ao menos suficiente. Dessa forma, as conclusões expostas nos resultados deste estudo não são inferenciais, mas sim meramente exploratórios e para aplicação das técnicas estatísticas.

Primeiramente, foram estabelecidas as variáveis comportamentais. Elas foram definidas em função de evidências de comportamentos dentre os levantados pelo estudo de Thayer, Newman e McClain (1994) que foram evidenciadas de alguma forma nos dados. As variáveis são apresentadas na Tabela 1.

Tabela 1: Variáveis do banco tratado relacionadas ao comportamento

Variável	Tipo de variável	Descrição
atividade_fisica	Nominal (booleana)	Relatou a realização de atividades físicas
interacao_social	Nominal (booleana)	Relatou a realização de interações sociais
passatempos	Nominal (booleana)	Relatou a realização de passatempos e <i>hobbies</i>
alcool_e_substancias	Nominal (booleana)	Relatou uso de álcool, medicações e outras substâncias
tecnicas_de_relaxamento	Nominal (booleana)	Relatou a realização de técnicas de relaxamento (e.g. meditação, práticas religiosas)
choro	Nominal (booleana)	Relatou o ato de choro

Segundamente, foram transformadas as variáveis demográficas para limpeza e agrupamento de categorias de forma que fosse mais relevante ao estudo. Esses agrupamentos foram feitos tendo em vista o isolamento de uma categoria social mais privilegiada que as outras. Estas variáveis estão apresentadas na Tabela 2.

Tabela 2: Variáveis do banco tratado relacionadas a características demográficas

Variável	Tipo de variável	Categorias	Descrição
idade	Quantitativa		A idade do respondente
trabalha	Nominal (booleana)		Indicativo para se o respondente trabalha
genero	Categórica	- Homem (cis ou trans) - Não homem ou não declarado	A identidade de gênero do respondente
etnia	Categórica	- Branca - Não branca ou não declarado	A etnia do respondente
sexualidade	Categórica	- Heterossexual - Não heterossexual ou não declarado	A sexualidade do respondente

Finalmente, foi estabelecida a variável para o escore psicométrico da depressão. Para isso, foi utilizado os itens do instrumento de pesquisa DASS-21 do banco original que retratavam o sofrimento psicológico e depressivo. Com base em uma das análises fatoriais de Góes (2022), que teve o foco de construir o fator da depressão a partir do mesmo banco de dados que este estudo utilizou, foram utilizadas as cargas fatoriais que as variáveis apresentaram para realizar uma média ponderada desses itens do questionário com suas respectivas cargas fatoriais, normalizada para ajustar o escore do intervalo de

[0, 3] para [0, 1].

3.2 Software e ferramentas

Para o processamento dos dados e execução das técnicas, foi utilizada a linguagem R versão 4.2.2 e fazendo uso dos pacotes `tidyverse`, `magrittr`, `GGally`, `haven`, `ggplot2`, `gamlss`, `knitr`, `clipr`, `qqplotr`, `gamlss.ggplots`, `hnp`, `cowplot` e `moments`.

3.3 Nível de significância

O nível de significância estabelecido para este estudo foi $\alpha = 5\%$ para as evidências estatísticas.

4 Resultados

4.1 Análise exploratória

Nesta seção são feitas análises estatísticas para compreender como se comportam as variáveis do banco e suas relações, bem como testar as hipóteses da existência de evidências estatísticas para esses relacionamentos. Assim, é possível escolher de forma mais fundamentada a modelagem a ser feita para o fator psicológico e interpretar com maior precisão seus resultados.

Ainda, em função de melhorar leitura dos gráficos, alguns rótulos em formato reduzido foram definidos para as variáveis:

Tabela 3: Rótulos de cada variável para os gráficos da análise exploratória

Variável	Rótulo nos gráficos
atividade_fisica	Ativ. Física
interacao_social	Int. Social
passatempos	Passatempos
alcool_e_substancias	Álcool
tecnicas_de_relaxamento	Relaxa
choro	Chora
idade	Idade
trabalha	Trabalha
genero	Gênero
etnia	Etnia
sexualidade	Sexualidade

4.1.1 Variáveis de comportamento

As variáveis de comportamento do estudo são as definidas com base no trabalho de Thayer, Newman e McClain (1994), conforme foram descritas na Tabela 1.

A Figura 1 é uma matriz de gráficos cruzando todas as variáveis de comportamento. Na diagonal, estão os gráficos de barra de frequência univariada de cada variável e no triângulo superior da matriz estão os gráficos de barra de frequência bivariada entre as variáveis da respectiva linha e coluna. No triângulo inferior estão os testes de hipótese, nesse caso todos de independência χ^2 , entre as variáveis da respectiva linha e coluna. Os textos em azul destacam os testes de hipótese que foram significativos para rejeição da

hipótese nula.

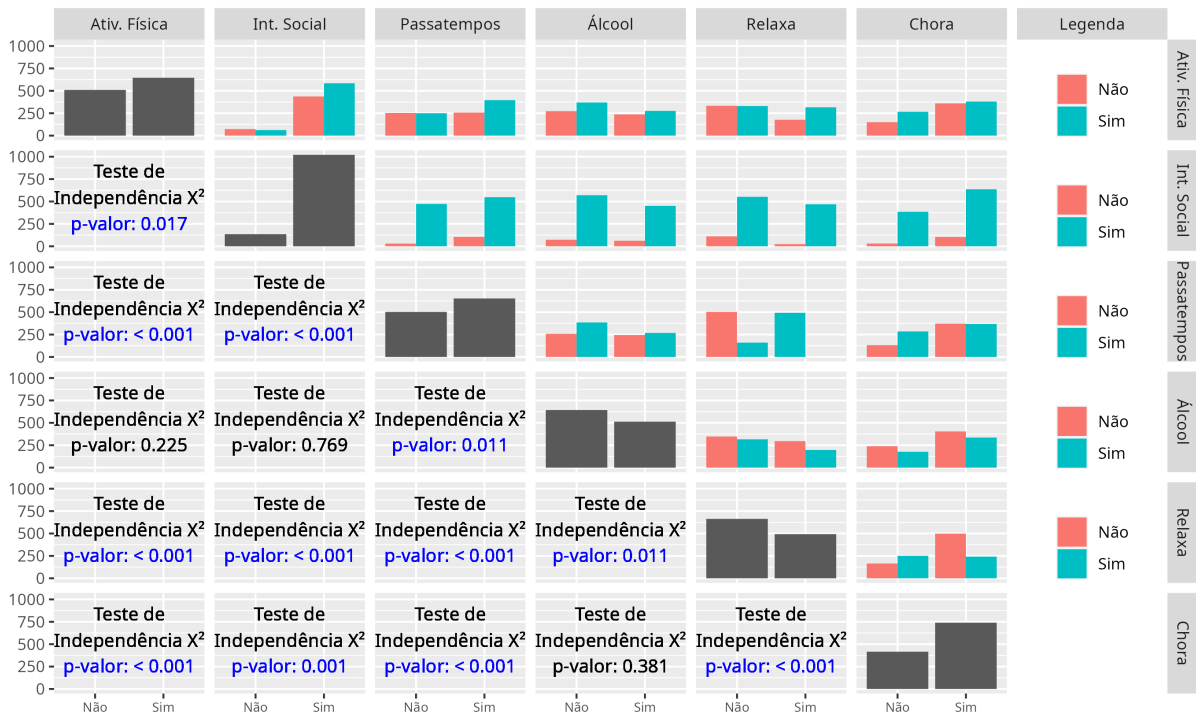


Figura 1: Matriz de gráficos cruzando todas as variáveis de comportamento dos universitários.

