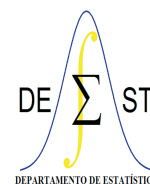




UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE ESTATÍSTICA
BACHARELADO EM ESTATÍSTICA



Utilização de Regressão Logística na Análise da Percepção da Comunidade da UFOP acerca da Pandemia de Covid-19

Flávia Eloí Miranda

Ouro Preto-MG
Março de 2023

Flávia Eloi Miranda

Utilização de Regressão Logística na Análise da
Percepção da Comunidade da UFOP acerca da
Pandemia de Covid-19

Monografia de Graduação apresentada ao
Departamento de Estatística do Instituto de
Ciências Exatas e Biológicas da Universidade
Federal de Ouro Preto como requisito parcial
para a obtenção do grau de bacharel em Es-
tatística.

Orientador

Dr. Eduardo Bearzoti

UNIVERSIDADE FEDERAL DE OURO PRETO – UFOP
DEPARTAMENTO DE ESTATÍSTICA – DEEST

Ouro Preto-MG

Março de 2023

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

M672u Miranda, Flavia Eloi.

Utilização de regressão logística na análise da percepção da comunidade da UFOP acerca da Pandemia de Covid-19. [manuscrito] / Flavia Eloi Miranda. - 2023.

35 f.: il.: gráf., tab..

Orientador: Prof. Dr. Eduardo Bearzoti.

Monografia (Bacharelado). Universidade Federal de Ouro Preto. Instituto de Ciências Exatas e Biológicas. Graduação em Estatística .

1. Modelos Lineares (Estatística). 2. Logística. 3. COVID-19 (Doença).
I. Bearzoti, Eduardo. II. Universidade Federal de Ouro Preto. III. Título.

CDU 31

Bibliotecário(a) Responsável: Luciana De Oliveira - SIAPE: 1.937.800



FOLHA DE APROVAÇÃO

Flávia Elói Miranda

Utilização de regressão logística na análise da percepção da comunidade da UFOP acerca da pandemia de Covid-19

Monografia apresentada ao Curso de Estatística da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Estatística

Aprovada em 31 de março de 2023

Membros da banca

Dr. Eduardo Bearzoti - Orientador (Universidade Federal de Ouro Preto)
Dr. Ricardo Tavares (Universidade Federal de Ouro Preto)
Dr. Diana Campos de Oliveira (Universidade Federal de Ouro Preto)

Professor Dr. Eduardo Bearzoti, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 31/03/2023



Documento assinado eletronicamente por **Eduardo Bearzoti**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 06/04/2023, às 11:11, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Ricardo Tavares**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 06/04/2023, às 12:44, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Diana Campos de Oliveira**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 06/04/2023, às 13:37, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0504098** e o código CRC **5CF46BF3**.

Agradecimentos

Em especial, a minha mãe Milian e minha irmã Carolina que sempre me incentivaram nos estudos, me dando suporte em todos os sentidos possíveis. Aos meus irmãos Frederico e Ramon por estarem comigo em todos os momentos. A minha melhor amiga Daniele por aguentar meus surtos. A todos os amigos que fiz em Ouro Preto. Por último, ao meu orientador Eduardo por ter me auxiliado nessa reta final do curso.

Sem vocês nada disso seria possível.

Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito.

Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes.

Marthin Luther King

Utilização de Regressão Logística na Análise da Percepção da Comunidade da UFOP acerca da Pandemia de Covid-19

Autor: Flávia Eloi Miranda

Orientador(a): Dr. Eduardo Bearzoti

RESUMO

Devido à pandemia do novo coronavírus, vários países do mundo todo, inclusive o Brasil, decidiram tomar medidas de isolamento. Diante disso, as instituições de ensino tiveram suas atividades suspensas, para a volta as aulas, como na Universidade Federal de Ouro Preto (UFOP). Durante a pandemia, pesquisadores desta instituição realizaram uma pesquisa de consulta à comunidade acadêmica, chamada “Percepções e perspectivas da comunidade universitária na pandemia de Covid-19: o que é importante para o retorno presencial?”. Neste trabalho foram utilizados modelos de regressão logística simples e múltipla, para elucidar relações entre as características dos respondentes e suas respostas na questão “sentimento em relação a pandemia”. Em tais análises, verificou-se que as variáveis campus e gênero foram significativas para “apreensivo”, enquanto que as variáveis explicativas: campus, gênero, etnia e estado civil foram significativas para “esperançoso”. Concluiu-se, pelas razões de chances, que as mulheres tendem a se sentirem mais apreensivas que os homens, da mesma forma que respondentes do campus de Mariana; da mesma forma, pessoas autodeclaradas pardas ou pretas apresentaram uma chance maior de se sentirem mais esperançosas. Pessoas casadas apresentaram maior chance de se sentirem esperançosas em relação a pessoas solteiras na análise marginal, porém esta tendência se inverteu no ajuste da regressão múltipla.

Palavras-chave: Modelos Lineares Generalizados, Regressão Logística, Covid-19.

Use of Logistic Regression in the Analysis of the UFOP Community's Perception of the Covid-19 Pandemic

Author: Flávia Eloi Miranda

Advisor: Dr. Eduardo Bearzoti

ABSTRACT

Due to the new coronavirus pandemic, several countries around the world, including Brazil, decided to take isolation measures. In view of this, educational institutions had their activities suspended, for back to school, as at the Federal University of Ouro Preto (UFOP). During the pandemic, researchers from this institution carried out a survey to consult the academic community, called “Perceptions and perspectives of the university community in the Covid-19 pandemic: what is important for face-to-face return?”. In this work, simple and multiple logistic regression models were used to elucidate relationships between the characteristics of the respondents and their answers to the question “feeling in relation to the pandemic”. In such analyses, it was found that the campus and gender variables were significant for “apprehensive”, while the explanatory variables: campus, gender, ethnicity and marital status were significant for “hopefulness”. It was concluded, by the odds ratios, that women tend to feel more apprehensive than men, in the same way as respondents from the Mariana campus; similarly, self-declared brown or black people were more likely to feel more hopeful. Married people were more likely to feel hopeful than single people in the marginal analysis, but this trend was reversed in the multiple regression adjustment.

Keywords: Generalized Linear Models, Logistic Regression, Covid-19.

