



FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em Estatística

**Avaliação sobre o papel da mídia
na promoção da adesão às medidas
de prevenção contra a COVID19
na Cidade de Quelimane**

Autor: Fernando Domingos Inácio

Maputo, 29 de Março de 2025



FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em Estatística

**Avaliação sobre o papel da mídia
na promoção da adesão às medidas
de prevenção contra a COVID19
na Cidade de Quelimane**

Autor: Fernando Domingos Inácio

Supervisor: Doutor Osvaldo Francisco André Loquiha

Maputo, 29 de Março de 2025

Dedicatória

Dedico este trabalho científico à memória da minha falecida mãe, Piedade Timóteo Machaieie. Ela foi a luz radiante da minha vida e continua a brilhar intensamente em minha memória. Sua presença amorosa e seu apoio incondicional foram fontes de inspiração constantes ao longo da minha jornada acadêmica.

Declaração de Honra

Declaro por minha honra que o presente trabalho de licenciatura é resultado da minha investigação e que o processo foi concebido para ser submetido apenas para a obtenção do grau de Licenciado em Estatística, na Faculdade de Ciências da Universidade Eduardo Mondlane.

Maputo, Março de 2025

(Fernando Domingos Inácio)

Agradecimentos

Gostaria de expressar meus sinceros agradecimentos a todas as pessoas que estiveram ao meu lado durante a jornada acadêmica e a produção deste trabalho científico. Seu apoio e contribuições foram inestimáveis, e sou imensamente grato por todo o suporte recebido.

Começo expressando minha eterna gratidão à minha querida mãe, Piedade Timóteo Machaieie. Seu amor e apoio incondicional nos últimos anos, mesmo em espírito, foram pilares fundamentais ao longo de toda a minha vida. Desde o início, acreditou em mim e me incentivou a perseguir meus sonhos. Suas palavras de encorajamento e orientação continuaram a me guiar durante toda essa jornada acadêmica. Sinto sua presença em cada conquista e desafio superado, e sou eternamente grato pelo legado de amor e sabedoria que você deixou.

Agradeço também aos meus irmãos, Hélio Machaieie e Timóteo Siteo, por estarem ao meu lado em todos os momentos desta jornada. Seu apoio inabalável e encorajamento sincero foram inestimáveis. Partilhamos laços profundos de cumplicidade, e sou grato por isso.

À minha amada namorada, Vilela Momola, agradeço por seu amor, paciência e compreensão ao longo do processo de redação deste trabalho. Seu apoio incondicional, encorajamento constante e palavras de motivação foram fundamentais para eu superar desafios e alcançar meus objetivos. Sua presença ao meu lado trouxe alegria, equilíbrio e motivação, sou profundamente grato por todo o suporte emocional e incentivo que me deu.

Expresso meu profundo agradecimento ao meu supervisor, Osvaldo Loquiha. Sua orientação, conhecimento e apoio foram essenciais para o sucesso deste trabalho. Sua *expertise* na área de pesquisa científica e sua dedicação em partilhar seu conhecimento foram uma verdadeira fonte de inspiração. Agradeço também por sua paciência durante o processo de produção deste trabalho, pois sua orientação cuidadosa e *feedback* construtivo foram de inestimável valor.

Ao professor Alfredo Muxlhanga, o meu co-supervisor, agradeço por seu suporte e contribuições valiosas na produção deste trabalho. Sua experiência e orientação foram fundamentais para aprimorar minha compreensão do assunto e para a qualidade deste estudo. Sua dedicação em partilhar conhecimento e sua disposição em me ajudar são verdadeiramente apreciadas.

Quero estender meu agradecimento aos docentes e staff do Departamento de Matemática e Informática (DMI) da Faculdade de Ciências na Universidade Eduardo Mondlane. Agradeço a todos os professores que contribuíram para minha formação acadêmica, em especial ao Professor Reinaldo Zezela. Sua mentoria e orientação foram essenciais para minha navegação

no mundo profissional. Agradeço também ao Professor Elísio Mabasso, cujos ensinamentos valiosos enriqueceram minha compreensão e aprofundaram meu conhecimento na área.

Agradeço aos meus amigos e colegas que, de diferentes formas, suportaram a caminhada e jornada acadêmica ao meu lado. Em especial, gostaria de agradecer ao Idilson Buce e ao Yves Baia por sua amizade verdadeira, apoio incondicional e pelos momentos partilhados ao longo desses anos. Agradeço também aos meus colegas David Mondlane, Sebastião Dengo, William Tamele, Alfredo Manhiça, Elsa da Graça, Tircia da Silva e a todos os outros colegas da turma de Estatística de 2017. Nossa jornada conjunta foi repleta de aprendizado, colaboração e superação. O apoio e encorajamento mútuos foram fundamentais para enfrentarmos os desafios acadêmicos juntos.

A todos vocês, meu profundo agradecimento. Seu suporte, amizade e contribuições foram fundamentais para o meu crescimento acadêmico e pessoal. Cada um de vocês deixou uma marca indelével em minha vida, e sou grato por ter tido a honra de contar com cada um de vocês em minha jornada.

Com gratidão sincera,

Fernando Inácio

O homem sábio é aquele que não se entristece com as coisas que não possui, mas se regozija com as coisas que possui

Epicteto

Quem conhece os outros é perspicaz; quem conhece a si mesmo é iluminado. Quem vence os outros é forte; quem vence a si mesmo é poderoso

Lao Tzu

Resumo

A demanda por informações durante crises de saúde, como a pandemia da COVID19, é alta, e as pessoas tendem a recorrer a diferentes fontes/meios para obtenção de informações. Diversas fontes de informação estão disponíveis, incluindo a tradicional (televisão, jornais), mídia social (*Facebook, WhatsApp, Instagram, Twitter*), além de familiares, amigos e colegas de trabalho.

Este trabalho tem como objetivo avaliar o papel da mídia na promoção de comportamentos preventivos contra a COVID-19 e identificar os factores associados à percepção sobre o papel da mídia nessa promoção. Além disso, busca estudar os factores relacionados às fontes de informação sobre a COVID-19, aos comportamentos socioculturais durante a pandemia, de forma especial a adesão a medidas restritivas de mobilidade da população. Este estudo transversal utilizou dados de fontes secundárias e um questionário aplicado a 768 residentes da cidade de Quelimane, em Moçambique, com o objetivo de investigar os comportamentos socioculturais e padrões de movimento da população em relação à pandemia da COVID-19. Os dados foram analisados usando técnicas descritivas, bivariadas e de regressão, linear e logística.

Os resultados deste estudo revelaram associações significativas entre variáveis sócio-demográficas e o uso de diferentes fontes de informação. Observou-se que participantes com maior nível de educação tendiam a atribuir maior importância ao papel da mídia na disseminação de comportamentos preventivos. Além disso, aqueles que relataram maior exposição à mídia também valorizavam mais seu papel. A análise estatística mostrou que a educação dos participantes e a exposição à mídia desempenham um papel importante na percepção do papel da mídia na divulgação de comportamentos preventivos. Além disso, o estado civil dos participantes e o uso de internet e mídia social foram associados às fontes de informação utilizadas. Indivíduos em relacionamentos tendiam a usar mais a internet e mídia social como fontes de informação.

Em conclusão, este estudo destaca a importância da mídia na promoção de comportamentos preventivos contra a COVID-19. A compreensão dos factores associados à percepção e ao uso de diferentes fontes de informação pode apoiar intervenções educacionais e aprimorar a gestão de crises de saúde pública.

Palavras-chave: COVID-19, Mídia, Comportamentos preventivos, Fontes de informação, Restrição de mobilidade, Regressão Logística

Abstract

The demand for information during health crises, such as the COVID-19 pandemic, is high, and people tend to rely on different sources/media to obtain information. Various sources of information are available, including traditional media (television, newspapers), social media (*Facebook, WhatsApp, Instagram, Twitter*), as well as family, friends, and colleagues.

This study aims to assess the role of the media in promoting preventive behaviors against COVID-19 and identify factors associated with perception of the media's role in this promotion. Additionally, it seeks to study factors related to sources of information about COVID-19 and sociocultural behaviors during the pandemic, particularly adherence to restrictive population mobility measures. This cross-sectional study used data from secondary sources and a questionnaire administered to 768 residents of Quelimane City, Mozambique, with the objective of investigating sociocultural behaviors and movement patterns of the population in relation to the COVID-19 pandemic. Data were analyzed using descriptive, bivariate, linear, and logistic regression techniques.

The results of this study revealed significant associations between sociodemographic variables and the use of different sources of information. It was observed that participants with higher levels of education tended to attribute greater importance to the media's role in disseminating preventive behaviors. Furthermore, those who reported higher media exposure also valued its role more. Statistical analysis showed that participants' education and media exposure play an important role in the perception of the media's role in disseminating preventive behaviors. Additionally, participants' marital status and use of the internet and social media were associated with the sources of information used. Individuals in relationships tended to use the internet and social media more as sources of information.

In conclusion, this study highlights the importance of the media in promoting preventive behaviors against COVID-19. Understanding the factors associated with perception and the use of different sources of information can support educational interventions and enhance the management of public health crises.

Keywords: COVID-19, Media, Preventive behaviors, Sources of information, Mobility restriction, Logistic Regression

Lista de Abreviaturas

ANOVA	Análise de Variância
COVID19	Doença do Coronavírus 19
EP	Erro padrão
OMS	Organização Mundial da Saúde
QME	Quadrado médio do erro
QMR	Quadrado médio de regressão
RC	Razão de chances

Índice

1	Introdução	1
1.1	Contextualização	1
1.2	Formulação do Problema	2
1.3	Objectivos	5
1.3.1	Geral	5
1.3.2	Específicos	5
1.4	Relevância do estudo	5
1.5	Estrutura do trabalho	6
2	Revisão da Literatura	7
2.1	Análise de Regressão	7
2.1.1	Análise de regressão múltipla	7
2.1.2	Regressão Logística	14
2.1.3	Regressão logística múltipla	19
2.2	Avaliação do papel da mídia na divulgação de comportamentos preventivos sobre a COVID19	23
3	Material e Métodos	26
3.1	Material	26
3.2	Métodos	28
3.2.1	Análise descritiva dos dados	28
3.2.2	Teste de independência	29
3.2.3	Análise de Regressão	30
3.2.4	Seleção de variáveis do modelo	30
3.2.5	Significância estatística	31
4	Resultados e Discussão	33
4.1	Análise Descritiva	33
4.2	Teste de independência das variáveis	39
4.3	Percepção sobre o papel da mídia na divulgação de comportamentos preventivos contra a COVID19	40

4.4	Fontes de informação sobre a COVID19	42
4.4.1	Número de fontes de informação	42
4.4.2	Internet, mídia social e tradicional como uma das principais fontes de informação da COVID19	44
4.5	Comportamentos sócio-culturais durante a pandemia da COVID19	45
4.6	Discussão	47
5	Conclusões e Recomendações	50
	Referências Bibliográficas	52

Lista de Figuras

2.1.1 Funções de regressão logística	16
5.0.1 Informação do participante	56
5.0.2 Comportamentos Sócio-culturais - Página 1	57
5.0.3 Mídia	58
5.0.4 Distribuição da percepção do papel da mídia	59
5.0.5 Distribuição do número de fontes/meios de informação	60
5.0.6 Distribuição por comprimento da medida de isolamento	61
5.0.7 Resíduos padronizados - Modelo da percepção do papel da mídia	62
5.0.8 Normal QQ-Plot - Modelo da percepção do papel da mídia	62
5.0.9 Resíduos padronizados - Modelo do número de fontes de informação	63
5.0.10 Normal QQ-Plot - Modelo do número de fontes de informação	63

Lista de Tabelas

2.1.1 Dados para regressão linear múltipla	8
2.1.2 Análise de variância para significância de regressão	11
3.1.1 Descrição das variáveis	27
4.1.1 Descrição da amostra em estudo	34
4.1.2 Descrição das variáveis em estudo	34
4.1.3 Percepção do papel da mídia na promoção de atitudes preventivas	36
4.1.4 Número médio de fontes/meios frequentes de obtenção de informações sobre a COVID19	37
4.1.5 Fontes frequentes de obtenção de informações sobre a COVID19	38
4.1.6 Deixou de frequentar lugar habitual devido a COVID19	39
4.2.1 Teste de Independência das variáveis	40
4.3.1 Factores associados a percepção sobre o papel da mídia na divulgação de comportamentos preventivos contra a COVID19	41
4.3.2 Análise de variância para modelo de percepção do papel da mídia	41
4.4.1 Factores relacionados ao número de fontes de informação sobre a COVID19	43
4.4.2 Análise de variância para modelo do número de fontes de informação	43
4.4.3 Factores relacionados as fontes de informação sobre a COVID19	45
4.4.4 Teste de Hosmer e Lemeshow para o modelo das fontes de informação da COVID19	45
4.5.1 Factores relacionados a comportamentos sócio-culturais durante a pandemia da COVID19	46
4.5.2 Teste de Hosmer e Lemeshow para o modelo dos comportamentos sócio-culturais durante a pandemia da COVID19	47
5.0.1 Estimação do modelo sobre as fontes de informação sobre a COVID19- Resultados	55
5.0.2 Estimação do modelo sobre comportamentos sócio-culturais durante a pandemia da COVID19	55

Capítulo 1

Introdução

1.1 Contextualização

COVID-19 é uma doença infecciosa de proporções globais que emergiu na cidade de Wuhan na província de Hubei na China em Dezembro de 2019. Ela espalhou-se por cerca de 210 países em todo mundo. A Organização Mundial da Saúde (OMS) declarou uma pandemia em 30 de Janeiro de 2020 o que levantou preocupações internacionais de saúde pública (Anwar et al, 2020).

A pandemia de COVID-19 já infetou mais de 30 milhões de pessoas no mundo, tendo grande impacto na saúde global com danos colaterais. Em Moçambique, foi declarado estado de emergência pública no final de Março de 2020 (das Neves et al., 2021).

Uma campanha nacional de informação e educação na mídia, ensinando medidas preventivas, foi transmitida pela televisão e rádio em todo o país. Estas medidas preventivas podem ter reduzido o número de infeções: no momento da redação deste artigo, Moçambique era o 9^o em número de casos na África (3.651), no qual ocorreram principalmente em homens (60%), maioritariamente na faixa etária de 25 a 44 anos (53%); 58% dos doentes eram assintomáticos, 34% apresentavam sintomas ligeiros e registaram-se 21 óbitos (taxa de mortalidade de 0,6%, a 7^a mais baixa de África) (das Neves et al., 2021).

O primeiro caso de COVID-19 em Moçambique foi detetado a 22 de março de 2020, e uma semana mais tarde, a 30 de Março, o Presidente da República de Moçambique declarou o Estado de Emergência, nível 3 (com níveis variando de 1 a 4, sendo 4 o nível mais rigoroso de distanciamento). A ordem incluiu o encerramento de escolas, restaurantes e bares, igrejas e instalações desportivas, aglomerações restritas a 10 pessoas ou menos, restringiu o número de pessoas no local de trabalho e restrições ao transporte público e à circulação. Este estado de emergência foi prorrogado mensalmente (nos dias 29 de abril, 28 de maio e 26 de junho) e foi mantido até 6 de setembro de 2020 (Júnior et al., 2021).