Apenas a variável de interação social apresentou grande disparidade de frequência, já que a grande maioria apresentou indícios deste comportamento: 1021(88%) indivíduos de 1156. Isto evidencia o balanceamento das categorias das demais variáveis.

Com relação aos cruzamentos, quase todos se apresentaram estatisticamente significativos para a rejeição do teste de independência. Em especial, os indicadores de passatempo e de técnicas de relaxamento apresentaram dependência com todos os outros comportamentos. Ao mesmo tempo a variável de consumo de álcool, medicamentos ou substâncias é a única que apresentou alguns cruzamentos que não rejeitam a hipótese independência, sendo estes os feitos com a prática de atividade física, com interagir socialmente e com chorar.

Para a modelagem que foi feita, essa análise fornece indícios de quais variáveis podem ser preditores interessantes para a resposta, como por exemplo o álcool.

4.1.2 Variáveis demográficas

As variáveis demográficas foram construídas de forma geral para avaliar a influência que características individuais poderiam ter no fator psicológico. Assim como para as

variáveis comportamentais, as demográficas também estão listadas e descritas na Tabela 2.

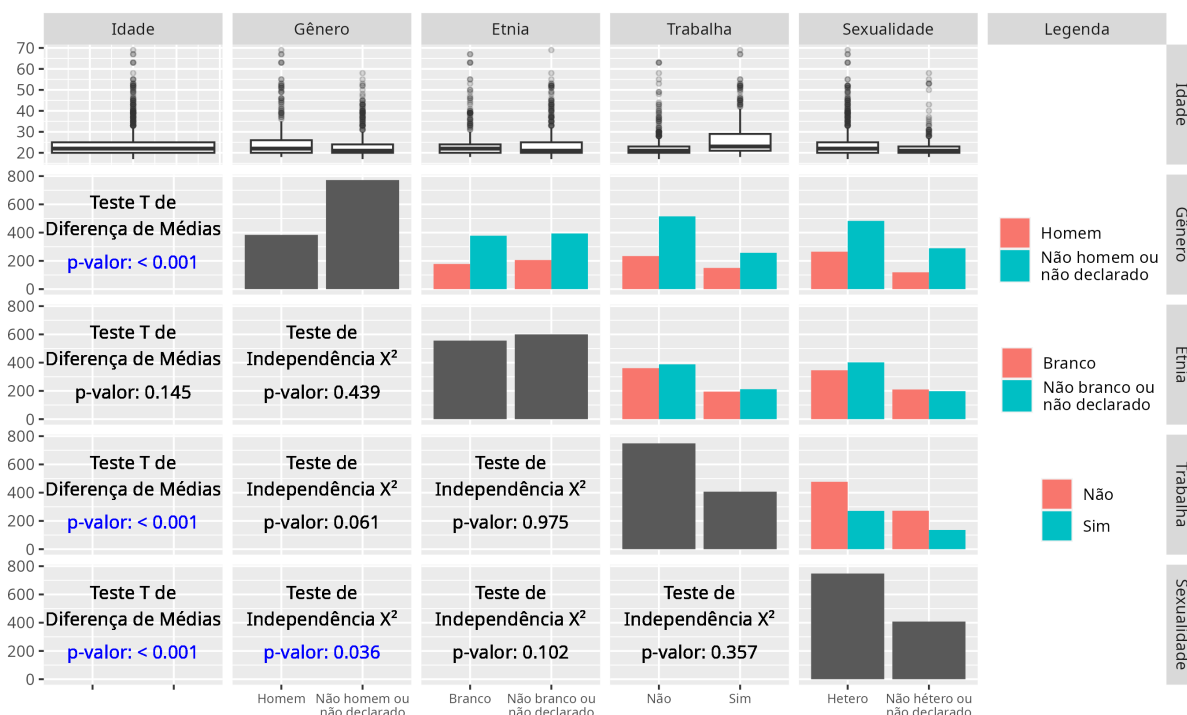


Figura 2: Matriz de gráficos cruzando todas as variáveis demográficas dos universitários.

Diferentemente das variáveis comportamentais, as demográficas apresentam disparidades entre as categorias da maioria das dimensões e poucas apresentaram significância nos testes de independência ou nos testes de diferença de médias.

A maioria dos universitários nesta amostra não se declararam como homem, não trabalham ou se declararam como heterossexuais. Parte dessas discrepâncias se devem ao agrupamento que foi feito das categorias, enquanto que algumas variáveis como o gênero se apresentou como um problema da amostra com relação à qualidade do instrumento de pesquisa, visto que uma parcela grande dos respondentes foram mulheres, fato que não se confere na realidade da universidade.

Alguns dos cruzamentos que foram significativos são coerentes, como por exemplo a relação entre trabalhar e a idade, em que naturalmente se espera que pessoas mais velhas possuam mais trabalhos do que as mais novas. Por outro lado, algumas associações surpreenderam um pouco em não haver dependência considerável, como no caso entre etnia e trabalho e entre gênero e trabalho, ainda que este último quase tenha apresentado evidências suficientes para ser significativa.

4.1.3 Variáveis comportamentais e demográficas

Após análise univariada e bivariada dentro de cada grupo de variáveis, foi avaliada a relação que existe entre cada variável comportamental e cada variável demográfica.



Figura 3: Matriz de gráficos cruzando as frequências ou distribuições das variáveis demográficas (nas colunas) com as de comportamento (nas linhas).

Gênero	Etnia	Trabalha	Sexualidade	Idade	
Teste de independência X ² p-valor: 0.033	Teste de independência X ² p-valor: 0.162	Teste de independência X ² p-valor: 0.615	Teste de independência X ² p-valor: 0.121	Teste T diferença de médias p-valor: 0.817	Ativ. Física
Teste de independência X ² p-valor: 0.365	Teste de independência X ² p-valor: 0.122	Teste de independência X ² p-valor: 0.021	Teste de independência X ² p-valor: 0.546	Teste T diferença de médias p-valor: 0.918	Int. Social
Teste de independência X ² p-valor: 0.859	Teste de independência X ² p-valor: 0.021	Teste de independência X ² p-valor: 0.296	Teste de independência X ² p-valor: < 0.001	Teste T diferença de médias p-valor: 0.371	Passatempos
Teste de independência X ² p-valor: 0.105	Teste de independência X ² p-valor: 0.928	Teste de independência X ² p-valor: 0.003	Teste de independência X ² p-valor: < 0.001	Teste T diferença de médias p-valor: 0.032	Álcool
Teste de independência X ² p-valor: 0.297	Teste de independência X ² p-valor: 0.003	Teste de independência X ² p-valor: 0.702	Teste de independência X ² p-valor: < 0.001	Teste T diferença de médias p-valor: 0.661	Relaxa
Teste de independência X ² p-valor: < 0.001	Teste de independência X ² p-valor: 0.862	Teste de independência X ² p-valor: 0.007	Teste de independência X ² p-valor: < 0.001	Teste T diferença de médias p-valor: < 0.001	Chora

Figura 4: Matriz cruzando as variáveis demográficas nas colunas com as de comportamento nas linhas, realizando os respectivos testes de comparação de médias e de independência.