Lista de figuras

| | | |
|---|---|-------|
| 1 | Gráfico da função logito inversa. | p. 15 |
|---|---|-------|

Lista de tabelas

- 1 Deviances e valores- p correspondentes ao ajuste de modelos de regressão logística simples, considerando diferentes variáveis explicativas, em separado, e as variáveis resposta dicotômicas (sim ou não): Apreensão e Esperança, acerca do sentimento de respondentes de um questionário em relação à pandemia de Covid-19. p. 22
- 2 Razões de Chances para a variável resposta Apreensão, considerando a variável explicativa campus (acima da diagonal) e intervalos de confiança destas mesmas razões, com um índice de confiança de 95% (abaixo da diagonal). p. 23
- 3 Razões de Chances para a variável resposta Apreensão, considerando a variável explicativa gênero (acima da diagonal) e intervalos de confiança destas mesmas razões, com um índice de confiança de 95% (abaixo da diagonal). p. 24
- 4 Razões de Chances para a variável resposta Esperança, considerando a variável explicativa campus (acima da diagonal) e intervalos de confiança destas mesmas razões, com um índice de confiança de 95% (abaixo da diagonal). p. 25
- 5 Razões de Chances para a variável resposta Esperança, considerando a variável explicativa estado civil (acima da diagonal) e intervalos de confiança destas mesmas razões, com um índice de confiança de 95% (abaixo da diagonal). p. 26
- 6 Razões de Chances para a variável resposta Esperança, considerando a variável explicativa etnia (acima da diagonal) e intervalos de confiança destas mesmas razões, com um índice de confiança de 95% (abaixo da diagonal). p. 26

| | | |
|---|--|-------|
| 7 | Ajuste de um modelo de regressão logística múltipla, após seleção de variáveis conforme o método backward, considerando a variável resposta “Apreensão”. | p. 28 |
| 8 | Ajuste de um modelo de regressão logística múltipla, após seleção de variáveis conforme o método backward, considerando a variável resposta “Esperança”. | p. 29 |

Sumário

| | | |
|----------|---|-------|
| 1 | Introdução | p. 11 |
| 2 | Referencial Teórico | p. 13 |
| 2.1 | GLM's e Regressão Logística | p. 13 |
| 2.2 | Razões de Chances | p. 15 |
| 2.3 | Ajuste do Modelo | p. 16 |
| 2.4 | Métodos de Seleção de Variáveis | p. 17 |
| 3 | Material e Métodos | p. 18 |
| 4 | Resultados e Discussão | p. 21 |
| 4.1 | Razões de Chances: “Apreensão” | p. 23 |
| 4.2 | Razões de Chances: “Esperança” | p. 25 |
| 4.3 | Regressão Logística Múltipla | p. 27 |
| 5 | Considerações finais | p. 31 |
| 6 | Referências | p. 32 |

1 Introdução

Em dezembro de 2019 surgiu na cidade de Wuhan (China) uma infecção pulmonar até então tendo a causa desconhecida. Posteriormente foi descoberto que a infecção era causada por um vírus da família coronavírus do qual foi chamado de novo coronavírus pela OMS (Organização Mundial de Saúde).

Devido a COVID-19 ser uma doença ser altamente contagiosa, em março de 2020 o isolamento social foi adotado em ampla escala no Brasil, para tentar diminuir a contaminação em massa. Para isso, apenas os serviços considerados essenciais não pararam de funcionar. Com isso, as escolas e universidades do país tiveram suas atividades paralisadas, sendo logo depois retomadas de forma remota.

No contexto da Universidade Federal de Ouro Preto (UFOP), alguns meses após a paralização, foi realizada a pesquisa: “Percepções e perspectivas da comunidade universitária na pandemia de Covid-19: o que é importante para o retorno presencial?”, na qual foram coletados dados de estudantes e servidores da Universidade acerca da percepção da comunidade acerca da pandemia, e do que seria necessário para um retorno presencial. Uma das questões abordadas nesta pesquisa foi o “sentimento em relação a pandemia”, na qual as respostas eram de múltipla escolha, podendo ser assinalada mais de uma opção, tais como: ansioso, confiante na ciência, esperançoso, tenso, apreensivo *etc.* O presente trabalho teve como principal objetivo a análise das respostas desta questão, considerando duas das opções mais frequentemente assinaladas, quais sejam, as respostas: “apreensivo” e “esperançoso”, a fim de investigar se haveria alguma relação entre estas respostas e outras características dos respondentes (outras questões).

Como as variáveis escolhidas foram de natureza dicotômica, este estudo de caso foi propício para ilustrar a técnica estatística da regressão logística. Segundo esta técnica, as respostas dicotômicas são expressas como funções de outras variáveis, que podem ser categóricas ou quantitativas. Assim, sendo um modelo preditivo, a regressão logística permite estimar a probabilidade do sucesso, para específicos valores de interesse referentes às

variáveis dependentes.

2 Referencial Teórico

A análise estatística padrão para dados dicotômicos corresponde à chamada *regressão logística*, embora, rigorosamente, esta terminologia também possa ser empregada para dados com distribuição normal, ao se ajustar curvas de formato “sigmoide” (Draper e Smith, 2014). Na regressão logística, a probabilidade de sucesso (por exemplo, marcar “esperançoso” em um questionário acerca da percepção sobre a pandemia de Covid-19) é modelado em função de variáveis explicativas, também chamadas independentes, sendo que a curva ajustada também apresentará este mesmo comportamento sigmoide, com a particularidade de que a imagem da função corresponde ao intervalo entre 0 e 1 (Agresti, 2002).

A regressão logística é um caso particular de um modelo linear generalizado, uma classe de modelos descrita brevemente a seguir.

2.1 GLM’s e Regressão Logística

Modelos lineares generalizados (comumente referenciados pela sua sigla inglesa, GLM’s) são uma classe de modelos estatísticos que estendem a estrutura de regressão linear para lidar com variáveis de resposta não normalmente distribuídas. Esta classe de modelos foi proposta por Nelder e Wedderburn (1972). A principal ideia por trás dos GLM’s é especificar uma função preditora linear que relaciona o valor esperado da variável de resposta com as variáveis explicativas por meio de uma função de ligação. O modelo então assume que a variável de resposta segue uma distribuição de probabilidade da família exponencial, que inclui distribuições comumente usadas, como as distribuições normal, de Poisson e binomial.

