



Universidade de Brasília
Departamento de Estatística

A relação entre consumo midiático e o conhecimento do inglês

Gustavo Alves Wanderley

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

Gustavo Alves Wanderley

A relação entre consumo midiático e o conhecimento do inglês

Orientadora: Prof^ª. Thais Carvalho Valadares Rodrigues

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

- À professora Thais, que me botou nesse curso e agora está me tirando dele.
- À professora Thais de novo, que, em todo o seu conhecimento, não imaginaria que teria que aguentar um homem adulto falando de joguinho por um ano, muito menos revisar um TCC gigante com essa linguagem alienígena (mas fez mesmo assim e fez muito bem).
- Ao meu amigo Gabriel Fruet, que fez o cartaz do questionário do modo mais chamativo possível e alavancou a amostragem de modo absurdo, conseguindo mais de 500 respostas só na UnB, fora as outras centenas na internet.
- A mim mesmo, pois consegui escrever 64 páginas de um TCC sobre videogames (antes dos apêndices!), o que é surpreendente, irônico, engraçado e meio triste.

Resumo

Neste trabalho, será estudada a relação entre o nível de inglês e o conteúdo midiático consumido. O estudo do tema é oportuno visto que uma das consequências de um mundo globalizado é a disponibilidade em abundância de conteúdo internacional, o qual nem sempre está integralmente traduzido para o português. Dentre essas mídias, os jogos digitais se destacam por possuírem menor disponibilidade no idioma, de modo que a variedade do meio pode enriquecer a pesquisa. Foi aplicada a regressão logística multinomial utilizando dados provenientes de um questionário divulgado a alunos da UnB, com auxílio de grafos, análise exploratória e estudo de correlações. Foram discutidos os fatores mais associados ao nível de inglês da pessoa e propostos agrupamentos entre os respondentes. Os resultados apontam a influência da condição socioeconômica no acesso ao conhecimento, bem como o fato de que videogames, apesar de refletirem bem o conhecimento de língua estrangeira, não necessariamente a predizem.

Palavras-chaves: regressão logística multinomial; língua inglesa; inglês; questionário; jogos digitais; videogames; tradução

Abstract

The present study researches the relation between the knowledge of the English language and the media consumed by Brazilian students. This topic is timely, as one of the consequences of a globalized world is the abundant availability of international content, which

is not always fully translated into Portuguese. Among other media, video games stand out due to their smaller availability in the idiom, in a way that the variety of this media can enrich the research. The multinomial logistic regression was applied, using data from a survey answered by the students of the University of Brasília, while being supported by methods such as graphs, exploratory analysis, and the study of correlations. The factors most associated with a person's English level were discussed, and groupings among respondents were proposed. The results point to the influence of socioeconomic status on the access to knowledge, as well as the fact that videogames, despite reflecting the knowledge in foreign languages, are not particularly influent in causing it.

Keywords: multinomial logistic regression; English language; survey; video games; translation

Lista de Tabelas

1	P-valores dos testes de Wilcoxon aplicados a cada grupo	49
2	Parâmetros estimados pela regressão logística multinomial para o nível de inglês 1 (β^1) e 3 (β^3), e seus respectivos erros padrão	53
3	Razões de chance estimadas pelo modelo logístico	55
4	Matriz de confusão utilizando banco de testes	59

Lista de Figuras

1	Protesto de jogadores brasileiros em MMORPG infantil (MERRIFIELD, 2017)	9
2	Contracapa de <i>Hades</i> (ARRAZ, 2021)	10
3	<i>Screenshot</i> de <i>Elite Beat Agents</i> (ELITE..., 2007)	10
4	<i>Screenshot</i> de <i>Advogados de Primeira: Fênix Veríssimo</i> (JACUTEM Sa- bão, 2022)	11
5	Imagem usada para divulgação	23
6	Distribuição do nível de inglês dos alunos da UnB respondentes	26
7	Distribuição do conteúdo em inglês consumido pelos respondentes	27
8	Associação entre conteúdo consumido em inglês e nível de inglês. As tonali- dades de lilás indicam o coeficiente Gama de Goodman-Kruskal, enquanto o tamanho das barras indicam o Tau-c de Stuart-Kendall	29
9	Associação entre nível de inglês e covariáveis diversas	31
10	Associação entre plataformas onde se joga videogame e o nível de inglês	32
11	Distribuição do gênero dos respondentes	33
12	Quantidade de horas diárias em que os respondentes jogam videogame	34
13	Distribuição das respostas quanto ao afastamento de conteúdo qualquer pela falta de conhecimento da língua inglesa	35
14	Distribuição das respostas quanto ao interesse em conteúdos quaisquer de- vido ao lançamento de uma versão em português	36
15	Distribuição do gosto dos respondentes por diferentes gêneros de jogo digital	38

16	Associação entre gosto em gêneros de videogame e nível de inglês	39
17	Distribuição do idioma no qual os respondentes consomem diferentes gêneros de jogos	41
18	<i>Pokémon UNITE</i> , sendo jogado com interface em Português (GOOGLE, 2023)	42
19	Associação entre idioma no qual se consomem jogos e nível de inglês	43
20	Tipos de contato dos respondentes com a língua inglesa na infância e o impacto no conhecimento adquirido	44
21	Associação entre formas de contato com inglês na infância e nível atual de inglês	45
22	Principais grupos de jogos entre os respondentes	47
23	Principais ligações entre jogos preferidos e nível de idioma	51
24	Gráfico quantil-quantil com envelope simulado	58

Sumário

1 Introdução	8
2 Objetivos	13
2.1 Objetivo Geral	13
2.2 Objetivos Específicos	13
3 Referencial Teórico	14
3.1 Regressão Logística	14
3.1.1 Regressão Logística Múltipla	15
3.2 Regressão Logística Multinomial	15
3.2.1 Odds e Odds Ratio	17
3.2.2 Estimador de Máxima Verossimilhança	18
3.2.3 Teste de Razão de Verossimilhança	19
3.2.4 Teste de Wald	20
3.2.5 Intervalos de confiança para Beta e Odds Ratio	21
4 Metodologia	22
4.1 Conjunto de dados	22
4.2 Metodologia	24
5 Resultados	26
5.1 Análise exploratória e correlações	26
5.1.1 Nível de inglês	26
5.1.2 Conteúdo em inglês consumido	27
5.1.3 Variáveis categóricas diversas	30
5.1.4 Gênero	32
5.1.5 Tempo jogado	33
5.1.6 Interesse e afastamento de conteúdo com base no idioma	34
5.1.7 Gosto por diferentes gêneros de jogo	36
5.1.8 Idioma no qual se joga diferentes gêneros de jogo	40

5.1.9	Contato com o inglês	44
5.1.10	Grafo dos principais grupos de jogadores	46
5.1.11	Grafo dos gostos por jogos de acordo com nível de inglês	50
5.2	Modelo Logístico Multinomial	52
5.2.1	Equação logística	52
5.2.2	Chance de sucesso	54
5.2.3	Diagnóstico do modelo	57
6	Conclusão	60
	Referências	62
	Apêndices	65

1 Introdução

Com o avanço da globalização, a disponibilidade de conteúdo internacional no Brasil está cada vez mais presente, graças a meios como a televisão, cinema e internet. Isso faz com que as gerações mais recentes possam ter maior contato com material vindo de outros países, seja para fins de educação ou entretenimento. Entretanto, há uma parcela deste conteúdo que, apesar de disponível, requer conhecimento da língua inglesa para ser desfrutado com êxito.

Esse problema era mais evidente quando a internet era menos presente na vida do cidadão brasileiro. Devido à comum ausência de filmes e séries traduzidas para o português, havia uma certa dependência de materiais não oficiais para se ter acesso antecipado a conteúdo midiático, ou para sequer ter algum tipo de acesso. Um exemplo desse material são as *fansubs*, legendas feitas por fãs para filmes, animações e outras mídias.

Em 2023, a agilidade e o número de opções oficiais e não oficiais de traduções de mídias como filmes, séries e animações é consideravelmente maior, graças ao *streaming* e ao maior número de usuários na internet, representando maior demanda, a qual é atendida por empresas ou grupos independentes como a *Luxia Subs* (PEREIRA, 2023).

Entretanto, os jogos digitais (videogames) não são disponíveis em português em mesma escala, mesmo que haja um avanço em relação ao passado. Ou seja, o conhecimento da língua inglesa é mais necessário nesse tipo de obra, apesar do fato de apenas 5% da população brasileira saber inglês (ANDRADE, 2022).

A título de exemplo, a Figura 1 mostra um reflexo desse aumento de demanda no contexto de jogos digitais, ilustrando um protesto de jogadores brasileiros realizado no MMORPG infantil *Club Penguin* em 2007, antes do jogo incluir o idioma português.



Figura 2: Contracapa de *Hades* (ARRAZ, 2021)

A ausência de jogos digitais com o idioma português faz com que alguns jogadores tenham que aprender outro idioma enquanto jogam para aproveitar o produto em questão. Apesar de haver argumentos a favor do uso desse tipo de produto como ferramenta para o desenvolvimento linguístico (SMOLINSKI, 2016; SILVA; TOASSI, 2020), essa raramente é a intenção dos *game designers*, que não otimizam seus jogos para fins educativos.

Há jogos onde a linguagem escrita e falada são pouco utilizadas, pois eles utilizam outros meios de comunicar objetivos e história ao jogador. A Figura 3 mostra o jogo *Elite Beat Agents*, que se encaixa nessa categoria, pois sua premissa é o uso de comandos simples no ritmo da música de fundo, reservando o uso da linguagem verbal para os menus, músicas de fundo e elementos mínimos da interface.



Figura 3: Screenshot de *Elite Beat Agents* (ELITE..., 2007)

Em contrapartida, também há jogos quase inteiramente apresentados por meio de texto, como *Danganronpa* e *VA-11 Hall-A*. Como existe demanda para esse tipo de conteúdo, originalmente inacessível para quem não tem domínio de idioma estrangeiro, e essa demanda é pouco atendida pelas produtoras, alguns fãs localizam os jogos por meio de *fan translations* para atender à demanda e adaptar elementos narrativos para a cultura brasileira.

Na Figura 4, pode ser observado um exemplo desse tipo de trabalho, realizado pela equipe *Jacutem Sabão*, que localiza jogos da franquia *Ace Attorney*, entre outros (JACUTEM Sabão, 2022). Nela, temos um *screenshot* ilustrando a localização para português do jogo *Phoenix Wright: Ace Attorney*, cujo título foi adaptado para *Advogados de Primeira: Fênix Veríssimo*. Nessa localização, elementos como roteiro e nomes de personagens também foram adaptados.



Figura 4: *Screenshot* de *Advogados de Primeira: Fênix Veríssimo* (JACUTEM Sabão, 2022)

Apesar de haver jogos onde a barreira linguística não necessariamente impede o proveito do jogador, é imperativo assumir que jogadores que tem domínio da língua inglesa, padrão em grande parte dos jogos atualmente, tenham maior facilidade no proveito e no ato de conhecer diferentes gêneros de jogos.

Portanto, jogos digitais são uma interessante fonte de estudo para associar conhecimento do inglês e conteúdo consumido: a variedade nesse tipo de obra, somada com sua relativa falta de presença em português, pode produzir um reflexo do impacto do conhecimento de línguas nas mídias consumidas. Se espera que, por exemplo, um jogador que consome vários tipos de jogos em inglês também consuma materiais como livros e

notícias nesse mesmo idioma.

Para realizar esse tipo de análise, a regressão múltipla torna-se útil, devido ao grande leque de fatores relativos ao conteúdo consumido, o qual inclui diversos gêneros de jogos. Tendo em vista os aspectos metodológicos da regressão, não será necessariamente buscada uma relação da causalidade, mas sim de associação.

Considerando que a variável resposta analisada é categórica (nível de conhecimento da língua inglesa), a regressão logística é uma opção mais adequada do que a regressão linear gaussiana. E como essa variável possui mais de dois níveis, torna-se necessário o uso do modelo multinomial. Portanto, a regressão logística múltipla multinomial será utilizada para o caso em questão, com informações extraídas de um questionário assumindo o papel de covariáveis.

Além do mais, como a discussão a respeito da ausência de tradução em videogames tornou-se mais fervorosa nos últimos anos (LAURA, 2022) e este tópico de estudo é relativamente novo, a ausência de estudos do gênero faz com que seja conveniente o estudo a partir de vários ângulos, esses representados pelas variáveis da regressão, que podem ou não ser descartadas após ajuste, e pelas demais análises realizadas com o conjunto de dados.

2 Objetivos

2.1 Objetivo Geral

Associar o conhecimento da língua inglesa por parte dos estudantes da UnB ao tipo de conteúdo midiático que este grupo consome, com foco em jogos digitais, considerando a falta de material traduzido no meio e as divergências entre diferentes gêneros.

2.2 Objetivos Específicos

- Avaliar o impacto do nível de contato com a língua inglesa na infância com o conteúdo midiático consumido atualmente;
- Realizar análises exploratórias para visualizar os gêneros de jogo que o grupo estudado gosta, bem como o idioma no qual eles são jogados;
- Associar o conhecimento da língua inglesa com o consumo de outras mídias, como filmes e livros, em inglês;
- Avaliar o impacto da presença ou não de traduções em indivíduos de diferentes níveis de inglês;
- Utilizar lista de jogos favoritos dos respondentes para avaliar o perfil dos jogadores e associar estes jogos e perfis ao nível de inglês.

3 Referencial Teórico

3.1 Regressão Logística

A análise de regressão é utilizada para investigar a relação existente entre uma variável dependente Y e uma ou mais variáveis independentes X , para que sejam identificadas aquelas variáveis X que tem maior relevância no modelo de forma a explicar a variável Y . Em casos como no de questionários, ela é particularmente útil para se analisar os fatores contribuintes para o tema estudado, explicando assim um fenômeno.

A regressão linear, que é tradicionalmente usada, assume que os valores de cada variável dependente quantitativa são contínuos e apresentam normalidade. Isso a torna inviável para casos onde as variáveis independentes são categóricas. Nesses casos, é utilizada a regressão logística.

A regressão logística busca estimar a probabilidade $\pi(x)$ de diferentes categorias da variável dependente Y , dado um conjunto X de variáveis independentes (AGRESTI, 2007).

O modelo logístico pressupõe que a variável aleatória Y segue uma distribuição binomial com parâmetros n e $\pi(x)$. Ou seja, ele modela a probabilidade $\pi(x)$ de sucesso binomial em função das covariáveis X em uma amostra de tamanho n . Para isto, é utilizada uma forma linear para a transformação *logit* desta probabilidade.

Considerando o modelo logístico simples, também chamado de modelo logístico binário, onde Y assume apenas dois níveis e há apenas uma covariável, temos que:

$$\text{logit}[\pi(x)] = \log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha + \beta x, \quad (3.1.1)$$

onde α e β denotam, respectivamente, o intercepto e o coeficiente angular da forma linear do logit da probabilidade. Adicionalmente, isolando $\pi(x)$, conclui-se que:

$$\pi(x) = \frac{\exp((\alpha + \beta x))}{1 + \exp((\alpha + \beta x))}, \quad (3.1.2)$$

o que também pode ser escrito como:

$$\pi(x) = \frac{1}{1 + \exp((-(\alpha + \beta x)))}. \quad (3.1.3)$$

Segundo este modelo, o *logit* se locomove em β unidades a cada aumento unitário em X , apresentando, assim, comportamento linear. Já a probabilidade $\pi(x)$ apresenta em X um comportamento não linear (curva em formato de "S"), com aumento proporcional à inclinação dessa curva.

3.1.1 Regressão Logística Múltipla

A regressão logística múltipla, é uma extensão da regressão logística usada quando há mais de uma variável independente, ou seja, quando há múltiplas covariáveis (HOSMER; LEMESHOW; STURDIVANT, 2013).

Neste tipo de regressão, considerando p covariáveis, temos o seguinte logit:

$$\text{logit}[\pi(\mathbf{x})] = \log\left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right) = \beta_0 + \sum_{j=1}^p \beta_j x_j. \quad (3.1.4)$$

E isolando $\pi(\mathbf{x})$ na fórmula acima, tem-se que:

$$\pi(\mathbf{x}) = \frac{\exp(g(\mathbf{x}))}{1 + \exp(g(\mathbf{x}))}, \quad (3.1.5)$$

que pode ser reescrito (TAKAHASHI; INOUE, 2019) como:

$$\pi(\mathbf{x}) = \frac{1}{1 + \exp(-g(\mathbf{x}))}, \quad (3.1.6)$$

onde $g(\mathbf{x})$ é um preditor linear, ou seja:

$$g(\mathbf{x}) = \beta_0 + \sum_{j=1}^p \beta_j x_j, \quad (3.1.7)$$

onde j representa o índice das covariáveis, podendo este ir de 1 a p (HOSMER; LEMESHOW; STURDIVANT, 2013). β_0 , neste caso, assume o papel de intercepto da forma linear do logit da probabilidade.

3.2 Regressão Logística Multinomial

Quando a variável aleatória \mathbf{Y} possui dois ou mais níveis, a regressão logística binomial ou binária (onde há apenas dois níveis) é estendida para a regressão logística

multinomial.

Para realizar uma análise com este tipo de regressão, é necessário que os dados satisfaçam as seguintes suposições (LUND; LUND, 2018):

- A variável dependente deve ser categórica, ao invés de ser uma variável ordinal ou quantitativa;
- Uma ou mais variáveis independentes devem ser contínuas, categóricas ordinais ou categóricas nominais;
- As observações devem ser independentes, assim como as categorias da variável dependente devem ser mutuamente exclusivas e exaustivas;
- Não deve haver multicolinearidade, isto é, duas ou mais variáveis independentes altamente correlacionadas entre si;
- Deve haver uma relação linear entre o preditor linear e a transformação *logit* da variável dependente;
- Não devem haver observações influentes, isto é, (*outliers*).

Como \mathbf{Y} não está restrito a dois níveis, a distribuição binomial não é mais suficiente para descrever essa variável. Então, neste caso, o modelo assume a seguinte distribuição multinomial em \mathbf{Y} para uma amostra de tamanho n :

$$\mathbf{Y} \sim Multinomial(n, K, \pi_1, \dots, \pi_K). \quad (3.2.1)$$

Assumindo que \mathbf{Y} possui K níveis, e considerando que há p variáveis independentes na estrutura de regressão, podemos calcular a probabilidade $P(Y = k | X_1 = x_1, \dots, X_p = x_p)$, ou seja $\pi_k(\mathbf{x})$, para cada nível k , com:

$$\pi_k(\mathbf{x}) = P(\mathbf{Y} = k | \mathbf{x}) = \frac{\exp(g_k(\mathbf{x}))}{1 + \sum_{j=1}^k \exp(g_j(\mathbf{x}))}, \quad (3.2.2)$$

onde $g_k(x)$ é novamente um preditor linear:

$$g_k(x) = \beta_{0k} + \sum_{j=1}^p \beta_{jk} X_j. \quad (3.2.3)$$

Novamente, j representa o índice das covariáveis, podendo este ir de 1 a p (HOSMER; LEMESHOW; STURDIVANT, 2013), enquanto β_0 assume o papel de intercepto da forma linear do logit da probabilidade de sucesso.

Ou seja, o caso multinomial é uma sequência de regressões logísticas múltiplas. Após realizadas as regressões, a análise resultante é uma série de comparações entre dois diferentes níveis de Y (HUA; CHOI; SHI, 2021).

3.2.1 Odds e Odds Ratio

No contexto da regressão logística, temos o conceito de *odds of success*, isto é, a chance de sucesso, que é um valor não-negativo que representa a proporção da probabilidade de sucesso em relação à probabilidade de fracasso (AGRESTI, 2007). Essa chance pode ser definida como:

$$odds = \frac{\pi}{1 - \pi}, \quad (3.2.4)$$

que é equivalente ao numerador das equações 3.1.2, 3.1.5 e 3.2.2 nos casos de regressão logística simples, multivariada e multinomial, respectivamente.

Uma *odds* de valor de 2, por exemplo, indica que o sucesso é duas vezes mais comum que o fracasso, enquanto um valor de 0.5 indica o oposto, ou seja, que a probabilidade de sucesso equivale à metade da probabilidade de fracasso.