Observou-se que algumas variáveis possuem muitas relações significativas. A primeira delas é o comportamento de choro, que rejeita a hipótese de independência ou de igualdade de médias para todas as variáveis exceto etnia, com ênfase na variável de gênero, em que se observou uma discrepância grande entre as proporções de pessoas que choram entre homens e entre os não declarados homens.

Outra que se destacou muito é a característica de sexualidade, em que se observou que o grupo que não declarou heterossexualidade apresentou maiores frequências relativas para o teor menos vantajoso da respectiva variável de comportamento. Por exemplo, para esse grupo há uma menor proporção de indícios de passatempos e de técnicas de relaxamento, enquanto há maiores proporções para o choro e para o consumo de álcool, medicamentos ou outras substâncias.

Ainda vale pontuar que algumas relações esperadas se confirmaram significativas, tal como na média de idade ser diferente entre os que consomem e os que não consomem álcool, medicamentos e substâncias, que também é diferente para o choro, considerando o conceito de que pessoas mais novas tendem a apresentar menos estabilidade emocional.

4.1.4 Fator depressão

Conforme explicado na metodologia, o fator depressão foi definido a partir da média ponderada com base nas cargas fatoriais de variáveis de escala psicométricas do DASS-21 do banco de dados original.

Esta variável está distribuída entre zero e um, como se pode observar na Figura 5 abaixo. Foi considerada a utilização da distribuição beta (2.1.2) para modelar esta variável, já que seu domínio é no intervalo unitário e dada a grande flexibilidade que a beta tem em formatar sua densidade.

Para solucionar a limitação da distribuição beta não assumir valores nos limites de $[0, 1]$, já que seu domínio é apenas em $(0, 1)$, foi utilizada sua versão inflacionada em zero e um (BIZU) (2.1.3) para que pudesse haver um ajuste adequado à variável.

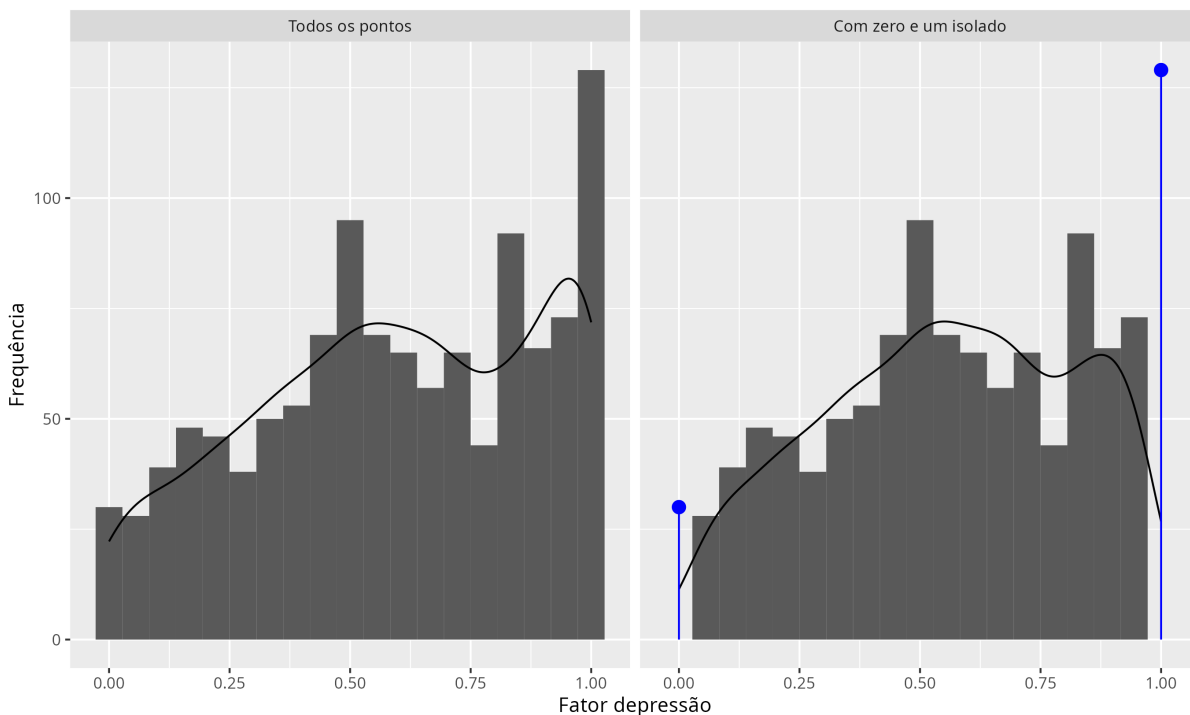


Figura 5: Gráficos de histograma e densidade empírica (escalonada para acompanhar o eixo de frequência do histograma), à esquerda com todos os pontos e à direita com os pontos de massa em zero e um (em azul) isolados dos outros pontos.

Ao isolar as frequências de zero e um da densidade empírica em $(0, 1)$, é perceptível que a curva sofre uma suavização que corrige uma possível indicação de bimodalidade da variável. Isto representa mais um ponto favorável para utilização da BIZU (2.1.3).

Ainda, há uma tendência crescente na frequência de universitários conforme o escore do fator cresce, significando diretamente que há poucos indivíduos com baixos

indícios de depressão dentro desta amostra.

Para avaliar diretamente as relações individuais entre o fator e as demais variáveis, foram feitas matrizes de gráficos com os grupos de comportamentos e demográficos.

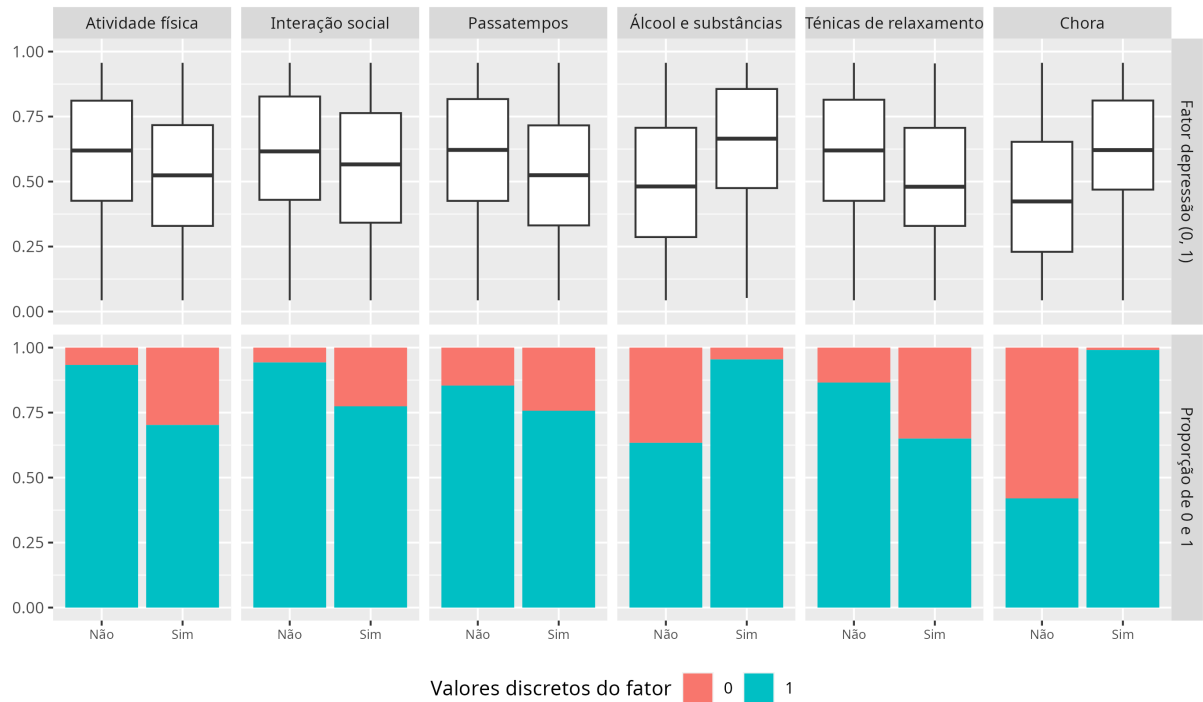


Figura 6: Primeira linha da grade: distribuição em boxplot dos pontos de fator ansiedade (eixo y) para cada categoria da variável na respectiva coluna. Segunda linha da grade: proporção de zeros e uns em cada categoria de cada variável.