A forma geral de um GLM pode ser expressa da seguinte forma:

$$g(\mu) = \beta_0 + \beta_1 x_1 + \dots + \beta_i x_i$$

sendo g uma função de ligação, conectando o valor esperado μ da variável resposta, a um preditor linear, função de parâmetros $\beta_0, \beta_1, \dots, \beta_i$, coeficientes de regressão referentes a variáveis explicativas x_1, x_2, \dots, x_i .

Os GLM's são amplamente utilizados em vários campos do conhecimento, incluindo biologia, ecologia, economia e ciências sociais, para modelar estruturas de dados complexas e entender as relações entre variáveis. Além da estimação dos coeficientes de regressão, a técnica também abrange a realização de testes de hipóteses, bem como a previsão de novos pontos, considerando valores futuros para as variáveis explicativas.

A regressão logística é um caso particular de um modelo linear generalizado, usado para modelar a probabilidade de um resultado dicotômico (ou seja, uma variável resposta com dois valores possíveis) com uma ou mais variáveis explicativas. O objetivo da regressão logística é modelar a probabilidade do resultado positivo em função de variáveis explicativas. A distribuição de probabilidade considerada para a variável resposta é assim, em geral, a distribuição Bernoulli (ou a binomial, dependendo de como a variável resposta é registrada), embora outras distribuições também possam ser considerados, por exemplo, em situações de sobredispersão.

As funções de ligação mais utilizadas para a regressão logística são as funções *probit* e *logit* (Agresti, 2002). A função *probit*, historicamente, foi a primeira a ser proposta, e é baseada na inversa da função distribuição (função de probabilidade acumulada) da distribuição normal padrão.

Embora a função *logit* tenha sido proposta posteriormente, trata-se da função de ligação mais popular, para modelar a relação entre a variável de resposta e as variáveis explicativas. A função *logit* modela as probabilidades de a variável resposta estar na categoria “sucesso”, tendo por domínio a reta real, permitindo assim a aplicação do modelo de regressão linear.

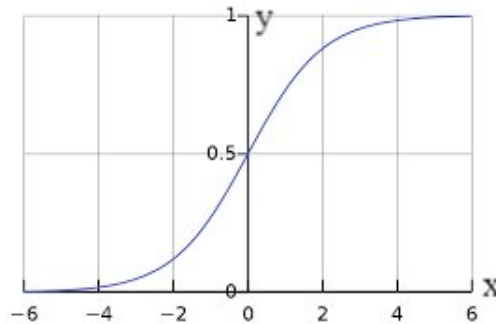
Em estatística, a função *logit* é uma função matemática usada para transformar probabilidades (que normalmente são limitadas entre 0 e 1) em valores que podem assumir qualquer valor na reta numérica real.

A função *logit* é definida como o logaritmo natural das chances de um evento ocorrer, onde as chances de um evento são definidas como a probabilidade do evento, dividida pela probabilidade de seu complemento. Ou seja, para um evento com probabilidade p , as chances de o evento ocorrer são dadas por: $p/(1 - p)$, e assim o *logit* de p é dado por:

$$\text{logit}(p) = \log \frac{p}{(1-p)}$$

O gráfico da inversa dessa função logito está apresentado na Figura 1, evidenciando seu formato sigmoide, quando p é expressa em termos da(s) variável(eis) independentes.

Figura 1: Gráfico da função logito inversa.



2.2 Razões de Chances

Ao se trabalhar com a regressão logística, é frequente o interesse na chamada *razão de chances*, justamente por se tratar de uma variável resposta dicotômica.

A razão de chances (também comumente referenciada pela sua terminologia inglesa, *odds ratio*) é definida como sendo a razão entre a probabilidade do sucesso e a probabilidade do fracasso. Matematicamente, a razão de chances pode ser expressa como:

$$\psi = \frac{p(1-q)}{q(1-p)}$$

A razão de chances pode assumir um valor maior ou menor que 1, sendo o valor 1 o valor de igualdade de chances. Um valor maior que 1 indica que a variável de exposição está associada a uma chance aumentada de o desfecho ocorrer, enquanto que um valor menor que 1 indica que a variável de exposição está associada a uma probabilidade diminuída de ocorrência do resultado de sucesso.

Uma razão de chances igual a 1 indica que não há associação entre a variável de exposição e a variável de resultado.

Na regressão logística, a razão de chances é frequentemente usada para quantificar o efeito de uma variável preditora nas chances de um resultado ocorrer, enquanto controlada

(ajustada) para outras variáveis preditoras.

2.3 Ajuste do Modelo

Existem alguns métodos para averiguar se o modelo de regressão logística foi bem ajustado. Dois critérios quase sempre considerados são: a chamada análise de *deviance*, e o critério de informação de Akaike (comumente abreviado por sua sigla inglesa, AIC).

A *deviance* é um termo estatístico referente à diferença entre o valor observado e o valor esperado em um modelo estatístico. É frequentemente usado no contexto de testes de hipóteses, para avaliar se um modelo GLM explica adequadamente os dados. A análise de deviance guarda uma analogia com a Análise de Variância (ANOVA), para aquelas situações em que a variável resposta é admitida como tendo distribuição normal.

Na regressão logística, a deviance é uma medida da falta de ajuste do modelo aos dados. É usada para avaliar a qualidade do ajuste do modelo, bem como sua comparação com outros modelos. A deviance de um modelo de regressão logística é calculada como a diferença entre a deviance do modelo nulo (um modelo sem preditores) e a deviance do modelo ajustado.

Já o critério de informação de Akaike (AIC) é uma medida da qualidade relativa de modelos estatísticos para um determinado conjunto de dados. Ele fornece um critério para comparar diferentes modelos, e selecionar aquele com maior probabilidade de representar a verdadeira relação subjacente entre as variáveis.

O AIC é baseado no princípio da máxima verossimilhança, que busca encontrar o modelo que maximiza a função verossimilhança dos dados. O AIC leva em consideração tanto a qualidade do ajuste do modelo quanto sua complexidade, no sentido de levar em conta o número de termos do modelo. Quanto menor o valor do AIC, melhor o modelo se ajusta aos dados, levando-se em conta também a sua complexidade. Um modelo com um valor de AIC mais baixo é preferível a um modelo com um valor de AIC mais alto.

O AIC é frequentemente usado na seleção de modelos, onde os pesquisadores comparam vários modelos para encontrar o melhor que explique os dados, como com os métodos de *forward*, *backward* e *stepwise*, discutidos brevemente a seguir.