Em tabelas de contingência 2 x 2, onde cada linha representa uma *odds* distinta, podemos interpretar a razão de chances, ou, do inglês, a *odds ratio* (denotada por θ), para avaliar a razão entre as diferentes *odds* de sucesso.

$$\theta = \frac{odds_1}{odds_2} = \frac{\pi_1 / (1 - \pi_1)}{\pi_2 / (1 - \pi_2)}. \quad (3.2.5)$$

Se $\theta > 1$, as *odds* de sucesso são maiores na linha 1, mais especificamente $\theta = 2$ indica que as *odds* de sucesso da linha 1 são duas vezes maiores que na linha 2. Em contraponto, se $\theta < 1$, as *odds* de sucesso são maiores na linha 2. Como exemplo, $\theta = 0,5$ informa que as *odds* da linha 1 são metade daquelas da linha 2. Por fim, se $\theta = 1$, fica implícito que $\pi_1 = \pi_2$, ou seja, que as variáveis X e Y do modelo são independentes.

No caso de uma regressão logística multinomial, a *odds ratio* é calculada para quaisquer combinações de níveis da variável dependente e dois valores de uma covariável, mesmo que a covariável tenha mais de dois possíveis valores. No caso de covariável

quantitativa, a *odds ratio* mostra a interferência em \mathbf{Y} quando se incrementa x em uma unidade.

Generalizando, se temos uma variável dependente \mathbf{Y} que pode assumir os valores k (a variável é do nível k) ou 0 (a variável não é do nível k), e também temos a variável independente \mathbf{x} que pode assumir os valores distintos a ou b , temos (HOSMER; LEMESHOW; STURDIVANT, 2013):

$$\theta_k(a, b) = \frac{odds_1}{odds_2} = \frac{P(Y = k|x = a)/P(Y = 0|x = a)}{P(Y = k|x = b)/P(Y = 0|x = b)}. \quad (3.2.6)$$

3.2.2 Estimador de Máxima Verossimilhança

A meta da regressão logística é estimar os $K - 1$ vetores β do modelo logístico multinomial, ou seja, um vetor de parâmetros para cada nível (K-1) da variável \mathbf{Y} . Aplicar o princípio da verossimilhança significa obter a estimação dos parâmetros β que maximizam a probabilidade de serem encontrados os mesmos valores de uma amostra observada (BUSSAB; MORETTIN, 2012).

Para estimar os parâmetros β desconhecidos a partir de uma variável aleatória observada y , o estimador de máxima verossimilhança (EMV) é uma possível solução. Para calculá-lo, o primeiro passo é usar a função de verossimilhança de β dado y :

$$L(\beta|\mathbf{y}) = \prod_{i=1}^N \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{(n_i - y_i)}. \quad (3.2.7)$$

A equação 3.2.7 é a função $f(\mathbf{y}|\beta)$ de densidade de probabilidade conjunta de \mathbf{Y} , que apresenta distribuição binomial. Nela, i representa a i -ésima observação do modelo, com cada população sendo um conjunto dos n_i resultados de uma covariável em uma amostra.

A combinação na equação 3.2.7 representa as diferentes formas de arranjar y_i sucessos no conjunto dos n_i resultados, de modo que π_i é a probabilidade de sucesso em n_i , e $\pi_i^{y_i}$ é a probabilidade de haverem y_i sucessos em n_i (CZEPIEL, 2022).

No caso multinomial, temos a seguinte fórmula para a função de verossimilhança:

$$L(\beta|\mathbf{y}) = \prod_{i=1}^N \left[\prod_{k=1}^K \pi_{ik}^{y_{ik}} \right], \quad (3.2.8)$$

onde agora temos k representando o k -ésimo nível de um total de K níveis da variável dependente (CZEPIEL, 2022).

Independentemente da função $L(\boldsymbol{\beta}|\mathbf{y})$ utilizada, é possível encontrar o estimador de máxima verossimilhança derivando a equação em relação aos parâmetros do modelo e igualando ela a zero (TAKAHASHI; INOUE, 2019):

$$\frac{\partial(l(\boldsymbol{\beta}|\mathbf{y}))}{\partial(l\boldsymbol{\beta})} = 0, \quad (3.2.9)$$

onde o $\hat{\boldsymbol{\beta}}$ resultante é o EMV para o $\boldsymbol{\beta}$ procurado.

Essa equação não tem solução analítica na regressão logística, sendo necessário recorrer a algoritmos numéricos, como o método score de Fisher, dos Modelos Lineares Generalizados. No caso binomial, o score de Fisher pode ser simplificado para o algoritmo de Newton-Raphson, que utiliza uma função polinomial de formato côncavo para aproximar a função de log-verossimilhança a um chute inicial (AGRESTI, 2007).

3.2.3 Teste de Razão de Verossimilhança

Ao encontrar o estimador de máxima verossimilhança em diferentes modelos, podemos comparar os valores encontrados para identificar o modelo que é melhor ajustado. No caso da regressão logística múltipla, essa comparação é utilizada para avaliar se o modelo com todas as variáveis independentes é significativamente melhor do que um modelo reduzido, isto é, onde nem todas as covariáveis inicialmente estabelecidas são utilizadas de fato.

Ao realizar o teste de razão de verossimilhança no caso de Y discreto, temos as seguintes hipóteses:

- H0 (Hipótese nula): O modelo completo não é significativamente distinto do modelo reduzido.
- H1 (Hipótese alternativa): O modelo completo é significativamente distinto do modelo reduzido.

E a seguinte estatística do teste:

$$\lambda = -2\ln\left(\frac{L_0}{L_1}\right) \sim \chi_v^2, \quad (3.2.10)$$

onde L_0 representa o valor máximo da função de verossimilhança sob a hipótese nula, e L_1 denota o valor máximo de modo generalizado, isto é, o valor máximo no modelo completo (AGRESTI, 2007). L_1 sempre será um valor igual ou maior a L_0 .

Essa estatística é um valor não-negativo que segue aproximadamente a distribuição qui-quadrado (χ^2) com v graus de liberdade, onde v é o número de parâmetros no modelo completo menos o número de parâmetros no modelo reduzido (NGUYEN, 2022a). Valores muito pequenos da estatística do teste favorecem H_1 , pois indicam uma grande diferença entre L_0 e L_1 .

Se, após a aplicação do teste, $P(\lambda > \chi^2)$ (ou seja, o *p-valor* do teste) for menor que α , com α sendo um nível de significância escolhido a priori, rejeitamos a hipótese nula.

3.2.4 Teste de Wald

Apesar do teste de razão de verossimilhança ser preferível, outra possível maneira de ajustar o modelo é utilizando o teste de Wald. O teste é aplicado para cada covariável x (NGUYEN, 2022b), testando as seguintes hipóteses:

- H0 (Hipótese nula): $\beta_j = 0$ (β_j não é significativo no modelo, podendo ser descartado sem grandes prejuízos).
- H1 (Hipótese alternativa): $\beta_j \neq 0$ (β_j é significativo no modelo).

Sua estatística é:

$$W = \frac{\hat{\beta}_j}{EP(\hat{\beta}_j)} \sim N(0, 1) \quad (3.2.11)$$

onde $\hat{\beta}$ é a estimativa de máxima verossimilhança de β , e $EP(\hat{\beta}_j)$ é o erro padrão de $\hat{\beta}_j$ (HOSMER; LEMESHOW; STURDIVANT, 2013). Esta estimativa pode ser encontrada a partir da inversa da matriz de informação de Fisher. No caso em questão, a mesma pode ser calculada com (TABOGA, 2021):

$$I(\hat{\beta}) = E_{\hat{\beta}}[\nabla_{\hat{\beta}} l(\hat{\beta}; y) \nabla_{\hat{\beta}} l(\hat{\beta}; y)^T], \quad (3.2.12)$$

com $l(\hat{\beta}; y)$ sendo a função de log-verossimilhança $\ln[L(\hat{\beta}; y)]$, de modo que o erro padrão seria a raiz quadrada dos elementos diagonais do inverso da matriz encontrada (MONO-

LIX, 2022).

Se, para um determinado β , $P(Z > z) + P(Z < -z)$ (ou seja, o *p-valor* do teste) for menor que o nível de significância α escolhido à priori, rejeitamos a hipótese nula (HOSMER; LEMESHOW; STURDIVANT, 2013).

3.2.5 Intervalos de confiança para Beta e Odds Ratio

Sabendo do erro padrão $EP(\hat{\beta})$, é possível calcular intervalos de confiança para β , e consequentemente também para θ .

Temos que o intervalo de confiança para cada β , com nível de confiança $1 - \alpha$ pode ser calculado com (AGRESTI, 2007):

$$I.C.(\beta) = \hat{\beta} \pm z_{\alpha/2} EP(\hat{\beta}), \quad (3.2.13)$$

onde $z_{\alpha/2}$ representa o valor de z dentro de uma distribuição normal padrão com nível de confiança $(1 - \alpha)\%$. Se considerarmos que o intervalo possui nível de confiança de 95%, o valor de $z_{\alpha/2}$ passa a ser 1,96.

Agora, sabendo que na regressão logística binária podemos escrever as *odds of success* como $\exp(\alpha + \beta x)$, é possível escrever a *odds ratio* em função de β , conforme a equação (CHANG, 2023):

$$\ln \left[\frac{\pi(x_{j+1})}{1 - \pi(x_{j+1})} \right] - \ln \left[\frac{\pi(x_j)}{1 - \pi(x_j)} \right] = \ln \left[\frac{\frac{\pi(x_{j+1})}{1 - \pi(x_{j+1})}}{\frac{\pi(x_j)}{1 - \pi(x_j)}} \right] = (\alpha + \beta) - \alpha = \beta, \quad (3.2.14)$$

onde $\pi(x_j)$ se refere à probabilidade de $Y = j$, e $\pi(x_{j+1})$ se refere à probabilidade de $Y = j + 1$, de modo que podemos inferir que $\exp(\hat{\beta})$ é a *odds ratio* estimado da probabilidade $P(Y = j + 1)$ contra a probabilidade $P(Y = j)$.

Tendo descrito θ em função de β , onde $\ln[\theta] = \beta$, chegamos enfim no intervalo de confiança para θ :

$$I.C.(\theta) = [\exp(LI(\beta)); \exp(LS(\beta))], \quad (3.2.15)$$

onde $L.I.(\beta)$ e $L.S.(\beta)$ se referem, respectivamente, aos limites inferior e superior do intervalo de confiança de β .

4 Metodologia

4.1 Conjunto de dados

Este trabalho utiliza dados primários que foram coletados por meio de um questionário (WANDERLEY, 2022) realizado via *Google Forms*, voltado para alunos da UnB. A pesquisa foi divulgada por diversos meios, entre eles estando:

- Exposição de cartazes no campus Darcy Ribeiro da UnB, sob a autorização da Prefeitura ou do Departamento onde eles foram expostos;
- Divulgação da pesquisa via e-mail, por parte de departamentos como o de Estatística e de Design;
- Mensagens em grupos privados em aplicativos como *WhatsApp* e *Telegram*, com foco em grupos específicos de cursos da UnB;
- Postagens em redes sociais.

Foi esperado alcance a um público diverso dentro da UnB, local com uma grande variedade de cursos e perfis socioculturais, evitando assim um viés excessivo.

O questionário, entretanto, teve como principal foco o público da UnB que consome jogos digitais, para que se pudesse avaliar o modo que os jogos são consumidos por este grupo, e relacioná-lo com a experiência dos respondentes no idioma inglês.

Ou seja, o público que não consome jogos digitais não é de interesse da pesquisa, pois o objetivo do trabalho não é diferenciar quem consome jogos digitais de quem não joga, mas sim diferenciar sub-grupos dentro de quem consome jogos digitais.



Figura 5: Imagem usada para divulgação

Apesar do público-alvo ser composto de alunos da universidade, outras pessoas também puderam responder o questionário, podendo inclusive informar o curso que estudam em suas respectivas faculdades, caso sejam estudantes.

O questionário possui diversas perguntas a respeito do conhecimento da língua inglesa do entrevistado, bem como de seu contato com a mesma no decorrer da vida. Além disso, a pesquisa aborda o tipo de conteúdo que a pessoa consome, aprofundando na área de jogos digitais mas também buscando saber os outros tipos de mídia consumidos. Também são feitas perguntas a respeito de idade, gênero e curso do respondente, entre outras.

No modelo logístico da pesquisa, a variável de interesse é o nível de conhecimento de inglês dos respondentes, e as variáveis explicativas são:

- Idade dos respondentes;
- Gênero dos respondentes;

- Número de banheiros na casa dos respondentes;
- Envolvimento dos respondentes com jogos digitais no último ano;
- Plataformas utilizadas pelos respondentes para jogar jogos digitais;
- Horas por dia em que os respondentes gastam em jogos digitais;
- Tipo de conteúdo em inglês que os respondentes consomem e compreendem que não seja jogos digitais;
- Gosto dos respondentes a respeito de diferentes gêneros de jogos digitais;

Outras informações obtidas do questionário serão também utilizadas neste trabalho, especialmente na análise exploratória dos dados, para que sejam feitas observações mais pontuais.

Apesar da pesquisa incluir variáveis referentes ao idioma no qual se joga diferentes gêneros de videogame, essas não serão inclusas no modelo, pois diferentemente de outras mídias mais “tradicionais” como séries e filmes, muitas vezes um indivíduo não consome videogames em inglês por escolha própria, de modo que essa informação não possui a mesma relevância numa regressão.

O questionário recebeu respostas até o mês de abril de 2023, contando com 802 respostas, 552 das quais serão efetivamente utilizadas por serem do grupo de interesse (alunos da UnB que jogam). Ele pode ser visualizado no apêndice A deste trabalho, disponível na página 65.

4.2 Metodologia

O questionário gerou uma planilha com todas as respostas, que foi posteriormente tratada para separar e transformar variáveis apropriadamente. Algumas respostas de itens de múltipla escolha, por exemplo, geraram uma única coluna de respostas. Essas colunas foram separadas em várias colunas com respostas binárias.

Após o tratamento dos dados, foi realizada uma sequência de análises exploratórias, utilizando métodos como grafos e diferentes tipos de gráfico de barras. Em seguida, foi ajustado um modelo de regressão logística com base nas variáveis do questionário, usando o software estatístico *RStudio*, na linguagem *R*, versão 4.2.1.

Na lista citada na seção “conjunto de dados”, alguns dos itens representam mais de uma variável. Por exemplo, a opinião dos respondentes a respeito de cada gênero

de jogo digital citado no questionário já engloba diversas variáveis. Ou seja, no estado mais puro do banco de dados, houveram dezenas de variáveis a serem analisadas. Para otimizar o modelo, foi utilizado o método *stepwise* para encontrar um número adequado de variáveis relevantes. Após isso, foram analisadas as *odds ratios* geradas a partir das equações logísticas do modelo, a fim de gerar conclusões a respeito dos dados coletados.

Também foram estudadas as correlações de diversas características dos respondentes com seus respectivos níveis de inglês, utilizando os coeficientes V de Cramer, Qui-quadrado de Pearson, Tau-c de Stuart-Kendall e Gama de Goodman-Kruskal, e associando os resultados ao modelo logístico.

Mesmo com 4 níveis da variável respostas presentes na pesquisa (avançado/fluyente, intermediário, básico e muito baixo), foi implementado um modelo considerando apenas 3 níveis, unindo as respostas “básico” e “muito baixo” num só nível chamado “básico”, considerando que essas opções são as menos populosas no grupo estudado.

5 Resultados

Apesar do modelo relacionar a variável dependente com múltiplas covariáveis simultaneamente, também foram construídas correlações individuais, a fim de observar o impacto isolado de certas características no nível de inglês dos respondentes. Foram estudados tanto a independência das covariáveis quanto o grau de associação das mesmas com a variável dependente, para que se pudesse concluir se há ou não associação e, posteriormente, o grau dessa associação.

5.1 Análise exploratória e correlações

5.1.1 Nível de inglês

A variável “nível de inglês” foi agrupada em 3 categorias: “baixo ou não sabe inglês”, “intermediário” e “avançado ou fluente”. A Figura 6 ilustra em um gráfico de barras a distribuição observada do nível de inglês dos respondentes, o qual será a variável dependente do modelo:

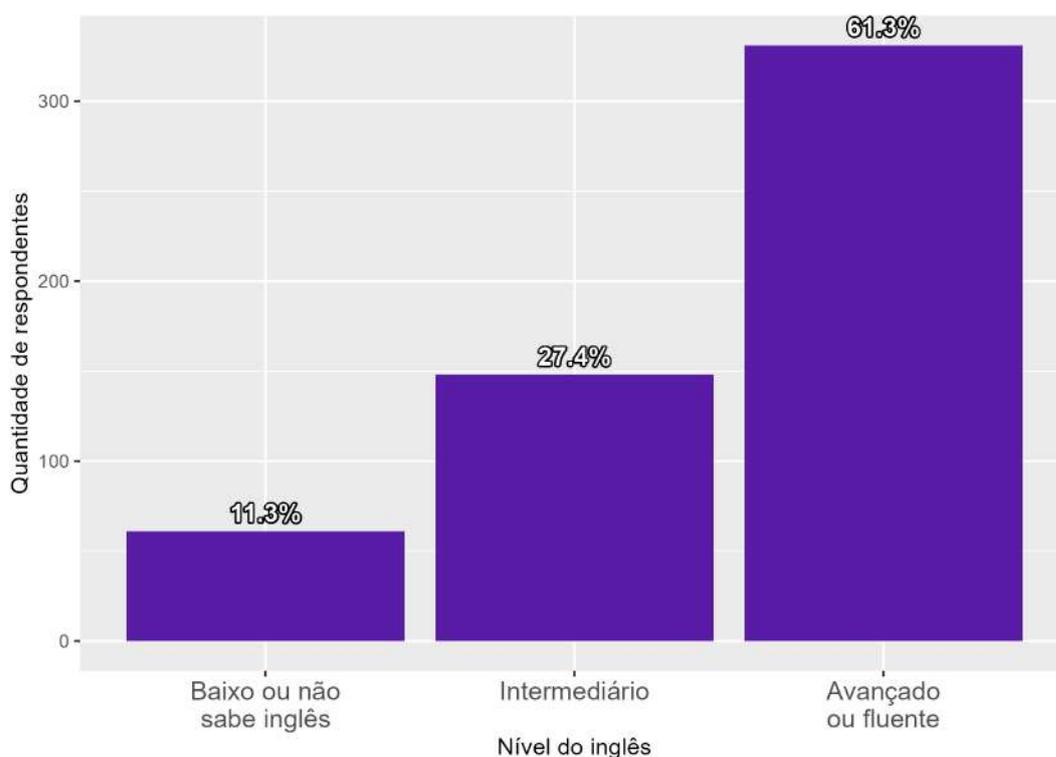


Figura 6: Distribuição do nível de inglês dos alunos da UnB respondentes

De acordo com a Figura 6, dentre os alunos da UnB, 11,3% responderam que o nível é baixo ou não sabem inglês, 27,4% afirmaram que o nível é intermediário, e 61,3% disseram que o próprio nível de inglês é avançado ou fluente. Esses serão os níveis 1, 2 e 3 da variável resposta, respectivamente.

Estes dados são divergentes da realidade brasileira, o que demonstra tanto um possível maior grau de instrução no quesito língua inglesa dentro dos universitários que jogam, quanto uma predisposição daqueles que responderam ao questionário a saber inglês, visto que não foi realizada uma amostragem probabilística na coleta dos dados devido às limitações da pesquisa.

5.1.2 Conteúdo em inglês consumido

No gráfico de barras da Figura 7, são mostrados os tipos de conteúdos em inglês consumidos pelos respondentes.