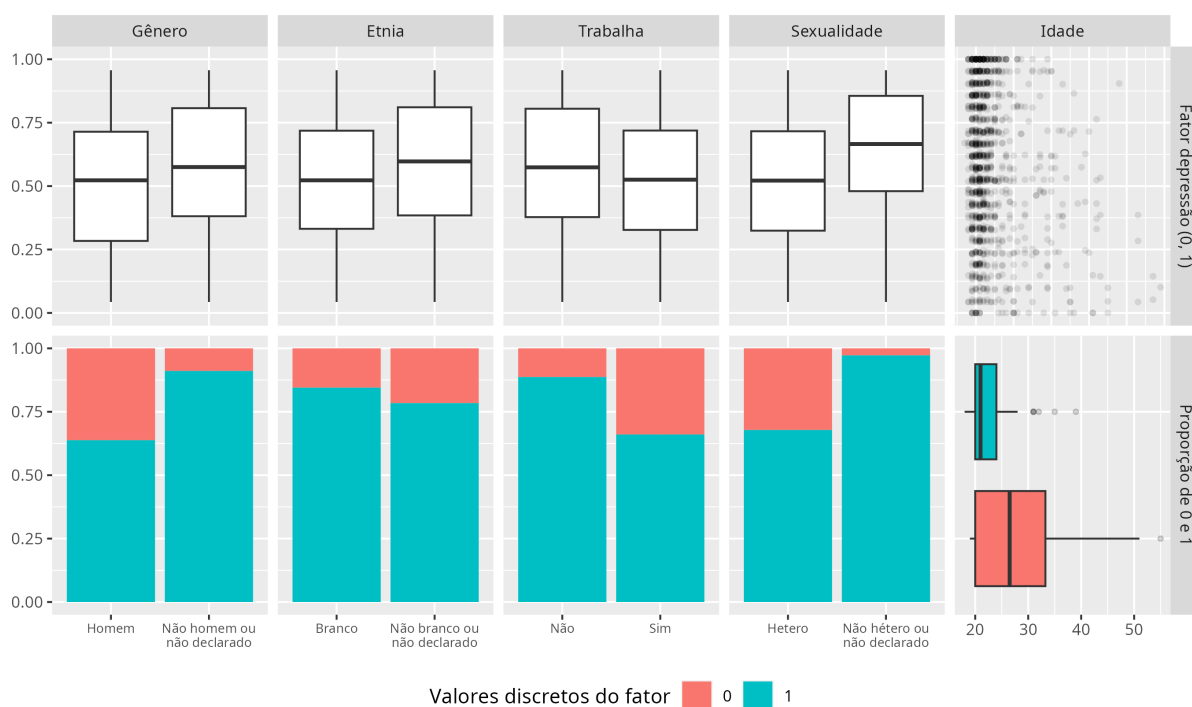


Figura 7: Primeira linha da grade: distribuição em boxplot e dispersão dos pontos de fator ansiedade (eixo y) para cada categoria da variável na respectiva coluna. Segunda linha da grade: proporção de zeros e uns em cada categoria de cada variável e distribuição de idade em boxplot nos casos zero e um.

Algumas variáveis apresentaram discrepância entre suas categorias na distribuição do fator, para a região contínua, e na proporção de zeros e uns, para a região discreta. Em destaque, a idade evidencia associação com o caso discreto, visto que existem mais universitários mais velhos no caso do fator igual 0 do que no caso do fator igual a 1.

O consumo de álcool, gênero, sexualidade e, em especial, chorar, apresentaram relações semelhantes: realizar a atividade ou ser do grupo demográfico mais frágil evidenciaram deslocamento da distribuição do fator para cima no caso contínuo, enquanto no caso discreto também aumentaram em muito a proporção de fatores iguais a um em comparação com a outra categoria.

A grosso modo, todas as variáveis agiram conforme o esperado no fator. Atitudes ou características conhecidamente benéficas ou maléficas para o humor, como a prática de atividades físicas e ingestão de álcool, tiveram o efeito em conformidade para o fator depressão.

Somente a variável de etnia não foi de encontro com as expectativas de efeitos no fator, visto que a categoria de estudantes que não se declararam brancos apresentou uma proporção maior de fator igual a zero do que os declarados brancos.

4.2 Variáveis significativas para o fator psicológico

Para chegar no objetivo principal do estudo, que é encontrar comportamentos que apresentam impacto no humor, foi analisado o ajuste do modelo de regressão logística beta inflacionado em zero e um (RLBIZU) (2.3.1), por conta do BIZU (2.1.3) ter sido o modelo escolhido para descrever o fator depressão.

O método foi baseado em testar quais dos parâmetros de comportamento e características demográficas foram significativos, seguido da avaliação de se seus impactos positivos ou negativos nos fatores condizem com o esperado ou com o estudo de Thayer, Newman e McClain (1994).

O modelo é composto pela formulação dos quatro parâmetros μ, σ, ν, τ da BIZU (2.1.3) que foram estimados a partir das variáveis explicativas. Para este estudo, fixou-se o parâmetro σ , que é referente à variância da distribuição beta, visto que não fez parte do interesse da análise. A fórmula de regressão completa para os demais parâmetros é

$$\begin{aligned} \mathit{link}(\theta) = & \alpha_0 + \alpha_1 \text{atividade_fisica} + \alpha_2 \text{interacao_social} + \alpha_3 \text{passatempos} + \\ & \alpha_4 \text{alcool_e_substancias} + \alpha_5 \text{tecnicas_de_relaxamento} + \alpha_6 \text{choro} + \\ & \alpha_7 \text{idade} + \alpha_8 \text{nao_eh_homem} + \alpha_9 \text{nao_eh_branco} + \alpha_{10} \text{trabalha} + \\ & \alpha_{11} \text{nao_eh_hetero}, \end{aligned} \quad (4.2.1)$$

onde

- θ : referente a um dos parâmetros a ser estimado pelo modelo, podendo ser μ, ν , ou τ ,
- $\mathit{link}(\theta)$: referente a função de ligação de θ entre o parâmetro estimado e a variável resposta da regressão, neste caso podendo ser *log* ou *logito*,
- α_0 : referente ao coeficiente estimado para o intercepto e denotado diferentemente para cada parâmetro estimado, podendo ser β, ρ ou γ
- $\alpha_i, i = 1, \dots, 11$: referente aos coeficientes estimados para cada *i*-ésima variável explicativa e denotado diferentemente para cada parâmetro estimado, podendo ser β, ρ ou γ ,
- A variável explicativa **nao_eh_homem** se refere ao impacto dos casos em que a resposta à variável **gênero** foi "Não homem ou não declarado" em comparação com a resposta "Homem",

- A variável explicativa **nao_eh_branco** se refere ao impacto dos casos em que a resposta à variável **etnia** foi "Não branco ou não declarado" em comparação com a resposta "Branco",
- A variável explicativa **nao_eh_hetero** se refere ao impacto dos casos em que a resposta à variável **sexualidade** foi "Não heterossexual ou não declarado" em comparação com a resposta "Heterossexual".