2.4 Métodos de Seleção de Variáveis

Os métodos de seleção a seguir, conhecidos por sua terminologia inglesa: *backward*, *forward* e *stepwise* são métodos de seleção de variáveis candidatas a ingressar em um modelo estatístico, sendo variações quanto à ordem de possível entrada e/ou saída de variáveis no modelo.

Embora geralmente utilizados para dados com distribuição normal (Draper e Smith, 2014), estes métodos podem ser perfeitamente adaptados para situações com outras distribuições, como no caso da regressão logística (ver, por exemplo, Bilder, 2015).

A seleção *backward* começa com todas as possíveis variáveis preditoras no modelo e, em seguida, procura-se removê-las uma de cada vez, até que o melhor modelo seja encontrado, com todas as variáveis significativas. Isso é feito, em cada etapa, julgando a inclusão da variável seja mediante o valor- p , seja mediante o AIC, removendo-a sempre que não satisfizer o critério adotado. Para que uso deste método de seleção seja possível, é necessário que o número máximo de parâmetros (de variáveis candidatas) não seja maior que o número de observações, para que haja um número de graus de liberdades residuais mínimo para a realização desses testes.

A seleção *forward* é o oposto do método anterior, começando com um modelo vazio e, em seguida, adicionando aquela variável preditora com valor- p mais baixo (se o valor- p for o critério), caso seja significativa. O método prossegue até que nenhuma outra variável possa ser adicionada.

A seleção *stepwise* é uma combinação dos métodos de seleção *forward* e *backward*. Também começa com um modelo vazio e, adicionando a variável preditora com valor- p mais baixo. A cada etapa que uma nova variável é adicionada ao modelo, as demais variáveis são reavaliadas, removendo-se aquela com o valor- p mais alto, caso seja não significativa. Esta alternância de passos *forward* com passos de *backward* é o que define o método de *stepwise*.

Embora estes métodos sejam definidos tendo por critério o valor- p , alguns recursos computacionais, como a linguagem **R**, fazem uso do AIC.

O objetivo dos três métodos é identificar as variáveis preditoras mais importantes que contribuem para a precisão do modelo e eliminar aquelas que não contribuem. No entanto, o método específico usado dependerá dos objetivos do pesquisador e da natureza dos dados que estão sendo analisados.

3 Material e Métodos

A base de dados usada neste trabalho correspondeu aos dados levantados pela pesquisa: “Percepções e perspectivas da comunidade universitária na pandemia de Covid-19: o que é importante para o retorno presencial?”, realizada por uma equipe de profissionais da UFOP, coordenada pela pesquisadora Mariza A. Costa Pena¹, a quem agradecemos a possibilidade do uso de parte dos dados do questionário.

Esta pesquisa consistiu de uma consulta à comunidade acadêmica da UFOP, sobre suas percepções e perspectivas em relação à crise do Covid-19, durante um período crítico da pandemia, em que ainda não havia disponibilidade de vacinas. Esta consulta foi realizada entre os dias 23 de novembro de 2020 e 31 de janeiro de 2021, através de questionário disponibilizado *online*.

Com uma ampla divulgação, era solicitado o preenchimento deste questionário a todos que se dispusessem, tendo sido assim uma amostragem de conveniência da comunidade da Ufop. Ao final do período, as questões foram respondidas por 1.146 pessoas, sendo 139 docentes, 105 técnicos administrativos e 902 estudantes.

Originalmente, a pesquisa continha 37 questões, sobre diversos aspectos, considerando o perfil dos respondentes, bem como a percepção e o estilo de vida da comunidade durante a pandemia. No presente estudo, foi considerada a questão: “Sentimento em Relação à Pandemia”, que correspondeu a uma questão de múltipla escolha, em que mais de uma escolha podia ser marcada, bem como uma última opção denominada ‘Outros’, em que o respondente poderia apontar algum outro fator.

Dentre as diversas opções fornecidas como respostas a esta questão, foram escolhidas duas, quais sejam, “esperançoso(a)” e “apreensivo(a)”, em função de terem sido duas das mais frequentes opções assinaladas nesta questão.

Posteriormente, foram escolhidas outras questões, referentes a outros fatores, em busca de possível associação com as respostas desta questão, tidas como realizações de uma variá-

¹https://saci2.ufop.br/data/pauta/22155_pesquisa_covid19_ufop_resultados_v6.pdf

vel dependente. Assim, foram escolhidos os fatores (questões): campus de origem, função acadêmica, estado civil, etnia, renda, idade, gênero e se, até o momento, havia testado positivo para COVID-19. Estes fatores foram considerados como variáveis potencialmente explicativas, compondo o preditor linear de uma regressão logística.

Dessa forma, a regressão logística foi utilizada tanto em sua forma simples (uma única variável de cada vez), como em sua forma múltipla. No primeiro caso, temos o seguinte preditor linear:

$$\ln \frac{\mu}{1 - \mu} = \beta_0 + \beta_1 X$$

Nesta expressão, μ é a esperança de um dado valor esperado, a qual, considerando a distribuição de Bernoulli, corresponde à probabilidade do evento de sucesso, que, no presente estudo, correspondia ao evento de o participante assinalar a resposta “apreensivo”, em um primeiro ajuste (não assinalar essa opção correspondia a um evento de fracasso), e, em um segundo ajuste, o evento de sucesso correspondia a se o participante assinalava a resposta “esperançoso”. O parâmetro β_0 corresponde ao intercepto do preditor linear, e o parâmetro β_1 corresponde ao efeito que a covariável X exerce sobre a esperança da variável resposta.

A regressão logística múltipla, no presente estudo, correspondeu a considerar várias das variáveis (fatores, ou questões) citadas acima, simultaneamente, no mesmo preditor linear:

$$\ln \frac{\mu}{1 - \mu} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

aqui genericamente representadas como sendo um conjunto de p variáveis.

Quando mais do que uma variável explicativa era considerada, simultaneamente, no mesmo modelo, foi utilizado o método backward de seleção de variáveis, tendo como critério de exclusão ou manutenção de variáveis o critério AIC, se este sofria ou não alguma redução, após a remoção ou não de alguma variável.

Como critério final de permanência no modelo, um teste de significância era realizado, para julgamento da hipótese H_0 , de a variável em questão não exercer nenhum efeito sobre a variável resposta, mantidas as demais variáveis explicativas constantes. Eram considerados como modelos finais apenas aqueles apenas coeficientes significativos, considerando um nível de significância de 5%.