Até 29 de novembro de 2020, mais de 15.000 casos confirmados de COVID-19 foram diagnosticados em Moçambique. A Cidade de Maputo como o epicentro da epidemia de

COVID-19 no país com cerca de metade dos casos diagnosticados em residentes; a seguir a Província de Maputo com cerca de 17% dos casos diagnosticados.

Um anterior inquérito de seroprevalência mostrou uma prevalência comunitária de aproximadamente 4% na Cidade de Maputo. Atualmente, não está claro quão bem as pessoas aderem a medidas preventivas ao longo do tempo e quais factores determinam a adesão. Foi conduzida uma pesquisa online para avaliar até que ponto as pessoas em Moçambique respeitam as estratégias preventivas da COVID-19, incluindo distanciamento social, ficar em casa e medidas de higiene pessoal e coletiva, e foi feita a avaliação dos factores associados à adesão reduzida (Júnior et al., 2021).

Para obter informações sobre o COVID-19, pontualidade e precisão, embora difíceis de alcançar e medir, são fundamentais para mitigar e curar a doença tanto para o público e a comunidade científica. Entre o público, a internet é a fonte mais popular de informação sobre as etiologias e modelos de intervenção de doenças médicas (Wang et. al., 2021).

De acordo com Wang et. al.,(2021) a informação tem vindo a proliferar nos meios tradicionais e mídias sociais desde o surto da COVID-19. Um estudo recente de 21 países descobriram que o número de pesquisas no Google por “lavar mãos” aumentou com menor velocidade de propagação do COVID-19. A desinformação sobre a COVID-19, no entanto, também foi proliferando na internet, especialmente nas mídias sociais.

Segundo Alshareef et. al.,(2021) durante a emergência de doenças infecciosas, a adoção de medidas preventivas é uma das intervenções mais cruciais para travar a propagação da infeção. Pesquisas anteriores demonstraram o papel da mídia em fornecer informações sobre saúde e promover ações e comportamentos preventivos.

O envolvimento em comportamentos preventivos é mais provável de ocorrer quando há informações claras, confiáveis e comunicação frequentes. Por outro lado, fraca ou inconsistente comunicação tem o potencial de reduzir a confiança do público e,consequentemente, aumentar a probabilidade de eventos sociais e impactos económicos. É evidente como a rápida comunicação e publicação contribuiu para o reconhecimento da gravidade e magnitude da COVID-19 pelo público em geral (Alshareef et al., 2021).

1.2 Formulação do Problema

De acordo com Alshareef et. al.,(2021) em situações de crise de saúde, a demanda por informações costuma ser alta; há muitas incógnitas e as pessoas tendem a recorrer a fontes que elas confiam .Várias fontes de informação estão atualmente disponíveis para obter informações relacionadas à saúde. Como consequência,a importância dessas fontes durante uma crise de saúde global intensifica.

Por exemplo, a mídia tradicional, como a televisão e jornais, desempenham um papel na comunicação de informações ao público baseada em evidências. As pessoas também podem confiar em seus familiares, amigos e colegas de trabalho para obter informações sobre COVID-

19. A mídia social também representa uma plataforma vital para que as pessoas expressem suas opiniões, percepções e atitudes em relação as políticas de saúde pública COVID-19.

Dada a ampla gama de canais de comunicação disponíveis, evidências mostram que as pessoas confiam em uma variedade de fontes de informação diferentes e que a utilização dessas fontes tende a mudar ao longo do tempo. Por exemplo, em 2017, Spence et al. citado por Alshareef et al.,(2021) mostraram que os meios de comunicação tradicionais eram os canais de informação preferidos durante uma crise de saúde pública, enquanto um estudo recente mostrou que plataformas de mídia social ou notícias online locais foram as fontes mais predominantes para obter informação durante momentos de crises.

A OMS declarou que a pandemia de COVID-19 foi acompanhada por uma chamada "infodemia" de desinformação que torna a descoberta de informação clara, orientação e fontes confiáveis difícil.

A desinformação sobre a pandemia pode representar um risco à saúde pública e às ações públicas, prejudicando esforços de saúde pública para mitigar o profundo impacto do vírus. Nesse sentido, um grande desafio é garantir que as pessoas tem acesso a informações precisas que lhes permitam agir adequadamente (Alshareef et al., 2021).

Segundo Eniroso et. al.,(2021) com o aumento da compreensão do SARS-CoV-2, as intervenções locais para a COVID-19 começaram a se concentrar sobre as fontes e a precisão percebida de informações sobre prevenção divulgadas, que prenuncia um importante processo na prevenção e controle da doença.

Potenciais fontes de informação durante surtos de doenças incluem a internet e plataformas de mídia social (*Facebook, WhatsApp, Instagram e Twitter*), mídia tradicional (televisão, rádio e jornais), locais de culto, prestadores de cuidados de saúde, familiares e amigos e locais de trabalho. Portanto, devido às diversas fontes de informação sobre COVID-19, identificar factores relacionados a essas fontes pode apoiar intervenções educacionais públicas (Eniroso et al., 2021).

Além da internet, as mídias tradicionais também são importantes fontes de informação durante surtos de doenças. No entanto, a exposição repetida da mídia a informações relacionadas à crise eleva respostas de ansiedade e estresse entre as pessoas. O público poderá também receber informações sobre o COVID-19 da equipe médica e leigos, como seus amigos, familiares e colegas de trabalho.

Como as pessoas obtêm informações sobre a COVID-19 de vários fontes, entendendo os factores relacionados a essas fontes de informações ajuda os profissionais de saúde na educação de indivíduos em particular e do público em geral. Assim, desenvolvendo sistemas de divulgação de informações que sejam transparentes e eficazes, a autoconfiança no combate da pandemia pode ser melhorada em sua audiência (Wang et al., 2020).

Até ao momento da redação do presente trabalho, em Moçambique a média semanal de casos positivos para a COVID-19 era de 8, o que se figurava num momento em que o estado de emergência já havia sido revogado e as medidas restritivas perdido o carácter obrigatório.

No entanto, o conhecimento sobre o impacto que a mídia, redes sociais e outros elementos exercem sobre a decisão de adesão a comportamentos e atitudes em relação a momentos de crise fornece bases para todos intervenientes na melhoria de gestão de crises em saúde pública, daí a pergunta de pesquisa:

Qual é o impacto da mídia na promoção das medidas de prevenção da COVID-19?

1.3 Objectivos

1.3.1 Geral

Avaliar o papel da mídia na promoção de comportamentos preventivos contra a COVID19

1.3.2 Específicos

- Descrever a amostra em estudo;
- Identificar os factores associados com a percepção sobre o papel da mídia na promoção de comportamentos preventivos contra a COVID19;
- Estudar os factores relacionados as fontes de informação relativas a COVID19;
- Descrever os factores associados aos comportamentos sócio-culturais durante a pandemia da COVID19;

1.4 Relevância do estudo

A pandemia da COVID 19 apresentou, não só para Moçambique, mas também para a comunidade internacional um desafio a definição de políticas e estratégias de contenção e promoção das práticas de prevenção, o que para além de muitos fundamentava-se no uso de meios de comunicação sociais para o seu auxílio.

Numa altura em que para além dos meios de comunicação tradicionais como o jornal, a rádio e a televisão, existe também em grande e contínua ascensão a internet e as redes sociais como meios de partilha e obtenção de informação, o que se configura como suficiente para justificar um ajuste nas estratégias de comunicação para a promoção de práticas preventivas e não só.

Deste modo este trabalho visa despertar a atenção ao papel das mídia no alcance dos objetivos de controle e prevenção de doenças infecciosas, assim como descrever os factores associados a escolha das diferentes fontes de informação; aos comportamentos adoptados durante o estado declarado de calamidade pública e ao grau de percepção, por parte da população, sobre o papel da mídia na divulgação de informações de comportamento preventivos para ajudar no controlo da COVID 19.

Até a altura da redação desse trabalho a COVID 19 em Moçambique e no mundo, não mais configurava uma situação de calamidade pública, tendo então sido levantadas maior parte das restrições que advinham da mesma. No entanto as ilações deste trabalho transcendem o caso específico da COVID 19 podendo servir para definição de estratégias e políticas de promoção a prevenção e controlo de doenças de um modo geral.

1.5 Estrutura do trabalho

Este trabalho compreenderá a seguinte organização:

O segundo capítulo da Revisão da Literatura, onde é feita uma abordagem referente à Análise de Regressão e a Regressão Logística Binária e também a revisão de estudos referentes a avaliação do papel da mídia no contexto da pandemia da COVID19; o terceiro capítulo do Material e Métodos, neste capítulo que é apresentada a fonte de obtenção de dados, as variáveis do estudo, aspectos à estas referentes e é feita uma abordagem referente à metodologia do trabalho e de seus passos aplicados de modo a alcançar os objetivos do mesmo; no quarto capítulo dos Resultados e Discussão, são apresentados os resultados do processo de análise dos modelos de regressão e logísticos estimados e algumas discussões acerca de sua aplicação; e por fim a Conclusões e Recomendações, neste capítulo são apresentadas as conclusões e algumas recomendações à luz dos objetivos estabelecidos e dos resultados obtidos.

Capítulo 2

Revisão da Literatura

2.1 Análise de Regressão

A análise de regressão é uma das técnicas mais amplamente utilizadas para analisar dados quantitativos. Seu amplo apelo e utilidade resultam do processo concetual lógico usar uma equação para expressar a relação entre uma variável de interesse (a resposta) e um conjunto de variáveis preditoras relacionadas.

A análise de regressão é também interessante teoricamente por causa da matemática subjacente elegante e uma teoria estatística bem desenvolvida. O uso bem-sucedido da regressão requer uma apreciação tanto da teoria quanto dos problemas práticos que normalmente surgem quando a técnica é empregada com dados do mundo real (Montgomery et. al., 2012).

2.1.1 Análise de regressão múltipla

De acordo com Hair et. al.,(2009), a análise de regressão múltipla é uma técnica estatística que pode ser usada para analisar a relação entre uma única variável dependente e várias variáveis independentes. O objetivo da análise de regressão múltipla é usar as variáveis independentes cujos valores são conhecidos para prever os valores da variável dependente selecionada pelo pesquisador. Cada variável independente é ponderada pelo procedimento da análise de regressão para garantir máxima previsão a partir do conjunto de variáveis independentes. Os pesos denotam a contribuição relativa das variáveis independentes para a previsão geral e facilitam a interpretação sobre a influência de cada variável em fazer a previsão, apesar de a correlação entre as variáveis independentes complicar o processo interpretativo. O conjunto de variáveis independentes ponderadas forma a variável estatística de regressão, uma combinação linear das variáveis independentes que melhor prevê a variável dependente.

A variável estatística de regressão, também conhecida como equação de regressão ou modelo de regressão, é o exemplo mais amplamente conhecido de uma variável estatística entre as técnicas multivariadas.

Segundo Hair et. al.,(2009), O objetivo da análise de regressão é prever uma única variável dependente a partir do conhecimento de uma ou mais variáveis independentes. Quando o problema envolve uma única variável independente, a técnica estatística é chamada de regressão simples. Quando o problema envolve duas ou mais variáveis independentes, chama-se regressão múltipla.

Modelo de regressão linear múltipla

Em geral, a resposta Y pode estar relacionada a k variáveis preditoras. O modelo:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon \quad (2.1)$$

é chamado de modelo de regressão linear múltiplo com k preditoras. Os parâmetros β_j , $j = 0, 1, \dots, k$, são chamados de coeficientes de regressão. Este modelo descreve um hiperplano no espaço k -dimensional das variáveis preditoras X_j . O parâmetro β_j representa a mudança esperada na resposta Y por unidade de mudança em X_j quando todas as demais variáveis preditoras X_i ($i \neq j$) são mantidas constantes. Por este motivo os parâmetros β_j , $j = 1, 2, \dots, k$, são frequentemente chamados de coeficientes de regressão parcial (Montgomery et. al., 2012).

Modelos de regressão linear múltipla são frequentemente usados como modelos empíricos ou funções de aproximação. Ou seja, a verdadeira relação funcional entre Y e X_1, X_2, \dots, X_k é desconhecido, mas em certos intervalos das variáveis do regressor, o modelo de regressão linear é uma aproximação adequada para a verdadeira função desconhecida (Montgomery et. al., 2012).

Estimativa de mínimos quadrados dos coeficientes de regressão

O método dos mínimos quadrados pode ser usado para estimar os coeficientes de regressão em 2.1. Suponha que $n > k$ observações estejam disponíveis, e deixe y_i denotar a i -ésima resposta observada e x_{ij} denotam a i -ésima observação ou nível do preditor x_j . Os dados aparecerão como na tabela abaixo. Assumimos que o termo de erro no modelo tem $E(\epsilon) = 0$, $Var(\epsilon) = \sigma$

Tabela 2.1.1: Dados para regressão linear múltipla

Observação, i	Resposta, y	Preditores			
		x_1	x_2	\dots	x_k
1	y_1	x_{11}	x_{12}	\dots	x_{1k}
2	y_2	x_{21}	x_{22}	\dots	x_{2k}
\vdots	\vdots	\vdots	\vdots		\vdots
n	y_n	x_{n1}	x_{n2}	\dots	x_{nk}

Podemos escrever o modelo de regressão amostral correspondente à 2.1 como:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon \\ &= \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i, i = 1, 2, \dots, n \end{aligned} \quad (2.2)$$

A função de mínimos quadrados é,

$$S(\beta_0, \beta_1, \dots, \beta_k) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 \quad (2.3)$$

A função S deve ser minimizada em relação a $\beta_0, \beta_1, \dots, \beta_k$. O mínimo quadrados estimadores de $\beta_0, \beta_1, \dots, \beta_k$ deve satisfazer

$$\left. \frac{\partial S}{\partial \beta_0} \right|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij} \right) = 0 \quad (2.4)$$

e

$$\left. \frac{\partial S}{\partial \beta_j} \right|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij} \right) x_{ij} = 0, j = 1, 2, \dots, k \quad (2.5)$$

É mais conveniente lidar com modelos de regressão múltipla se eles forem expressos em notação matricial. Isso permite uma exibição muito compacta do modelo, dados e resultados.

Em notação matricial, o modelo dado pela 2.2 é:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

onde:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_k \end{bmatrix}$$

Em geral, y é um vetor $n \times 1$ das observações, X é uma matriz $n \times p$ dos níveis das variáveis preditoras, β é um vetor $p \times 1$ dos coeficientes de regressão, e ϵ é um vetor $n \times 1$ de erros aleatórios.