Elencando cada parâmetro com sua função de ligação e sua notação dos coeficientes das variáveis, conforme também explicitado no referencial teórico (2.3.1), temos:

Tabela 4: Relação entre cada parâmetro estimado com sua função de ligação e sua notação de coeficiente

Parâmetro	Função de ligação	Notação do coeficiente
μ	$\log\left(\frac{\mu}{1-\mu}\right)$ (logito)	β
ν	$\log(\nu)$	ρ
τ	$\log(\tau)$	γ

Por fim, ao interpretar o impacto que certo preditor possui na variável resposta do modelo para $\theta \in \{\mu, \nu, \tau\}$, o impacto se refere na verdade ao valor do respectivo $link(\theta)$, conforme elencado na Tabela 4.

4.2.1 Interpretação do modelo

Primeiramente, para μ , que representa a média do fator depressão no intervalo $(0, 1)$, observou-se que todas as variáveis no modelo se mostraram significativas, exceto *interacao_social*, *passatempos*, *tecnicas_de_relaxamento* e *trabalha*.

Tabela 5: Coeficientes de regressão estimados para μ do fator depressão

Variável	Parâmetro	Estimativa	Erro Padrão	Estatística T	p-valor
Intercepto	β_0	0,005	0,185	0,025	0,98
atividade_fisica	β_1	-0,181	0,064	-2,822	0,005
interacao_social	β_2	-0,045	0,115	-0,392	0,695
passatempos	β_3	0,028	0,108	0,257	0,797
alcool_e_substancias	β_4	0,526	0,065	8,054	< 0,001
tecnicas_de_relaxamento	β_5	-0,139	0,109	-1,273	0,203
choro	β_6	0,494	0,069	7,203	< 0,001
idade	β_7	-0,016	0,005	-3,584	< 0,001
nao_eh_homem	β_8	0,145	0,067	2,157	0,031
nao_eh_branco	β_9	0,161	0,063	2,559	0,011
trabalha	β_{10}	-0,08	0,068	-1,164	0,245
nao_eh_hetero	β_{11}	0,292	0,069	4,258	< 0,001

Dentre as significativas, temos que somente ser mais velho e a prática de atividade física se demonstraram como variáveis que contribuíram para seu decréscimo, o que faz sentido com o esperado. Cada ano de vida contribui no decréscimo médio de μ por 0,016, que é elevado considerando por exemplo que um indivíduo com 30 anos possuiria em média um fator depressão diminuído em 0,48 unidades. Já quem pratica atividade física em comparação com quem não pratica apresentou em média um fator menor em 0,181 unidades.

Quanto às demais variáveis significativas, que possuem efeito de acréscimo da variável resposta, destacou-se que indivíduos que evidenciaram o consumo de álcool, medicamentos ou outras substâncias possuem em média um fator maior em 0,524 unidades do que os que não consomem. Ainda, quanto às variáveis demográficas, dentre as relativas ao gênero, etnia e sexualidade, há um destaque no impacto desta última, em que indivíduos que não se declararam como heterossexuais possuem um acréscimo médio de 0,303 unidades no fator depressão.

A figura a seguir relaciona o efeito fixo de cada preditor no parâmetro, cada um com seus intervalos de confiança.

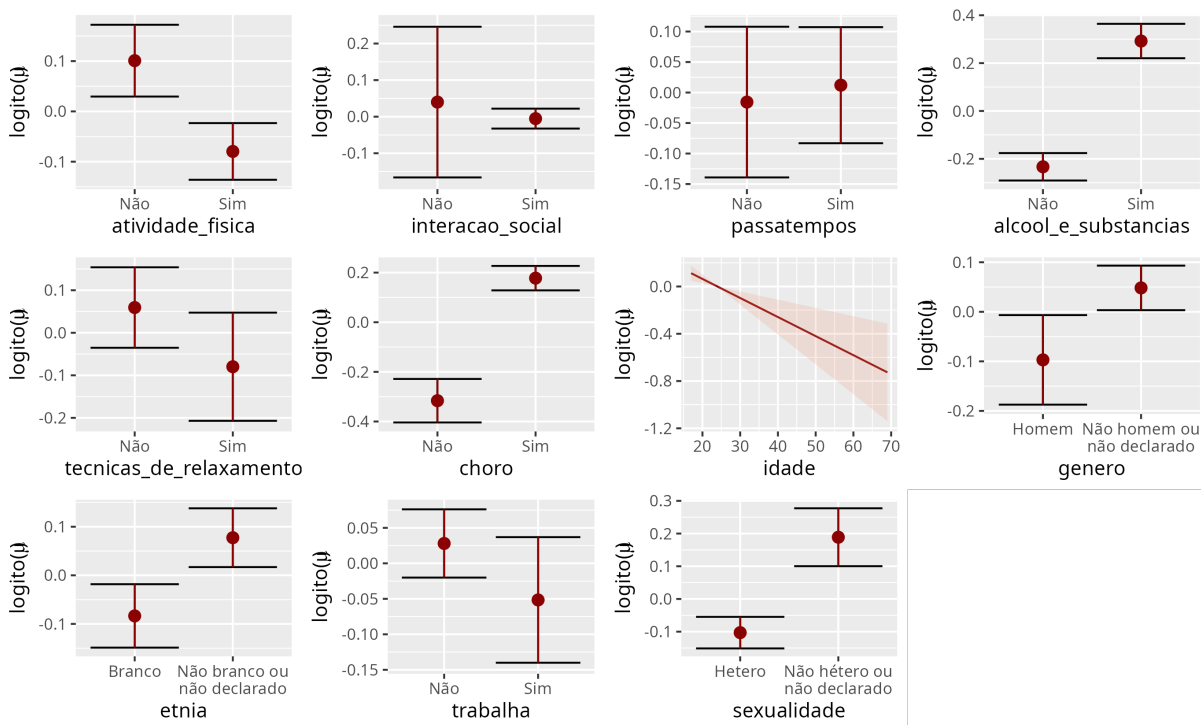


Figura 8: Efeitos fixos das variáveis para o parâmetro μ para o fator depressão

Em seguida, com relação ao parâmetro ν referente à chance do fator depressão ser igual a zero, é importante ressaltar que os preditores que possuem um efeito de acréscimo refletem num efeito positivo para o indivíduo, já que aumentam a chance dele não ter evidências de depressão.

Tendo isso em vista, notou-se que somente três variáveis foram estatisticamente significativas, das quais somente a *atividade_fisica* reflete num acréscimo para ν , adicionando 115,5% a sua chance em comparação com os que não praticam. As outras duas, *alcool_e_substancias* e *choro*, apresentaram um forte impacto negativo de decréscimo em 160,1% e 362,8%, respectivamente.