Estes testes de significância se constituíram basicamente de testes oriundos de análises

de deviance, ditas parciais, ou seja, em que cada variável sendo testada aparece em último lugar no modelo (e portanto, tendo-se removido o efeito das outras variáveis).

4 Resultados e Discussão

Conforme apontado no capítulo anterior, modelos de regressão logística foram ajustados para duas variáveis resposta dicotômicas, conforme o participante do questionário ter ou não marcada as opções: “Apreensivo” e “Esperançoso” na questão referente ao sentimento do respondente em relação à pandemia.

Estas respostas binárias foram modeladas em função de outros fatores relativos a cada respondente, conforme suas respostas em outras questões: campus de origem, função acadêmica, estado civil, etnia, renda, idade, gênero e se, até o momento, havia testado positivo para COVID-19. Em um primeiro momento, foram ajustados modelos contendo apenas cada uma dessas variáveis explicativas, em separado. A Tabela 1 apresenta as deviances em relação ao modelo nulo, para cada um desses modelos, bem como seus respectivos valores- p . As deviances (ou diferenças de deviances, mais precisamente) são logaritmos de razões de verossimilhanças, multiplicadas por -2 , e os testes de significância correspondentes são baseados na distribuição assintótica de qui-quadrado, sob H_0 .

Adotou-se aqui um nível de significância de 5%, e assim são dignas de atenção aquelas variáveis com valor- p abaixo de 0,05, as quais são consideradas estatisticamente significantes, com a rejeição de H_0 .

Pode-se observar na Tabela 1 que, para a variável resposta “Apreensão” (que corresponde ao respondente ter marcado ou não “apreensivo” como um dos sentimentos em relação à pandemia), dentre as oito variáveis explicativas consideradas, houve apenas duas significativas, que foram: campus e gênero. Para a variável resposta “Esperança” (que corresponde ao respondente ter marcado ou não “esperançoso” como um dos sentimentos em relação à pandemia), também houve duas variáveis significativas: estado civil e etnia, embora a variável estado civil tenha apresentado um valor- p muito próximo ao limiar de 0,05.

Para variáveis com apenas duas categorias (e um único grau de liberdade no teste da razão de verossimilhanças), a simples rejeição de H_0 implicaria que as duas categorias são

Tabela 1: Deviances e valores- p correspondentes ao ajuste de modelos de regressão logística simples, considerando diferentes variáveis explicativas, em separado, e as variáveis resposta dicotômicas (sim ou não): Apreensão e Esperança, acerca do sentimento de respondentes de um questionário em relação à pandemia de Covid-19.

| | GL | Apreensão RV (valor-p) | Esperança RV (valor-p) |
|----------------------------------|----|---------------------------|---------------------------|
| Campus | 2 | 8,2433 (0,0162)* | 7,2145 (0,0271)* |
| Função | 2 | 0,0129 (0,9935) | 4,5888 (0,1008) |
| Estado Civil | 2 | 0,5294 (0,7674) | 5,6545 (0,0591) |
| Etnia | 5 | 8,5683 (0,1276) | 12,334 (0,0304)* |
| Renda | 4 | 4,3768 (0,3574) | 3,734 (0,4432) |
| Gênero | 2 | 10,039 (0,0066)** | 4,4919 (0,1058) |
| Idade | 2 | 2,0667 (0,3558) | 1,3382 (0,5122) |
| Testou positivo pra COVID | 1 | 0,0293 (0,864) | 2,8824 (0,0895) |

estatisticamente diferentes. Por exemplo, caso a variável explicativa: “testou positivo para COVID” tivesse sido significativa (Tabela 1) para uma das duas variáveis resposta, isto bastaria para indicar uma diferença significativa, quanto as probabilidades de sucesso, entre as categorias: testou positivo, ou não testou positivo.

Quando a variável explicativa é de natureza categórica (ou seja, é um fator), com mais de duas categorias (2 ou mais graus de liberdade), então seria interessante ir além dos resultados da Tabela 1, procurando investigar onde residem as diferenças, para aquela variável explicativa. Poderia acontecer, por exemplo, de a primeira categoria ser diferente da segunda, que por sua vez seria estatisticamente igual à terceira.

Quando os dados têm distribuição normal, esta investigação posterior geralmente é feita mediante testes *post hoc* de comparações múltiplas, como o teste de Tukey. Quando os dados são binários, também é possível a realização de comparações múltiplas de médias, embora seja mais comum o cálculo de razões de chances (ver Capítulo anterior), considerando uma das categorias como sendo uma categoria de referência, a qual é comparada com todas as demais categorias. Na linguagem **R**, por padrão utiliza-se o primeiro nível de um fator como a categoria de referência, embora esta pode ser facilmente alterada, bastando mudar a ordem de especificação dos níveis do fator em questão.

Na presente pesquisa, optou-se pela utilização do cálculo de razões de chances, embora procedimentos de comparações múltipla para a regressão logística possam ser utilizados, por exemplo com o pacote **emmeans**, da linguagem **R**. Dessa forma, são apresentadas a seguir as razões de chances para aquelas variáveis explicativas significativas, conside-

rando a variável resposta “Apreensão”, em um primeiro momento, seguida da variável “Esperança”.

4.1 Razões de Chances: “Apreensão”

Para a variável resposta “Apreensão”, duas variáveis explicativas foram significativas, o campus de origem (categorias: João Monlevade, Ouro Preto e Mariana) e gênero (feminino, masculino e outro), cada qual com 2 graus de liberdade.

A Tabela 2 apresenta as razões de chances para as 3 categorias da variável explicativa campus de origem. Nesta Tabela, a categoria de referência se encontra na coluna indicadora, enquanto que as categorias sendo comparadas com esta categoria de referência estão dispostas no cabeçalho da Tabela. Por exemplo, na primeira linha da Tabela, temos as razões de chances considerando o campus de João Monlevade como categoria de referência.

Além dos valores das razões de chances (apresentadas acima da diagonal), a Tabela 2 apresenta (abaixo da diagonal) intervalos de confiança para cada razão de chances, com um índice de confiança de 95%, intervalos esses construídos conforme a aproximação normal (ver Agresti, 2002). Se os limites de um dado intervalo de confiança são ambos superior ao valor 1 (ou seja, o intervalo não passa pelo valor 1), isto indica que a categoria em questão apresenta uma probabilidade de sucesso (no caso, marcar a opção “apreensivo”) significativamente maior do que a categoria de referência.