Queremos encontrar o vetor de estimadores de mínimos quadrados, $\hat{\beta}$, que minimiza:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n \epsilon_i^2 = \boldsymbol{\epsilon}^T \boldsymbol{\epsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Note que $S(\boldsymbol{\beta})$ pode ser expresso como

$$\begin{aligned} S(\boldsymbol{\beta}) &= \mathbf{y}^T \mathbf{y} - \boldsymbol{\beta}^T \mathbf{X} \mathbf{y} - \mathbf{y}^T \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} \\ &= \mathbf{y}^T \mathbf{y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} \end{aligned}$$

como $\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y}$ é uma matriz 1×1 ou escalar, e o seu transposto $(\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y})^T = \mathbf{y}^T \mathbf{X} \boldsymbol{\beta}$ é o mesmo escalar. Os estimadores mínimos quadrados devem satisfazer:

$$\left. \frac{\partial S}{\partial \boldsymbol{\beta}} \right|_{\hat{\boldsymbol{\beta}}} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} = 0$$

que simplifica para

$$\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{y} \quad (2.6)$$

As equações acima são chamadas equações normais mínimos quadrados, para resolver as equações, multiplicam-se ambos lados de 2.6 pelo inverso de $\mathbf{X}^T \mathbf{X}$. Daí o estimador de mínimos quadrados de $\boldsymbol{\beta}$ é

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (2.7)$$

Teste de significância da regressão linear

Segundo Montgomery et. al.,(2012) O teste de significância da regressão é um teste para determinar se existe uma relação linear entre a resposta y e qualquer uma das variáveis preditoras x_1, x_2, \dots, x_k . Esse procedimento geralmente é considerado um teste geral ou global de adequação do modelo.

As hipóteses apropriadas são:

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = \dots = \beta_k = 0 \\ H_1 &: \beta_j \neq 0 \text{ para pelo menos um } j \end{aligned}$$

A rejeição desta hipótese nula implica que pelo menos um dos preditores x_1, x_2, \dots, x_k contribui significativamente para o modelo.

O procedimento de teste é uma generalização da análise de variância usada em regressão linear. A soma total dos quadrados SQT é particionada em uma soma dos quadrados devido à regressão, SQR , e uma soma residual dos quadrados, SQE . Por isso,

$$SQ_{Total} = SQ_{Reg} + SQ_{Res}$$

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.8)$$

em que:

- SQ_{Total} é a variabilidade total dos dados (corrigida pela média);
- SQ_{Reg} é a variabilidade dos dados explicada pela regressão;
- SQ_{Res} é a variabilidade dos dados não explicada pela regressão (variação residual).

Dessa forma, quanto maior SQ_{Reg} em detrimento a SQ_{Res} , maior a parcela da variação total dos dados explicada pela regressão. Associado a cada componente dessa composição temos:

- $n - 1$ graus de liberdade para SQ_{Total} (perda de um grau devido a estimação da média);
- k graus de liberdade para SQ_{Reg} (perda devido a estimação de k coeficientes β)
- $n - k - 1$ graus de liberdade para SQ_{Res} .

Por definição a estatística F

$$F_0 = \frac{\frac{SQ_{Reg}}{k}}{\frac{SQ_{Res}}{n-k-1}} = \frac{QM_{Reg}}{QM_{Res}} \quad (2.9)$$

segue distribuição $F_{k,n-k-1}$

O procedimento de teste geralmente é resumido em uma tabela de análise de variância, como

Tabela 2.1.2: Análise de variância para significância de regressão

Fonte de variação	Soma de quadrados	Graus de liberdade	Quadrado médio	F
Regressão	SQ_{Reg}	k	QM_{Reg}	$\frac{QM_{Reg}}{QM_{Res}}$
Residual	SQ_{Res}	$n - k - 1$	QM_{Res}	
Total	SQ_{Total}	$n - 1$		

Para testar a hipótese $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$, calcula-se a estatística de teste F_0 e rejeita-se H_0 se:

$$F_0 > F_{\alpha,k,n-k-1}$$

Teste de significância dos coeficientes de regressão

Uma vez que tenhamos determinado que pelo menos um dos regressores é importante, a questão torna-se qual (s) dele (s). Adicionar uma variável a um modelo de regressão sempre faz com que a soma dos quadrados para a regressão aumente e a soma dos quadrados

residuais diminua. Devemos decidir se o aumento na soma dos quadrados da regressão é suficiente para justificar o uso do regressor adicional no modelo. A adição de um regressor também aumenta a variância do valor ajustado \hat{y} , então devemos ter cuidado para incluir apenas regressores que são de valor real para explicar a resposta. Além disso, adicionar um regressor sem importância pode aumentar o quadrado médio residual, o que pode diminuir a utilidade do modelo (Montgomery et. al., 2012).

As hipóteses para testar individualmente a significância dos coeficientes de regressão, como β_j , são:

$$H_0 : \beta_j = 0, H_1 : \beta_j \neq 0 \quad (2.10)$$

Se $H_0 : \beta_j = 0$ não é rejeitada, indica então que a variável preditora x_j pode ser removida do modelo. A estatística do teste para a hipótese é:

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{ij}}} \quad (2.11)$$

Onde C_{ij} é o elemento na diagonal de $(\mathbf{X}^T \mathbf{X})^{-1}$ correspondente a $\hat{\beta}_j$. A hipótese nula $H_0 : \beta_j = 0$ é rejeitada se $|t_0| > t_{\alpha/2, n-k-1}$. Nota que trata-se de teste marginal ou parcial porque o coeficiente de regressão $\hat{\beta}_j$ depende de todas as variáveis predictoras $x_i (i \neq j)$ que encontram-se no modelo. Por isso trata-se de um teste de contribuição de x_j dado outras variáveis regressoras presentes no modelo.

Adequação do modelo

De acordo com Montgomery et al. (2012) as principais suposições que fizemos até agora em nosso estudo de análise de regressão são como segue:

1. A relação entre a resposta y e as variáveis predictoras é linear, pelo menos aproximadamente;
2. O termo de erro ε tem média zero;
3. O termo de erro ε tem variância constante σ^2 ;
4. Os erros não estão correlacionados;
5. Os erros são normalmente distribuídos.

Tomados em conjunto, os pressupostos 4 e 5 implicam que os erros são variáveis aleatórias independentes. A suposição 5 é necessária para testes de hipótese e estimativa de intervalo.

Devemos sempre considerar a validade dessas suposições como duvidosa e conduzimos análises para examinar a adequação do modelo que cogitamos provisoriamente. Violações

grosseiras das suposições podem produzir um modelo instável no sentido de que uma amostra diferente poderia levar a um modelo totalmente diferente com conclusões opostas. Geralmente não podemos detectar desvios do subjacente suposições por exame das estatísticas de resumo padrão, como as estatísticas t ou F , ou R^2 . Estas são propriedades do modelo “globais” e, como tal, não garantem adequação do modelo.

Análise residual

Os resíduos podem ser definidos como:

$$e_i = y_i - \hat{y}_i, i = 1, 2, \dots, n \quad (2.12)$$

onde y_i é uma observação e \hat{y}_i é o valor ajustado correspondente. Como um resíduo pode ser visto como o desvio entre os dados e o ajuste, é também uma medida de variabilidade na variável resposta não explicada pelo modelo de regressão.

É também conveniente pensar nos resíduos como os valores realizados ou observados do erros de modelo. Assim, quaisquer desvios das suposições sobre os erros devem mostrar nos residuais. A análise dos resíduos é uma maneira eficaz de descobrir vários tipos de inadequações do modelo. Plotar os resíduos é um método muito eficaz para investigar o quão bem o modelo de regressão se ajusta aos dados e verificar a suposições listadas acima.

Os resíduos têm várias propriedades importantes. Eles têm média zero e sua variância média aproximada é estimada por

$$\frac{\sum_{i=1}^n (e_i - \bar{e})^2}{n - p} = \frac{\sum_{i=1}^n e_i^2}{n - p} = \frac{SS_{Res}}{n - p} = QM_{Res}$$

Resíduos escalonados são úteis em encontrar observações que são outliers, ou valores extremos, isto é, observações que são separados de alguma forma do resto dos dados.

- Resíduos padronizados, Como a variância média aproximada de um resíduo é estimado por QM_{Res} , uma escala lógica para os resíduos seria a padronização de resíduos;
- Resíduos estudentizados, Usando QM_{Res} como a variância do i – *esimo* resíduo, e_i é apenas uma aproximação. Podemos melhorar a escala residual dividindo e_i pelo exato desvio padrão do i – *esimo* resíduo;

Gráficos residuais são úteis para investigar os pressupostos de análise de regressão, desde o gráfico qq-normal para normalidade, resíduos vs \mathbf{X} para linearidade e o pressuposto da variância constante,

2.1.2 Regressão Logística

Segundo Hair et. al., (2009) A regressão logística é uma forma especializada de regressão que é formulada para prever e explicar uma variável categórica binária (dois grupos), e não uma medida dependente métrica. A forma da variável estatística de regressão logística é semelhante à da variável estatística da regressão múltipla. A variável estatística representa uma relação multivariada com coeficientes como os da regressão indicando o impacto relativo de cada variável preditora.

A natureza binária da variável dependente (0 ou 1) tem propriedades que violam as suposições da regressão múltipla. Primeiro, o termo de erro de uma variável discreta segue a distribuição binomial ao invés da normal, invalidando assim todos os testes estatísticos que se sustentam nas suposições de normalidade. Segundo, a variância de uma variável dicotômica não é constante, criando casos de heteroscedasticidade também. Além disso, nenhuma violação pode ser remediada por meio de transformações das variáveis dependente ou independentes.

A regressão logística foi desenvolvida para lidar especificamente com essas questões. Não obstante, sua relação única entre variáveis dependente e independentes exige uma abordagem um tanto diferente para estimar a variável estatística, avaliar adequação de ajuste e interpretar os coeficientes, quando comparada com regressão múltipla.

Hosmer e Lemeshow,(2000) afirmam que o que distingue o modelo de regressão logística do linear é que a variável resposta da regressão logística é *binária*. Essa diferença entre regressão logística e linear é refletida tanto na escolha do modelo paramétrico bem como os pressupostos. Uma vez levada em consideração essa diferença, os métodos empregues em uma análise usando a regressão logística seguem os mesmos princípios gerais usados na regressão linear. Portanto, as técnicas usadas na análise de regressão linear irão motivar a abordagem de regressão logística.

Em sua forma tradicional, a regressão logística binária consiste em um modelo que relaciona uma ou um conjunto de p variáveis independentes x_1, x_2, \dots, x_p a uma variável dependente Y que assume apenas dois possíveis estados, 0 ou 1, ou seja, insucesso e sucesso.

Em qualquer problema de regressão a quantidade chave é o valor médio da variável resposta dado o valor da variável independente. Essa quantidade é chamada de média condicional e é expressa como " $E(Y|x)$ " onde Y denota a variável resposta e x denota o valor da variável independente. A quantidade " $E(Y|x)$ " lê-se "*o valor esperado de Y dado o valor de x.*" Na regressão linear assume-se que a média pode ser expressa como uma equação linear de x (ou alguma transformação de x ou Y), como

$$E(Y|x) = \beta_0 + \beta_1 x$$

Essa expressão implica que é possível que $E(Y|x)$ assumira qualquer valor uma vez que os valores de x variam de $-\infty$ e $+\infty$. Com dados binários, a média condicional deve ser maior

ou igual a zero e menor ou igual a 1 i.e $[0 \leq E(Y|x) \leq 1]$.

Muitas funções de distribuição foram propostas para o uso em análise de variáveis resposta binária. Existem duas principais razões para a escolha da *distribuição logística*. Primeiro, do ponto de vista matemático, é uma função extremamente flexível e fácil de usar, e segundo, interpretação significativa. É usada a notação $\pi(x)$ para representar $E(Y|x)$ quando a regressão logística é usada. A forma específica de modelo de regressão logística é:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (2.13)$$

A transformação de $\pi(x)$ que é central para o estudo da regressão logística é a *transformação logit*. A transformação é definida, em termos de $\pi(x)$, como:

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x$$

A importância dessa transformação é que $g(x)$ têm várias propriedades desejadas de um modelo de regressão linear. O *logit*, $g(x)$, é linear nos seus parâmetros, pode ser contínuo, e pode variar de $-\infty$ a $+\infty$ dependendo do intervalo de x .

No modelo de regressão linear, assume-se que uma observação da variável resposta pode ser expressa como $y = E(Y|x) + \varepsilon$. A quantidade ε é chamada de *erro* e expressa o desvio de uma observação da média condicional. O pressuposto mais comum é que ε segue uma distribuição normal com média zero e alguma variância que é constante entre os níveis da variável dependente. Segue que a distribuição condicional da variável resposta dado x será normal com média $E(Y|x)$, e variância constante. O que não é o caso com variáveis resposta binárias. Nesse caso pode-se expressar o valor da variável resposta dado x como:

$$y = \pi(x) + \varepsilon$$

Aqui ε pode assumir um de dois possíveis valores. Se $y = 1$ então $\varepsilon = 1 - \pi(x)$ com probabilidade $\pi(x)$, e se $y = 0$ então $\varepsilon = -\pi(x)$ com probabilidade $1 - \pi(x)$. Então ε tem uma distribuição com média zero e variância igual a $\pi(x)[1 - \pi(x)]$, e segue uma distribuição binomial com probabilidade dada pela média condicional, $\pi(x)$.

De acordo com Agresti (2007) muitas variáveis resposta categóricas têm apenas duas categoria. Denote uma variável resposta binária por Y e seus dois resultados possíveis por $\mathbf{1}$ (“sucesso”) e $\mathbf{0}$ (“fracasso”).

A distribuição de Y é especificada pelas probabilidades $P(Y = 1) = \pi$ de sucesso e $P(Y = 0) = (1 - \pi)$ de falha. Sua média é $E(Y) = \pi$. Para n observações independentes, o número de sucessos tem a distribuição binomial especificada pelo índice n e parâmetro π .

As relações entre $\pi(x)$ e x são geralmente não lineares em vez de lineares. Uma mudança fixa em x pode ter menos impacto quando π está próximo de 0 ou 1 do que quando π está

próximo do meio de intervalo.

Na prática, $\pi(x)$ geralmente aumenta ou diminui continuamente conforme x aumenta. As curvas em forma de S exibidas na figura 2.1.2 geralmente são formas realistas para o relacionamento. A função matemática mais importante com esta forma tem a fórmula:

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

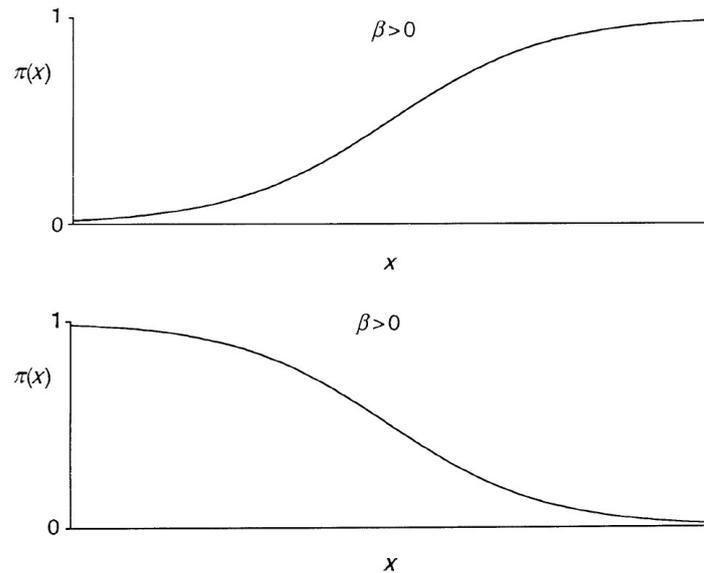


Figura 2.1.1: Funções de regressão logística

De acordo com Agresti,(2007), os modelos de regressão logística podem ser classificados segundo o número de variável(is) independente(s). Quando a variável independente é única sendo esta qualitativa ou quantitativa então pode se estabelecer um Modelo de Regressão Logística Simples, mas se o modelo contiver mais de uma variáveis qualitativa ou quantitativa o modelo passa ser Modelo de Regressão Logística Múltiplo.