Quanto à coerência dos resultados dos coeficientes, vale um adendo de que os que apresentaram efeitos fora do esperado (aumentam a chance ao invés de diminuir, por exemplo) são justamente as que apresentaram os maiores p-valores, estando coerente pelo menos com a consistência das evidências estatísticas.

Tabela 6: Coeficientes de regressão estimados para ν do fator depressão

Variável	Parâmetro	Estimativa	Erro Padrão	Estatística T	p-valor
Intercepto	ρ_0	-3,09	1,106	-2,793	0,005
atividade_fisica	ρ_1	1,155	0,518	2,232	0,026
interacao_social	ρ_2	-0,001	0,814	-0,001	0,999
passatempos	ρ_3	-0,346	0,745	-0,465	0,642
alcool_e_substancias	ρ_4	-1,601	0,576	-2,778	0,006
tecnicas_de_relaxamento	ρ_5	-0,347	0,722	-0,48	0,631
choro	ρ_6	-3,628	1,038	-3,495	< 0,001
idade	ρ_7	0,02	0,019	1,054	0,292
nao_eh_homem	ρ_8	-0,779	0,445	-1,751	0,08
nao_eh_branco	ρ_9	0,348	0,418	0,833	0,405
trabalha	ρ_{10}	0,724	0,438	1,653	0,099
nao_eh_hetero	ρ_{11}	-1,351	0,765	-1,767	0,078

A seguir, o efeito fixo de cada preditor e seus intervalos de confiança no parâmetro

ν .

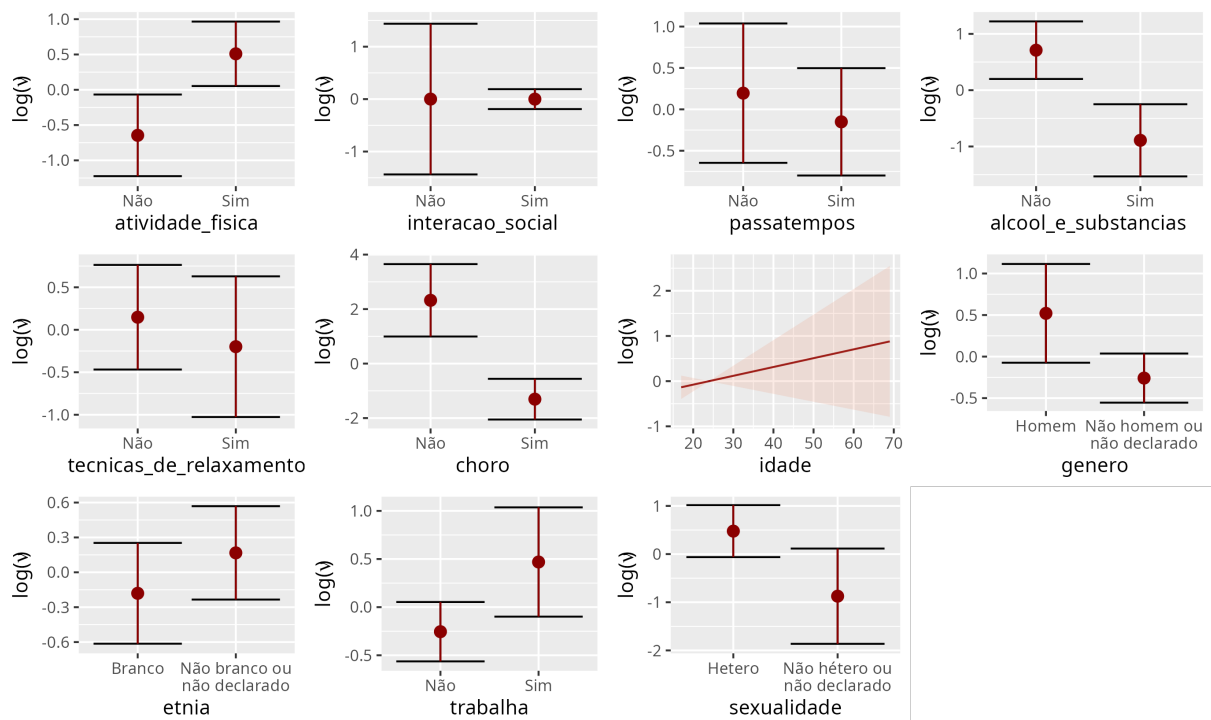


Figura 9: Efeitos fixos das variáveis para o parâmetro ν para o fator depressão

Por fim, para o parâmetro τ , que é referente à chance do fator depressão ser 1, é retomada a interpretação de que quão maior o coeficiente da variável, maior a chance do indivíduo possuir todas as evidências para o fator e, portanto, pior o impacto para si.

Tal qual no modelo para ν , poucos preditores foram consideráveis para este modelo, assim como que os que são significativos em comum para ambos são concordantes, ou seja, para ν essas variáveis reduzem sua chance enquanto para τ aumentam.

Tabela 7: Coeficientes de regressão estimados para τ do fator depressão

Variável	Parâmetro	Estimativa	Erro Padrão	Estatística T	p-valor
Intercepto	γ_0	-0,991	0,686	-1,444	0,149
atividade_fisica	γ_1	-0,159	0,202	-0,787	0,431
interacao_social	γ_2	-1,159	0,301	-3,85	< 0,001
passatempos	γ_3	-0,454	0,322	-1,41	0,159
alcool_e_substancias	γ_4	0,922	0,207	4,456	< 0,001
tecnicas_de_relaxamento	γ_5	-0,404	0,365	-1,105	0,269
choro	γ_6	0,735	0,263	2,793	0,005
idade	γ_7	-0,04	0,022	-1,809	0,071
nao_oh_homem	γ_8	0,068	0,22	0,311	0,756
nao_oh_branco	γ_9	-0,023	0,2	-0,117	0,907
trabalha	γ_{10}	-0,272	0,227	-1,197	0,232
nao_oh_hetero	γ_{11}	0,575	0,203	2,834	0,005

Sobre os preditores que foram concordantes para os modelos de ν e τ , que são os referentes ao consumo de álcool, medicamentos ou outras substâncias e ao choro, tem-se que praticar essas atividades acrescentou em média 92,2% e 73,5% na variável resposta, respectivamente.

Já para as demais explicativas, que são as relativas à interagir socialmente e a não se declarar heterossexual, ambas se relacionaram com o fator conforme o esperado, visto que elas o impactam respectivamente no decréscimo da chance em 115,9% e no seu acréscimo em 57,5%. Em destaque, observou-se que essa primeira variável foi a única das significativas deste modelo a ter efeito positivo para o indivíduo, ou seja, diminui a chance do mesmo ter o fator depressão máximo.