Por outro lado, se ambos os limites do intervalo de confiança são inferiores a 1 (também aqui, o intervalo não passando pelo valor 1), isto indica que a categoria em questão apresenta uma probabilidade de sucesso significativamente menor do que a categoria de referência. Finalmente, se o intervalo passa pelo valor 1, isto seria um indicativo de que não há evidência de que as probabilidades de sucesso das duas categorias sejam diferentes.

Tabela 2: Razões de Chances para a variável resposta Apreensão, considerando a variável explicativa campus (acima da diagonal) e intervalos de confiança destas mesmas razões, com um índice de confiança de 95% (abaixo da diagonal).

| | João Monlevade | Ouro Preto | Mariana |
|-----------------------|-----------------------|-------------------|----------------|
| João Monlevade | 1 | 1,69 | 2,03 |
| Ouro Preto | [1,1 ; 2,6] | 1 | 1,20 |
| Mariana | [1,25 ; 3,29] | [0,89 ; 1,61] | 1 |

Observa-se na Tabela 2 que os respondentes de Mariana têm uma chance de ter respondido que se sentem apreensivos em relação à pandemia 2,03 vezes maior que a correspondente chance dos respondentes de João Monlevade. Já os respondentes do campus de Ouro Preto teriam 1,69 mais chance de terem respondido que se sentem apreensivos, em relação aos de João Monlevade.

Já em relação à comparação entre os campus de Ouro Preto e Mariana, como o intervalo de confiança passou pelo valor 1, isto indica que a razão de chances observada (1,20), não foi estatisticamente diferente de 1, indicando chances semelhantes, quanto à Apreensão em relação à pandemia, entre os campus de Ouro Preto e Mariana.

A segunda variável explicativa que apresentou efeito significativo no ajuste da regressão logística foi a variável gênero. As razões de chances, bem como seus intervalos de confiança, estão apresentados na Tabela 3.

Tabela 3: Razões de Chances para a variável resposta Apreensão, considerando a variável explicativa gênero (acima da diagonal) e intervalos de confiança destas mesmas razões, com um índice de confiança de 95% (abaixo da diagonal).

| | Feminino | Masculino | Outro |
|------------------|-----------------|------------------|--------------|
| Feminino | 1 | 0,68 | 0,33 |
| Masculino | [0,54 ; 0,87] | 1 | 0,49 |
| Outro | [0,03 ; 3,67] | [0,04 ; 5,45] | 1 |

Conforme pode-se observar nesta Tabela 3, a categoria de gênero outro não apresentou razões de chances estatisticamente diferentes de 1, em relação aos sexos feminino e masculino, conforme se pode constatar pelos dois intervalos de confiança, ambos abrangendo o valor 1. Por outro lado, conforme se observa nesta mesma Tabela, as chances da categoria masculino foram significativamente menores que as da categoria feminino, dado que o intervalo não passou pelo valor 1. Isto é equivalente a dizer que a categoria feminino apresentou uma probabilidade de marcar a opção “apreensivo” maior, em relação à categoria masculino. A estimativa pontual (masculino em relação ao feminino) da razão de chances foi igual a 0,68. Havendo interesse no cálculo da razão de chances de feminino em relação ao masculino, basta inverter esta estimativa pontual, obtendo-se $1/0,68 = 1,47$.

4.2 Razões de Chances: “Esperança”

Conforme comentado anteriormente, para a variável resposta: “Esperança”, foram significativas as variáveis explicativas: campus de origem e etnia. Embora rigorosamente falando a variável estado civil não tenha sido significativa, seu valor- p foi muito próximo do nível nominal de significância (0,0591, conforme Tabela 1), e assim optou-se, nesse estudo, em também calcular-se suas razões de chances, na eventualidade de alguma ser significativa.

Assim, para a variável campus, as razões de chances e seus respectivos intervalos de confiança estão apresentados na Tabela 4.

Tabela 4: Razões de Chances para a variável resposta Esperança, considerando a variável explicativa campus (acima da diagonal) e intervalos de confiança destas mesmas razões, com um índice de confiança de 95% (abaixo da diagonal).

| | João Monlevade | Ouro Preto | Mariana |
|-----------------------|-----------------------|-------------------|----------------|
| João Monlevade | 1 | 0,86 | 0,58 |
| Ouro Preto | [0,55 ; 1,33] | 1 | 0,67 |
| Mariana | [0,35 ; 0,95] | [0,49 ; 0,93] | 1 |

Observa-se na Tabela 4 que, também para essa variável resposta, não houve diferença significativa entre os campus de Ouro Preto e Mariana, tendo o correspondente intervalo de confiança passado pelo valor 1.

Por outro lado, considerando o campus de João Monlevade como categoria de referência, observa-se na Tabela 4 que o campus Mariana apresentaram chances significativamente menores, em relação ao campus de João Monlevade, quanto à marcação da opção “esperançoso”, no questionário. Isto pode ser apreciado devido ao fato de o intervalo não terem abrangido o valor 1. As chances de marcar “esperançoso” no campus de Ouro Preto correspondeu a 0,86 das mesmas chances, no campus de João Monlevade. Para as chances de Mariana em relação ao campus de João Monlevade, tem-se uma razão de chances igual a 0,58.

Este resultado pode ser considerado coerente com os resultados referentes à resposta Apreensão. Isto porque, associado aos níveis mais elevados de apreensão em Ouro Preto e Mariana, coerentemente observou-se menores níveis de esperança para estes campus, em relação aos campus de João Monlevade.

Conforme apontado acima, neste estudo optou-se por calcular as razões de chances

também para o modelo tendo o estado civil como variável explicativa, embora o valor- p tenha estado um pouco acima do valor nominal do nível de significância. A Tabela 5 apresenta as razões de chances, bem como seus intervalos de confiança, para a variável resposta esperança (sim ou não), em função da variável explicativa estado civil.