Estimação do modelo de regressão logístico

Supondo uma amostra de n observações independentes de par $(x_i, y_i), i = 1, 2, \dots, n$ onde y_i denota o valor de uma variável resposta binária e x_i o valor da variável independente para a i -ésima observação. Assumindo que a variável resposta tenha sido codificada em 0 ou 1, representando a ausência ou presença de determinada característica, respetivamente.

Para ajustar um modelo de regressão logística 2.13 a um conjunto de dados requer a estimativa dos valores de β_0 e β_1 (Hosmer e Lemeshow, 2000).

De acordo com Hosmer e Lemeshow,(2000) de forma geral o método de máxima verosimilhança produz valores dos parâmetros β que maximizam a probabilidade de obter os valores observados no conjunto de dados. De forma a aplicar o método é necessário, primeiro, construir a função de verosimilhança. Essa função expressa a probabilidade dos dados observados em função dos parâmetros desconhecidos. Os *estimadores de máxima verosimilhança* desses parâmetros são escolhidos por serem os valores que maximizam a função.

Se Y é codificado em 0 ou 1 então a expressão para $\pi(x)$ dada em 2.13 providencia a probabilidade condicional que Y é igual a 1 dado x . Será denotado como $P(Y = 1|x)$. Segue que $1 - \pi(x)$ oferece a probabilidade condicional de Y igual a zero dado x , $P(Y = 0|x)$. Por isso, para os pares (x_i, y_i) , onde $y_i = 1$, a contribuição para a função de verosimilhança é $\pi(x)$ e para os pares onde $y_i = 0$, a contribuição para a função de verosimilhança é $1 - \pi(x)$. Uma forma conveniente de expressar a contribuição do par (x_i, y_i) é através de:

$$\pi(x_i)^{y_i} [1 - \pi(x)]^{1-y_i} \quad (2.14)$$

Uma vez que assume-se que as observações são independentes, a função de verosimilhança é obtida através do produto dos termos dados pela expressão 2.14 :

$$L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x)]^{1-y_i} \quad (2.15)$$

O princípio de máxima verosimilhança afirma que é usada a estimativa de β que maximiza 2.15 . Entretanto, é mais fácil trabalhar com o *log* da equação 2.15 . Essa expressão, *log de verosimilhança* é definida como

$$L(\beta) = \ln[L(\beta)] = \sum_{i=1}^n \left\{ y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)] \right\} \quad (2.16)$$

Para encontrar o valor de β que maximiza $L(\beta)$ encontra-se a diferencial de $L(\beta)$ em ordem de β_0 e β_1 e iguala-se a zero as expressões resultantes.

Essas equações são conhecidas como *equações de verosimilhança*

$$\sum [y_i - \pi(x_i)] = 0 \quad (2.17)$$

e

$$\sum x_i [y_i - \pi(x_i)] = 0 \quad (2.18)$$

O valor de β dado pela solução das equações 2.17 e 2.18 é chamada de estimativa de máxima verosimilhança e será denotada por $\hat{\beta}$.

Teste de significância dos coeficientes do modelo logístico

Razão de verosimilhança Em regressão logística, comparação dos valores observados para os valores previstos é baseada na função log de verosimilhança definida na equação 2.16. Ajuda conceptualmente a pensar em um valor observado da variável resposta como sendo também um valor previsto resultante de um modelo saturado. Um modelo saturado é aquele que contém tantos parâmetros como observações.

A comparação dos valores observados para os previstos usando a função de verosimilhança é baseada na seguinte expressão:

$$D = -2 \ln \left[\frac{\text{verosimilhança do modelo ajustado}}{\text{verosimilhança do modelo saturado}} \right] \quad (2.19)$$

Designada por razão de verosimilhança, o respetivo teste é chamado de *teste de razão de verosimilhança*. Usando a equação 2.16, temos:

$$D = -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{\pi}_i}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}_i}{1 - y_i} \right) \right] \quad (2.20)$$

onde $\hat{\pi}_i = \pi(\hat{x}_i)$.

A estatística D , tem papel similar que a soma dos quadrados dos resíduos (SQ_{Res}) na regressão linear.

Com o propósito de avaliar a significância de uma variável independente compara-se o valor da estatística D com e sem a variável independente na equação. A variação em D devido a inclusão de uma variável independente no modelo é obtida como:

$$G = D(\text{modelo sem a variável}) - D(\text{modelo com a variável})$$

A estatística tem papel similar com o numerador do teste parcial F na regressão linear. Como a verosimilhança do modelo saturado é comum para ambos valores de D sendo diferenciado para computar G , pode ser expresso como

$$G = -2 \ln \left[\frac{(\text{verosimilhança sem a variável})}{(\text{verosimilhança com a variável})} \right] \quad (2.21)$$

Para o caso específico de uma única variável independente, é fácil mostrar que quando a variável não está no modelo, a estimativa de máxima verosimilhança de β_0 é $\ln(n_1/n_0)$ onde $n_1 = \sum y_i$ e $n_0 = \sum(1 - y_i)$ e o valor previsto é constante, n_1/n . Nesse caso o valor de G é:

$$G = -2 \ln \left[\frac{\binom{n_1}{n}^{n_1} \binom{n_0}{n}^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{1-y_i}} \right] \quad (2.22)$$

Sob a hipótese de β_1 é igual a zero, a estatística G segue a distribuição de *Qui-quadrado* (χ) com 1 grau de liberdade.

Teste de Wald O teste de Wald é obtido pela comparação da estimativa de máxima verosimilhança do parâmetro de declive, $\hat{\beta}_1$ pela estimativa de seu erro padrão. A razão resultante, sob a hipótese que $\beta_1 = 0$, irá seguir a distribuição normal padrão.

$$W_j = \frac{\hat{\beta}_j}{\text{var}(\hat{\beta}_j)} \cap N(0, 1) \quad (2.23)$$

Avaliação da qualidade de ajuste do modelo de estimação

Segundo Hair et. al.,(2009) Hosmer e Lemeshow desenvolveram um teste de classificação no qual os casos são primeiramente divididos em aproximadamente 10 classes iguais. Em seguida, os números de eventos reais e previstos são comparados em cada classe com a estatística qui-quadrado. Esse teste fornece uma medida ampla de precisão preditiva que é baseada não no valor de verossimilhança, mas sim na real previsão da variável dependente.

O uso apropriado desse teste requer um tamanho de amostra de pelo menos 50 casos para garantir que cada classe tenha pelo menos cinco observações e geralmente até mesmo uma amostra maior, uma vez que o número de eventos previstos nunca fica abaixo de 1. Além disso, a estatística qui-quadrado é sensível a tamanho da amostra, permitindo assim que essa medida encontre diferenças muito pequenas, estatisticamente significantes, quando o tamanho da amostra se torna grande.

$$\begin{aligned} H_0 &: \text{Valores observados} = \text{Valores esperados} \\ H_1 &: \text{Valores observados} \neq \text{Valores esperados} \end{aligned} \quad (2.24)$$

2.1.3 Regressão logística múltipla

De acordo com Hosmer e Lemeshow (2000) como no modelo de regressão linear, a força da técnica está na sua capacidade de modelar inúmeras variáveis, algumas das quais podem estar em diferentes escalas de medida. O modelo logístico múltiplo é uma extensão do modelo simples para o caso de mais de uma variável independente, referida como caso multivariado.

O modelo de regressão logística múltipla

Hosmer e Lemeshow,(2000) levam a considerar uma coleção de p variáveis independentes denotadas pelo vetor $\mathbf{x}^T = (x_1, x_2, \dots, x_p)$. A probabilidade condicional que a resposta é presente é denotada por $P(Y = 1|x) = \pi(x)$. O *logit* do modelo de regressão logística múltipla é dada pela equação

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p, \quad (2.25)$$

que no caso o modelo de regressão logística é

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad (2.26)$$

Se algumas das variáveis independentes são discretas ou de escala nominal, é inapropriado incluir no modelo como se de variáveis de escala intervalar se tratassem. Os números usados para representar os diferentes níveis dessas variáveis de escala nominal são apenas identificadores, e não têm significado numérico. Nessa situação a escolha recai no uso de variáveis *dummy*.

De forma geral, se uma variável de escala nominal tem k possíveis valores, então $k - 1$ variáveis *dummy* serão necessárias.

Supondo que a j -ésima variável x_j tenha k_j níveis. As $k_j - 1$ variáveis *dummy* serão denotadas como D_{jl} e os coeficientes para essas variáveis serão denotadas como $\beta_{jl}, l = 1, 2, \dots, k_j - 1$. Por isso o *logit* para o modelo com p variáveis e a j -ésima variável sendo discreta será

$$g(x) = \beta_0 + \beta_1 x_1 + \dots + \sum_{l=1}^{k_j-1} \beta_{jl} D_{jl} + \beta_p x_p.$$

Estimação do modelo de regressão logística múltipla

De acordo com Hosmer e Lemeshow,(2000) assumindo que temos uma amostra de n observações independentes $(x_i, y_i), i = 1, 2, \dots, n$. Assim como no caso univariado, ajuste do modelo requer a obtenção do vector de estimativas $\beta^T = (\beta_0, \beta_1, \dots, \beta_p)$.

O método de estimação será o de máxima verossimilhança. A função de verossimilhança é quase idêntica ao dado em 2.15 com a única alteração sendo que $\pi(x)$ é agora definido na equação 2.26. Serão obtidas $p + 1$ equações de verossimilhança diferenciando a função log de verossimilhança em ordem dos $p + 1$ coeficientes. As equações resultantes podem ser expressas como:

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

e

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

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

Para estimar os parâmetros da regressão logística múltiplo por máxima verosimilhança encontra-se o valor de β que maximiza $L(\beta)$, o que exige um processo iterativo e que faz necessário derivar $L(\beta)$ em relação a cada parâmetro:

$$\frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=1}^n \left[y_i x_{ij} - \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}} x_{ij} \right] \quad (2.27)$$

Onde $L(\beta)$ é a função 2.16.

A matriz de covariância dos coeficientes estimados é obtida a partir das derivadas parciais de segunda ordem do logaritmo da função de verosimilhança:

$$\frac{\partial^2 \ln[L(\beta)]}{\partial \beta_j^2} = - \sum_{i=1}^n [x_{ij}^2 \pi_i (1 - \pi_i)] \quad (2.28)$$

$$\frac{\partial^2 \ln[L(\beta)]}{\partial \beta_j \partial \beta_k} = - \sum_{i=1}^n [x_{ij} x_{ik} \pi_i (1 - \pi_i)] \quad (2.29)$$

Onde, $j, k=0,1,2,\dots,p$ e π_i representa $\pi(\mathbf{x}_i)$.

Se for formada uma matriz quadrada de dimensão $(p+1)$, constituída pelo simétrico dos valores médios dos termos referidos nas equações 2.28 e 2.29, obtém-se $I(\beta)$, a chamada Matriz de Informação.

A matriz de informação de Fisher é dada por:

$$I(\beta) = E \left(- \frac{\partial^2 L(\beta)}{\partial \beta \partial \beta^T} \right) = X^T Q X \quad (2.30)$$

em que, $Q = \text{diag}[\pi_i(1 - \pi_i)]$, $i = 1, \dots, n$ e X a matriz de dados e a sua inversa $[I(\beta)^{-1}]$, é a matriz de variância e covariância das estimativas de máxima verosimilhança dos parâmetros.

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

Segundo Hair et. al.,(2009) A regressão logística tem uma única variável estatística composta de coeficientes estimados para cada variável independente – como na regressão múltipla. Tal variável estatística é estimada de uma maneira diferente. A regressão logística deriva seu nome da transformação logit usada com a variável dependente, criando diversas diferenças no processo de estimação.

Interpretação dos parâmetros do modelo

Hosmer e Lemeshow,(2000) advogam que na regressão logística, o interesse não só consiste em modelar a probabilidade de ocorrência de um dado evento de interesse em função de um conjunto de variáveis explicativas, mas também em compreender quais dessas variáveis contribuem para a variação da tal probabilidade e em quanto é essa variação. Essa análise é feita com base no conceito de razão de chances, que é uma medida de intensidade de associação que proporciona informação sobre a probabilidade de ocorrência do evento de interesse quando sob influência de uma dada variável explicativa.

De acordo com Agresti,(2002), para uma probabilidade de sucesso π , a chance é definida, de acordo com a equação 2.31, como sendo:

$$g(x) = \frac{\pi(x)}{1 - \pi(x)} \quad (2.31)$$

A interpretação dos parâmetros de um modelo de regressão logística é obtida comparando a probabilidade de sucesso com a probabilidade de fracasso, usando a função de Chances. Assim, ao se tomar dois valores distintos da variável explicativa, x_i e x_{i+1} , obtém-se:

$$OR = \frac{g(x_{j+1})}{g(x_j)} = \frac{e^{\beta_0 + \beta_1 x_{j+1}}}{e^{\beta_0 + \beta_1 x_j}} \quad (2.32)$$

Tem-se ainda que:

$$\ln(OR) = \ln \left[\frac{g(x_{j+1})}{g(x_j)} \right] = \ln[g(x_{j+1})] - \ln[g(x_j)] = \beta_0 + \beta_1 x_{j+1} - \beta_0 - \beta_1 x_j = \beta_1(x_{j+1} - x_j) \quad (2.33)$$

Admitindo que $(x_{j+1} - x_j) = 1$, então:

$$\ln(OR) = \ln(e^{\beta_1}) = \beta_1 \quad (2.34)$$

Assim, obtém-se o quão provável o resultado ocorrerá entre os indivíduos x_{j+1} em relação aos indivíduos x_j , fazendo, portanto, algumas análises:

$$\beta_j > 0 \implies OR > 1 \implies \pi_i(x_{j+1}) > \pi_i(x_j)$$

$$\beta_j < 0 \implies OR < 1 \implies \pi_i(x_{j+1}) < \pi_i(x_j)$$

2.2 Avaliação do papel da mídia na divulgação de comportamentos preventivos sobre a COVID19

O artigo escrito e publicado por Alshareef et. al.,(2021) tinha como objetivo investigar o impacto de diferentes fontes de informação sobre COVID-19 nas atitudes e práticas do público em geral na Arábia Saudita. O estudo utilizou uma pesquisa transversal online para coletar dados de participantes na Arábia Saudita. Dados coletados foram analisados por meio de análises descritivas, bivariadas e por meio da aplicação da regressão logística multivariada.

O estudo constatou que a maioria dos participantes confiou nas mídias sociais como sua principal fonte de informação sobre COVID-19. No entanto, os participantes que confiaram em fontes de mídia tradicionais, como televisão e jornais, foram mais propensos a relatar atitudes positivas em relação às medidas preventivas para COVID-19 e a praticar essas medidas.