A seguir, os efeitos fixos das variáveis explicativas neste modelo:

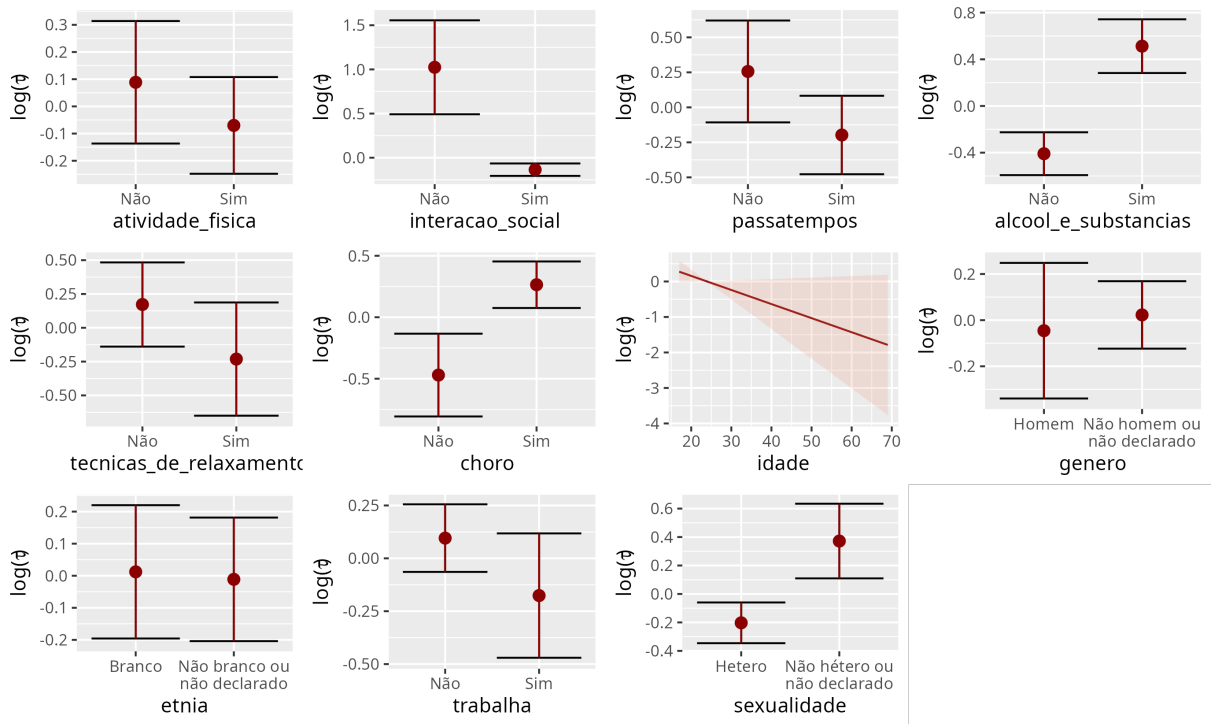


Figura 10: Efeitos fixos das variáveis para o parâmetro τ para o fator depressão

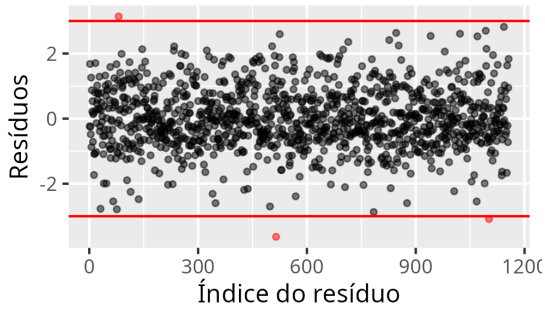
Diagnóstico

Inicialmente, analisou-se a escolha estrutural do modelo quanto às funções de ligação utilizadas e à presença de pontos influentes. Quanto ao gráfico 11b, não se observou uma tendência evidente entre os valores ajustados de μ e seus respectivos resíduos, indicando que as funções de ligação escolhidas não estão inadequadas.

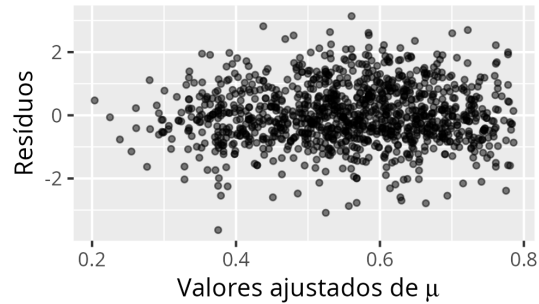
Já para o gráfico 11a, são observados alguns pontos (em vermelho) fora do intervalo comumente utilizado de $[-3, 3]$ para este diagnóstico, evidenciando a presença de pontos influentes na amostra. Contudo, mesmo com a retirada deles, o modelo não apresentou melhora nos todos os ajustes e também são aceitáveis dado que foram apenas três pontos, o que levou à decisão de permanecer com eles.

Figura 11: Diagnóstico de estrutura das funções de ligação e de pontos influentes do modelo para o fator depressão

(a) Relação entre índice da observação versus Resíduo quantil aleatorizado. Pontos que ultrapassam o intervalo $[-3, 3]$ estão destacados em vermelho.



(b) Relação entre valores ajustados de μ versus Resíduo quantil aleatorizado.

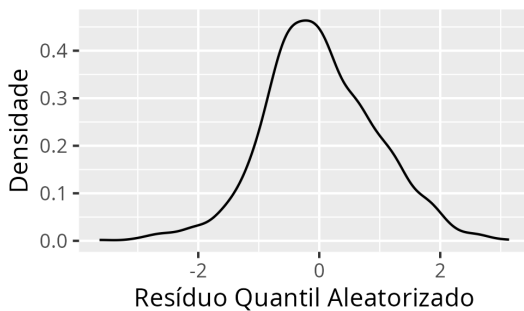


Com relação ao ajuste dos resíduos do modelo à distribuição normal, foram observados alguns problemas. O teste de Shapiro-Wilk, apresentado na Tabela 8, conclui que há evidências estatísticas para a rejeição da hipótese de que os resíduos aderem à distribuição.

Contudo, os gráficos 12a e 12b evidenciaram um desvio bem mais suave da normalidade, visto que a densidade se apresentou um pouco assimétrica e sua cauda à direita pesada, assim como alguns pontos do *qqplot* foram observados fora da região de confiança dos quantis teóricos.

Figura 12: Diagnóstico de normalidade dos resíduos do modelo para o fator depressão