Tabela 5: Razões de Chances para a variável resposta Esperança, considerando a variável explicativa estado civil (acima da diagonal) e intervalos de confiança destas mesmas razões, com um índice de confiança de 95% (abaixo da diagonal).

| | Solteiro | Casado | Outro |
|-----------------|-----------------|---------------|--------------|
| Solteiro | 1 | 1,40 | 1,09 |
| Casado | [1,01 ; 1,93] | 1 | 0,94 |
| Outro | [0,57 ; 2,10] | [0,44 ; 2,01] | 1 |

Conforme pode-se observar na Tabela 5, de fato, apesar de o teste de qui-quadrado na análise de deviance ter apresentado um valor- p acima do nível de significância de 0,05 (Tabela 1), embora muito próximo, observou-se aqui na Tabela 5 uma razão de chances significativamente diferente de 1. Esta correspondeu à razão referente às categorias casado e solteiro, com estimativa pontual de 1,40, tendo-se utilizado a categoria solteiro como aquela de referência. Dessa forma, pode-se dizer que os respondentes casados desta pesquisa apresentaram chances de responder “esperançoso” 1,40 vez superior, quando comparadas com as dos respondentes solteiros.

Finalmente, a última variável explicativa que foi significativa foi a etnia, cujas razões de chances para a variável resposta Esperança estão apresentadas na Tabela 6.

Tabela 6: Razões de Chances para a variável resposta Esperança, considerando a variável explicativa etnia (acima da diagonal) e intervalos de confiança destas mesmas razões, com um índice de confiança de 95% (abaixo da diagonal).

| | Branco | Amarelo | Indígena | Pardo | Preto |
|-----------------|---------------|----------------|-----------------|--------------|--------------|
| Branco | 1 | 1,14 | 1,07 | 1,38 | 1,61 |
| Amarelo | [0,48; 2,74] | 1 | 0,94 | 1,21 | 1,41 |
| Indígena | [0,10; 11,89] | [0,07; 12,00] | 1 | 1,29 | 1,50 |
| Pardo | [1,05 ; 1,82] | [0,50; 2,93] | [0,12 ; 14,35] | 1 | 1,16 |
| Preto | [1,10; 2,34] | [0,56; 3,53] | [0,13; 16,93] | [0,58; 1,28] | 1 |

A variável etnia apresentou 5 categorias (4 graus de liberdade). Particularmente nestes casos (com maior número de categorias) é interessante algum procedimento *post hoc* para identificar onde residem as diferenças que levaram à rejeição da hipótese H_0 , na análise de deviance.

Conforme pode-se observar na Tabela 6 que houve apenas duas razões de chances significativamente diferentes de 1. Tomando-se (arbitrariamente) a categoria branco como categoria de referência, tem-se que estas razões significativas são aquelas referente às categorias pardo e preto. Assim, com base nestes resultados, estas etnias tenderiam a marcar, com maiores chances, a opção “esperançoso”. Este resultado pode causar alguma surpresa, no sentido de que tais etnias, por terem sofrido historicamente maior vulnerabilidade social, teriam maiores razões para não se declararem “esperançosos”. No entanto, estes foram os resultados observados para este universo em questão.

4.3 Regressão Logística Múltipla

Todas os resultados anteriores se referiram a ajustes de modelos de regressão logística *simples*, no sentido de que era considerada uma variável explicativa, de cada vez, em separado.

Neste item, é avaliado o ajuste de um modelo de regressão logística *múltipla*, ou seja, considerando diferentes variáveis explicativas simultaneamente. A mudança, em termos de explicação do fenômeno, consiste basicamente na mudança de interpretação das estimativas dos parâmetros do modelo, ou coeficientes de regressão. Em um modelo múltiplo, o efeito significativo de uma variável deve ser entendido como sendo um efeito desta variável, mantidas as demais variáveis constantes. Ou seja, trata-se de um comportamento *condicional*, que pode ser drasticamente diferente do comportamento *marginal* que esta variável explicativa exerça sobre esta variável resposta. Este aspecto faz sentido tanto no contexto da regressão múltipla usual, quando se considera distribuição normal (Draper e Smith, 2014), como na regressão logística múltipla.

Em outras palavras, uma variável significativa na regressão logística simples pode deixar de sê-lo na regressão logística múltipla, devido à ocorrência, por exemplo, de multicolinearidade. E vice-versa. Uma variável não significativa no ajuste simples pode passar a sê-lo no ajuste múltiplo.

O enfoque de regressão logística múltipla foi aqui empregado, através de um método de seleção de variáveis, qual seja, o método de *backward* (Draper e Smith, 2014). O método de backward parte de um modelo completo (contendo todas as variáveis explicativas candidatas possíveis). Para isso, é necessário que o tamanho da amostra seja superior ao número de parâmetros do modelo completo. No presente estudo, o número de respondentes era muito maior que o número de variáveis explicativas, e assim o método pôde ser

utilizado.

Aqui, o método de backward foi empregado utilizando a linguagem **R**, que, por padrão, faz uso do método baseando-se no Critério de Informação de Akaike (AIC). Ou seja, o método é interrompido quando a exclusão de uma determinada variável explicativa não resultar em um modelo ajustado com menor AIC. Este critério costuma ter um bom desempenho, apesar de não se tratar de um procedimento baseado em testes de significância.

No presente estudo, tendo-se chegado a um modelo final, conforme o método de backward, utilizando a linguagem **R**, um modelo com as variáveis explicativas desse modelo final era ajustado. Caso a variável com o maior valor- p não fosse significativa, conforme o nível de significância de 5%, era removida. E, assim, caso necessário, este processo continuaria, iterativamente, semelhantemente ao método de backward.

Em relação à primeira variável resposta (Apreensão), o modelo final para o ajuste da regressão logística múltipla, utilizando o método de backward, está apresentado na Tabela 7. Esta Tabela apresenta as estimativas dos parâmetros, seus erros padrões, bem como os valores- p .

Tabela 7: Ajuste de um modelo de regressão logística múltipla, após seleção de variáveis conforme o método backward, considerando a variável resposta “Apreensão”.

| | Estimativa | Erro Padrão | Valor-p |
|--------------------------|-------------------|--------------------|-----------------------------|
| Intercepto | -0,0555 | 0,2229 | 0,8032 |
| Campus Mariana | 0,5866 | 0,2338 | 0,0208 * |
| Campus Ouro Preto | 0,4581 | 0,2221 | 0,0392 * |
| Gênero Masculino | -0,3227 | 0,1269 | 0,0109 * |
| Gênero Outro | -1,1819 | 1,2294 | 0,3364 |

Conforme fica implícito nesta Tabela 7, as variáveis explicativas selecionadas, conforme o método de backward, foram as variáveis campus e gênero. Nesta Tabela, as categorias ausentes correspondem às categorias (arbitrariamente) tidas como categorias de referência. Assim, para a variável explicativa campus, a categoria de referência correspondeu ao campus de João Monlevade. Já para a variável explicativa gênero, o gênero feminino correspondeu à categoria de referência.