Além disso, o estudo constatou que os participantes que confiaram em fontes oficiais do governo sobre COVID-19 foram mais propensos a ter atitudes positivas em relação às medidas preventivas para COVID-19 e a praticar essas medidas. O estudo sugere que as fontes oficiais do governo podem desempenhar um papel importante na moldagem das atitudes e comportamentos públicos durante emergências de saúde pública.

Além disso, o estudo constatou que os participantes que relataram níveis mais altos de ansiedade e medo relacionados ao COVID-19 eram menos propensos a praticar medidas preventivas. O estudo destaca a necessidade de estratégias eficazes de comunicação de risco que não apenas forneçam informações precisas, mas também abordem o impacto emocional e psicológico da pandemia sobre o público em geral.

Em conclusão, o estudo enfatiza a importância de estratégias de comunicação de risco direcionadas e eficazes que levem em conta as diversas fontes de informação sobre COVID-19 e o impacto emocional e psicológico da pandemia sobre o público em geral na Arábia Saudita.

Um estudo por Wang et. al.,(2020) teve como objetivo identificar as fontes de informação relacionadas à COVID-19 mais utilizadas em Taiwan e avaliar a relação entre o uso dessas fontes de informação e a confiança das pessoas em lidar com a COVID-19. Os dados foram

coletados por meio de uma pesquisa no *Facebook*.

Os resultados indicaram que a maioria dos participantes confiou nas informações fornecidas pelo Centro de Controle de Doenças de Taiwan e em fontes de notícias tradicionais, como jornais e televisão. Por outro lado, as redes sociais foram consideradas menos confiáveis para obter informações sobre a COVID-19.

Além disso, os participantes que usaram fontes de informações confiáveis sobre a COVID-19 foram mais propensos a relatar maior confiança em sua capacidade de lidar com a pandemia. Por outro lado, aqueles que confiavam mais em informações imprecisas ou em fontes menos confiáveis sobre a COVID-19 tinham menos confiança em sua capacidade de lidar com a pandemia.

Os resultados do estudo sugerem a importância de fontes de informação confiáveis para aumentar a confiança das pessoas em sua capacidade de lidar com a COVID-19. Além disso, os autores enfatizam a importância de educar o público sobre a importância de verificar a veracidade das informações antes de partilhá-las ou agir com base nelas.

Em resumo, o estudo destaca a importância de fontes de informação confiáveis para aumentar a confiança das pessoas em lidar com a COVID-19 em Taiwan e destaca a necessidade de educação sobre a verificação de informações antes de partilhá-las ou agir com base nelas.

Estudo realizado por Erinoso et. al.,(2021) é um estudo transversal online que investigou os preditores das fontes de informação sobre a COVID-19 na Nigéria e percepção de sua precisão. O estudo teve como objetivo compreender como os nigerianos obtiveram informações sobre a COVID-19 e identificar os factores que influenciaram sua confiança nessas fontes.

O estudo utilizou uma abordagem de questionário online para coletar dados de uma amostra representativa de indivíduos na Nigéria. Os participantes foram questionados sobre suas principais fontes de informação sobre a COVID-19 e avaliaram a precisão percebida dessas fontes. Além disso, foram investigados os factores demográficos e socioeconómicos que poderiam influenciar as escolhas das fontes de informação e a percepção de sua precisão.

Os resultados do estudo revelaram que a televisão foi a principal fonte de informação sobre a COVID-19 na Nigéria, seguida pelas redes sociais e os jornais. Além disso, a percepção de precisão variou entre as diferentes fontes, com os profissionais de saúde e os órgãos governamentais sendo considerados as fontes mais confiáveis. Por outro lado, as redes sociais foram percebidas como menos precisas.

Quanto aos factores preditores, o estudo encontrou uma associação significativa entre o nível de educação e o uso de fontes de informação específicas. Os participantes com maior nível de educação tinham maior probabilidade de utilizar fontes como a internet e os jornais, enquanto aqueles com menor nível de educação tinham maior probabilidade de confiar nas redes sociais. Além disso, a renda e a ocupação também mostraram influenciar a escolha das fontes de informação.

Em conclusão, o estudo destacou a importância da televisão, das redes sociais e dos jornais como fontes de informação sobre a COVID-19 na Nigéria. Além disso, identificou factores

demográficos e socioeconómicos como preditores significativos das escolhas das fontes de informação e da percepção de sua precisão. Esses resultados podem ajudar os formuladores de políticas e os profissionais de saúde a desenvolver estratégias eficazes de comunicação e divulgação de informações precisas sobre a COVID-19 na Nigéria.

Capítulo 3

Material e Métodos

3.1 Material

O trabalho é de carácter transversal, em que os dados são coletados em um único momento no tempo, sem acompanhamento ao longo do tempo, e os dados que sustentam o estudo são de fonte secundária e são resultado de um inquérito administrado a 768 residentes da Cidade de Quelimane, Província da Zambézia na região centro de Moçambique, no mês de Julho de 2022, cujo principal objetivo passava por estudar os comportamentos sócio-culturais e padrões de movimento da população de modo a ajudar os órgãos competentes na tomada de decisões locais sobre as medidas da mitigação e contenção da propagação da pandemia da COVID-19 em Moçambique.

Para o presente trabalho o foco incide sobre a componente da comunicação social e da forma como o papel da mídia na promoção de medidas preventivas é percebida no seio da comunidade.

A população incluía todos residentes da cidade de Quelimane, tendo como critério de inclusão a presença na residência no momento da entrevista, membros dos agregados não presentes durante as entrevistas seriam excluídos do estudo.

Durante a realização do inquérito questões sobre informações dos participantes, sobre a percepção dos mesmos sobre o papel da mídia na divulgação e promoção de medidas preventivas contra a COVID, e sobre os comportamentos sócio-culturais foram levantadas.

Abaixo apresentam-se as variáveis alvo do estudo.

Tabela 3.1.1: Descrição das variáveis

#	Descrição	Classificação
1. Informação do Participante		
1.1	Sexo	Categórica
1.2	Estado Civil	Categórica
1.3	Faixa Etária(em anos)	Categórica
1.5	Estado Ocupacional	Categórica
1.6	Nível mais alto de educação	Categórica
2. Mídia		
2.2	Considera que a comunicação social divulga comportamentos preventivos para ajudar no controlo da COVID-19?	Categórica
2.3	Considera que os meios de comunicação social conscientizam sobre o COVID-19?	Categórica
2.8	Com que frequência recebe informação relativa a COVID-19 dos seguintes meios/fontes?	
a	Internet	Categórica
b	Mídia Social	Categórica
c	Mídia Tradicional	Categórica
d	Amigos	Categórica
e	Colegas	Categórica
f	Familiares	Categórica
g	Agentes de saúde	Categórica
h	Órgãos oficiais	Categórica
3. Comportamentos sócio-culturais		
3.10	Já faltou a algum local de frequência habitual por causa da COVID-19?	Categórica

- **Variáveis demográficas**

Dados demográficos dos participantes foram extraídos do inquérito em anexo 5.0.1, nomeadamente o Sexo, Estado civil, Faixa Etária, Estado ocupacional e o Nível académico, de forma a fornecer melhor entendimento sobre as características da amostra.

- **Percepção sobre o papel da mídia**

De forma a medir a percepção sobre o papel da mídia na promoção de comportamentos preventivos foi computada a avaliação média dos participantes com base nas respostas as questões 2.2 e 2.3. em anexo 5.0.3 referenciadas na tabela 3.1.

- **Fonte de obtenção de informações relacionadas a COVID19**

A questão 2.8 em anexo 5.0.3 capta a frequência de uso dos diferentes meios/fontes de obtenção de informações relacionadas a COVID19 por parte dos respondentes.

Nova variável, total de fontes, foi computada como número de principais fontes/meios de informação, contabilizando as fontes/meios em que os participantes responderam como "Quase sempre". Podendo variar na escala de 0 (nenhuma) até 8 (todas fontes/meios).

De forma especial serão analisadas as diferentes relações em comparação com o uso da internet, mídia social e a mídia tradicional de forma independente.

- **Comportamentos sócio-culturais**

A variável 3.10 em anexo 5.0.2 foi selecionada com o intuito de descrever os comportamentos sócio-culturais, especificamente relacionada a adesão a medida de redução de mobilidade durante a vigência do estado de calamidade pública declarada para o controle da pandemia da COVID19;

A variável será também referenciada como *Ficou em casa*.

Com vista a limpeza preliminar da base de dados recorrer-se-á ao pacote *Microsoft Excel* pela facilidade de manipulação de base de dados de pequena dimensão;

Para a produção das análises descritivas recorrer-se-a ao *IBM SPSS Statistics* na versão 25, que é um programa de computador do tipo científico, que inclui: aplicação analítica, mineração de dados e texto e estatísticas que transformam os dados em informações importantes.

Com objetivo de realizar testes sobre os dados e desenvolvimento de modelos estatísticos para responder os objetivos do trabalho, recorrer-se-a a ferramenta estatística *R*, *R* disponibiliza uma ampla variedade de técnicas estatísticas e gráficas, incluindo modelação linear e não linear, testes estatísticos clássicos, análise de séries temporais, classificação, agrupamento e outras.

A *R* é facilmente extensível através de funções e extensões que permitem a integração com outras ferramentas sejam elas de análise estatística como não.

3.2 Métodos

Para o alcance dos objectivos do trabalho serão analisados dados das respostas dos residentes da Cidade de Quelimane em contraste com os grupos em que se inserem. Inicialmente através de uma análise exploratória dos dados com o propósito de examinar os dados previamente à aplicação das técnicas estatísticas; desta forma a alcançar um entendimento preliminar de dados e das relações existentes entre as variáveis analisadas. Técnicas gráficas e tabulares serão aplicadas.

Esta etapa é fundamental, pois uma análise descritiva detalhada permite familiarizar-se com os dados, organizá-los e sintetizá-los de forma a obter as informações necessárias do conjunto de dados para responder as questões em estudo.

3.2.1 Análise descritiva dos dados

Antes de iniciar qualquer tipo de análise estatística, é imperativo fazer um exame preliminar dos dados com três objetivos principais em mente:

1. Verificar para erros e anomalias;
2. Entender a distribuição de cada uma das variáveis por conta própria; e
3. Começar a entender a natureza e força das relações entre as variáveis.

Erros devem, é claro, ser corrigidos, uma vez que mesmo uma pequena percentagem de valores de dados errados pode drasticamente influenciar os resultados. Compreendendo a distribuição das variáveis, especialmente as respostas, é crucial para escolher o apropriado método de regressão. Finalmente, compreender a natureza e a força das relações é o primeiro passo na construção de um modelo estatístico mais formal a partir do qual tirar conclusões (Vittinghoff et. al., 2005).

De acordo com Myers e Randolph, (2013) métodos estatísticos descritivos são usados para resumir um conjunto de dados existente em menos números, tornando os dados mais fáceis de visualizar e entender. Faulkner e Faulkner, (2009) definem métodos estatísticos como “maneiras de organizar, descrever e apresentar dados quantitativos (numéricos) de maneira concisa, gerenciável, e compreensível”. Estatísticas descritivas utilizam métodos estatísticos univariados para examinar e resumir os dados uma variável de cada vez. Podemos calcular valores numéricos que descrevem amostras ou populações.

Valores numéricos que descrevem amostras são chamados de estatísticas, enquanto valores numéricos que descrevem populações são chamados de parâmetros.

Myers e Randolph, (2013) destacam três tipos de métodos estatísticos univariados: 1) distribuições de frequência e percentagem, 2) medidas de tendência central, e 3) medidas de variabilidade.

3.2.2 Teste de independência

Segundo Agresti, (2007), utiliza-se o teste de independência quando os dados da pesquisa se apresentam sob forma de frequências em categorias discretas. Para se aplicar este teste de hipótese, recorre-se à estatística de Qui-quadrado (χ^2) para determinar se duas variáveis são independentes, usando a seguinte estatística:

$$\chi^2 = \sum_{i=1}^l \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi_{(L-1)(C-1)}^2 \quad (3.1)$$

Onde: O_{ij} - número de casos observados na linha “i” da coluna “j”.

E_{ij} = número de casos esperados, na linha “i” da coluna “j”.

Para obter a frequência esperada E_{ij} em cada célula, multiplicam-se os totais marginais comuns a uma determinada célula e divide-se pelo total de casos.

Os valores de χ^2 obtidos pela equação 3.1, tem distribuição aproximadamente Qui-quadrado com $gl = (L - 1)(C - 1)$, onde L é o número de linhas e C é o número de colunas, cujas hipóteses a testar são:

$$\begin{cases} H_0 : \text{As variáveis são independentes.} \\ H_1 : \text{As variáveis são dependentes.} \end{cases} \quad (3.2)$$

Rejeita-se a hipótese nula ao nível de significância α se $\chi^2 > \chi_{(l-1)(c-1)}^2$.

3.2.3 Análise de Regressão

A aplicação da Regressão linear e Regressão Logística no presente trabalho é baseada no intuito de identificar factores que significativamente explicam o uso das diferentes fontes de informação sobre a COVID19, os seus comportamentos e o papel percebido da mídia no processo de promoção de comportamentos preventivos;

Hair et. al.,(2009) advoga que a análise de regressão múltipla, uma forma de modelagem linear geral, é uma técnica estatística multivariada usada para examinar a relação entre uma única variável dependente e um conjunto de variáveis independentes. O ponto de partida necessário na regressão múltipla, como ocorre em todas as técnicas estatísticas multivariadas, é o problema de pesquisa. A flexibilidade e a adaptabilidade da regressão múltipla permitem seu uso em quase toda relação de dependência. Ao selecionar aplicações adequadas de regressão múltipla, o pesquisador deve considerar três questões principais:

1. Adequação do problema de pesquisa;
2. Especificação de uma relação estatística;
3. Seleção das variáveis dependentes e independentes;

No presente trabalho o problema de pesquisa é de explicação dos fenómenos de estudo;

O pesquisador que se defronta com uma variável dependente dicotômica não precisa apelar para métodos elaborados para acomodar as limitações da regressão múltipla, e nem precisa ser forçado a empregar a análise discriminante, especialmente se suas suposições estatísticas são violadas. A regressão logística aborda esses problemas e fornece um método desenvolvido para lidar diretamente com essa situação da maneira mais eficiente possível (Hair et. al., 2009).

3.2.4 Seleção de variáveis do modelo

A seleção das variáveis no modelo será feita com base no método *backward*.

De acordo com Montgomery et. al.,(2012) a seleção *forward* começa sem regressores no modelo e tenta inserir variáveis até que um modelo adequado seja obtido. O procedimento

backward tenta encontrar um bom modelo trabalhando na direção oposta. Isto é, começamos com um modelo que inclui todos os K regressores candidatos. Então a estatística F parcial (ou equivalentemente, uma estatística t) é calculada para cada regressor como se fosse a última variável a entrar no modelo.

A menor dessas estatísticas parciais F (ou t) é comparado com um valor pré-selecionado, F_{fora} (ou t_{fora}), por exemplo, e se o menor F parcial (ou t), é menor que F_{fora} (ou t_{fora}), esse regressor é removido do modelo. Agora um modelo de regressão com $K - 1$ regressores é adequado, as estatísticas parciais F (ou t) para este novo modelo calculado e o procedimento repetido. O procedimento *backward* termina quando o menor valor F parcial (ou t) não é menor do que o valor de corte pré-selecionado F_{fora} (ou t_{fora}).