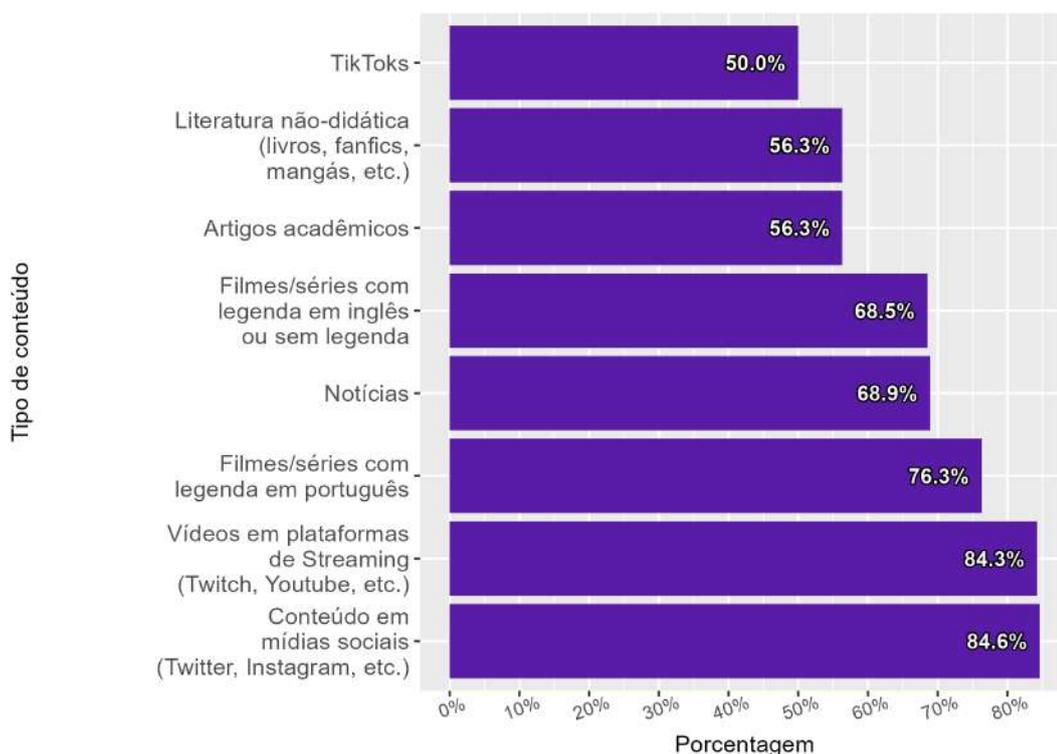


Figura 7: Distribuição do conteúdo em inglês consumido pelos respondentes

Segundo o gráfico, o conteúdo em inglês mais consumido é aquele das mídias sociais como *Twitter* e *Instagram*, com 84,6% de representatividade. Esse resultado era esperado, pois as redes sociais são uma constante na vida de milhões de brasileiros, es-

pecialmente o público jovem, que compõe grande parte do público-alvo da pesquisa. Nas redes, é comum entrar em contato com conteúdo do estrangeiro, mesmo que não intencionalmente, graças aos algoritmos que regem seu funcionamento.

O mesmo ocorre com plataformas como *Youtube* e *Twitch*, que ocupam a segunda colocação com porcentagem próxima. Essas plataformas são muito utilizadas não só para se conhecer jogos e assistir jogatinas, mas também para a criação e consumo de conteúdo relacionado a videogames, como *video essays* sobre o tema e paródias musicais, conforme pode ser visto nos canais *Game Maker's Toolkit* e *SiIvagunner*, respectivamente (BROWN, 2006) (SIIVAGUNNER, 2016). Grande parte desse conteúdo está em inglês.

Em terceiro lugar, estão filmes/séries com legenda em português, demonstrando como esse tipo de acessibilidade contribui para o acesso do público a conteúdos diversos, e, portanto, à cultura. Há uma diferença de 7,8% entre essa opção e o consumo de filmes/séries com legenda em inglês ou sem legenda. Na análise do modelo de regressão, será mostrada a relação do consumo desses conteúdos em particular com o nível de inglês dos indivíduos.

Como os tipos de conteúdo e o nível de inglês são variáveis categóricas ordinais, é possível avaliar a presença, a direção e a força da associação entre ambas por meio do Gama de Goodman-Kruskal e do Tau-c de Stuart-Kendall, o qual é otimizado para casos onde o número de valores das duas variáveis difere (BOTSCH, 2011). No gráfico de barras da Figura 8, é apresentada a associação da variável dependente com a covariável indicadora do conteúdo consumido em inglês pelos entrevistados.

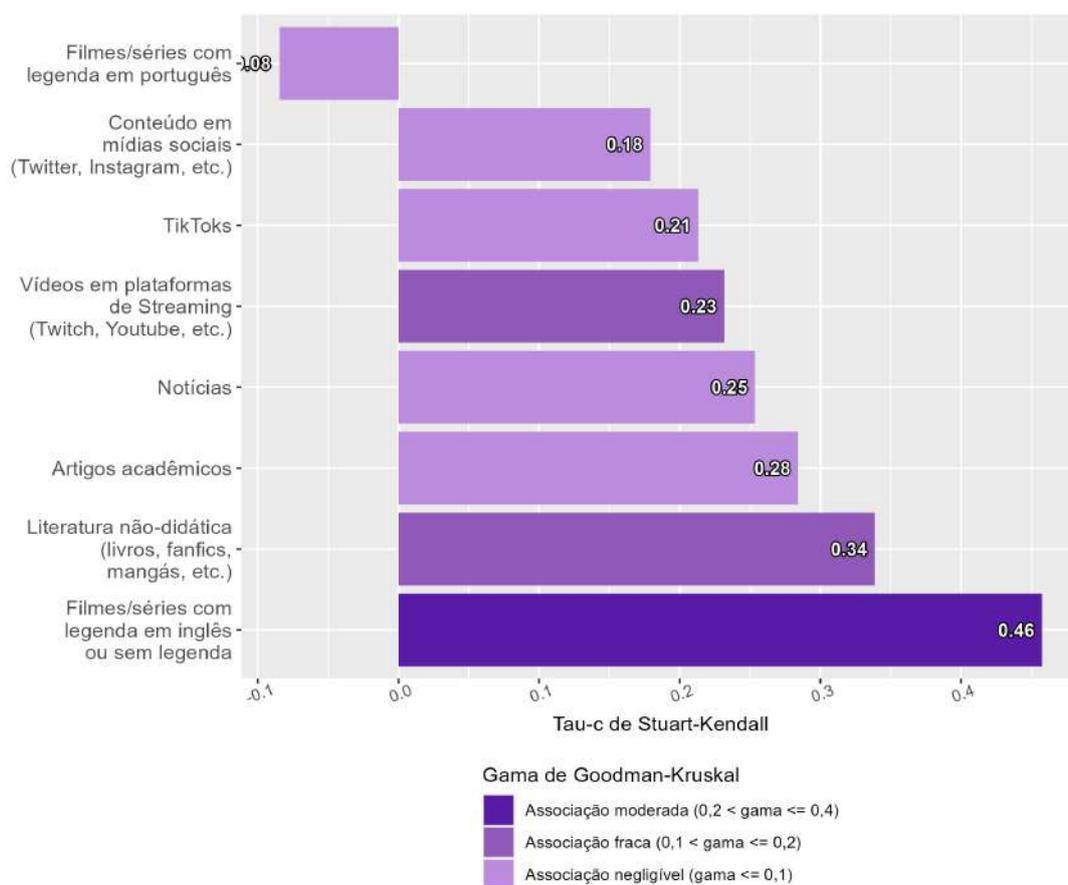


Figura 8: Associação entre conteúdo consumido em inglês e nível de inglês. As tonalidades de lilás indicam o coeficiente Gama de Goodman-Kruskal, enquanto o tamanho das barras indicam o Tau-c de Stuart-Kendall

O coeficiente Gama de Goodman-Kruskal está resumido na Figura 8 por meio das 3 diferentes tonalidades da cor roxa. Valores inferiores à 0,1 indicam associação negligível, valores entre 0,1 e 0,2 indicam associação fraca, e valores entre 0,2 e 0,4 indicam associação moderada. Já no caso do Tau-c de Stuart-Kendall, valores cujo módulo é superior à 0,3 mostram associação forte, que fica mais fraca na medida que o módulo decresce. No geral, o Tau-c apresentou valores maiores, possivelmente por sua otimização ao caso em questão, onde o número de níveis da variável dependente (3) e o número de níveis das covariáveis (2) difere.

A maior associação observada foi com filmes/séries com legenda em inglês ou sem legenda, variável que também apresenta altas razões de chance no modelo logístico segundo a Tabela 3, sendo um bom reflexo do nível do idioma do indivíduo. Outras variáveis com associação comparativamente maiores são os consumos de artigos acadêmicos e de literatura na língua inglesa. De fato, essas mídias envolvem leituras mais extensas,

necessitando assim de conhecimento mais avançado do inglês.

Segundo a estatística Tau-c, a única medida de direção negativa é a da variável “filmes/séries com legenda em português”, ressaltando novamente como consumir mídia audiovisual desta maneira pode indicar que o inglês do espectador não é avançado, mesmo que seja intermediário.

A covariável “conteúdo em mídias sociais” possui associação baixa segundo ambas as estatísticas. Esse resultado mostra que as mídias sociais são muito difundidas na população como um todo, inclusive com conteúdos em inglês. Ela não foi incluída no modelo logístico por não ter sido significativa para explicar o nível de inglês, o que está em concordância com o que foi observado aqui.

5.1.3 Variáveis categóricas diversas

No caso de variáveis categóricas, o Qui-quadrado de Pearson pode ser utilizado para avaliar se há independências em relação variável dependente (nível de inglês), pois neste teste a hipótese alternativa indica que as variáveis são dependentes (GOYAL, 2020). Entretanto, para medir esta relação, a estatística V de Cramer mostra-se adequada, considerando a nominalidade de uma das variáveis. Nela, a intensidade da associação pode ser medida de modo similar ao Tau-c de Stuart Kendall, de modo que valores acima de 0,3 indicam associação forte (BOTSCH, 2011). Essa será a análise realizada nas etapas seguintes desta pesquisa.

Na Figura 9, essas estatísticas são apresentadas para diversas variáveis da pesquisa.

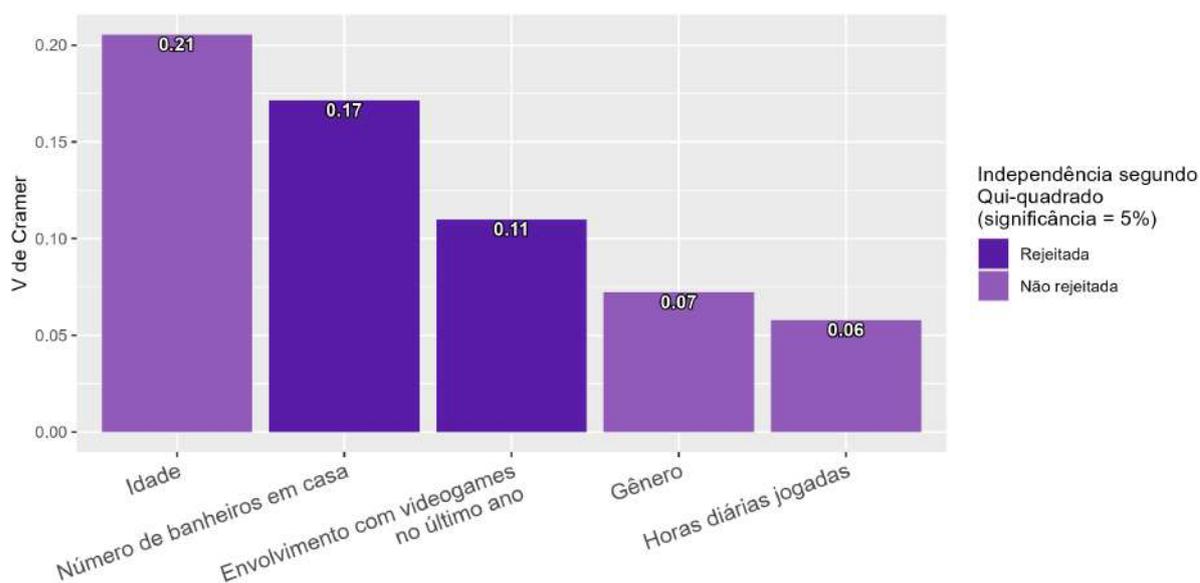


Figura 9: Associação entre nível de inglês e covariáveis diversas

Apesar da variável “idade” ter obtido uma associação moderada segundo o V de Cramer, a sua independência não foi rejeitada, indicando que idade não necessariamente tem relação com habilidade na língua inglesa.

Conforme apresentado na Figura 9, nenhuma das variáveis obteve uma associação particularmente alta segundo a estatística V de Cramer, mas o número de banheiros em casa se destaca entre aquelas que tiveram a independência rejeitada, indicando a relevância do fator socioeconômico no acesso ao conhecimento. Isso é reforçado no gráfico de barras da Figura 10, que mede a associação do nível de inglês com as covariáveis relativas às plataformas onde os entrevistados jogam jogos digitais.

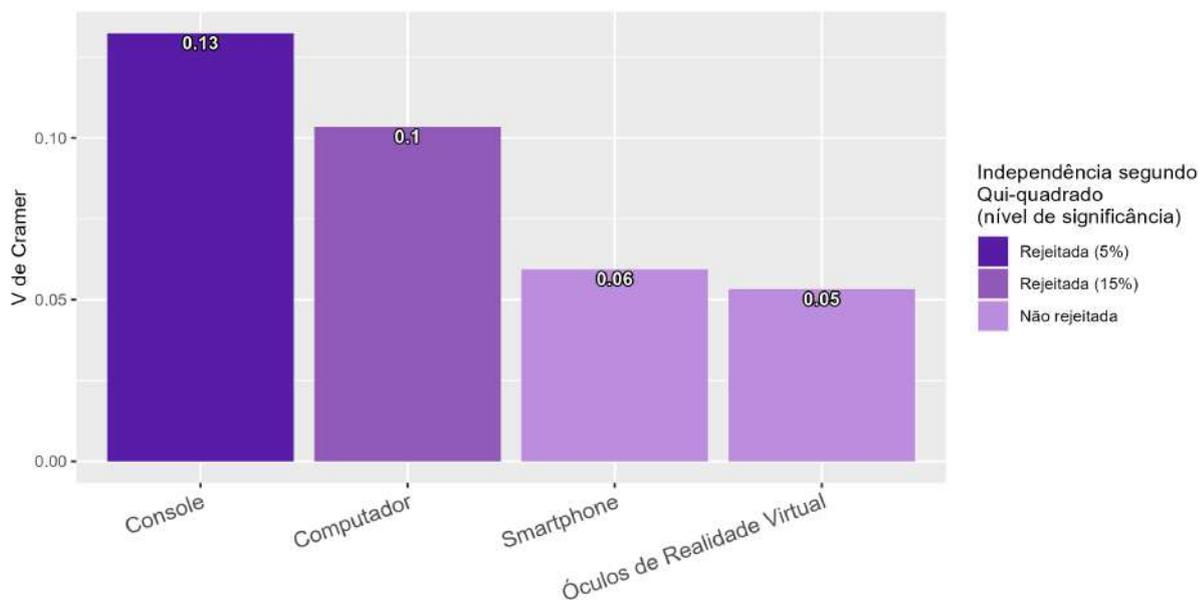


Figura 10: Associação entre plataformas onde se joga videogame e o nível de inglês

No gráfico da Figura 10, a única variável com independência rejeitada a um nível de significância de 5%, bem como a que apresenta maior V de Cramer, é a relativa ao uso de consoles de videogame. As plataformas “computador” e “smartphone” não obtiveram resultados tão positivos, possivelmente por serem mais amplamente acessíveis hoje em dia, enquanto consoles ainda são caros e de certa forma restritos a quem possui classe social mais elevada. Esta covariável foi a única inclusa no modelo logístico dentre essas.

Óculos de Realidade Virtual, por sua vez, não são tão utilizados, inclusive por quem possui classe social elevada, e portanto não há dados concretos o suficiente para se chegar em uma conclusão a respeito de sua utilização, conforme visto no gráfico.

5.1.4 Gênero

O gênero dos alunos da UnB respondentes foi distribuído em 71,5% masculino e 25% feminino, sendo que 3,7% dos respondentes tem outra identificação ou preferiram não informar o próprio gênero, conforme o gráfico de barras da Figura 11.

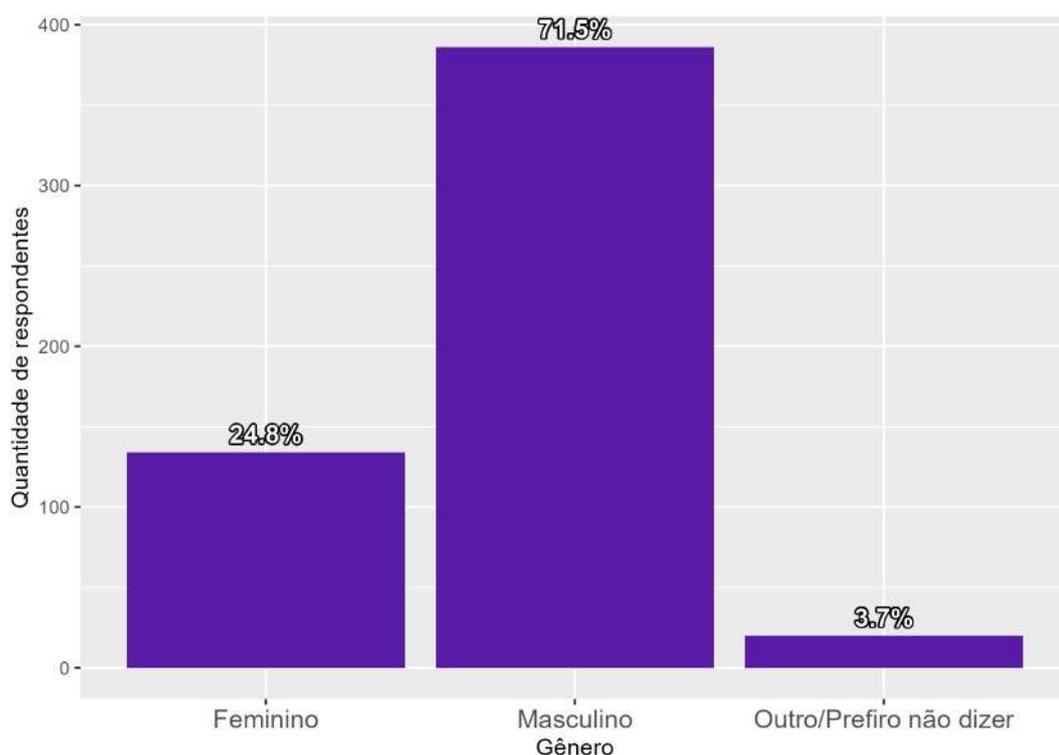


Figura 11: Distribuição do gênero dos respondentes

O resultado observado reforça a ideia de que, atualmente, jogos digitais são vistos como um *hobby* masculino, excluindo o restante do público de diversas maneiras, seja em um processo de desenvolvimento de jogos que não visa mulheres como público-alvo significativo ou no comportamento dos jogadores em jogos online, conforme apontado no artigo *Sexism in video games* (BURNETT; WEGMANN, 2020).

5.1.5 Tempo jogado

Dentre as opções relativas ao tempo diário de jogatina digital dos respondentes, a mais comum foi a referente ao tempo de 1 a 2 horas diárias, conforme a Figura 12.