Nesta Tabela 7, os campus de Ouro Preto e Mariana apresentaram coeficientes de regressão significativos, ou seja, demonstraram efeitos significativos, em relação ao campus de João Monlevade. Em relação ao gênero, o masculino foi estatisticamente diferente do

feminino, o qual não foi estatisticamente diferente do gênero outro.

Considerando os sinais desses coeficientes, verifica-se que não houve mudanças de interpretação em relação aos ajustes anteriores, considerando modelos de regressão logística simples (uma variável explicativa de cada vez). Ou seja, em relação ao campus de João Monlevade, os campus de Ouro Preto e Mariana representaram um aumento nas chances de apreensão, enquanto que o gênero masculino representou uma diminuição, para esta mesma variável resposta.

A Tabela 8 apresenta os resultados do ajuste do modelo final, conforme o método backward, para a variável resposta “Esperança”. Nesta Tabela, são apresentados as estimativas dos parâmetros, seus erros padrões, bem como os valores- p .

Tabela 8: Ajuste de um modelo de regressão logística múltipla, após seleção de variáveis conforme o método backward, considerando a variável resposta “Esperança”.

| | Estimativa | Erro Padrão | Valor-p |
|------------------------------|-------------------|--------------------|-----------------------------|
| Intercepto | -0,69077 | 0,27426 | 0,0028 ** |
| Campus Ouro Preto | -0,20033 | 0,22758 | 0,37873 |
| Campus Mariana | -0,69333 | 0,26584 | 0,00911 ** |
| Estado Civil Outro | -0,05161 | 0,39725 | 0,89663 |
| Estado Civil Solteiro | 0,39091 | 0,17110 | 0,02233 * |
| Etnia Amarelo | 0,33017 | 0,46208 | 0,47489 |
| Etnia Indígena | 1,16340 | 1,43996 | 0,41912 |
| Etnia Não Respondeu | -0,39719 | 0,36187 | 0,27238 |
| Etnia Pardo | 0,33932 | 0,14449 | 0,01886 * |
| Etnia Preto | 0,53649 | 0,19752 | 0,00661 ** |
| Gênero Masculino | -0,27489 | 0,13478 | 0,04140 * |
| Gênero Outro | -13,4769 | 486,2706 | 0,977 |

Pode-se observar, na Tabela 8, que as variáveis explicativas que foram selecionadas conforme o método de backward corresponderam ao campus, estado civil, etnia e gênero.

Em relação às variáveis: campus, etnia e gênero, houve poucas mudanças de interpretação, em relação aos ajustes de modelos de regressão logística simples. Em relação ao campus de João Monlevade, o campus de Ouro Preto deixou de apresentar uma diferença significativa. Por outro lado, o campus de Mariana continuou apresentando uma tendência de menores chances de marcação da opção de “esperançoso”.

Mas em relação à etnia e ao gênero, as tendências se mantiveram: as categorias pardo e preto apresentaram uma tendência significativa de maiores chances de esperança, em relação à categoria (arbitrariamente) colocada como referência, branco. Também em re-

lação ao gênero, o gênero masculino apresentou uma tendência de menores chances para a esperança, em relação ao sexo feminino.

Foi somente na variável explicativa estado civil que houve uma mudança de sinal do coeficiente, e conseqüentemente no sentido da interpretação. Na regressão logística múltipla ajustada (Tabela 8) a categoria solteiro apresentou uma tendência de chances para a marcação da opção de “esperançoso”, ao contrário do ajuste da regressão logística simples, em que estas chances foram reduzidas, para a categoria solteiro, conforme se pôde observar na Tabela 5. Este fato ilustra como a interpretação e o sentido de um parâmetro, em um ajuste de regressão simples, podem se alterar, com o ajuste de uma regressão múltipla.

5 Considerações finais

Essa monografia apresentou uma aplicação de regressão logística na análise da percepção da comunidade da UFOP acerca da pandemia de covid-19, com o intuito em descobrir se as características dos respondentes tinham relação com suas respostas ao questionário.

Para “apreensivo” as variáveis que foram significativas foram campus e gênero, não tendo diferenças na análise univariada da multivariada, e concluímos que o campus de mariana e o gênero feminino tem chances maiores de se sentirem apreensivos em relação às demais categorias.

Já para ”esperançoso” as variáveis significativas foram: campus, estado civil e etnia. Na regressão logística múltipla, o estado civil apresentou uma mudança de interpretação, com maiores valores de esperança para a categoria solteiro, ao contrário do que ocorreu no ajuste da regressão logística simples.

Um aspecto interessante se refere ao fato de se tratar de uma questão de múltipla escolha, com possibilidade de marcação tanto de “apreensivo”, como de ”esperançoso”. Portanto, uma modelagem estatística elegante consistiria não em tratar essas opções como variáveis resposta em separado, mas como uma resposta bivariada conjunta. Ou seja, poder-se-ia pensar no ajuste de um modelo de regressão logístico bivariado. Aparentemente, havia um pacote da linguagem **R** para este tipo de análise, que deixou de ser mantido. Isto limitou a utilização deste enfoque para este estudo, mas, servindo, contudo, como sugestão para trabalhos futuros.

6 Referências

- AGRESTI, A. **Categorical Data Analysis**. 2.ed. New Jersey, Editora John Wiley and Sons, 721p. 2002.
- BARRETO, A. S. Modelos de Regressão: Teoria e Aplicação com o Programa Estatístico R, Edição do Autor, 1 Edição, Brasília 2011.
- Bilder, C.R. **Analysis of Categorical Data With R**. Boca Raton: CRC Press, 2015.
- DRAPER, N. R.; SMITH, H. **Applied Regression Analysis**. John Wiley & Sons, 2014.
- GIOLO, S. R. Introdução à Análise de Dados Categóricos com Aplicações, Paraná 2012
- Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). Applied logistic regression (Vol. 398). John Wiley and Sons.
- Nelder, J.A.; Wedderburn, R.W.M. Generalized linear models. **Journal of the Royal Statistical Society A**, v135, n3, 1972.