(a) Densidade estimada dos resíduos



(b) QQplot resíduos

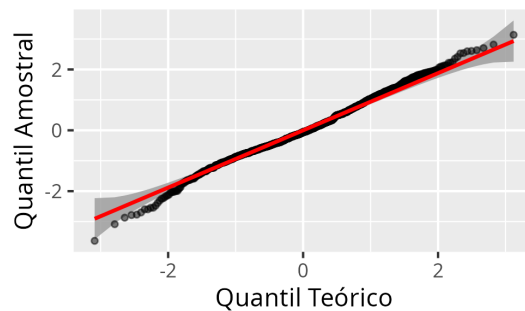


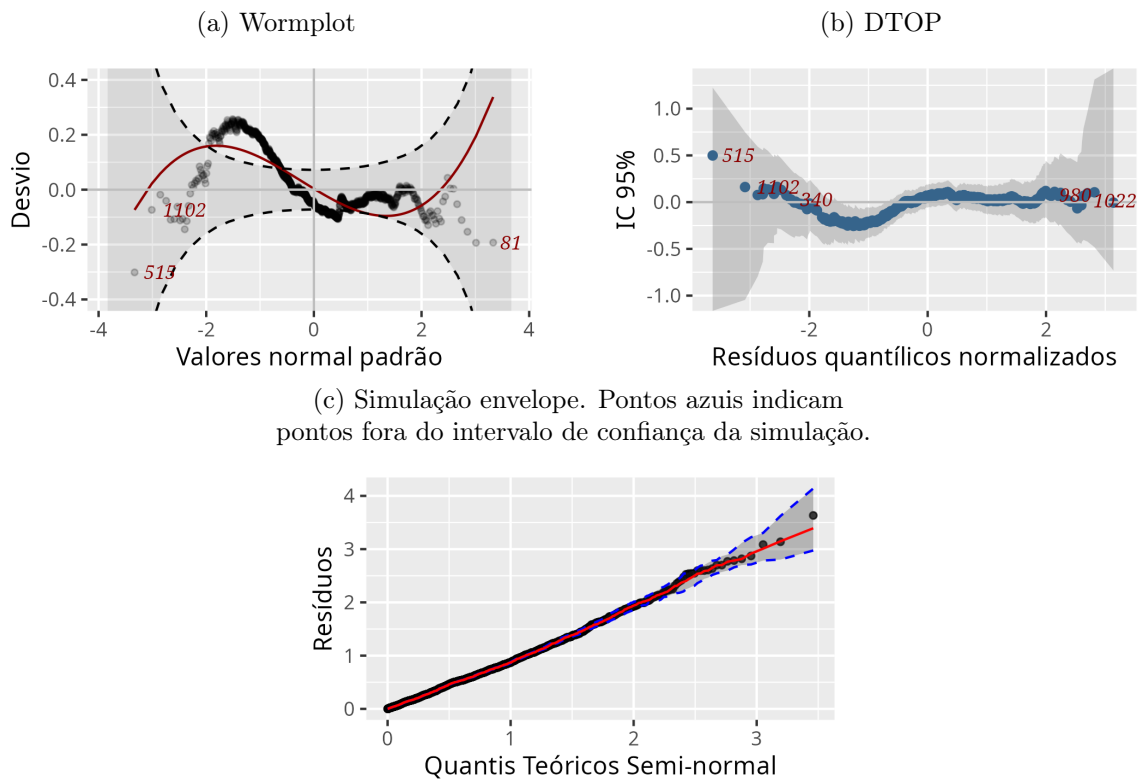
Tabela 8: Medidas resumo dos resíduos do modelo para o fator depressão

Medidas	Valores
Média	0.015
Mediana	-0.049
Assimetria	0.058
p-valor teste de normalidade Shapiro-Wilk	< 0.001

Ainda, no *wormplot* (13a) também foi evidenciado problemas da adequabilidade do modelo, visto que há uma quantidade relevante de pontos fora dos intervalos. Além disso, a curva em *S* crescente e acentuada indica que a distribuição dos resíduos possui caudas muito pesadas. A não adequação também se apresenta no gráfico *DTOP* (13b), em que se observou que parte de seus resíduos possuem intervalo de confiança empírico que não inclui o eixo horizontal 0.

Contudo, no gráfico de simulação envelope (13c) não se encontraram evidências da não adequação (não há pontos azuis, que indicariam a presença de resíduo além do intervalo de confiança), dando mais um indício de que este desajuste não é severo.

Figura 13: Diagnóstico de ajuste e adequabilidade do modelo para o fator depressão com *wormplot*, *DTOP* e simulação envelope



5 Conclusão

Inicialmente, é importante ressaltar o caráter meramente exploratório deste estudo. Isso se deve principalmente por conta das limitações do banco de dados, que incluem questões como o evidente viés amostral, o descompasso do banco original para ser adaptado ao estudo e as poucas possibilidades disponíveis para permitir a evidência da prática dos comportamentos elencados por Thayer, Newman e McClain (1994).

Também, se incluem nos problemas as evidências de inadequação de ajuste do modelo, ainda que suavizados pelas análises gráficas do diagnóstico de resíduos. Enfim, todos esses fatores contribuem para não ser pertinente uma conclusão inferencial dos resultados.

Apesar disso, a maioria das relações entre o fator depressão e as variáveis de comportamentos e características demográficas se revelaram conforme o esperado. Na Tabela 9 abaixo, se destacam os comportamentos de praticar atividades físicas, chorar e ingerir álcool, medicamentos e substâncias como os que apresentaram consistentemente impacto nas três medidas do fator depressão. Também em especial, os indícios encontrados para passatempos e para a prática de técnicas de relaxamento não apresentaram impacto significativo de nenhuma forma para o fator.

Tabela 9: Variáveis que foram significativas para cada esfera do modelo. Em cores, se apresenta em verde e vermelho as significâncias que representam respectivamente um impacto positivo e negativo para o indivíduo

Variável	Apresentou impacto significativo no		
	fator de depressão na		
	Média do escore (μ)	Chance do escore ser zero (ν)	Chance do escore ser um (τ)
atividade_fisica	✓	✓	
interacao_social			✓
passatempos			
alcool_e_substancias	✓	✓	✓
tecnicas_de_relaxamento			
choro	✓	✓	✓
idade	✓		
trabalha			
genero	✓		
etnia	✓		
sexualidade	✓		✓

Ainda, há um contraste principalmente em relação ao consumo de álcool, medicamentos e outras substâncias entre os resultados deste trabalho e aos declarados dentro do estudo de Thayer, Newman e McClain (1994). Lá, esse consumo foi adotado como sendo uma estratégia para que haja a melhora do humor, enquanto aqui ele foi bastante significativo em se relacionar com o aumento do fator depressão e, portanto, piora do humor e bem estar do estudante.

Por fim, conclui-se a partir das análises que há evidências para relacionar a prática de alguns comportamentos e algumas características demográficas com a incidência ou gravidade melhor ou pior da depressão, segundo mensuração psicométrica DASS-21 (LOVIBOND, 1983).

Referências

CERCHIARI, E. A. N. et al. Saúde mental e qualidade de vida em estudantes universitários. *Campinas: Faculdade de Ciências Médicas da Universidade Estadual de Campinas*, 2004.

GÓES, P. H. M. Avaliação de uma escala de mensuração de sofrimento psicológico: aplicação de análise fatorial confirmatória. 2022.

JOHNSON, R. A.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. [S.l.]: Pearson Education, Inc., 2007.

LOVIBOND, S. The nature and measurement of anxiety, stress and depression. In: *18th Annual Conference of the Australian Psychological Society, University of Western Australia, May*. [S.l.: s.n.], 1983.

MARINHO, F. et al. Burden of disease in brazil, 1990–2016: a systematic subnational analysis for the global burden of disease study 2016. *The Lancet*, Elsevier, v. 392, n. 10149, p. 760–775, 2018.

MARTINEZ, R. O. *Modelos de regressão beta inflacionados*. Tese (Doutorado) — Universidade de São Paulo, 2008.

ROCHA, L. Pesquisas apontam aumento nos casos de depressão no brasil. In: CNN BRASIL. São Paulo, 2022. Disponível em: <<https://www.cnnbrasil.com.br/saude/pesquisas-apontam-aumento-nos-casos-de-depressao-no-brasil/>>. Acesso em: 20 jan. 2023.

THAYER, R. E.; NEWMAN, J. R.; MCCLAIN, T. M. Self-regulation of mood: Strategies for changing a bad mood, raising energy, and reducing tension. *Journal of Personality and Social Psychology*, v. 67(5), p. 910–925, 1994.

WORLD HEALTH ORGANIZATION. Depressive disorder (depression). In: _____. 2022. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/depression>>. Acesso em: 26 jul. 2023.

WORLD HEALTH ORGANIZATION. Mental disorders. In: _____. 2022. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/mental-disorders>>. Acesso em: 12 fev. 2023.