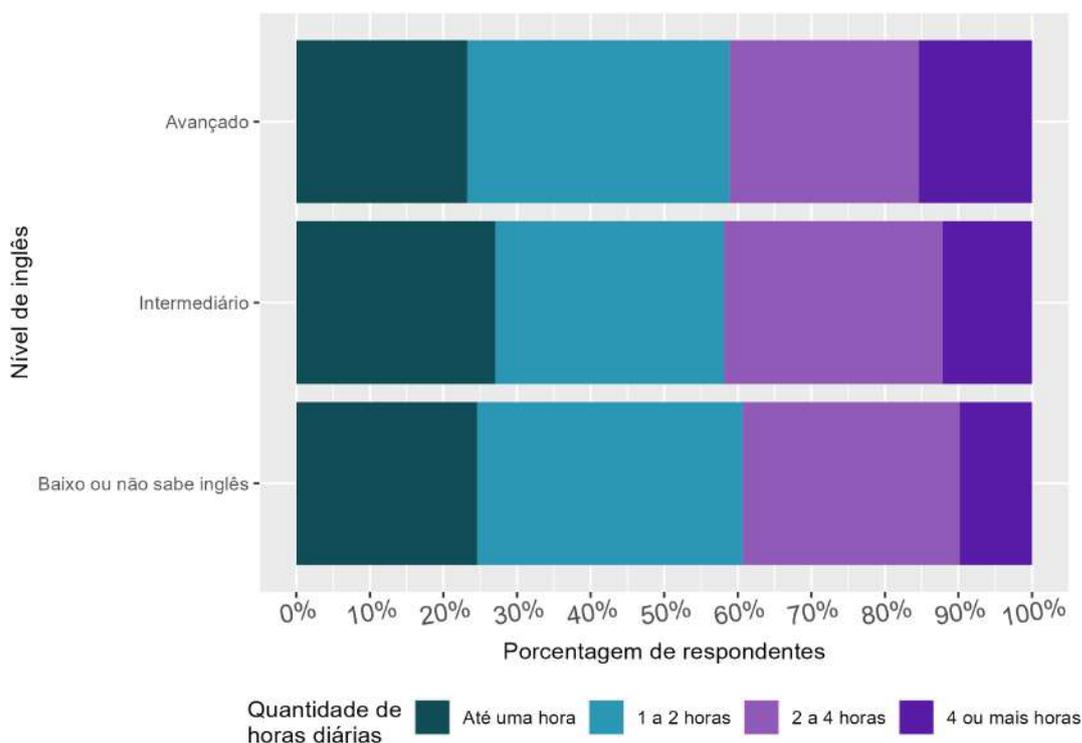


Figura 12: Quantidade de horas diárias em que os respondentes jogam videogame

Nenhuma das outras opções possui representatividade consideravelmente baixa, inclusive as opções que apresentam tempos maiores de dedicação a jogos digitais. Isso reforça como os jogos são parte considerável da vida de muitos jovens e, portanto, refletem o acesso à cultura de quem os consome. Também é evidente como a distribuição da quantidade de horas é similar nos três grupos de nível de inglês, indicando que o conhecimento do idioma não precede o tempo diário de jogatina de um indivíduo. Esta informação será confirmada pelo modelo logístico, que excluirá a variável em questão por não considerá-la significativa o bastante para indicar o nível de inglês.

5.1.6 Interesse e afastamento de conteúdo com base no idioma

No questionário, foi perguntado se, nos últimos anos, os respondentes já se afastaram de algum tipo de conteúdo por não ter domínio da língua inglesa. No gráfico de barras da Figura 13, é apresentada a distribuição das respostas dentro dos diferentes grupos de domínio do inglês.

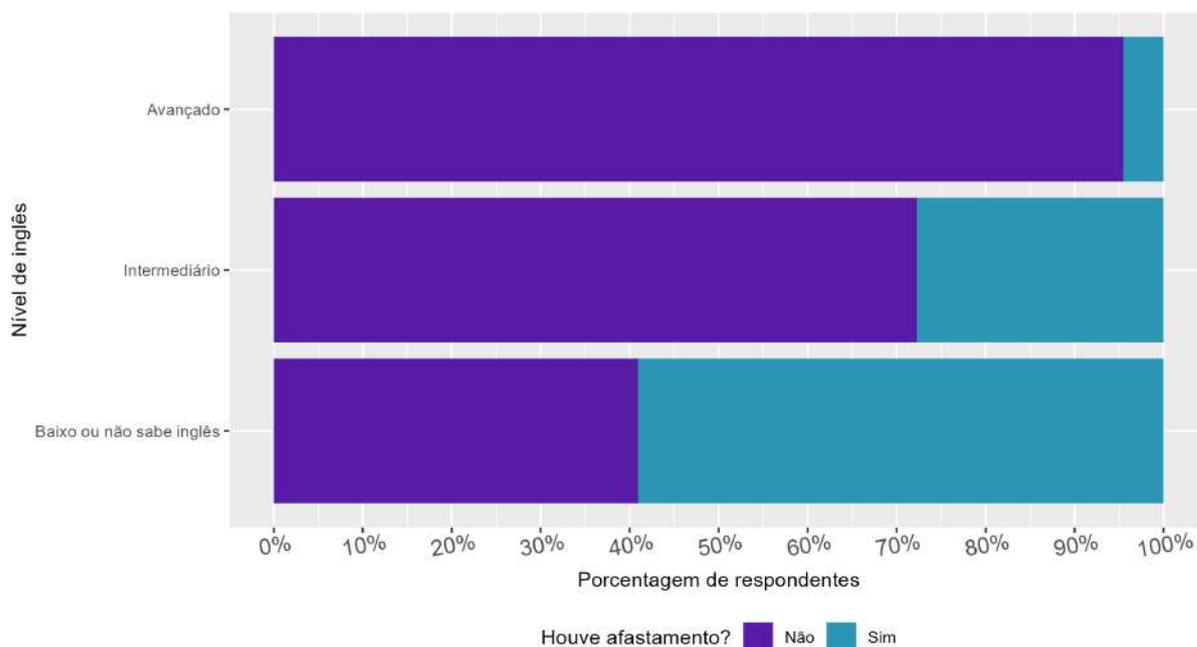


Figura 13: Distribuição das respostas quanto ao afastamento de conteúdo qualquer pela falta de conhecimento da língua inglesa

Pode ser observado que, naturalmente, maior domínio do idioma implica menor índice de afastamento. Entretanto, aproximadamente 40% dos respondentes com nível baixo de inglês não se afastaram de nenhum conteúdo, mesmo sem ter um conhecimento que poderia ser essencial para o aproveitamento de alguma obra. Essa informação pode indicar que o conteúdo consumido por esse grupo não necessita da língua inglesa. Um exemplo de conteúdo desse tipo seria, por exemplo, séries e filmes legendados. Ou seja, a acessibilidade contempla essa parcela do grupo.

Em contrapartida, foi perguntado se os respondentes já se interessaram por algum conteúdo nos últimos anos por ter sido lançada uma versão em português. Casos desse tipo de fenômeno incluem a tradução de livros e jogos que anteriormente não estavam disponíveis no idioma, de modo oficial ou não. Como exemplo não oficial, há as traduções da *Brazil Alliance*, grupo aberto dedicado a traduzir primariamente os jogos da série *Like a Dragon* (BRAZIL Alliance, 2023). O gráfico de barras da Figura 14 mostra a distribuição das respostas dentro dos diferentes grupos de domínio do inglês.

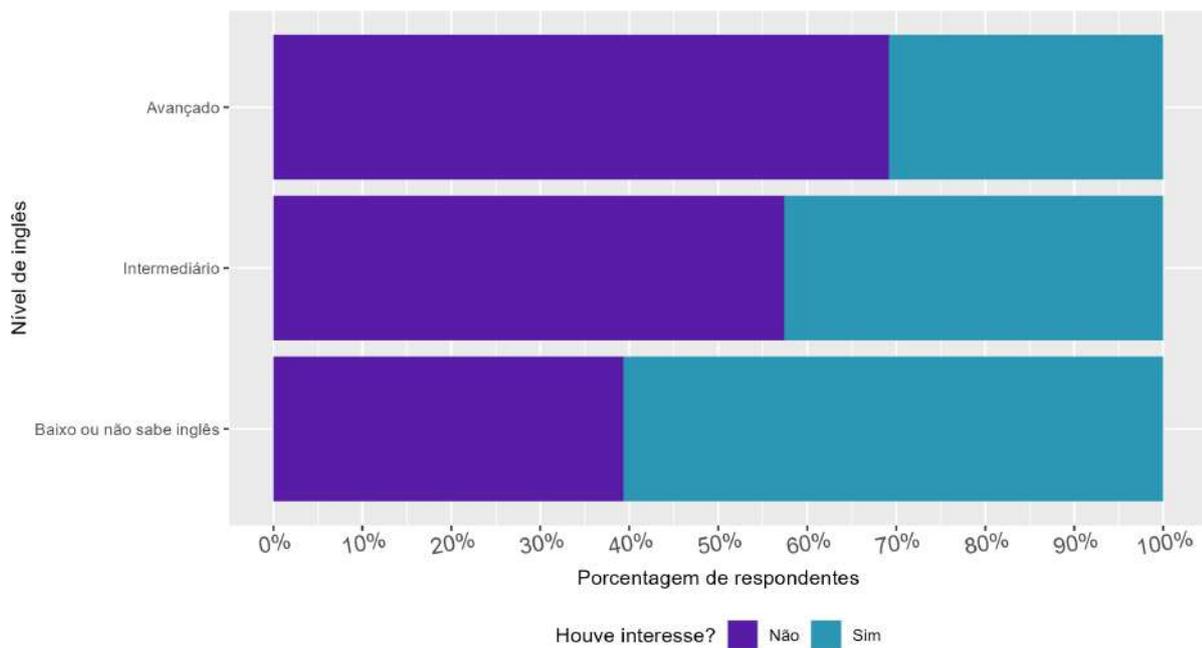


Figura 14: Distribuição das respostas quanto ao interesse em conteúdos quaisquer devido ao lançamento de uma versão em português

Assim como no gráfico da Figura 13, a porcentagem de interesse decai com o crescimento do nível de inglês. Ou seja, esse tipo de lançamento interessa mais a quem não tem domínio da língua inglesa, demonstrando a importância de se contemplar esse grupo, que é maioria no país (ANDRADE, 2022).

5.1.7 Gosto por diferentes gêneros de jogo

Algumas das perguntas do questionário coletavam informações a respeito dos gêneros dos jogos consumidos pelos respondentes. Para maior esclarecimento, foram apresentados também alguns exemplos de jogos de cada gênero, conforme lista abaixo:

- Jogos de tiro no estilo FPS (*Call of Duty*, *Counter Strike*, *Overwatch*, etc.);
- Jogos de luta (*Mortal Kombat*, *Super Smash Bros.*, *Street Fighter*, etc.);
- Visual Novels/Investigação (*Ace Attorney*, *Danganronpa*, *L.A. Noire*, *Doki Doki Literature Club*, *Outer Wilds*, etc.);
- Jogos de plataforma (*Super Mario*, *Sonic*, *Celeste*, *Crash Bandicoot*, etc.);
- Jogos de ritmo (*OSU!*, *Guitar Hero*, *Just Dance*, etc.);

- Jogos de corrida/esporte (*FIFA, Forza, Mario Kart, Rocket League*, etc.);
- Jogos de puzzle (*Tetris, Candy Crush, Portal*, etc.);
- Jogos no estilo RPG/estratégia (*Persona, Final Fantasy, Pokémon, Teamfight Tactics*, etc.);
- Jogos de ação (*Devil May Cry, God of War, Nier:Automata*, etc.);
- Jogos de simulação (*Stardew Valley, Animal Crossing, Harvest Moon*, etc.);
- MOBAs (*League of Legends, DOTA 2, Pokémon UNITE*, etc.);
- Jogos com foco em exploração/aventura (*Metroid, Genshin Impact, Minecraft, The Legend of Zelda*, etc.);

Vale ressaltar que não só existem mais gêneros, como também foi feito um agrupamento de vários gêneros em algumas opções. A opção “Jogos com foco em exploração/aventura”, por exemplo, contempla gêneros como “Sandbox”, “Metroidvania”, “mundo aberto”, entre outros.

No universo dos jogos digitais, é possível que um só jogo tenha mais de um gênero, assim como ocorre em filmes. Portanto, a simplificação realizada no questionário foi feita para manter a objetividade na pesquisa. Com esse mesmo propósito, os nomes dos diferentes gêneros foram simplificados nos gráficos a seguir para mantê-los concisos.

No apêndice B deste trabalho, disponível na página 75, estão disponíveis descrições de cada gênero apresentado no questionário, considerando as particularidades descritas acima, auxiliando assim na compreensão dos resultados da pesquisa.

A Figura 15 mostra um gráfico de barras com a distribuição do gosto dos respondentes pelos diferentes gêneros de jogos de videogame.

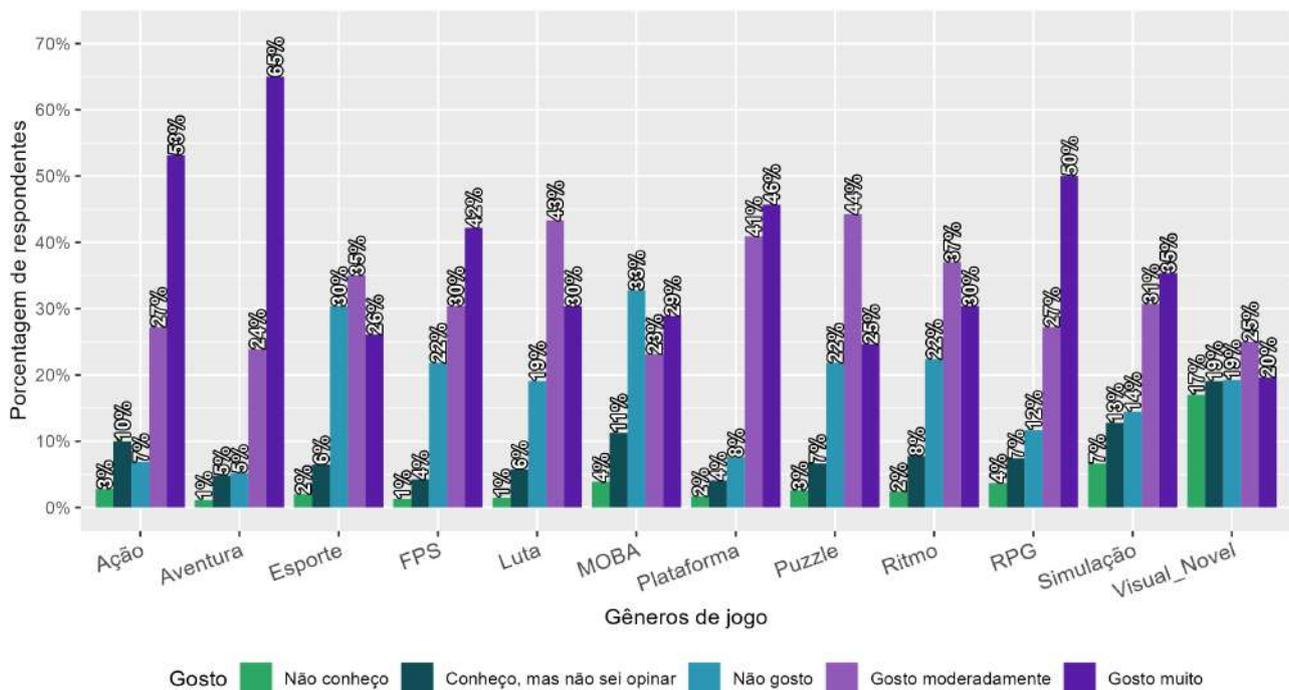


Figura 15: Distribuição do gosto dos respondentes por diferentes gêneros de jogo digital

Entre as informações notáveis que podem ser obtidas no gráfico da Figura 15, estão:

- Os gêneros “Ação” e “Aventura” possuem o maior índice da resposta “Gosto muito”. O gênero “Aventura”, em particular, possui o menor índice de “Não gosto”;
- O gênero “FPS” possui a menor ocorrência da opção “Conheço, mas não sei opinar”, e um dos menores índices da opção “Não conheço”. Ou seja, é possível concluir que esse é um dos gêneros mais populares, onde os respondentes tem mais opiniões formadas. Tamaña popularidade pode ser observada nas vendas de jogos desse gênero, como no caso de *Call of Duty: Modern Warfare II*, que obteve mais de 1 bilhão de dólares em vendas em menos de duas semanas após seu lançamento (KAIN, 2022).
- Os gêneros “Esporte”, “Luta”, “Puzzle” e “Ritmo” possuem mais ocorrências da opção “Gosto moderadamente” em relação à opção “Gosto muito”. Esses gêneros são conhecidos por terem alto *skill ceiling*, isto é, um alto nível de dedicação necessário para ter uma melhor performance. Isso é refletido no cenário competitivo dos jogos, onde há, por exemplo, torneios profissionais de jogos de luta como *Street Fighter* e disputas pela maior pontuação em músicas de jogos de ritmo como *Osu!*. Ambas

as situações requerem um nível de habilidade que não está presente em um jogador casual.

- O gênero “MOBA” é o único onde a opção “Não gosto” possui representatividade maior que a opção “Gosto muito”, tendo também a maior ocorrência de “Não gosto”. Jogos como *League of Legends* são conhecidos por apresentarem um ambiente hostil para muitos jogadores, em especial por causa de comportamento tóxico no *chat* do jogo, onde é possível interagir tanto com aliados quanto com oponentes. Em 2016, um jogador profissional foi multado por apresentar esse tipo de comportamento (GACH, 2016).
- O gênero “Visual Novel” não só é o mais desconhecido, como também é o gênero onde os respondentes que conhecem tem menos opinião formada. Por jogos desse tipo terem o maior foco na leitura, e não em outros métodos de interação mais tradicionais, esse gênero é comumente categorizado como “de nicho”. Vale ressaltar que esse é o gênero que mais depende do usuário saber o idioma no qual o jogo está disponível.

O gráfico de barras da Figura 16 mostra a associação entre o nível de inglês e o gosto dos respondentes por gêneros de jogos.

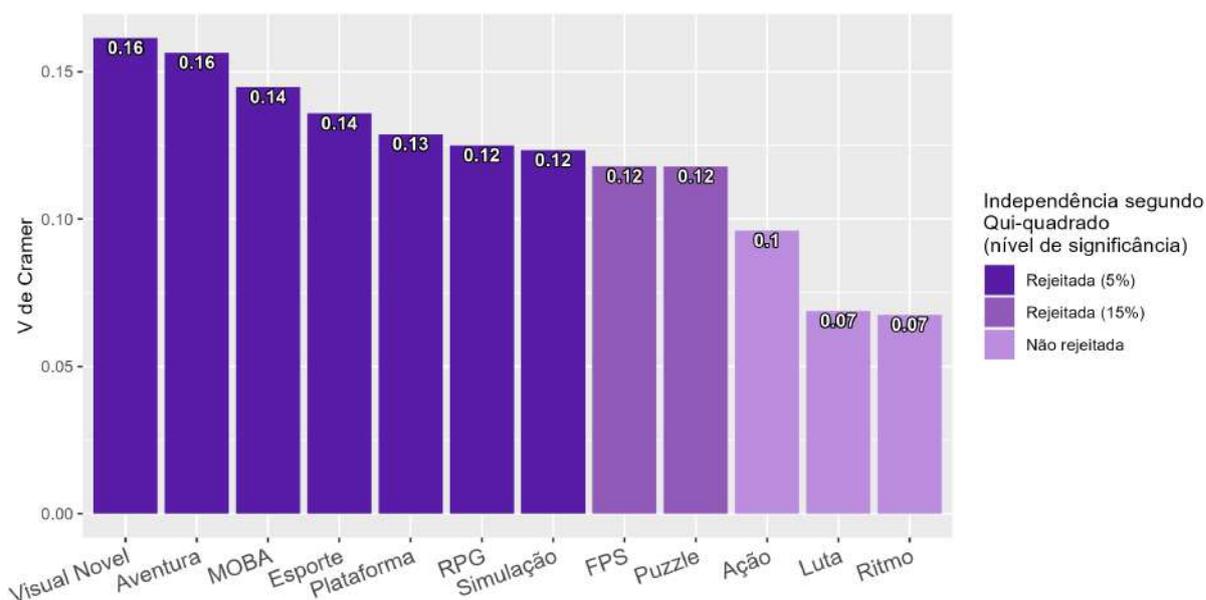


Figura 16: Associação entre gosto em gêneros de videogame e nível de inglês

Apesar de nenhum gênero apresentar, individualmente, uma associação particu-

larmente alta, o gênero “Visual Novel” se destaca, possivelmente por ser um grande “divisor de águas” dentro da mídia, devido ao seu conteúdo nichado e dependente da leitura.

Vale lembrar que, nesta análise da Figura 16, as covariáveis não são binárias. Ou seja, as covariáveis não são tão simples quanto “gostar ou não”, estando também inclusas opiniões como “não conheço”. Este pode ser o motivo pelo qual o gênero MOBA encontra-se entre os principais desta análise, mesmo mostrando ter relação inversamente proporcional ao nível de inglês no modelo logístico. O fato do V de Cramer não indicar direção da associação também contribui para este fato.