O algoritmo *backward* costuma ser um procedimento de seleção de variável muito bom. Isso é particularmente favorecido por analistas que gostam de ver o efeito de incluir todos os candidatos regressores, apenas para que nada "óbvio" seja esquecido.

3.2.5 Significância estatística

De acordo com Hair et. al.,(2009) todas as técnicas multivariadas, exceto análise de agrupamentos e mapeamento perceptual, são baseadas na inferência estatística dos valores ou relações entre variáveis de uma população a partir de uma amostra aleatória extraída daquela população.

Um censo da população inteira torna a inferência estatística desnecessária, pois qualquer diferença ou relação, não importa quão pequena, é "verdadeira" e existe. Entretanto, raramente, ou nunca, um censo é realizado. Logo, o pesquisador é obrigado a fazer inferências a partir de uma amostra.

Interpretar inferências estatísticas requer que o pesquisador especifique os níveis de erro estatístico aceitáveis devido ao uso de uma amostra (conhecidos como erro amostral). A abordagem mais comum é especificar o nível do erro *Tipo I*, também conhecido como alfa (α). O erro Tipo I é a probabilidade de rejeitar a hipótese nula quando a mesma é verdadeira, ou, em termos simples, a chance de o teste exibir significância estatística quando na verdade esta não está presente - o caso de um "falso positivo". Ao especificar um nível alfa, o pesquisador estabelece os limites permitidos para erro, especificando a probabilidade de se concluir que a significância existe quando na realidade esta não ocorre (Hair et. al., 2009).

O erro Tipo II é a probabilidade de não rejeitar a hipótese nula quando na realidade esta é falsa. Uma probabilidade ainda mais interessante é $1 - \beta$, chamada de poder do teste de inferência estatística.

Ainda de acordo com Hair et. al. (2009) a verdadeira magnitude do efeito de interesse na população é denominado tamanho do efeito.

Assumindo tamanho do efeito como moderado e tamanho de amostra suficientemente grande, um nível de α de 5% garante um poder estatístico acima de 0.8 que segundo Hair

et. al. (2009) é desejado.

Capítulo 4

Resultados e Discussão

4.1 Análise Descritiva

A tabela 4.1.1 descreve a amostra em estudo pelas características demográficas. de onde pode-se observar que cerca de 56% da amostra é constituída por indivíduos do sexo feminino. aproximadamente 66% da amostra não se encontra em relacionamento amoroso contra 34% que declarou estar em um relacionamento. A amostra é constituída na sua maioria por indivíduos dos 20 a 29 e 10 a 19 anos com cerca 29% e 27% do total respetivamente. Aproximadamente 37% da amostra é estudante, seguida de cerca de 29% que declarou não ter nenhuma ocupação formal ou informal, apenas 11% revelou ser empregada. Observando pelo nível de educação, ela é representada maioritariamente por indivíduos cujo nível de educação mais alto é o ensino secundário com cerca de 41% e por último apenas 7% aproximados declarou não ter nenhum grau de ensino concluído.

A tabela 4.1.2 apresenta a descrição das variáveis de comportamento, fontes de informação e de percepção, de onde é possível observar que mais de 70% afirmou que com a pandemia da COVID19 deixou de frequentar um lugar habitual.

Quase 40% dos entrevistados afirmou que a mídia tradicional quase sempre foi a fonte de obtenção de informações sobre a COVID19. Cerca de 20% afirmou que quase sempre recorreu a internet como fonte principal de informações sobre a COVID19. De forma similar, aproximadamente 23% afirmou que quase sempre recorreu a mídia social (*Facebook, Twitter, etc*) como fonte de informações contra COVID19.

A tabela mostra que em os participantes reportaram uma média de duas (2) fontes/meios de informação como sendo as suas principais e de maior frequência para obtenção de informações sobre a COVID19, verificou-se grande dispersão dos dados em relação a média.

Verificou-se que a percepção dos participantes em relação a mídia é que ela quase sempre divulga comportamentos preventivos em relação a COVID19, com um escore aproximado de quatro (4), indicando uma percepção positiva sobre o papel da mídia, entretanto, o valor do desvio padrão de quase 1 sugere grande dispersão das repostas em relação a média.

Tabela 4.1.1: Descrição da amostra em estudo

Variável	Estatística
Sexo, n(%)	
Feminino	404 (55.5)
Masculino	324 (45.5)
Estado Civil, n(%)	
Não-relacionamento	479 (65.8)
Em relacionamento	249 (34.2)
Faixa Etária, n(%)	
9 anos ou menos	61 (8.4)
10 a 19	195 (26.8)
20 a 29	210 (28.8)
30 a 39	100 (13.7)
40 a 49	65 (8.9)
50 a 59	51 (7)
60 anos ou mais	46 (6.3)
Estado Ocupacional, n(%)	
Nenhuma	209 (28.7)
Estudante	270 (37.1)
Empregado	82 (11.3)
Sazonal	167 (22.9)
Nível de Educação, n(%)	
Nenhum	48 (6.6)
Alfabetização	53 (7.3)
Primário	268 (36.8)
Secundário	301 (41.3)
Superior	58 (8)

Tabela 4.1.2: Descrição das variáveis em estudo

Variável	Estatística
Deixou de frequentar lugar habitual devido a COVID19, n(%)	
Sim	541 (74.3)
Não	187 (25.7)
Internet, n(%)	
Sim	152 (20.9)
Não	576 (79.1)
Mídia Social, n(%)	
Sim	171 (23.5)
Não	557 (76.5)
Mídia Tradicional, n(%)	
Sim	282 (38.7)
Não	446 (61.3)
Total de fontes, média(desvio padrão)	2.3 (2.2)
Percepção, média(desvio padrão)	3.7 (0.9)

Os dados da tabela 4.1.3 apresentam a percepção do papel da mídia na promoção

de comportamentos preventivos contra a COVID19 pelas diferentes características sócio-demográficas de estudo.

Verificou-se que os grupos dos menores de 10 anos e 50 a 59 anos percebem que a mídia somente as vezes promove tais comportamentos, um escore aproximado de 3 enquanto que os demais grupos acredita que a mídia o faz quase sempre, com escore aproximado de 4.

Ao analisar o factor nível de educação notou-se que os participantes com o nível primário percebem que a mídia as vezes promove atitudes preventivas, com um escore aproximado de 3, contrariamente aos demais grupos que acreditam a mídia quase sempre promove, com escore aproximado de 4.

Participantes que listaram internet, mídia social e tradicional como umas das principais fontes/meios de obtenção de informação contra a COVID 19 percebem que a mídia quase sempre divulga comportamentos preventivos com escores aproximados de 4, contra os que não reportaram as fontes como principais, que percebem que a mídia somente as vezes divulga boas práticas de prevenção e controle da COVID19, com escore aproximado de 3.

A partir da figura 5.0.4 em anexo, pode-se notar que os dados encontram-se distribuídos de forma assimétrica negativamente ou a esquerda, com maiores concentrações em torno da média e junto do valor máximo do escore.

De forma geral e pelos diferentes grupos, em relação a percepção sobre o papel da mídia verificou-se grande dispersão de respostas em relação a média, sugerindo um certo grau de incerteza.

A partir da tabela 4.1.4 verificou-se que os participantes na faixa dos 30 a 39 anos ou com estado ocupacional empregado ou com nível mais alto de educação superior, reportaram em média cerca de 3 fontes principais de obtenção de informações relacionadas a COVID19 contra a média de 2 fontes/meios de informação.

O desvio padrão aproximado de 2 fontes em relação a média sugere grande dispersão de dados. A figura 5.0.5 em anexo, ilustra a distribuição da variável resposta, de onde verificou-se que os dados encontram-se distribuídos de forma assimétrica à direita.

Verificou-se pela tabela 4.1.5 que indivíduos dos 20 a 29 e 50 a 59 recorrem quase sempre a internet como uma das fontes principais de obtenção de informações da COVID19, 27% e 26% respetivamente, quando comparado aos demais grupos etários. Indivíduos que reportaram ser empregados ou estudantes quase sempre recorrem a internet para obtenção de informações, representado por 34% e 22% respetivamente, contra os indivíduos com nenhuma ou ocupação sazonal. Ao observar-se para característica nível de educação verificou-se que os indivíduos de nível superior(38%) em comparação com os demais níveis, recorrem quase sempre a internet para busca de informações relacionadas a COVID. Indivíduos em relacionamento (26%) recorrem mais a internet para busca de informações em relação aos que não se encontram em relacionamento (18%), segundo dados da tabela4.1.5.

Na perspetiva da mídia social, verificou-se que indivíduos dos 20 a 29 (29%) e 30 a 39(28%) anos de idade quase sempre obtêm informações pelas mídias sociais em comparação

Tabela 4.1.3: Percepção do papel da mídia na promoção de atitudes preventivas

Variável	Estatística
Sexo, média(desvio padrão)	
Feminino	3.75 (0.99)
Masculino	3.74 (0.98)
Estado Civil, média(desvio padrão)	
Não-relacionamento	3.70 (0.99)
Em relacionamento	3.83 (0.98)
Faixa Etária, média(desvio padrão)	
9 anos ou menos	3.45 (1.18)
10 a 19	3.69 (0.93)
20 a 29	3.83 (0.93)
30 a 39	3.94 (1)
40 a 49	3.78 (1)
50 a 59	3.56 (0.92)
60 anos ou mais	3.71 (1)
Estado Ocupacional, média(desvio padrão)	
Nenhuma	3.74 (0.93)
Estudante	3.77 (1)
Empregado	3.95 (0.92)
Sazonal	3.61 (1)
Nível de Educação, média(desvio padrão)	
Nenhum	3.84 (1)
Alfabetização	3.93 (1)
Primário	3.48 (1)
Secundário	3.87 (0.89)
Superior	4 (0.8)
Internet, média(desvio padrão)	
Sim	4.4 (0.71)
Não	3.57 (0.97)
Mídia Social, média(desvio padrão)	
Sim	4.49 (0.63)
Não	3.52 (0.96)
Mídia tradicional, média(desvio padrão)	
Sim	4.26 (0.83)
Não	3.42 (0.94)

com demais grupos etários. De forma similar ao uso da internet, verificou-se que indivíduos empregados (35%) quase sempre recorrem a mídia social como fonte de informação em relação aos não empregados. Indivíduos com nível superior (43%) em comparação com os demais níveis de educação, recorrem mais as mídias sociais como fonte de informações. Verificou-se que os indivíduos em relacionamento (29%) contra os que não (20%) quase sempre recorrem a mídia social para obtenção de informações relacionadas a COVID19.

Analisando os dados de frequência de uso da mídia tradicional, verificou-se que os indivíduos em relacionamento(44%) recorrem mais a esta fonte em relação aos indivíduos que

Tabela 4.1.4: Número médio de fontes/meios frequentes de obtenção de informações sobre a COVID19

Variável	Estatística
Sexo, média(desvio padrão)	
Feminino	2.35 (2.32)
Masculino	2.26 (2.15)
Estado Civil, média(desvio padrão)	
Não-relacionamento	2.24 (2.14)
Em relacionamento	2.44 (2.33)
Faixa Etária, média(desvio padrão)	
9 anos ou menos	1.85 (2.27)
10 a 19	2.10 (2.11)
20 a 29	2.5 (2.25)
30 a 39	2.79 (2.45)
40 a 49	2.22 (2.07)
50 a 59	2.18 (2.18)
60 anos ou mais	2.20 (2.46)
Estado Ocupacional, média(desvio padrão)	
Nenhuma	2.02 (2.13)
Estudante	2.38 (2.28)
Empregado	2.9 (2.4)
Sazonal	2.26 (2.21)
Nível de Educação,média(desvio padrão)	
Nenhum	2.23 (2.02)
Alfabetização	2.40 (1.97)
Primário	1.96 (2.18)
Secundário	2.45 (2.35)
Superior	3.17 (2.14)

declararam não relacionamento (36%). Segundo dados da tabela 4.1.5 verificou-se também que 40 a 49 (46%) e 30 a 39 (44%) recorrem mais a mídia tradicional como fonte principal de informações da COVID em relação aos demais grupos etários. Indivíduos empregados (45%) e os sem nenhuma ocupação(45%) formal ou informal quase sempre recorrem a mídia tradicional como uma das principais fontes de informação quando comparados com os estudantes(34%) e os de ocupação sazonal(34%). Os indivíduos com o nível de alfabetização(64%) quase sempre obtêm a informações da COVID19 pelos meios de informação tradicionais quando comparado com indivíduos dos demais níveis de educação.

Verificou-se através dos dados da tabela 4.1.6 que indivíduos em não relacionamento (77%) em comparação com os em relacionamento (69%) deixaram mais de frequentar algum lugar habitual devido a pandemia da COVID19. Observou-se também que indivíduos dos 10 a 19 anos (81%), 60 anos ou mais(80%) e 50 a 59 anos (78%) foram mais propensos a deixar de frequentar pelo menos um lugar habitual em conformidade com as recomendações de redução de mobilidade pelos órgão de comunicação sociais e oficiais. Os estudantes (82%) mais deixaram de frequentar algum lugar habitual devido a COVID19 em comparação

Tabela 4.1.5: Fontes frequentes de obtenção de informações sobre a COVID19

Variável	Internet	Mídia Social	Mídia Tradicional
Sexo, n(%)			
Feminino	87 (21.5)	94 (23.3)	158 (39.1)
Masculino	65 (20.1)	77 (23.8)	124 (38.3)
Estado Civil, n(%)			
Não-relacionamento	87 (18.2)	98 (20.5)	173 (36.1)
Em relacionamento	65 (26.1)	73 (29.3)	109 (43.8)
Faixa Etária, n(%)			
9 anos ou menos	9 (14.8)	13 (21.3)	15 (24.6)
10 a 19	32 (16.4)	36 (18.5)	76 (39)
20 a 29	56 (26.7)	61 (29)	80 (38.1)
30 a 39	19 (19)	28 (28)	44 (44)
40 a 49	13 (20)	13 (20)	30 (46.2)
50 a 59	13 (25.5)	12 (23.5)	18 (35.3)
60 anos ou mais	10 (21.7)	8 (17.4)	19 (41.3)
Estado Ocupacional, n(%)			
Nenhuma	37 (17.7)	45 (21.5)	94 (45)
Estudante	59 (21.9)	67 (24.8)	94 (34.8)
Empregado	28 (34.1)	29 (35.4)	37 (45.1)
Sazonal	28 (16.8)	30 (18)	57 (34.1)
Nível de Educação, n(%)			
Nenhum	11 (22.9)	13 (27.1)	21 (43.8)
Alfabetização	10 (18.9)	12 (22.6)	34 (64.2)
Primário	42 (15.7)	48 (17.9)	92 (34.3)
Secundário	67 (22.3)	73 (24.3)	109 (36.2)
Superior	22 (37.9)	25 (43.1)	26 (44.8)

com demais grupos. Indivíduos que quase nunca ou as vezes recorriam a internet, mídia social ou/e tradicional deixaram mais de frequentar pelo menos um local habitual devido a COVID19 quando comparados com os indivíduos que quase sempre recorriam as mesmas fontes para obtenção de informações da COVID19, sugerindo que maior exposição as mídias influenciou com que os indivíduos não deixassem de frequentar lugares habituais por causa da COVID19.