Os gêneros “Ação”, “Luta” e “Ritmo” apresentam independência rejeitada pelo Qui-quadrado e associação fraca pelo V de Cramer. De fato, são gêneros pouco dependentes do idioma, facilmente aproveitados por jogadores de diversos níveis de inglês.

5.1.8 Idioma no qual se joga diferentes gêneros de jogo

A Figura 17 demonstra o idioma no qual os respondentes consomem os diferentes gêneros de jogos, por meio de um gráfico de barras.

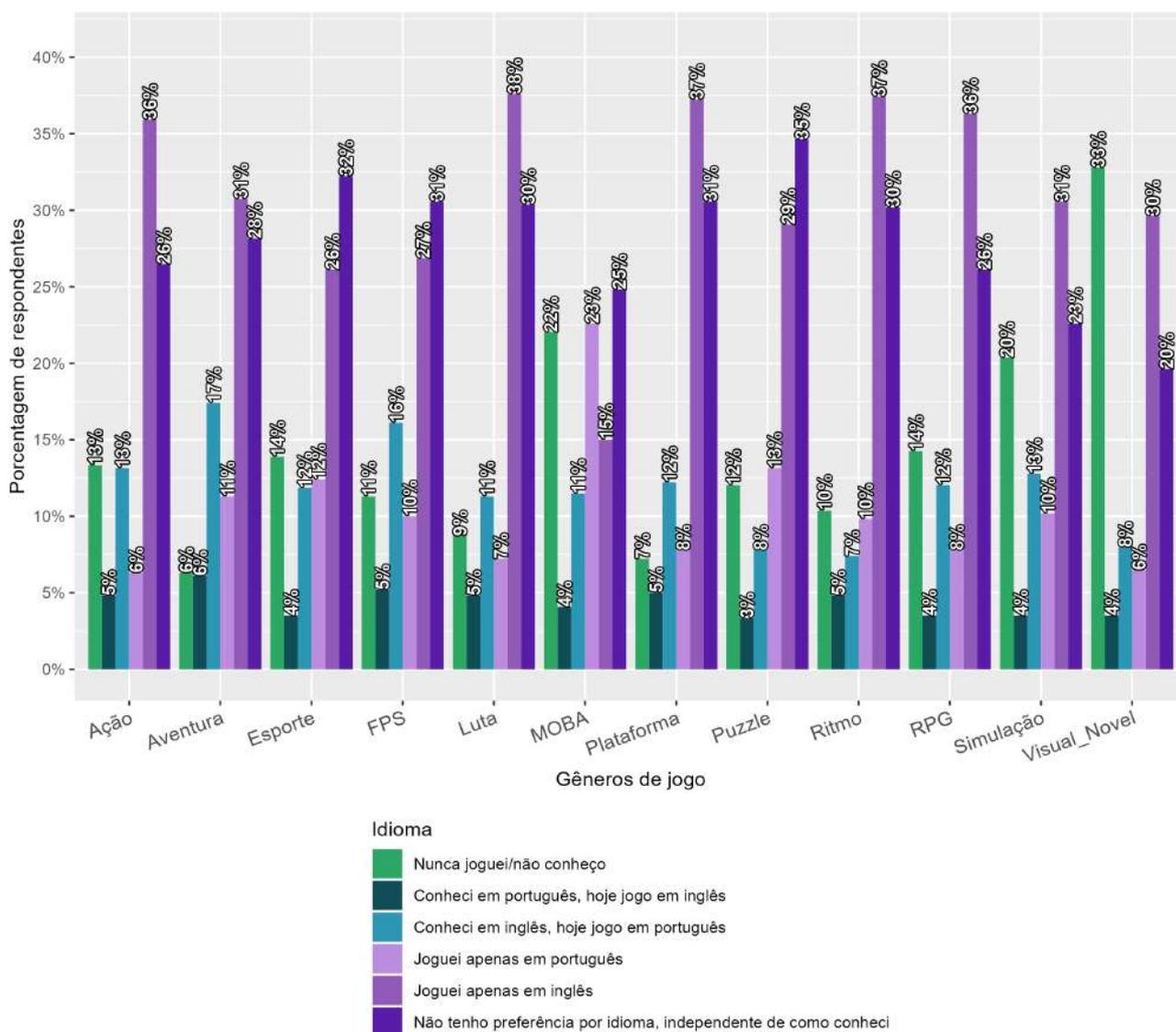


Figura 17: Distribuição do idioma no qual os respondentes consomem diferentes gêneros de jogos

Segundo o gráfico da Figura 17, a grande maioria dos gêneros possui a opção “Joguei apenas em inglês” como majoritária, demonstrando como a indústria não prioriza o idioma português em suas produções.

Também pode ser percebido que a opção “Conheci em inglês, hoje jogo em português” é sempre mais comum que “Conheci em português, hoje jogo em inglês”. Esta diferença deixa claro que a transição para o idioma materno, se possível, é mais prevalente que o processo oposto, deixando evidentes os reflexos da acessibilidade no meio.

Nos gêneros “Esporte” e “Puzzle”, a resposta mais comum foi “Não tenho preferência por idioma, independente de como conheci”. É possível afirmar que, nesses gêneros de jogo, o conhecimento do inglês não é tão necessário, pois um conhecimento básico das

regras é o suficiente para que um jogador possa progredir. Por exemplo, as regras de uma corrida são facilmente compreendidas por quem já conhece o esporte fora dos videogames, assim como as regras de *Tetris* podem ser aprendidas apenas observando o jogo. Muitos desses jogos utilizam linguagem não-verbal para demonstrar suas regras, e mesmo quando esse não é o caso, é possível aprender na base da tentativa e erro enquanto o jogador manuseia o controle.

A maior ocorrência da opção “Joguei apenas em português”, assim como a menor ocorrência de “Joguei apenas em inglês” está no gênero “MOBA”, onde os jogos mais populares como *DOTA 2* e *Pokémon UNITE* possuem o idioma português, como pode ser observado na Figura 18.



Figura 18: *Pokémon UNITE*, sendo jogado com interface em Português (GOOGLE, 2023)

Em contrapartida, o gênero “Visual Novel” possui a menor ocorrência de “Joguei apenas em português” e de “Conheci em inglês, hoje jogo em português”, demonstrando como dificilmente um jogo desse tipo é lançado em português, considerando sua falta de popularidade e o trabalho que seria necessário para localizar um jogo desse tipo para outro idioma.

Nesses jogos, o inglês frequentemente não é sequer o idioma original, e a localização para o idioma já é repleta de desafios. E no gênero, é ainda mais imperativo inserir a ideia original do roteiro de modo que o jogador compreenda, como ocorre na localização oficial de japonês para inglês nos jogos da franquia *Zero Escape* (GÓES, 2015). Por causa

disso, é comum o inglês ser o único idioma disponível além do original.

Segundo a Figura 19, que apresenta um gráfico de barras associando nível de inglês com idioma no qual se consomem jogos, o idioma tem associação relativamente maior que o gosto individual por certos gêneros.

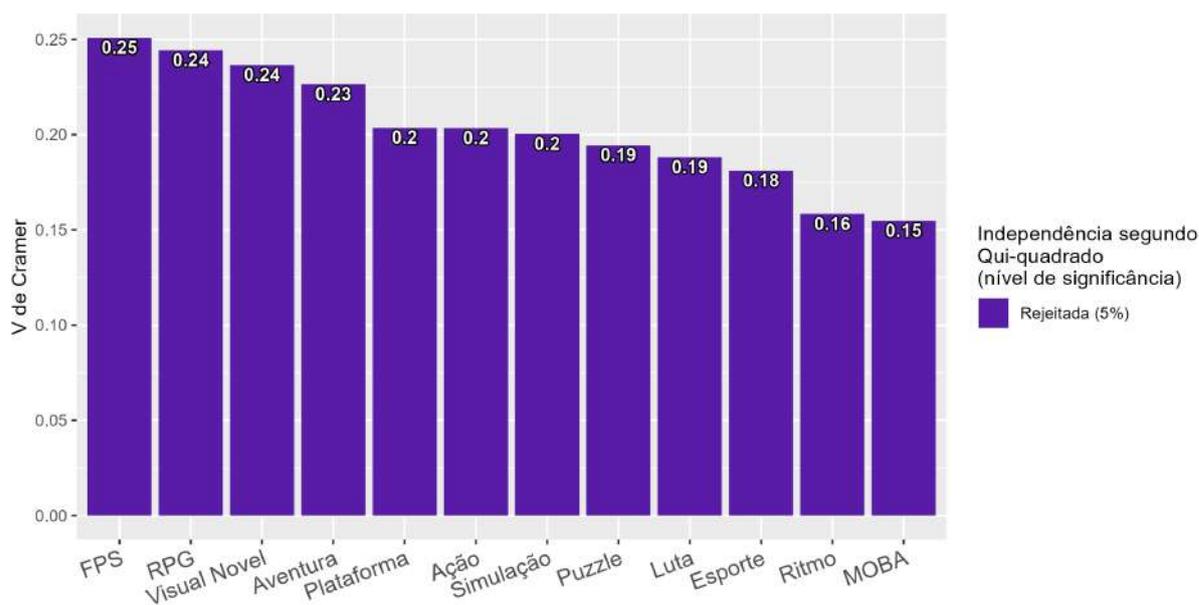


Figura 19: Associação entre idioma no qual se consomem jogos e nível de inglês

Percebe-se na Figura 19 que a independência foi rejeitada em todos os gêneros, mostrando que há sim uma associação, mesmo que baixa. Naturalmente, os gêneros “RPG” e “Visual Novel” apresentam algumas das maiores associações, considerando que esses gêneros dependem bastante de leitura.

Surpreendentemente, o gênero “FPS” possui a maior das associações, mesmo sem depender tanto do idioma no qual se joga. Isso pode ter acontecido por diversos motivos, entre os quais estão a falta de direção da estatística V de Cramer, e o fato da covariável não ser simplesmente “jogar ou não em inglês”, conforme explicado anteriormente. Possivelmente, como o gênero é popular e acessível em português, jogar em inglês por escolha pode refletir bastante o conhecimento do jogador no idioma. Isto em particular pode ser investigado em trabalhos futuros.

5.1.9 Contato com o inglês

Foram obtidos dados relativos ao contato dos respondentes com a língua inglesa na infância, representados no gráfico de barras da Figura 20:

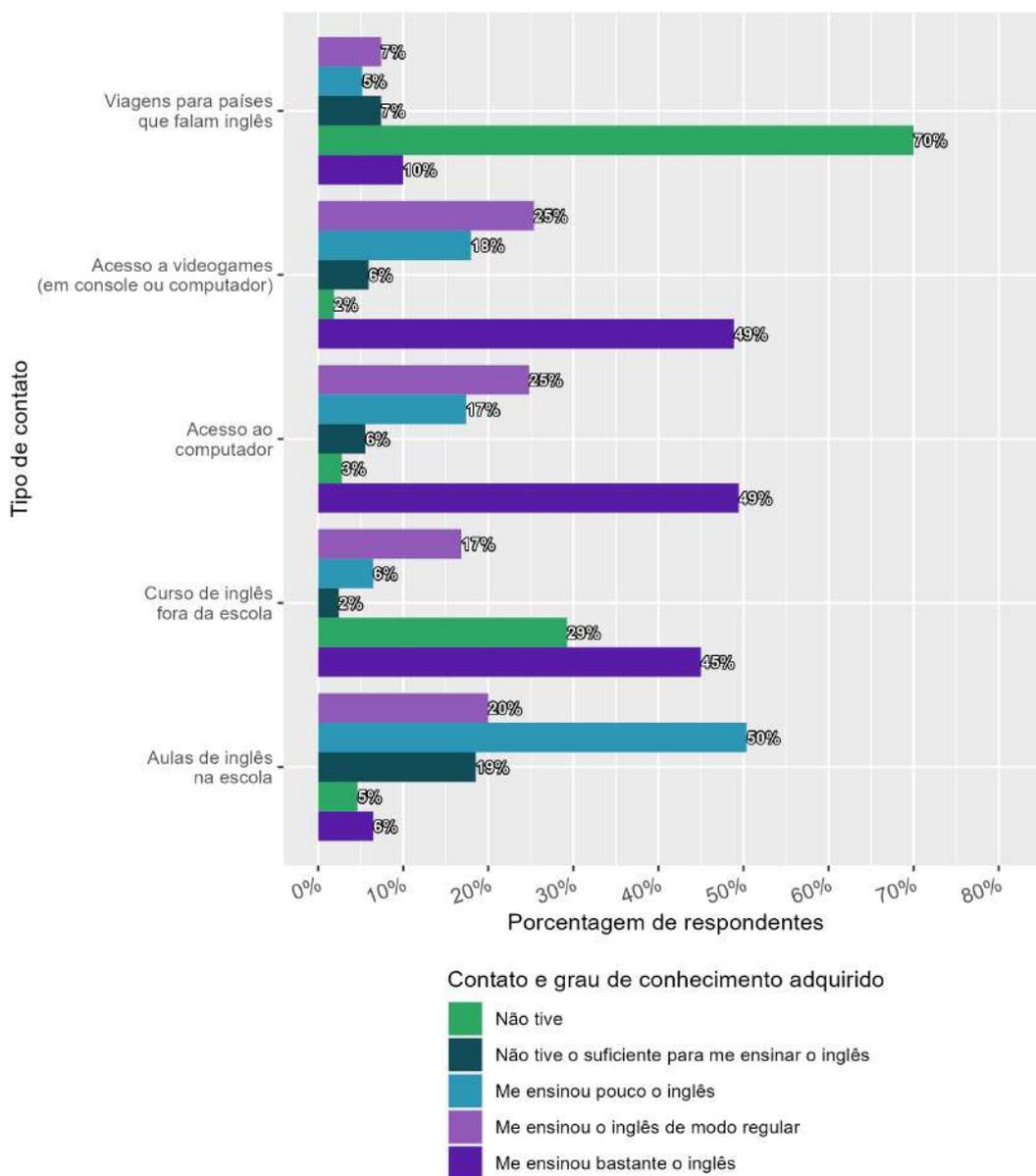


Figura 20: Tipos de contato dos respondentes com a língua inglesa na infância e o impacto no conhecimento adquirido

Considerando as opções apresentadas no questionário, observando a Figura 20, conclui-se que:

- Aulas de inglês na escola ensinaram pouco o idioma inglês aos respondentes.

- Acesso ao computador, acesso a videogames, e cursos de inglês fora da escola ensinaram bastante o idioma inglês aos respondentes.
- Viagens para países que falam inglês foram pouco experienciadas pelos respondentes, mas o efeito mais comum dentre aqueles que viajaram foi um bom ensino da língua. Algo similar ocorre com cursos de inglês fora da escola, onde o índice das respostas “não tive” ainda é significativo.

As maiores associações com o nível de inglês na pesquisa foram encontradas com as covariáveis deste tópico, conforme gráfico de barras da Figura 21.

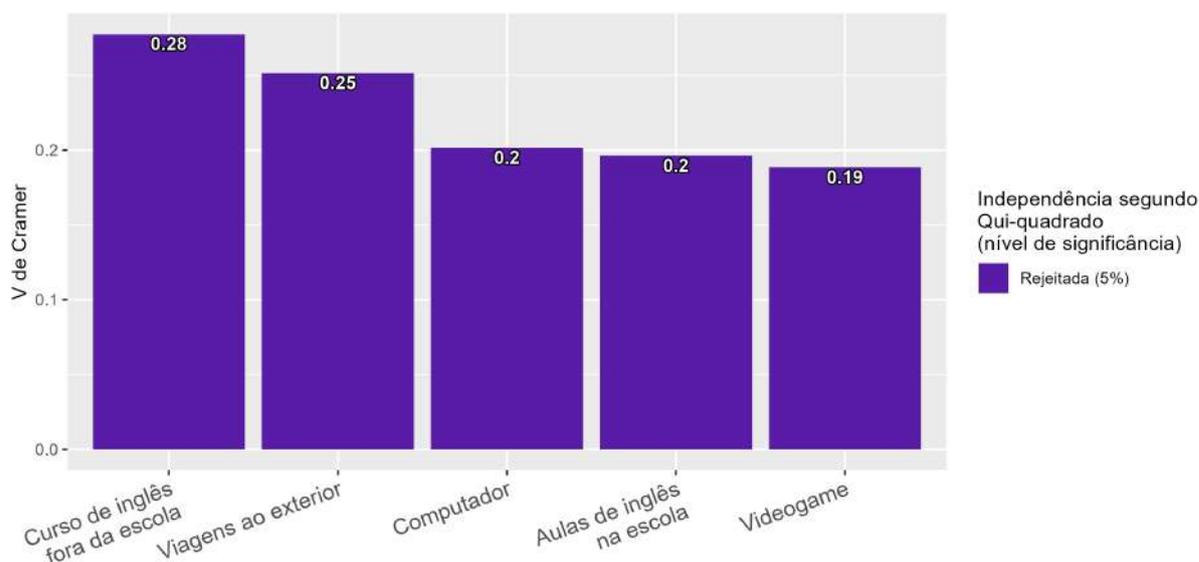


Figura 21: Associação entre formas de contato com inglês na infância e nível atual de inglês

Conforme esperado, as opções com maior associação são as mais fortemente relacionadas a fatores socioeconômicos, isto é, “acesso a curso de inglês fora da escola” e “viagens ao exterior”. Novamente, isso reforça a ideia de que, no Brasil, o acesso ao conhecimento e as condições socioeconômico de um indivíduo estão relacionados.

Todas as covariáveis tiveram independência rejeitada, e a única cuja associação é considerada fraca é “acesso a videogame”. Portanto, apesar de haver sim uma relação, a mesma não é tão forte quanto as outras. Isso em particular é de grande relevância, pois atualmente uma parcela da população defende a ideia de que videogames não precisam estar em português, pois eles podem ensinar outros idiomas.

Entretanto, a relação comparativamente fraca encontrada reforça que outros fa-

tores são muito mais influentes, de modo que é desproporcional imaginar que, sozinho, um videogame possa realizar esta tarefa, quando existem meios mais eficazes como aulas formais, contato presencial com quem fala outro idioma, e *softwares* desenvolvidos para de fato ensinar.

5.1.10 Grafo dos principais grupos de jogadores

Ao final do questionário, os respondentes eram incentivados a compartilhar até 10 de seus jogos favoritos de maneira livre e aberta. Assim, houve um total de 4781 jogos citados, considerando os 802 respondentes (alunos ou não da UnB).

Esses dados foram coletados para que fossem feitas duas análises:

- O agrupamento de jogos com base nos gostos em comum de seus jogadores;
- A visualização dos principais jogos consumidos entre pessoas com diferentes níveis de inglês.

Como o objetivo desta análise do grafo é categorizar jogadores e identificar seus jogos preferidos sem restrição a apenas alunos da UnB, todas as respostas do questionário foram utilizadas para a construção de grafos por meio do *software Gephi*, versão 0.10.

Além disso, para realizar a análise, as respostas foram agrupadas por franquia de jogos. Por exemplo, as respostas “*Kirby: Planet Robobot*” e “*Kirby Super Star Ultra*” se tornaram apenas “Kirby”, por se tratarem de jogos da franquia de mesmo nome.

No grafo da Figura 22, cada aresta representa o gosto simultâneo de um respondente por dois jogos, que são os nós nos extremos da aresta. A espessura da aresta representa o número de vezes que a relação ocorreu, de modo que quanto mais grossa a aresta, mais recorrente é o fato de um respondente gostar de duas franquias. O tamanho dos nós representa o peso de cada franquia, isto é, o número de vezes que ela foi citada no questionário. As franquias foram separadas em grupos, identificados por cores, e foram posteriormente organizadas de acordo com os grupos.

O grafo mostra apenas as arestas cujo peso é maior ou igual a 4. Ou seja, uma franquia de jogos só foi conectada a outra se pelo menos 4 respondentes citarem ambas as franquias em suas respostas. Os jogos que não possuem arestas também foram ocultados.

O *Gephi* possui a capacidade de identificar comunidades com base na modularidade utilizando um algoritmo aleatório (BLONDEL et al., 2008). Após algumas execuções desse algoritmo, foi observada a tendência de se obter 5 comunidades principais, com pequenas diferenças a cada execução. Com base nessa informação, e observando a frequência com as quais os jogos se relacionam, propusemos a distribuição com 5 comunidades ilustrada na Figura 22. Mais adiante será apresentado um teste estatístico para validar a existência dos grupos aqui propostos. Em outras palavras, os jogos foram agrupados com base na frequência com os quais os mesmos se relacionam. A posição dos nós no grafo foi então ajustada com base nesse critério.

As comunidades, identificadas na Figura 22 pelas cores dos nós, são:

- Grupo 1 (Verde): Jogos geralmente acessíveis, tanto em questão de preço quanto de requisitos técnicos. São jogos normalmente disponíveis em computadores, de simples entendimento, e com grande apelo *mainstream*, permitindo que pessoas não familiares com jogos digitais facilmente usem esses jogos como introdução. Jogos feitos por desenvolvedores independentes são comuns nesta categoria, em parte por causa de sua menor escala. Exemplos representativos incluem *Minecraft* e *Stardew Valley*. É relativamente mais comum jogos deste grupo terem o idioma português disponível.
- Grupo 2 (Rosa): Jogos que comumente possuem um nível de entrada maior, sendo menos acessíveis a iniciantes, mas que possuem um grande foco em suas comunidades, ou seja, nos jogadores em si. Isso pode ocorrer por meio de um aspecto competitivo ou cooperativo: seja pelo afeto dos jogadores aos personagens ser fortemente reforçado pelos desenvolvedores; ou por serem jogos que são bem encaixados em plataformas de *streaming* como a *Twitch*. Também é comum que jogos deste grupo estejam disponíveis em computadores, porém com maiores requisitos técnicos. *Genshin Impact* e *Valorant* exemplificam bem os aspectos citados.
- Grupo 3 (Amarelo): Jogos com grande foco na narrativa, muitas vezes desviando de aspectos tradicionais de interatividade; que podem focar mais em temas adultos que jogos de outros grupos; e que tipicamente possuem apenas um jogador. Suas naturezas não ortodoxas podem afastar até jogadores que apreciam uma diversidade

de gêneros, mas é comum que o público-alvo seja altamente dedicado. As séries *Dark Souls* e *Shin Megami Tensei* possuem aspectos citados.

- Grupo 4 (Vermelho): Jogos com grande foco na interatividade, que podem explorar com bastante profundidade até mesmo conceitos simples, e que mais comumente se afastam do foco em narrativa, se necessário. Jogos deste grupo, apesar de terem apelo a um público amplo, podem não ser tão acessíveis em questões de disponibilidade e preço, por causa das plataformas nas quais são lançados e pelo valor agregado a cada produto, pois não costumam ser produzidos por. Visuais vibrantes e trilhas sonoras cativantes também são comuns. As franquias *Super Mario* e *Pokémon* bem exemplificam estas características.
- Grupo 5 (Azul): Jogos com maior foco no espetáculo e no realismo. É comum os jogos deste grupo serem grandes demonstrações de tecnologia devido a seus visuais detalhados, de mesmo modo que momentos bombásticos e empolgantes sejam inseridos nos produtos. Essas características também são comuns em filmes *blockbusters*, mas ao contrário deles, é comum que um indivíduo não possa consumi-los por não atender a requisitos técnicos. É normal jogos deste grupo possuírem foco em narrativas, mas com apelo mais amplo que, por exemplo, o grupo amarelo. Estes detalhes podem fazer com que os jogos sejam mais curtos por questões de recursos de desenvolvimento, mas em contrapartida tenham uma maior “densidade de conteúdo” que outros grupos. As franquias *The Last of Us* e *God of War* são bons exemplos do que foi citado.

Foi utilizado o teste de Wilcoxon para avaliar a significância das 5 comunidades propostas no grafo, por meio da biblioteca *dnet* (FANG; GOUGH, 2023) da linguagem R. O teste de Wilcoxon compara a distância média entre uma franquia e as outras do mesmo grupo com a distância média entre uma franquia e as demais fora do grupo. Sua hipótese nula é a igualdade das médias das distâncias entre os dois grupos, enquanto a hipótese alternativa seria a existência de uma diferença significativa entre essas distâncias.

Após a realização dos testes, foram obtidos p-valores, conforme a Tabela 1:

	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
P-valor	0,0842	0,00773	0,0902	0,053	0,147

Tabela 1: P-valores dos testes de Wilcoxon aplicados a cada grupo

Sabendo que p-valores baixos indicam que uma comunidade é suficientemente

distinta das demais, e considerando um nível de significância de 15%, pode ser afirmado que todos os grupos são significativamente distintos. Se desconsiderado o grupo 5, os demais grupos são considerados distintos com nível de significância de 10%.

Vale ressaltar que, no questionário, a pergunta desta sessão era aberta, de modo que um respondente poderia escolher qualquer jogo, ao invés de escolher entre opções pré-selecionadas. Ainda assim, foram obtidos p-valores baixos após a padronização das respostas.

5.1.11 Grafo dos gostos por jogos de acordo com nível de inglês

Após separadas as principais comunidades, foi também analisado o nível de inglês dos respondentes em relação às franquias de sua preferência. O grafo da Figura 23 mostra essas interações com os 3 níveis de inglês estudados. Nele, cada aresta representa que um jogador que alegou gostar de um jogo possui determinado nível de inglês.

Para contornar a distribuição desigual dos níveis de inglês, respostas foram devidamente escalonadas. Cada aresta conectando um jogo ao nível básico possui um peso de 7,1; as arestas que conectam um jogo ao nível intermediário possuem peso 3,03; e as arestas ligadas ao nível avançado possuem peso 1. A espessura da aresta representa o número de vezes que a mesma relação ocorreu, de modo que os pesos de interações idênticas são somados. Foram mostradas apenas interações com peso maior que 15. Nós sem arestas foram ocultados.

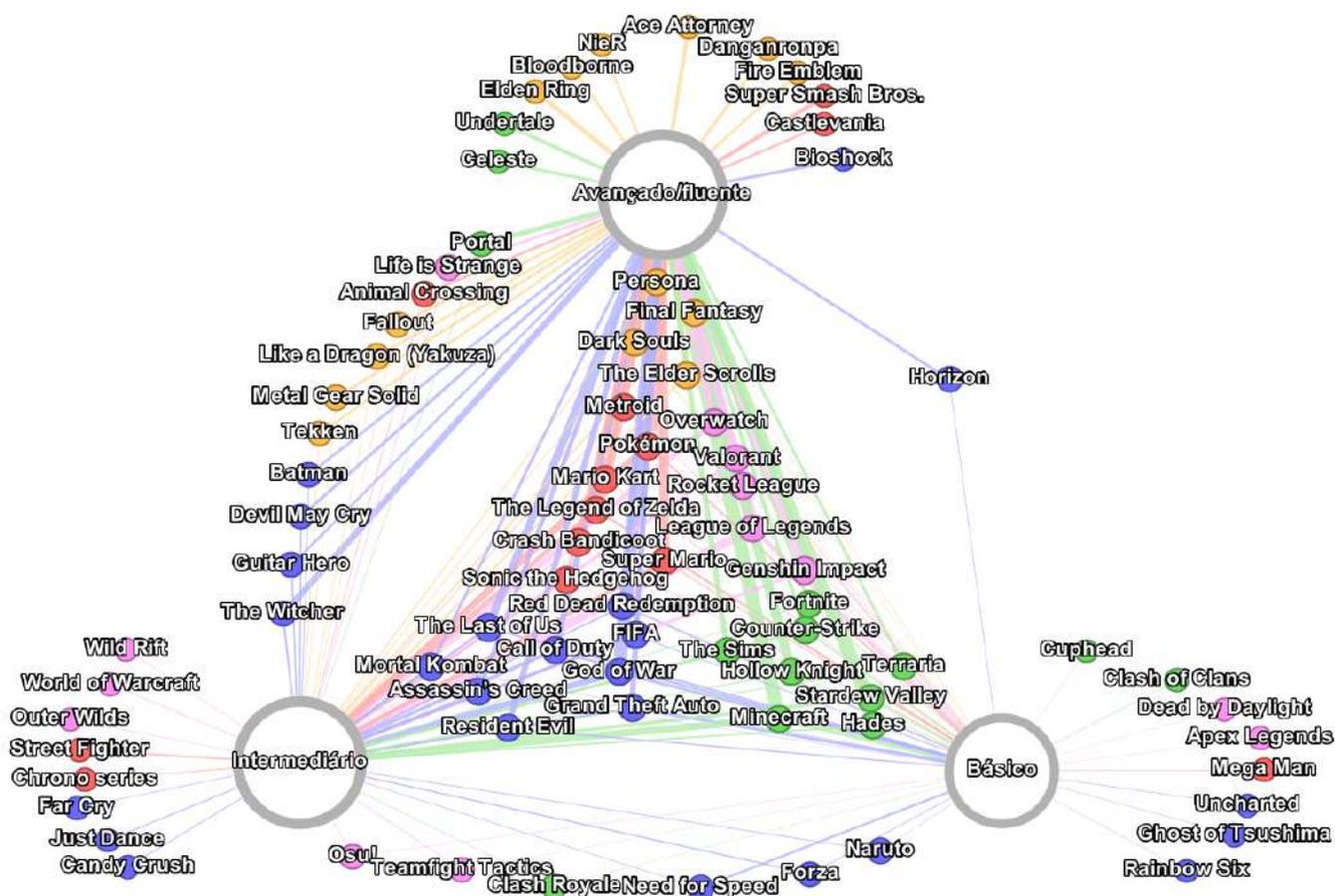


Figura 23: Principais ligações entre jogos preferidos e nível de idioma

No grafo, pode ser percebido que:

- Mais da metade dos jogos do Grupo 3 (amarelo) não estão associados ao nível de inglês básico. Todos esses jogos estão associados ao nível avançado;
- Dentre todas as interseções, a mais comum é um jogo estar na preferência de todos os níveis de inglês;
- A grande maioria dos jogos está ligada ao nível avançado de inglês;
- A maioria dos jogos ligados exclusivamente ao nível avançado de inglês não possuem o idioma português disponível;
- Há dificuldade de uma franquia de jogos ser associada aos níveis avançado e básico sem também estar associada ao nível intermediário. A presença de um nó nessa categoria pode ser considerada um *outlier*.

Considerando que, no total de respondentes não restrito a alunos da UnB, apenas 84 respondentes (10,4% do total) alegaram ter nível de inglês básico ou menor, não é possível chegar a conclusões precisas quanto a esse grupo, apesar dos pesos reescalados.

De forma exploratória, o grafo sugere que há uma relação entre o nível de inglês e os tipos de jogos consumidos numa esfera caso-a-caso.

5.2 Modelo Logístico Multinomial

5.2.1 Equação logística

Nesta etapa, foi utilizado um conjunto de dados de treino e outro de testes, separados aleatoriamente a partir do conjunto original. O conjunto de treino possui 80% dos dados originais, enquanto o de testes possui 20%.

A variável dependente do modelo é o nível de inglês dos respondentes com 3 níveis: baixo, intermediário e avançado. O nível 2 (intermediário) foi utilizado como nível de referência. Portanto, as equações geradas irão comparar os níveis 1 (nível baixo de inglês) e 3 (nível avançado de inglês) com ele.

Para otimizar o modelo, considerando a quantidade limitada de dados, as covariáveis que indicam o gosto dos respondentes por diferentes gêneros de videogame foram transformadas em binárias, unindo as respostas “Gosto moderadamente” e “Gosto muito” no valor “1” (“Gosto”), enquanto as opções “Não gosto”, “Não conheço” e “Não sei opinar” foram unidas no valor “0” (“Não gosto”).

Como o modelo inicial possui um grande número de covariáveis, há inúmeras possibilidades de combinações entre as mesmas, de modo que é difícil manualmente saber quais devem ser retiradas ou mantidas no modelo para mantê-lo mais verossímil. Por isso, ao invés de uma seleção manual de covariáveis, foi utilizada a seleção *stepwise* em sua variação bidirecional, que repetidamente compara diferentes combinações por meio do Critério de Informação de Akaike (AIC), o qual por sua vez utiliza a log-verossimilhança, mantendo ou retirando covariáveis no modelo conforme os resultados (HOSMER; LEMESHOW; STURDIVANT, 2013).

O banco de dados é relativamente pequeno, possuindo menos de 1000 observações. Sabendo-se disso, a seleção de covariáveis foi feita no conjunto de dados completo, para minimizar as perdas decorrentes da divisão em bancos de treino e de testes. Após a seleção das covariáveis, o modelo foi ajustado no banco de treinamento e sua performance

foi avaliada no banco de teste.

Após a implementação da regressão, foram mantidas no modelo apenas as covariáveis pertinentes as seguintes informações dos respondentes:

- Número de banheiros na casa;
- Jogar ou não videogames em console;
- Conteúdo consumido em inglês (incluindo todas as opções, exceto “Conteúdo em mídias sociais”);
- Gosto por diferentes gêneros de jogos digitais (incluindo apenas os gêneros “Visual Novel”, “Plataforma”, “Puzzle”, “RPG” e “MOBA”);

As informações acima totalizam 14 covariáveis. Para cada uma delas, foram calculados $\hat{\beta}^1$ e $\hat{\beta}^3$, referentes ao modelo logístico multinomial para os níveis 1 e 3, bem como os seus respectivos erros padrão, conforme a Tabela 2.

Tabela 2: Parâmetros estimados pela regressão logística multinomial para o nível de inglês 1 (β^1) e 3 (β^3), e seus respectivos erros padrão

	$\hat{\beta}^1$	$\hat{\beta}^3$	$EP(\hat{\beta}^1)$	$EP(\hat{\beta}^3)$
x_0 (Intercepto)	2,192	-1,575	0,921	0,676
x_1 (Número de banheiros)	-0,140	0,251	0,198	0,120
x_2 (Jogar em console)	-0,026	0,548	0,412	0,266
x_3 (Assistir a filmes/séries em inglês com legenda em português)	-0,628	-0,947	0,495	0,351
x_4 (Assistir a filmes/séries em inglês com legenda em inglês ou sem legenda)	-1,176	1,387	0,452	0,283
x_5 (Consumir notícias em inglês)	-1,151	0,066	0,411	0,294
x_6 (Consumir literatura não-didática em inglês)	-0,367	0,638	0,443	0,272
x_7 (Consumir artigos em inglês)	-0,236	0,351	0,419	0,266
x_8 (Assistir a vídeos em plataformas de streaming como Twitch e Youtube em inglês)	-0,033	0,463	0,440	0,382
x_9 (Assistir a TikToks em inglês)	-1,485	0,241	0,468	0,268
x_{10} (Gostar do gênero visual novel)	-0,776	0,134	0,440	0,266
x_{11} (Gostar do gênero plataforma)	-0,655	-0,24	0,523	0,414
x_{12} (Gostar do gênero puzzle)	0,242	-0,560	0,426	0,297
x_{13} (Gostar do gênero RPG)	-0,174	0,688	0,444	0,317
x_{14} (Gostar do gênero MOBA)	0,189	-0,590	0,427	0,271

Portanto, se aplicados na equação 3.2.2, esses valores podem ser utilizados para calcular a probabilidade de um indivíduo ter nível de inglês baixo ou avançado.

Na probabilidade estimada $\hat{\pi}_1(\mathbf{x})$ de um indivíduo ter nível baixo de inglês, temos como $g_1(\mathbf{x})$:

$$g_1(\mathbf{x}) = 2.192x_0 - 0.140x_1 - 0.026x_2 - 0.628x_3 + \dots + 0.189x_{14}, \quad (5.2.1)$$

enquanto na probabilidade estimada $\hat{\pi}_3(\mathbf{x})$ de um indivíduo ter nível avançado de inglês, temos como $g_3(\mathbf{x})$:

$$g_3(\mathbf{x}) = -1.575x_0 + 0.251x_1 + 0.548x_2 - 0.947x_3 + \dots - 0.590x_{14}. \quad (5.2.2)$$

Em ambos os casos, a probabilidade complementar (ou seja, $1 - \pi_k(x)$) é a probabilidade estimada do indivíduo ter nível intermediário de inglês.

5.2.2 Chance de sucesso

Com os parâmetros estimados, é possível estimar também razões de chance de sucesso (*odds ratio*) para cada covariável incluída no modelo. Para auxiliar a interpretação, utilizaremos as seguintes notações:

$$\theta_{21} = \frac{\frac{\pi_2(\mathbf{x})}{1-\pi_2(\mathbf{x})}}{\frac{\pi_1(\mathbf{x})}{1-\pi_1(\mathbf{x})}} = \frac{1}{\exp(\beta^1)}. \quad (5.2.3)$$

$$\theta_{32} = \frac{\frac{\pi_3(\mathbf{x})}{1-\pi_3(\mathbf{x})}}{\frac{\pi_2(\mathbf{x})}{1-\pi_2(\mathbf{x})}} = \exp(\beta^3), \quad (5.2.4)$$

onde foi utilizado θ_{21} ao invés de θ_{12} para que se mantenha um padrão crescente na relação entre níveis.

Na Tabela 3, serão apresentadas estimativas para as razões de chance θ_{21} e θ_{32} .

Tabela 3: Razões de chance estimadas pelo modelo logístico

	$\hat{\theta}_{21}$	$\hat{\theta}_{32}$
x_1 (Número de banheiros em casa)	1,151	1,286
x_2 (Jogar em console)	1,026	1,730
x_3 (Assistir a filmes/séries em inglês com legenda em português)	1,873	0,388
x_4 (Assistir a filmes/séries em inglês com legenda em inglês ou sem legenda)	3,242	4,001
x_5 (Consumir notícias em inglês)	3,162	1,068
x_6 (Consumir literatura não-didática em inglês)	1,444	1,893
x_7 (Consumir artigos em inglês)	1,266	1,421
x_8 (Assistir a vídeos em plataformas de streaming como Twitch e Youtube em inglês)	1,034	1,589
x_9 (Assistir a TikToks em inglês)	4,416	1,273
x_{10} (Gostar do gênero Visual Novel)	2,172	1,144
x_{11} (Gostar do gênero Plataforma)	1,926	1,276
x_{12} (Gostar do gênero Puzzle)	0,785	0,571
x_{13} (Gostar do gênero RPG)	1,190	1,990
x_{14} (Gostar do gênero MOBA)	0,827	0,554

Em resumo, θ_{21} indica o impacto de determinada covariável na razão de chances entre o nível de inglês intermediário e o nível de inglês básico. Valores maiores que 1 indicam que a chance do indivíduo ter inglês intermediário é maior que a chance de se apresentar inglês básico na presença desta covariável (no caso de covariáveis binárias).

De forma equivalente, θ_{32} indica o impacto de determinada covariável na razão de chances entre o nível de inglês avançado e o nível de inglês intermediário. Valores maiores que 1 indicam que a chance do indivíduo ter inglês intermediário é maior do que a do indivíduo apresentar inglês básico quando esta covariável está presente (no caso de covariáveis binárias).

Com essas informações, é possível chegar às seguintes conclusões:

- Covariáveis relativas a fatores socioeconômicos (número de banheiros em casa, jogar videogames em um console) aumentam a chance da pessoa apresentar nível de inglês mais elevado. Por exemplo, jogar em console aumenta em 73% a chance do nível de inglês ser avançado em comparação com o nível de inglês intermediário.