A figura 5.0.6 em anexo, ilustra que a maior parte do participantes reportou ter deixado de frequentar um lugar habitual por conta da pandemia da COVID19.

Tabela 4.1.6: Deixou de frequentar lugar habitual devido a COVID19

Variável	Estatística
Sexo, n(%)	
Feminino	299 (74)
Masculino	242 (74.7)
Estado Civil, n(%)	
Não-relacionamento	368 (76.8)
Em relacionamento	173 (69.5)
Faixa Etária, n(%)	
9 anos ou menos	40 (65.6)
10 a 19	158 (81)
20 a 29	150 (71.4)
30 a 39	73 (73)
40 a 49	43 (66.2)
50 a 59	40 (78.4)
60 anos ou mais	37 (80.4)
Estado Ocupacional, n(%)	
Nenhuma	144 (68.9)
Estudante	221 (81.9)
Empregado	59 (72)
Sazonal	117 (70.1)
Nível de Educação,n(%)	
Nenhum	31 (64.6)
Alfabetização	25 (47.2)
Primário	207 (77.2)
Secundário	231 (76.7)
Superior	47 (81)
Internet,n(%)	
Sim	104 (19.2)
Não	437 (80.2)
Mídia social,n(%)	
Sim	131 (24.2)
Não	410 (75.8)
Mídia tradicional,n(%)	
Sim	213 (39.4)
Não	328 (60.6)

4.2 Teste de independência das variáveis

Aplicando o teste de Qui-quadrado de pearson sob as variáveis de estudo obteve-se os resultados da tabela 4.2.1 de onde verificou-se:

- A variável sexo é significativamente independente de todas as variáveis resposta;
- De todas variáveis resposta, a variável estado civil mostrou-se independente das variações do número de fontes;

- A variável Faixa etária mostrou-se apenas dependente das variações da questão se deixou de frequentar um lugar habitual devido a frequência devido a COVID19;
- As variações das variáveis respostas estão significativamente dependentes da variável estado ocupacional e também do nível de educação;
- As variáveis número de fontes de informação e deixou de frequentar um lugar habitual devido a COVID19 são significativamente dependentes, segundo dados da tabela 4.2.1

Tabela 4.2.1: Teste de Independência das variáveis

	Internet	Mídia So- cial	Mídia Tra- dicional	Número de fontes	<i>Ficou em casa*</i>
Sexo	0.63	0.87	0.82	0.35	0.83
Estado Civil	0.012	0.007	0.04	0.66	0.031
Faixa Etária	0.178	0.171	0.22	0.46	0.07
Estado Ocupacional	0.008	0.019	0.04	0.021	0.004
Nível de educação	0.004	0.002	0.001	0.011	<0.001
Número de fontes	-	-	-	-	0.077

* : Deixou de frequentar um lugar habitual devido a COVID19

4.3 Percepção sobre o papel da mídia na divulgação de comportamentos preventivos contra a COVID19

Realizou-se a aplicação da análise de regressão linear múltipla com o intuito de identificar os factores associados a percepção pública do papel da mídia na promoção e divulgação de atitudes e comportamentos preventivos em relação a COVID19. Para a seleção das variáveis explicativas recorreu-se ao método *backward*, resultando no modelo que inclui o nível de educação, número de fontes/meios principais de obtenção de informações da COVID19, uso da mídia social e tradicional como uma das fontes principais de informações sobre a COVID19 como factores explicativos da variabilidade na variável resposta.

A tabela 4.3.1 ilustra os resultados da aplicação da regressão linear sobre os dados, de onde constatou-se que existem evidências suficientes para rejeitar a hipótese que tanto o nível de educação, número de fontes, uso da mídia social e tradicional têm efeito nulo sobre a percepção do papel da mídia na divulgação de comportamentos preventivos.

As evidências sugerem que de forma significativa ($p = 0.04$) quando comparado com indivíduos com nenhum nível de educação formal concluído, os indivíduos com o ensino primário como nível mais alto de educação percebem que a mídia divulga atitudes preventivas com menos frequência.

Os dados sugerem ainda que o aumento unitário no número de fontes de informação principais aumenta de forma significativa ($p < 0.001$) também em média 0.11 no escore de percepção da frequência da divulgação de comportamentos preventivos, i.e, o número de fontes está positivamente associado a percepção do papel da mídia.

O uso da mídia social e tradicional como uma das fontes principais de informações da COVID19 aumenta de forma significativa ($p < 0.001$) o escore de percepção do papel da mídia na divulgação em 0.4 e 0.37 respectivamente.

Tabela 4.3.1: Factores associados a percepção sobre o papel da mídia na divulgação de comportamentos preventivos contra a COVID19

Variável	β	EP (β)	$t_{calculado}$	$p-value$
Interceto	3.32	0.12	26.72	< 0.001
Nível de Educação				
Nenhum *	-	-	-	-
Alfabetização	0.01	0.16	0.07	0.93
Primário	-0.25	0.13	-1.98	0.04
Secundário	0.03	0.12	0.29	0.76
Superior	0.06	0.16	0.37	0.71
Total de fontes	0.11	0.02	5.39	< 0.001
Mídia social				
Não *	-	-	-	-
Sim	0.4	0.09	4.30	< 0.001
Mídia tradicional				
Não *	-	-	-	-
Sim	0.37	0.08	4.65	< 0.001

* : Categoria de referência

Com o intuito de avaliar a significância do modelo de regressão linear proposto, foi aplicada a análise de variância (ANOVA), os resultados são apresentados na tabela 4.3.2, onde observou-se que a variabilidade explicada pelo modelo é igual a 50.52 contra 0.7 de variabilidade residual, através da aplicação do teste F obtiveram-se evidências suficientes para rejeitar a hipótese que os efeitos dos factores explicativos presentes no modelo são iguais e nulos. Isto sugere que o modelo com os factores nível de educação, total de fontes de informação, uso da mídia social e tradicional é adequado para explicar a variabilidade na percepção pública sobre o papel da mídia na divulgação de atitudes e comportamentos preventivos e de controlo da COVID19.

Tabela 4.3.2: Análise de variância para modelo de percepção do papel da mídia

Fonte de variação	SQ	GL	QM	F	$p - value$
Regressão	202.10	4	50.52	72.17	< 0.001
Residual	508.87	723	0.7		
Total	710.98	727			

A figura 5.0.8 mostra ligeiros desvios na suposição de distribuição normal do termo de erro, uma vez que os pontos não recaem sobre a reta como expectável. A figura 5.0.7 dos resíduos versus valores ajustados mostra que a distribuição dos pontos ao longo dos valores ajustados têm uma padrão do tipo funil, ou seja não é constante, o que sugere violação do pressuposto de homocedasticidade.

Os desvios observados nas suposições de normalidade e variância constante dos termos de erro podem implicar a existência de viés nos coeficientes estimados. Os testes de significância podem produzir resultados enganosos, levando a conclusões erróneas sobre a significância estatística das variáveis explicativas. Os resultados ilustrados devem ser interpretados com atenção as possíveis implicações que tais desvios a suposição de normalidade podem encerrar.

4.4 Fontes de informação sobre a COVID19

Com o intuito de identificar os factores relacionados as fontes de informação sobre a COVID19, foram realizadas análises de regressão linear para explicar o número de fontes/meios de informação e a aplicação da regressão logística como forma de identificar factores que expliquem a frequência de uso de meios como a internet incluindo *blogs* e notícias da internet, redes sociais como o *Facebook, Twitter, WhatsApp* e os meios de comunicação tradicionais como a rádio, televisão e Jornais, para obtenção de informações relacionadas a COVID19.

Os resultados são descritos nas secções subsequentes.

4.4.1 Número de fontes de informação

Os dados da tabela 4.4.1 mostram os resultados da aplicação da regressão linear múltipla, tendo as características sócio-demográficas como variáveis explicativas e o número de fonte/meios de informação principais sobre a COVID19 como a variável resposta.

Aplicou-se o método *backward* para a seleção das variáveis de onde verificou-se através da aplicação do teste *t* para os coeficiente β , evidências estatísticas suficientes para rejeitar a hipótese que o nível de educação tem um efeito igual a zero, por outra é um factor significativo ($p = 0.03$) para explicar o número de fontes de informação principais sobre a COVID19.

Verificou-se que um indivíduo com nível superior de educação em média recorre a aproximadamente mais uma fonte de informação em relação aos indivíduos que não têm nenhum nível de educação formal. Em relação aos demais grupos as diferenças comparadas aos indivíduos com nenhum nível de educação, não foram constatadas evidências suficientes para rejeitar a hipótese de que os seus efeitos sejam significativos, i.e, diferentes de zero.

A tabela 4.4.2 ilustra os resultados da aplicação da análise de variância (ANOVA) para testar a significância do modelo de regressão linear contendo o nível de educação como variável explicativa. Verificou-se que a variabilidade explicada pelo modelo é igual 31.45

Tabela 4.4.1: Factores relacionados ao número de fontes de informação sobre a COVID19

Variável	β	EP (β)	$t_{calculado}$	$p-value$
Interceto	2.23	0.32	6.95	< 0.001
Nível de Educação				
Nenhum *	-	-	-	-
Alfabetização	0.16	0.44	0.38	0.71
Primário	-0.27	0.35	-0.77	0.44
Secundário	0.22	0.34	0.64	0.52
Superior	0.94	0.43	2.17	0.03

* : Categoria de referência

comparando com a variabilidade não explicada ou residual que é igual a 4.99.

Aplicou-se o teste F para medir a relação entre a variação explicada pelo modelo e não explicada, sob a hipótese nula que os efeitos das variáveis explicativas do modelo são iguais a zero, verificou-se que existem evidências suficientes para rejeitar tal hipótese, sugerindo que o modelo proposto é significativo ($p = 0.012$) para explicar a variabilidade do número de fontes/meios que os indivíduos recorrem para obtenção de informações relacionadas com a COVID19.

Tabela 4.4.2: Análise de variância para modelo do número de fontes de informação

Fonte de variação	SQ	GL	QM	F	$p - value$
Regressão	31.45	1	31.45	6.3	0.012
Residual	3626	726	4.99		
Total	3657	727			

A figura 5.0.10 mostra moderados desvios na suposição de distribuição normal do termo de erro, uma vez que os pontos não recaem sobre a reta como expectável. A figura 5.0.9 dos resíduos versus valores ajustados mostra que a distribuição dos pontos é constante ao longo dos valores ajustados o que sugere que o modelo capta adequadamente a variabilidade da variável resposta, número de fontes/meios de informação, em relação a variável explicativa, nível de educação.

Os desvios de normalidade podem implicar a existência de viés nos coeficientes estimados, na medida que as estimativas podem estar distorcidas e não refletir com precisão o real efeito do nível de educação sobre a determinação do número de fontes de informação da COVID19. Os testes de significância podem produzir resultados enganosos, levando a conclusões erróneas sobre a significância estatística das variáveis explicativas. Os resultados ilustrados devem ser interpretados com atenção as possíveis implicações que tais desvios a suposição de normalidade podem encarregar.

4.4.2 Internet, mídia social e tradicional como uma das principais fontes de informação da COVID19

De forma especial, é intuito avaliar de forma paralela quais factores sócio-demográficos são determinantes no uso da internet, mídia social e tradicional como uma fonte principal de obtenção de informações relacionadas a COVID19. Como forma de seleção das variáveis dos três (3) modelos explicativos foi usado o método *backward*, resultando em dois modelos que incluem as variáveis estado civil, nível de educação para explicar o uso da internet e também da mídia tradicional, e um último modelo que para além do estado civil e o nível de educação incorpora o variável estado ocupacional para explicar o uso da mídia social.

A tabela 4.4.3 inclui os resultados dos três modelos, da qual verificou-se evidências para rejeitar a hipótese que o real efeito do factor estado civil é nulo, sugerindo que as chances de uso da internet como uma fonte principal de obtenção de informações sobre a COVID19 aumenta em 1.52 vezes o equivalente a 52% quando o individuo encontra-se em um relacionamento em comparação com quem não.

Verificou-se que as chances de uso da mídia social como fonte principal de obtenção de informações da COVID19 aumentam 1.82 vezes ou em 82% quando os indivíduos encontram-se em relacionamento em comparação com os que não reportaram nenhum relacionamento.

Observou-se que chances de uso da mídia tradicional como fonte principal de obtenção de informações da COVID19 aumentam 1.37 vezes ou em 37% quando os indivíduos encontram-se em relacionamento em comparação com os que não reportaram nenhum relacionamento, segundo dados da tabela.

Não foram encontradas evidências suficientes para rejeitar a hipótese que o efeito do estado ocupacional para a mídia social, e o nível de educação para tanto a internet como mídia social e tradicional é igual a zero.

Com o objetivo de avaliar a ajuste dos modelos estimados foi aplicado o teste de Hosmer-Lemeshow, de onde sob a hipótese nula de que as frequências observadas e as esperadas são iguais, não foram encontradas evidências suficientes para rejeitar tal hipótese, como ilustrado na tabela 4.4.4, sugerindo que os modelos estimados para uso de internet, mídia social e internet são significativos e ajustam-se de forma adequada aos dados.

Tabela 4.4.3: Factores relacionados as fontes de informação sobre a COVID19

Variável	Internet		Mídia Social		Mídia Tradicional	
	Exp(β)	p-value	Exp(β)	p-value	Exp(β)	p-value
Estado Civil						
Não-relacionamento*	-	-	-	-	-	-
Em relacionamento	1.52	0.03	1.82	0.004	1.37	0.05
Nível de Educação						
Nenhum*	-	-	-	-	-	-
Alfabetização	0.81	0.66	0.89	0.81	2.36	0.36
Primário	0.64	0.25	0.56	0.13	0.68	0.27
Secundário	0.99	0.99	0.75	0.45	0.75	0.35
Superior	1.99	0.1	1.51	0.37	1.02	0.97
Estado Ocupacional						
Nenhuma*	-	-	-	-	-	-
Estudante	-	-	1.5	0.11	-	-
Empregado	-	-	1.38	0.31	-	-
Sazonal	-	-	0.76	0.34	-	-

* : Categoria de referência

Tabela 4.4.4: Teste de Hosmer e Lemeshow para o modelo das fontes de informação da COVID19

χ^2	Internet		χ^2	Mídia Social		χ^2	Mídia Tradicional	
	gl	p-value		gl	p-value		gl	p-value
0.47	4	0.99	4.86	7	0.68	2.44	4	0.65

4.5 Comportamentos sócio-culturais durante a pandemia da COVID19

Os dados da tabela 4.5.1 ilustram resultados da aplicação da regressão logística múltipla, cujo objectivo de identificar factores associados a propensão à aderência a medidas restritivas de mobilidade, de forma específica a atitude de deixar de frequentar lugares habituais, como escolas, locais de trabalho, lazer e diversão em conformidade com as recomendações feitas no âmbito da gestão da calamidade da COVID19.