- Quase todas as covariáveis relativas ao conteúdo consumido em inglês possuem razão de chances acima de 1, ou seja, estão mais presentes conforme se avança o nível de inglês. Isso é esperado, pois naturalmente o consumo de conteúdo em uma língua é reflexo do seu entendimento da mesma.
- A exceção para a informação acima é a covariável “assistir conteúdo legendado em português”, pois sua razão estimada $\hat{\theta}_{32}$ está abaixo de 1. Isso ainda faz bastante sentido, pois mesmo que seja normal uma pessoa assistir conteúdo legendado se tiver algum conhecimento de inglês (conforme $\hat{\theta}_{21}$), pode haver um afastamento deste conteúdo na medida em que se conhece mais da língua inglesa, favorecendo mídias em seus idiomas originais.
- Complementando o item acima, os θ estimados da covariável “assistir conteúdo com legenda em inglês ou sem legenda” são particularmente altos. Isto é esperado, pois assistir conteúdo audiovisual em outra língua requer mais habilidade no idioma, de modo que o espectador deve estar apto a compreender diálogo em tempo real, seja por meio da compreensão da fala, da leitura, ou por ambos simultaneamente.
- As covariáveis dos gostos pelos gêneros “Visual Novel”, “Plataforma” e “RPG” de videogame possuem θ acima de 1. No caso de “RPG” e “Visual Novel”, estes números correspondem à habilidade necessária do jogador em entender inglês, pois leitura é uma parte essencial de jogos deste tipo. Considerando que jogos de investigação também estão inclusos na categoria “Visual Novel” desta análise, também é necessário raciocínio lógico em língua estrangeira.
- No caso do gênero “Plataforma”, mesmo que jogos deste tipo não precisem tanto do inglês, os θ acima de 1 podem ser justificadas. Os jogos mais populares do gênero (como as franquias *Super Mario* e *Crash Bandicoot*) costumam ser mais inacessíveis, seja por estarem exclusivamente em consoles, ou por terem preços mais elevados. Isto ocorre em parte por serem jogos de empresas tradicionais e “*high budget*”. Ou seja, há um fator socioeconômico envolvido.
- Os gêneros “Puzzle” e “MOBA” aparentam possuir chances maiores em níveis menores de inglês. De fato, são jogos que não requerem tanta leitura e estão altamente disponíveis em português, respectivamente. Isso não necessariamente indica que jogadores de inglês mais avançado não gostem dos gêneros, podendo também indicar que, conforme se conhece mais da língua, maior contato se tem com outros gêneros, fazendo o jogador perceber qual realmente é seu gosto.

Utilizando os erros padrão estimados de cada β , é possível calcular intervalos de confiança de θ . Podemos afirmar com 95% de confiança que intervalos que contêm apenas valores acima de 1 representam as covariáveis mais positivamente impactantes de cada comparação, pois estas nunca apresentarão queda.

No caso de $\hat{\theta}_{21}$, esses intervalos seriam os das covariáveis “assistir a filmes/séries com legenda em inglês ou sem legenda” (1,336; 7,869), “consumir notícias em inglês” (1,312; 7,080), e “assistir a TikToks em inglês” (1,764; 11,050).

Já em $\hat{\theta}_{21}$, os intervalos são os de “número de banheiros em casa” (1,016; 1,627), “jogar em console” (1,027; 2,916), “assistir a filmes/séries com legenda em inglês ou sem legenda” (2,295; 6,974), e “gostar do gênero RPG” (1,070; 3,701).

5.2.3 Diagnóstico do modelo

Foi utilizado um gráfico do tipo HNP (*Half Normal Plot*), também conhecido como “gráfico quantil-quantil com envelope simulado”, para comparar os quantis dos resíduos do modelo ajustado com os quantis do modelo teórico correspondente, que podem ser obtidos via simulação da variável dependente (MORAL; HINDE; DEMÉTRIO, 2017).

No gráfico, é apresentado também o intervalo de confiança para os resíduos. Entretanto, nenhum deles está fora deste intervalo com nível de confiança de 95%, conforme a Figura 24.

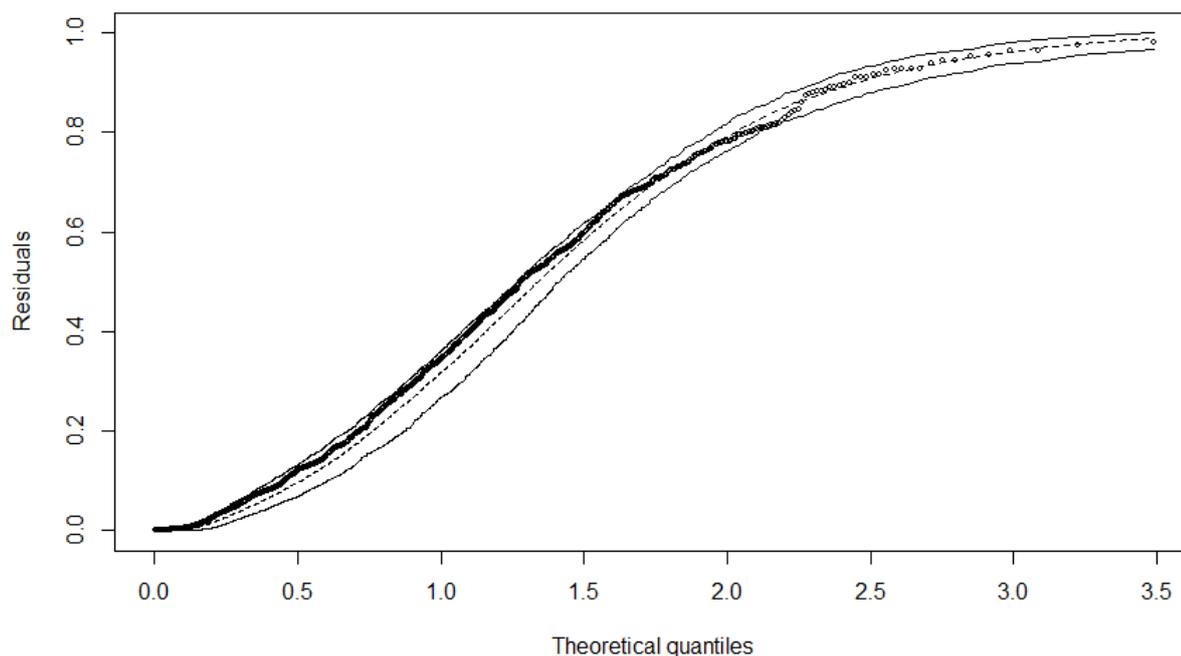


Figura 24: Gráfico quantil-quantil com envelope simulado

Portanto, como os dois conjuntos de resíduos são suficientemente similares, o modelo logístico encontrado neste trabalho mostra-se adequado.

Um diagnóstico completo dos resíduos não foi realizado devido às limitações de disponibilidade dos softwares utilizados, considerando que a regressão logística binária é mais amplamente utilizada, e portanto, possui mais recursos implementados que aqueles voltados para o caso multinomial.

Para avaliar a qualidade do ajuste do modelo, foi utilizado também o teste Qui-quadrado de Pearson para comparar os valores da variável dependente com os valores preditos no banco de dados de treino. Neste teste, a hipótese nula é a independência dos dois conjuntos de dados comparados (GOYAL, 2020). No caso em questão, isso significaria que os valores preditos são significativamente diferentes dos valores reais. Como o p-valor encontrado foi inferior à $2,2 * 10^{16}$, que por sua vez é inferior à uma significância de 0,01, pode ser afirmado que os conjuntos são suficientemente similares a uma significância próxima de 0%, indicando que o modelo logístico prevê os valores de modo verossímil.

Ao utilizar o modelo para prever o nível de inglês dos 110 indivíduos do conjunto de dados de testes, 86,4% das previsões foram corretas, conforme a matriz de confusão da Tabela 4.

		Nível previsto		
		1	2	3
Nível real	1	9	0	0
	2	10	11	5
	3	0	0	75

Tabela 4: Matriz de confusão utilizando banco de testes

Além disso, observa-se na Tabela 4 que todos os respondentes com nível de inglês básico ou avançado tiveram seus níveis previstos com sucesso, enquanto todas as previsões erradas ocorreram em indivíduos com nível intermediário de inglês. Isto pode ter ocorrido pelo fato da definição de “intermediário” ser bastante subjetiva, causando flutuação no nível de inglês daqueles que alegaram estar neste grupo. Possíveis soluções para este problema seriam a redução da subjetividade no questionário ou uso de regressão logística binária, ambas as quais podem ser investigadas em trabalhos futuros.

6 Conclusão

O principal objetivo deste trabalho foi a associação do nível de inglês de estudantes da UnB que jogam jogos digitais ao conteúdo midiático consumido, visto que videogames são mídias variadas com usos diversos da linguagem verbal e pouca disponibilidade em português, de modo que o gosto de um respondente por jogos é útil para a pesquisa. Foram coletados dados a partir de um questionário, o que possibilitou a realização de análises exploratórias, a modelagem de uma regressão logística multinomial e o estudo de correlações.

Um efeito colateral imediato da amostragem foi a percepção de que, dentre os estudantes que jogam videogame, saber o inglês é bem mais comum do que não saber, o que não condiz com a realidade brasileira. Portanto, no início da pesquisa, já surgiram indícios de que existe de fato uma associação. Mais adiante, foi observado que o inglês é a língua dominante na qual os respondentes consomem diferentes gêneros de jogo.

Foi utilizada uma amostragem por conveniência, incluindo observações sem probabilidade especificada ou conhecida na amostra, priorizando a disponibilidade dos respondentes. Apesar disso, houve uma preocupação em atingir uma variedade de perfis de estudantes, e não houve vieses aparentes nos resultados obtidos. Entretanto, para estudo mais aprofundado, a amostragem probabilística poderá ser utilizada em trabalhos futuros. Outro método que pode ser utilizado para maior desenvolvimento do tema é a regressão logística ordinal, considerando que o nível de inglês, apesar de qualitativo, ainda segue uma ordem.

O modelo logístico multinomial apontou os principais fatores relacionados ao nível de inglês dos respondentes, incluindo gostar de gêneros que dependem muito ou muito pouco do idioma, consumir diversas outras mídias em inglês, e fatores socioeconômicos. Foi concluído que a associação do inglês com o consumo de mídias existe, mas não é um fenômeno exclusivo de videogames, apenas refletindo outros fatores. No modelo, o consumo de outras mídias em inglês como livros e artigos apresentaram, num geral, *odds ratios* próximos, se não maiores, ao consumo de jogos digitais.

Os dados indicam que o consumo de videogames, de modo independente, não costuma ser a principal fonte de conhecimento do inglês. Não só as horas diárias jogadas aparentam ter baixa influência no nível do idioma, como também o consumo de videogames na infância apresentou a menor correlação com o nível atual de inglês se comparado a outros meios de contato com a língua, como aulas de inglês fora da escola e viagens

ao exterior, ambos os quais tiveram as maiores correlações na análise, e ambos os quais refletem fatores socioeconômicos. Isso não descarta a efetividade do uso do meio para o aprendizado, desde que haja o devido acompanhamento, como demonstrado em demais trabalhos.

O trabalho também ressalta a importância da acessibilidade, visto que o lançamento de versões traduzidas de mídias gera alto interesse em indivíduos com níveis mais baixos de inglês, e que a maior parte desse grupo já se afastou de alguma mídia justamente por não saber inglês. Também foi constatado que, no meio dos videogames, é mais comum conhecer um jogo em inglês e passar a se jogar em português do que o oposto. Conforme reforçado pelos grafos e pelas análises dos gêneros de videogame, há certos grupos de jogos que dependem fortemente do inglês.

Referências

AGRESTI, A. *An introduction to categorical data analysis*. 2. ed. [S.l.]: Wiley-Blackwell, 2007.

ANDRADE, M. *Brasil figura no Grupo de países com Baixo Domínio da Língua Inglesa*. Correio Braziliense, 2022. Acesso em: 20 de dez. 2022. Disponível em: <<https://www.correiobraziliense.com.br/euestudante/educacao-basica/2022/04/5003937-brasil-figura-no-grupo-de-paises-com-baixo-dominio-da-lingua-inglesa.html>>.

ARRAZ, L. *Hades Tem peculiares Erros de Tradução Na Capa Brasileira*. 2021. Acesso em: 03 de fev. 2023. Disponível em: <<https://canaltech.com.br/games/hades-tem-peculiares-erros-de-traducao-na-capa-brasileira-196689/>>.

BLONDEL, V. et al. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics Theory and Experiment*, v. 2008, 04 2008.

BOTSCH, R. *Significance and Measures of Association*. 2011. Acesso em: 29 de jun. 2023. Disponível em: <<https://polisci.usca.edu/apls301/Text/Chapter%2012.%20Significance%20and%20Measures%20of%20Association.htm>>.

BRAZIL Alliance, p. *Brazil Alliance*. 2023. Acesso em: 04 de jun. 2023. Disponível em: <<https://brazilalliance.com.br>>.

BROWN, M. *Game Maker's Toolkit*. YouTube, 2006. Acesso em: 01 de jul. 2023. Disponível em: <<https://www.youtube.com/@GMTK/about>>.

BURNETT, E.; WEGMANN, M. E. *Sexism in video games*. Delta State University, 2020. Acesso em: 06 de mar. 2023. Disponível em: <<https://deltastatement.com/6769/news/sexism-in-video-games/>>.

BUSSAB, W. d. O.; MORETTIN, P. A. *Estatística Básica*. 9. ed. [S.l.]: Saraiva, 2012. 370 p.

CHANG, G.-H. Youngstown State University, 2023. Acesso em: 11 de fev. 2023. Disponível em: <http://gchang.people.yzu.edu/class/mps/note/08_11_LogisticRegression.pdf>.

CZEPIEL, S. A. Maximum likelihood estimation of logistic regression models: Theory and implementation. 2022. Disponível em: <<https://czep.net/stat/mlelr.pdf>>.

ELITE Beat Agents (DS). Nintendo Life, 2007. Acesso em: 09 de jul. 2023. Disponível em: <https://www.nintendolife.com/games/ds/elite_beat_agents>.

FANG, H.; GOUGH, J. *dCommSignif*. 2023. Acesso em: 24 de jun. 2023. Disponível em: <<https://rdr.io/cran/dnet/man/dCommSignif.html>>.

GACH, E. *League of Legends Pro Suspended During Worlds For Toxic Chat, Will Be Fined \$2,000*. 2016. Acesso em: 04 de jun. 2023. Disponível em: <<https://kotaku.com/league-of-legends-pro-suspended-during-worlds-for-toxic-1787589483>>.

GOOGLE, p. *Pokémon UNITE*. 2023. Acesso em: 04 de jun. 2023. Disponível em: <<https://play.google.com/store/apps/details?id=jp.pokemon.pokemonunite&hl=pt>>.

- GOYAL, P. *Chi-Square-Goodness-Fit*. 2020. Acesso em: 28 de jun. 2023. Disponível em: <<https://rpubs.com/pg2000in/ChiSquareGoodnessFit>>.
- GÓES, O. H. C. K. D. G. Tradução: Adaptação além da substituição de palavras. *Universidade Federal do Paraná.*, 2015.
- HOSMER, D. W.; LEMESHOW, S.; STURDIVANT, R. X. *Applied Logistic Regression*. 3. ed. [S.l.]: Wiley, 2013.
- HUA, C.; CHOI, Y.-J.; SHI, Q. *Companion to BER 642: Advanced regression methods*. 2021. Capítulo 11 - Multinomial Logistic Regression, Acesso em: 30/01/2023. Disponível em: <https://bookdown.org/chua/ber642_advanced_regression/multinomial-logistic-regression.html#introduction-to-multinomial-logistic-regression>.
- JACUTEM Sabão, p. *Jacutem Sabão*. 2022. Acesso em: 20 de dez. 2022. Disponível em: <<https://jacutemsabao.bitbucket.io/>>.
- KAIN, E. 'Call Of Duty: Modern Warfare II' Passes \$1 Billion In Sales In Less Than Two Weeks. 2022. Acesso em: 04 de jun. 2023. Disponível em: <<https://www.forbes.com/sites/erikkain/2022/11/07/call-of-duty-modern-warfare-ii-passes-1-billion-in-revenue-in-less-than-two-weeks/?sh=2d85567c4614>>.
- LAURA [gasseruto]. *Fãs fazem Campanha Por Localização em Português de 'Pokémon Scarlet & Violet'*. 2022. Disponível em: <<https://www.jbox.com.br/2022/03/02/fas-fazem-campanha-por-localizacao-em-portugues-de-pokemon-scarlet-violet/>>.
- LUND, A.; LUND, M. *Multinomial logistic regression using SPSS statistics*. Laerd Statistics, 2018. Acesso em: 30 de jan. 2023. Disponível em: <<https://statistics.laerd.com/spss-tutorials/multinomial-logistic-regression-using-spss-statistics.php>>.
- MERRIFIELD, L. *What we learned from 350 million kids | Lane Merrifield | tedxkelowna*. YouTube, 2017. Acesso em: 03 de fev. 2023. Disponível em: <<https://www.youtube.com/watch?v=UtGC-kcrq2s>>.
- MONOLIX. *Standard error using the Fisher information matrix*. Monolix, 2022. Acesso em: 03 de fev. 2023. Disponível em: <<https://monolix.lixoft.com/tasks/standard-error-using-the-fisher-information-matrix/>>.
- MORAL, R. A.; HINDE, J.; DEMÉTRIO, C. G. B. Half-normal plots and overdispersed models in r: the hnp package. *Journal of Statistical Software*, 2017.
- NGUYEN, M. *A guide on Data Analysis*. 2022. Acesso em: 20 de dez. 2022. Disponível em: <https://bookdown.org/mike/data_analysis/the-likelihood-ratio-test.html>.
- NGUYEN, M. *A guide on Data Analysis*. 2022. Acesso em: 20 de dez. 2022. Disponível em: <https://bookdown.org/mike/data_analysis/wald-test.html>.
- NINTENDO. *Fire Emblem Three Houses (Comercial Dublado PT-BR)*. YouTube, 2021. Acesso em: 20 de dez. 2022. Disponível em: <<https://www.youtube.com/watch?v=toIvtpkgn5g>>.

- NINTENDO. *Surpresas e encantos te aguardam em cada canto na próxima evolução da diversão em um jogo de plataforma 2D do Mario! SuperMarioBrosWonder será lançado para o console NintendoSwitch no dia 20 de outubro. Disponível em Português do Brasil.* 2023. Twitter: @NintendoBrasil. Acesso em: 01 de jul. 2023. Disponível em: <<https://twitter.com/NintendoBrasil/status/1671539229401272320>>.
- PEREIRA, G. S. *Luxia Subs.* 2023. Acesso em: 01 de jul. 2023. Disponível em: <<https://luxialegendas.blogspot.com>>.
- SIIVAGUNNER. *SiIvaGunner.* YouTube, 2016. Acesso em: 01 de jul. 2023. Disponível em: <<https://www.youtube.com/watch?v=JqwcUhuz9a0>>.
- SILVA, F. W. d. C.; TOASSI, P. F. P. O papel dos jogos eletrônicos na aquisição da Língua inglesa. *Revista do GEL*, v. 17, n. 1, p. 284–307, jul. 2020. Disponível em: <<https://revistadogel.emnuvens.com.br/rg/article/view/2757>>.
- SMOLINSKI, C. Aprendendo inglês com pinguins: O desenvolvimento linguístico através do jogo eletrônico club penguin. *Revista Prâksis*, v. 2, p. 73–82, nov. 2016. Disponível em: <<https://periodicos.feevale.br/seer/index.php/revistapraksis/article/view/762>>.
- TABOGA, M. *Information matrix.* Kindle Direct Publishing, 2021. Acesso em: 03 de fev. 2023. Disponível em: <<https://www.statlect.com/glossary/information-matrix>>.
- TAKAHASHI, S.; INOUE, I. *Guia Mangá Análise de Regressão.* 1. ed. [S.l.]: Novatec Editora, 2019. 163 p. (Guia Mangá).
- WANDERLEY, G. A. *A relação entre consumo midiático e o conhecimento do Inglês.* 2022. Acesso em: 20 de dez. 2022. Disponível em: <<https://forms.gle/9B3V3Ebb5Vp78iBb6>>.

Apêndice A - Questionário que gerou os dados do trabalho

Section 1 of 8

A relação entre consumo midiático e o conhecimento do inglês

Neste questionário de 16 perguntas, queremos saber mais sobre o seu nível de inglês e o tipo de jogos digitais que você consome, a fim de associar ambas as informações!

Suas respostas serão anônimas.

1 - Qual é a sua idade? *

Short answer text

2 - Qual é o seu gênero? *

Masculino

Feminino

Outro/Prefiro não dizer

3 - Quantos banheiros há em sua casa?

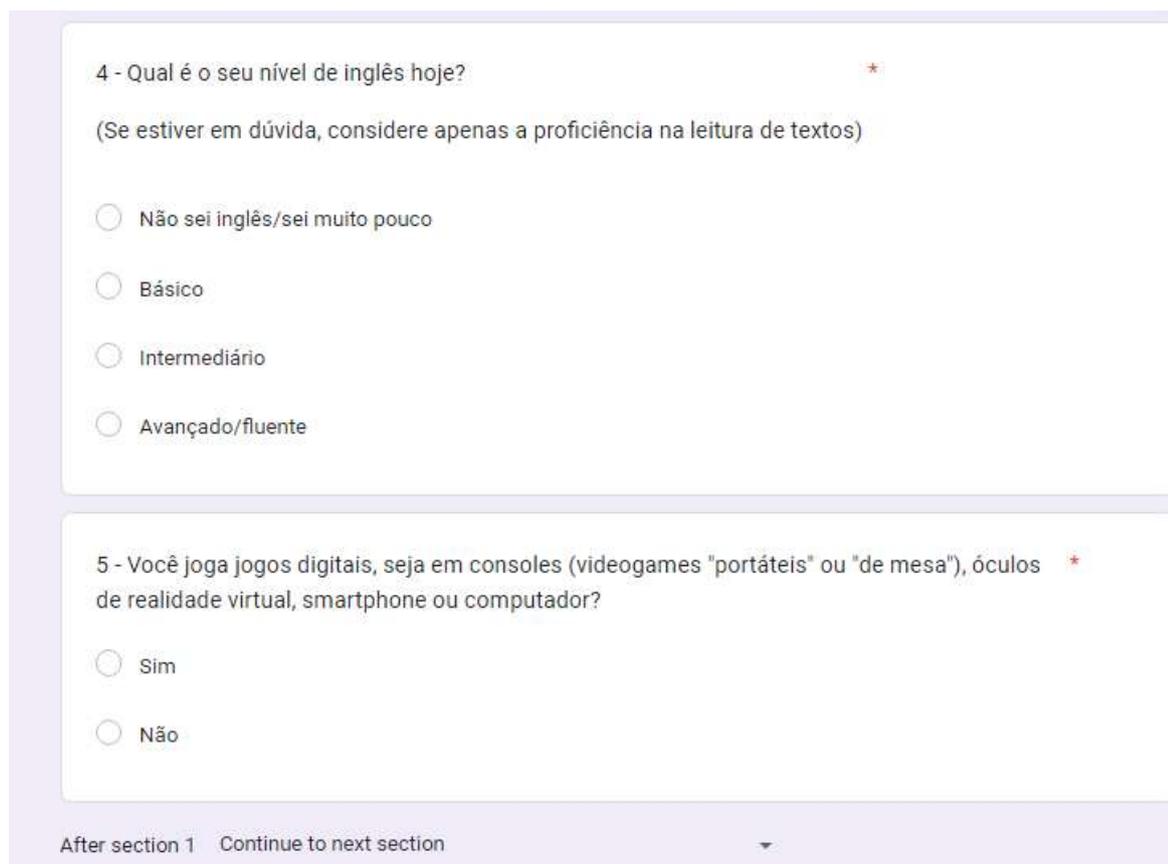
1

2

3

4 ou mais

Figura 25: Questionário (Parte 1)



4 - Qual é o seu nível de inglês hoje? *

(Se estiver em dúvida, considere apenas a proficiência na leitura de textos)

Não sei inglês/sei muito pouco

Básico

Intermediário

Avançado/fluyente

5 - Você joga jogos digitais, seja em consoles (videogames "portáteis" ou "de mesa"), óculos de realidade virtual, smartphone ou computador? *

Sim

Não

After section 1 Continue to next section ▼

Figura 26: Questionário (Parte 2)

Section 2 of 8

Seu curso

Description (optional)

6 - Qual é o seu curso na UnB? *

(Se não for aluno, selecione a primeira opção, "Não estudo na UnB")

1. Não estudo na UnB
2. Administração
3. Agronomia
4. Arquitetura e Urbanismo
5. Arquivologia
6. Artes Cênicas

Figura 27: Questionário (Parte 3). A pergunta possui mais respostas, mas a imagem foi cortada considerando o número alto de cursos a se escolher.

After section 2 Go to section 3 (Seus modos de jogar hoje) ▾

Section 3 of 8

Seus modos de jogar hoje ✕ ⋮

Description (optional)

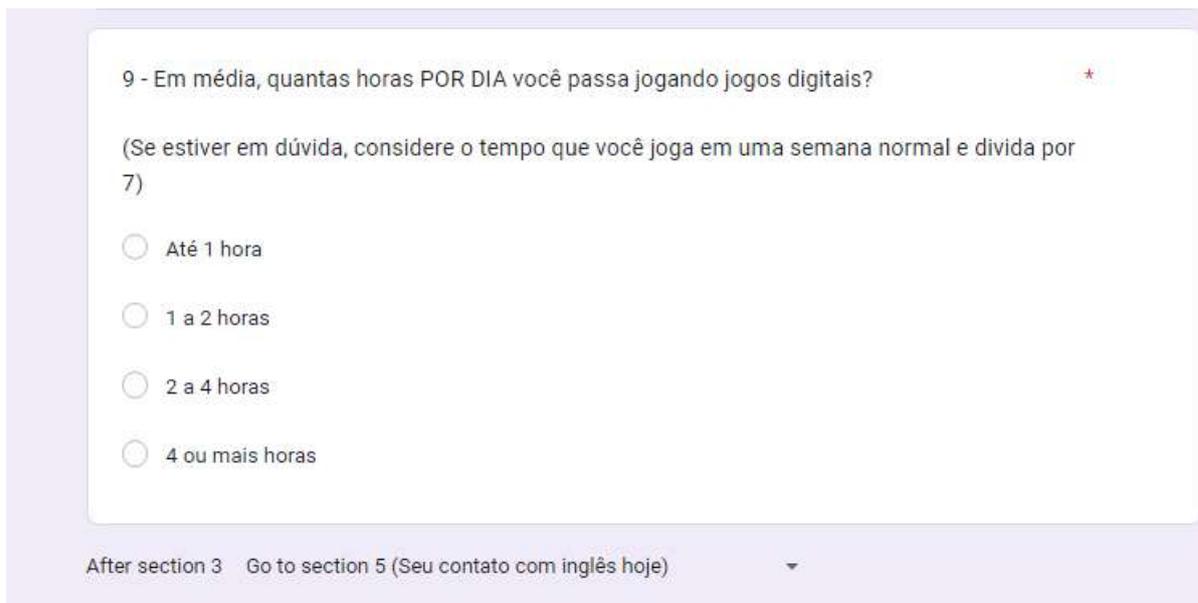
7 - Como você julga o seu envolvimento com jogos digitais no último ano?

- Eu joguei pouco, apenas ocasionalmente de modo casual.
- Eu joguei bem menos do que gostaria, por questões externas (financeiras, tecnológicas, tempo, etc.)
- Eu joguei moderadamente por escolha própria, e estou satisfeito com isso.
- Eu joguei bastante, e me dedico muito a esse hobby.

8 - Quais plataformas você usa para jogar jogos digitais hoje? Marque todas as opções que se aplicam.

- Smartphone
- Computador
- Console
- Óculos de Realidade Virtual

Figura 28: Questionário (Parte 4)



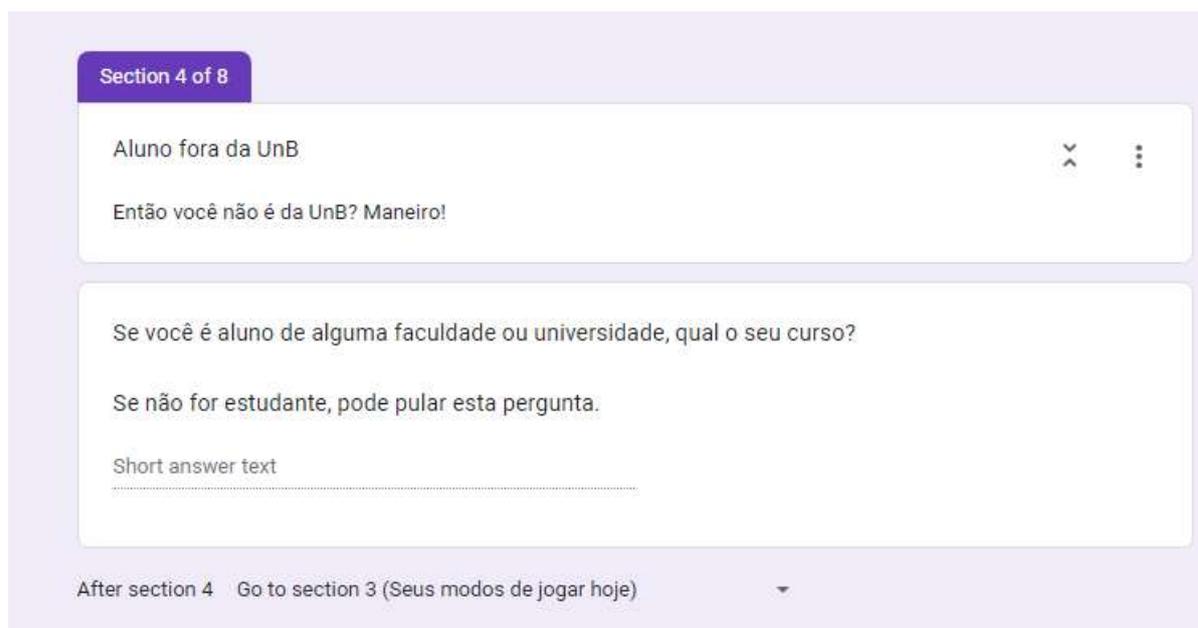
9 - Em média, quantas horas POR DIA você passa jogando jogos digitais? *

(Se estiver em dúvida, considere o tempo que você joga em uma semana normal e divida por 7)

- Até 1 hora
- 1 a 2 horas
- 2 a 4 horas
- 4 ou mais horas

After section 3 Go to section 5 (Seu contato com inglês hoje) ▾

Figura 29: Questionário (Parte 5)



Section 4 of 8

Aluno fora da UnB

Então você não é da UnB? Maneiro!

Se você é aluno de alguma faculdade ou universidade, qual o seu curso?

Se não for estudante, pode pular esta pergunta.

Short answer text

After section 4 Go to section 3 (Seus modos de jogar hoje) ▾

Figura 30: Questionário (Parte 6)

Section 5 of 8

Seu contato com inglês hoje

Description (optional)

10 - Fora jogos digitais, que tipo de conteúdo em inglês você consome (e compreende)?
Marque todas as opções que se aplicam.

- Filmes/séries em inglês com legendas em português
- Filmes/séries em inglês com legendas em inglês ou sem legenda
- Notícias em inglês
- Livros não-didáticos/fanfics/mangás em inglês
- Livros didáticos/Artigos Acadêmicos em inglês
- Conteúdo em inglês em sites de streaming como Twitch e Youtube
- Conteúdo em inglês de mídias sociais como Twitter e Instagram
- TikToks em inglês
- Other...

After section 5 Continue to next section

Figura 31: Questionário (Parte 7)

Section 6 of 8

Os jogos que você joga ✕ ⋮

Description (optional)

⋮
🖨

🎯 Multiple choice grid ▾

11 - Qual é a sua opinião sobre os seguintes gêneros de jogos digitais?

Rows		Columns	
1. Jogos de tiro no estilo FPS (Call of Duty, ...	✕	<input type="radio"/> Não conheço	✕
2. Jogos de luta (Mortal Kombat, Super Sm...	✕	<input type="radio"/> Conheço, mas não sei opinar	✕
3. Visual Novels/Investigação (Ace Attorne...	✕	<input type="radio"/> Não gosto	✕
4. Jogos de plataforma (Super Mario, Sonic...	✕	<input type="radio"/> Gosto moderadamente	✕
5. Jogos de ritmo (OSU!, Guitar Hero, Just ...	✕	<input type="radio"/> Gosto muito	✕
6. Jogos de corrida/esporte (FIFA, Forza, M...	✕	<input type="radio"/> Add column	
7. Jogos de puzzle (Tetris, Candy Crush, Po...	✕		
8. Jogos no estilo RPG/estratégia (Persona...	✕		
9. Jogos de ação (Devil May Cry, God of Wa...	✕		
10. Jogos de simulação (Stardew Valley, An...	✕		
11. MOBAs (League of Legends, DOTA 2, P...	✕		
12. Jogos com foco em exploração/aventu...	✕		
13. Add row			

Figura 32: Questionário (Parte 8)

12 - Quando você joga/jogou os jogos dos gêneros abaixo, qual é/era o idioma?

Multiple choice grid

Rows	Columns
1. Jogos de tiro no estilo FPS (Call of Duty, ...)	<input type="radio"/> Nunca joguei/não conheço
2. Jogos de luta (Mortal Kombat, Super Sm...)	<input type="radio"/> Joguei apenas em inglês
3. Visual Novels/Investigação (Ace Attorne...)	<input type="radio"/> Joguei apenas em português
4. Jogos de plataforma (Super Mario, Sonic...)	<input type="radio"/> Conheci em inglês, hoje jogo em portug...
5. Jogos de ritmo (OSU!, Guitar Hero, Just ...)	<input type="radio"/> Conheci em português, hoje jogo em in...
6. Jogos de corrida/esporte (FIFA, Forza, M...)	<input type="radio"/> Não tenho preferência por idioma, inde...
7. Jogos de puzzle (Tetris, Candy Crush, Po...)	<input type="radio"/> Add column
8. Jogos no estilo RPG/estratégia (Persona...)	
9. Jogos de ação (Devil May Cry, God of Wa...)	
10. Jogos de simulação (Stardew Valley, An...)	
11. MOBAs (League of Legends, DOTA 2, P...)	
12. Jogos com foco em exploração/aventu...	
13. Add row	

Require a response in each row

Figura 33: Questionário (Parte 9)

Section 7 of 8

Seu inglês no passado

Description (optional)

13 - Nos últimos anos, você já se afastou de algum tipo de conteúdo por não ter o domínio da *
língua inglesa?

Sim

Não

14 - Você já se interessou em um jogo apenas por ter sido lançada uma versão em *
português? (seja ela oficial ou não)

Sim

Não

Figura 34: Questionário (Parte 10)

15 - Classifique o seu contato com a língua inglesa durante a infância

Multiple choice grid

Rows	Columns
1. Aulas de inglês na escola	<input type="radio"/> Não tive
2. Acesso a computador	<input type="radio"/> Não tive o suficiente para me ensinar o ...
3. Acesso a videogames (em console ou co...	<input type="radio"/> Me ensinou pouco o inglês
4. Curso de inglês fora da escola	<input type="radio"/> Me ensinou o inglês de modo regular
5. Viagens para países que falam inglês	<input type="radio"/> Me ensinou bastante o inglês
6. Add row	<input type="radio"/> Add column

Require a response in each row

After section 7 Continue to next section

Figura 35: Questionário (Parte 11)

Section 8 of 8

Encerramento

Description (optional)

16 - Para encerrar, liste alguns de seus jogos favoritos!

(No máximo 10, separados por vírgula e espaço. Evite erros de digitação, se possível!)

(Este item é opcional)

Long answer text

Figura 36: Questionário (Parte 12)

Apêndice B - Descrição dos gêneros de jogos digitais utilizados no trabalho

- Jogos de tiro no estilo FPS (do inglês “*First Person Shooter*”) são centrados em combate por meio de armas de fogo, onde o jogador observa o jogo em primeira pessoa, isto é, na mesma visão do personagem controlado. Jogos desse gênero possuem bastante apelo *mainstream* e, quando são multijogador, podem incluir elementos competitivos ou cooperativos. Há alta disponibilidade do gênero em português.
- Jogos de luta, ao invés de usar como obstáculo oponentes propositalmente mais ou menos fortes que o jogador, focam no combate em tempo real entre (geralmente dois) personagens de potencial similar, que podem utilizar artes marciais, armas brancas, magias, etc. A competição entre jogadores humanos é especialmente encorajada nesse gênero. Sabendo das regras, há pouca necessidade de compreender o idioma do jogo.
- Visual Novels/Investigação (abreviados na pesquisa como “Visual Novels”) possuem grande foco em elementos textuais para que se acompanhe uma narrativa ou solucione mistérios. Portanto, há um pressuposto de que o jogador entenda o idioma do jogo. Esses jogos, por não seguirem uma estrutura tradicional de interatividade, não são tão populares, o que torna a localização desses títulos mais rara.
- Jogos de plataforma testam a noção espacial do jogador, e normalmente o maior oponente do jogador é o próprio cenário, com inimigos orgânicos sendo introduzidos pensando nessa dinâmica. Verticalidade é bastante comum, com o meio mais comum de interação sendo o pulo. Muitos mascotes clássicos de videogames surgiram em jogos desse gênero, em uma época que jogos em português quase não existiam.
- Jogos de ritmo são aqueles onde o jogador deve executar alguma ação no ritmo da música de fundo. Essa ação pode ser abstrata (como fazer um personagem realizar uma ação), ou não ter abstração nenhuma (como apertar um botão no momento que um círculo chega no local indicado). Usar o movimento do próprio jogador é menos incomum nesse gênero, dando ênfase em habilidades físicas. Compreender o idioma do jogo é pouco essencial.
- Jogos de corrida/esporte (abreviados na pesquisa como “Esporte”) adaptam esportes já existentes no mundo real para o mundo eletrônico, sendo mais facilmente compreendido por leigos. Essa adaptação pode ser feita de modo tradicional, sem muitos desvios em relação ao esporte de origem, assim como também pode usar elementos fantásticos e fantasiosos para criar uma variação do esporte real que não

seria possível fora dos videogames. Por eles adaptarem regras existentes, saber o idioma dos jogos não costuma ser grande obstáculo.

- Jogos de puzzle (“*quebra-cabeças*”, na tradução literal) testam a habilidade lógica do jogador, tirando o foco dos reflexos, do improviso e da memória muscular, mesmo que possam haver elementos dos três. São diferentes de jogos de investigação por não haver foco na parte textual para a solução de problemas. Após o conhecimento das regras, os jogos costumam apresentar seus objetivos com linguagem não-verbal.
- Jogos no estilo RPG/estratégia (abreviados na pesquisa como “RPG”, do termo “*Role-playing Game*”) possuem definições variadas, mas costumam envolver a abstração de ações reais (por exemplo, comandar um personagem a atacar ao invés de haver uma ação e reação instantânea), customização do arsenal de personagens, e o uso de valores numéricos que representam diversas características de personagens. A leitura é prevalente nesse tipo de jogo, de modo que compreender seu idioma é essencial.
- Jogos de ação testam o reflexo do jogador por meio de combate frenético, usando uma variedade de golpes chamativos. Se diferenciam de jogos de luta não só pelo jogador normalmente lutar contra mais de um inimigo, mas também pelo fato de que os inimigos não possuem o mesmo potencial do jogador, sendo meramente obstáculos e tendo golpes mais facilmente telegrafados, favorecendo apenas ao jogador. Noção espacial é menos importante no gênero. Apesar do foco em ação, é comum haver pausas entre momentos frenéticos onde há mais linguagem verbal.
- Jogos de simulação são jogos propositalmente mais lentos, onde o jogador realiza algum tipo de atividade similar à do mundo real, como gerenciar uma loja, interagir com uma comunidade ou realizar trabalhos reais sem muitos elementos fantasiosos. O propósito desses jogos normalmente é o relaxamento. Por estarem mais ligados ao mundo real, a linguagem verbal costuma ser prevalente.
- MOBAs (do inglês “*Multiplayer Online Battle Arena*”) são jogos onde duas equipes competem entre si em uma arena pré-estabelecida, com o objetivo variando a depender do jogo. Nesse gênero, o jogador observa seu personagem em terceira pessoa, podendo também enxergar seus arredores, enquanto usa uma variedade de poderes para superar obstáculos, que podem ou não ser outros jogadores. Apesar da linguagem verbal ser muito usada de modo similar a RPGs, esses jogos são altamente disponíveis em português.

-
- Jogos com foco em exploração/aventura (abreviados na pesquisa como “Aventura”) englobam diversos gêneros como *metroidvania* e *sandbox*, que possuem como elemento comum o incentivo à exploração de um cenário, podendo ou não ter um objetivo ou narrativa por trás da exploração. O uso da linguagem verbal varia bastante, dada a diversidade desses jogos, que também podem estar misturados com outros gêneros.