Verificou-se que as chances de deixar de frequentar um lugar habitual por conta da COVID19 aumentam 2.29 vezes ou 129% em indivíduos na faixa dos 10 a 19 anos quando comparados aos indivíduos menores de 10 anos de idade. As chances de deixar de frequentar um lugar habitual por conta da COVID19 aumentam 2.85 vezes ou 185% em indivíduos na faixa dos 50 a 59 anos quando comparados aos indivíduos menores de 10 anos de idade. As chances de deixar de frequentar um lugar habitual por conta da COVID19 aumentam em cerca de 3 vezes ou 200% em indivíduos na faixa dos 60 anos ou mais quando comparados

aos indivíduos menores de 10 anos de idade.

Foi possível observar também que as chances de deixar de frequentar lugares habituais por conta da COVID19 aumentam em cerca de 2.11 vezes ou 111% quando o indivíduo tem ensino primário em relação aos que não têm nenhum ensino formal concluído. As chances de aderência a medida restritiva de mobilidade aumenta em quase 3 vezes quando o indivíduo tem um nível superior concluído em relação aos que nenhum nível formal têm.

Aumenta em 1 vez a chance de adoção da medida restritiva de mobilidade quando o número de fontes de informação aumenta em uma unidade.

A significância dos parâmetros foi verificada pela aplicação do teste de Wald.

Tabela 4.5.1: Factores relacionados a comportamentos sócio-culturais durante a pandemia da COVID19

Variável	Exp(β)	<i>p-value</i>
Estado Civil		
Não-relacionamento*	-	-
Em relacionamento	0.67	0.06
Faixa Etária		
9 anos ou menos*	-	-
10 a 19	2.29	0.02
20 a 29	1.24	0.55
30 a 39	1.36	0.45
40 a 49	1.33	0.5
50 a 59	2.85	0.03
60 anos ou mais	3.1	0.02
Nível de Educação		
Nenhum *	-	-
Alfabetização	0.45	0.07
Primário	2.11	0.04
Secundário	2.04	0.06
Superior	2.98	0.03
Total de fontes	1.08	0.05

* : Categoria de referência

Com o objetivo de avaliar a ajuste do modelos estimado sobre os factores associados a adesão a medidas de restrição de mobilidade por conta da COVID19, foi aplicado o teste de Hosmer-Lemeshow. Sob a hipótese nula de que as frequências observadas e as esperadas são iguais não foram encontradas evidências suficientes para rejeitar tal hipótese, como ilustrado na tabela 4.5.2, sugerindo que o modelo estimados é significativo e ajusta-se de forma adequada aos dados.

Tabela 4.5.2: Teste de Hosmer e Lemeshow para o modelo dos comportamentos sócio-culturais durante a pandemia da COVID19

χ^2	gl	p-value
4.2	8	0.88

4.6 Discussão

Analisando os dados coletados na Cidade de Quelimane, na Província da Zambézia, no âmbito do estudo que visava compreender padrões de movimentos da população e comportamentos sócio-culturais no contexto da COVID19, verificou-se associações significativas entre as variáveis de estudos em relação aos factores sócio-demográficos.

A variável estado civil para além de significativamente estar associada a variável deixou de frequentar um lugar habitual devido a COVID19, estava associada também as variáveis do uso de Internet, Mídia Social e Mídia Tradicional. Resultado similar foi encontrado no estudo realizado por Alshareef et. al. (2021) que indicou associação entre estado marital e o uso da mídia tradicional.

As variáveis estado ocupacional e o nível de educação estavam associadas de forma significativa com as fontes de informação de estudo, a variável número de fontes de informação e a variável deixou de frequentar um lugar habitual por conta da COVID19. O estudo realizado por Alshareef et. al. (2021) indicou associações significativas entre o estado ocupacional e o uso da mídia social e tradicional similar aos resultados encontrados no presente estudo. De forma contrária o estudo do Alshareef et. al. (2021) relevou falta de evidências estatísticas sugerindo associação significativa entre o nível de educação e as fontes de informação como mídia social e tradicional.

Neste estudo, utilizou-se a técnica de regressão linear para identificar os factores associados à percepção do papel da mídia na divulgação de comportamentos preventivos contra a COVID-19. A análise da regressão linear nos permitiu investigar a relação entre variáveis independentes selecionadas e a percepção dos participantes sobre o papel da mídia na disseminação de informações relacionadas à pandemia.

Verificou-se que a educação dos participantes desempenhou um papel significativo na percepção do papel da mídia. Os participantes com maior nível de educação tendiam a atribuir uma maior importância ao papel da mídia na difusão de informações preventivas. Isso pode ser explicado pelo fato de que indivíduos com maior nível de educação tendem a ter maior acesso a fontes de informação e são capazes de avaliar melhor a qualidade e a precisão das informações divulgadas pela mídia. De forma particular os indivíduos com nível de educação primário atribuíram menor importância quando comparados ao indivíduos com nenhuma educação formal concluída.

Outra descoberta interessante foi a associação entre a exposição à mídia e a percepção do seu papel. Os participantes que relataram maior exposição à mídia, como assistir regu-

larmente a programas de notícias e ler jornais, *blogs*, e publicações nas redes sociais tendiam a atribuir maior importância ao papel da mídia na divulgação de comportamentos preventivos contra a COVID-19. Isso sugere que a exposição frequente à mídia pode aumentar a confiança nas informações divulgadas por ela.

Através da aplicação da regressão linear verificou-se também que nível de educação é um factor significativo na identificação do número de fontes/meios que o participante recorre para obtenção de informações da COVID19, dados revelaram que os indivíduos de nível superior tendem a recorrer a mais fontes de informação quando comparado com os indivíduos sem nenhum nível concluído, pelo contrário os estudos realizados por Wang et. al. (2021) e Erinoso et. al. (2021) revelaram que o nível de educação não é significativo para explicar as variações no número de meios usados para obtenção de informações da COVID19. O que pode-se justificar pelo facto dos estudos terem usado como referência algum nível de educação como base de comparação em suas análises enquanto que no presente estudo a base de referência é nenhum nível concluído.

Verificou-se, através da aplicação da regressão logística múltipla, que os indivíduos em relacionamento têm 1.5 e 1.82 vezes mais chances de usar internet e a mídia social, respectivamente, como fonte de informações em relação aos indivíduos que não se encontram em relacionamento. O estudo realizado por Erinoso et. al. (2021) encontrou resultado similar, mas para o uso da mídia tradicional, algo que pôde se verificar no presente estudo entretanto não houve evidências suficientes para rejeitar a hipótese que o real efeito é nulo. Em conformidade com os resultados verificados no presente estudo Alshareef et. al. (2021) verificou que indivíduos em não relacionamento tinham menores chances de uso da mídia social para buscar informações em comparação com os indivíduos em relacionamento.

Em relação a aderência a medidas restritivas de mobilidade da população, verificou-se através da aplicação da regressão logística múltipla, que indivíduos dos 60 anos ou mais e 50 a 59 têm aproximadamente 3 vezes mais chances de aderência a medidas restritivas a COVID19 quando comparadas aos indivíduos menores de 10 anos de idade, tal facto pode se justificar pelo facto de indivíduos de maior idade terem menos mobilidade em comparação aos mais novos. Indivíduos dos 10 a 19 anos tem mais de 2 vezes mais chances de aderir a medidas restritivas em relação a COVID19 quando comparados aos menores de 10 anos de idade.

No entanto, é importante ressaltar que o estudo possui algumas limitações. A natureza transversal do estudo impede a inferência de causalidade entre as variáveis estudadas. Além disso, a amostra foi limitada a uma determinada região geográfica, o que pode restringir a generalização dos resultados para outras populações.

Apesar dessas limitações, os resultados deste estudo fornecem insights importantes sobre os factores associados à percepção do papel da mídia na divulgação de comportamentos preventivos contra a COVID-19, as fontes de informações e a compreensão de atitudes associadas a comportamentos sócio-culturais face a medidas restritivas de mobilidade. Essas descobertas

podem ser utilizadas para orientar estratégias de comunicação durante a pandemia, destacando a importância da mídia na disseminação de informações confiáveis e promovendo a conscientização sobre comportamentos preventivos para combater a propagação do vírus.

Capítulo 5

Conclusões e Recomendações

Neste estudo, investigaram-se os factores associados à percepção do papel da mídia na divulgação de comportamentos preventivos contra a COVID-19, as fontes de informações e a comportamentos sócio-culturais face a pandemia da COVID19. Com base em nossos resultados, podemos tirar as seguintes conclusões e fornecer recomendações relevantes:

1. A mídia desempenha um papel significativo na divulgação de comportamentos preventivos contra a COVID-19. Os participantes com maior nível de educação e aqueles com maior exposição à mídia têm uma percepção mais positiva sobre o papel da mídia na disseminação de informações preventivas. Isso destaca a importância de utilizar diferentes canais de mídia para alcançar efetivamente diferentes grupos demográficos, adaptando as mensagens de acordo com suas preferências e necessidades.
2. A educação dos indivíduos está associada à sua percepção da mídia como uma fonte confiável de informações preventivas. É crucial investir em programas de alfabetização em saúde e mídia, especialmente para grupos com menor nível educacional. Isso pode capacitar os indivíduos a avaliar criticamente as informações da mídia e tomar decisões informadas sobre comportamentos preventivos.
3. Grupos com menor nível educacional e acesso limitado a diferentes fontes de informação também apresentaram menor aderência às medidas restritivas. Isso pode ser resultado de menor conscientização sobre a importância das medidas e suas consequências, bem como de dificuldades na obtenção de informações confiáveis e atualizadas.
4. É essencial implementar campanhas de conscientização direcionadas a esses grupos, enfatizando a importância das medidas restritivas de mobilidade na prevenção da disseminação da COVID-19. Essas campanhas devem ser adaptadas às diferentes faixas etárias e níveis de compreensão, utilizando abordagens adequadas para as crianças e estratégias de comunicação claras e acessíveis para os adultos.
5. São necessárias pesquisas adicionais para compreender melhor o impacto da mídia na adoção de comportamentos preventivos contra a COVID19. Estudos longitudinais e

qualitativos podem fornecer insights mais aprofundados sobre as atitudes, percepções e barreiras enfrentadas pela população em relação às mensagens da mídia e sua influência na adoção de comportamentos preventivos.

As recomendações visam abordar as limitações identificadas em nosso estudo e promover uma compreensão mais abrangente da aderência às medidas restritivas de mobilidade. Ao implementar essas recomendações em estudos futuros, podemos obter uma visão mais aprofundada dos factores que influenciam os comportamentos de conformidade e desenvolver estratégias mais eficazes para controlar a disseminação da COVID-19.

Em resumo, a mídia desempenha um papel fundamental na divulgação de comportamentos preventivos contra a COVID-19. Ao considerar os factores demográficos, como estado marital, ocupacional e o nível educacional, e promover a confiança e a transparência na mídia, pode-se otimizar a eficácia das campanhas de informação e educação. Essas recomendações podem ajudar a orientar políticas e estratégias de comunicação durante a pandemia, visando reduzir a propagação do vírus e proteger a saúde pública.

Referências Bibliográficas

- [1] Agresti, A. (2007). *An Introduction to Categorical Data Analysis* (Wiley Series in Probability and Statistics) (2nd ed.). Wiley.
- [2] Alshareef, N., Yunusa, I., & Al-Hanawi, M. K. (2021). The Influence of COVID-19 Information Sources on the Attitudes and Practices Toward COVID-19 Among the General Public of Saudi Arabia: Cross-sectional Online Survey Study. *JMIR Public Health Surveillance*, 7(7), 15. 10.2196/28888
- [3] Anwar, A., Malik, M., Raees, V., & Anwar, A. (2020). Role of Mass Media and Public Health Communications in the COVID-19 Pandemic. *Cureus*, 12(9), e10453. <https://doi.org/10.7759/cureus.10453>
- [4] das Neves Martins Pires, P.H., Macaringue, C., Abdirazak, A. et al. (2021). Covid-19 pandemic impact on maternal and child health services access in Nampula, Mozambique: a mixed methods research. *BMC Health Serv Res* 21, 860. <https://doi.org/10.1186/s12913-021-06878-3>
- [5] Erinoso, O., Wright, K. O., Anya, S., Kuyinu, Y., Abdur-Razzaq, H., & Adewuya, A. (2021, 1 25). Predictors of COVID-19 Information Sources and Their Perceived Accuracy in Nigeria: Online Cross-sectional Study. *JMIR Public Health Surveillance*, 7(1), 1-7. 10.2196/22273
- [6] Faulkner, C. A., & Faulkner, S. S. (2009). *Research Methods for Social Workers: A Practice-Based Approach*. Chicago, IL: Lyceum Books, Inc.
- [7] Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. T. (2009). *Análise Multivariada de dados*. São Paulo: bookman.
- [8] Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Wiley.
- [9] Júnior, A., Dula, J., Mahumane, S., Koole, O., Enosse, S., Fodjo, J. N. S., & Colebunders, R. (2021). Adherence to COVID-19 Preventive Measures in Mozambique: Two Consecutive Online Surveys. *International Journal of Environmental Research and Public Health*, 18(3), 1091. MDPI AG. Retrieved from <http://dx.doi.org/10.3390/ijerph18031091>

- [10] Montgomery, D. C., Peck, E. A., & Vinning, G. G. (2012). *Introduction to Linear Regression Analysis* (5th ed.). Wiley.
- [11] Myers, L. L., & Randolph, K. A. (2013). *Basic Statistics in Multivariate Analysis*. OUP USA.
- [12] Vittinghoff, E., Shiboski, S. C., McCulloch, C. E., & Glidden, D. V. (2005). *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. Springer.
- [13] Wang, P.-W., Lu, W.-H., Ko, N.-Y., Chen, Y.-L., Li, D.-J., Chang, Y.-P., & Yen, C.-F. (2020, May 9). COVID-19-Related Information Sources and the Relationship With Confidence in People Coping with COVID-19: Facebook Survey Study in Taiwan. *J Med Internet Res*, 22(6), 6-10. 10.2196/20021

Anexos

Variável	Internet			Mídia Social			Mídia Tradicional		
	β	EP(β)	Z	β	EP(β)	Z	β	EP(β)	Z
Estado Civil									
Não-relacionamento*	-	-	-	-	-	-	-	-	-
Em relacionamento	0.42	0.19	2.22	0.6	0.21	2.85	0.31	0.16	1.94
Nível de Educação									
Nenhum*	-	-	-	-	-	-	-	-	-
Alfabetização	-0.22	0.49	-0.43	-0.11	0.47	-0.24	0.86	0.4	2.1
Primário	-0.44	0.38	-1.16	-0.58	0.39	-1.51	-0.38	0.32	-1.18
Secundário	-0.006	0.37	-0.02	-0.29	0.38	-0.75	-0.29	0.31	-0.93
Superior	0.69	0.44	1.57	0.41	0.49	0.9	0.02	0.39	0.04
Estado Ocupacional									
Nenhuma*	-	-	-	-	-	-	-	-	-
Estudante	-	-	-	0.4	0.25	1.6	-	-	-
Empregado	-	-	-	0.32	0.32	1	-	-	-
Sazonal	-	-	-	-0.27	0.28	-0.96	-	-	-

Tabela 5.0.1: Estimação do modelo sobre as fontes de informação sobre a COVID19- Resultados

* : Categoria de referência

Variável	β	EP (β)	Z
Estado Civil			
Não-relacionamento*	-	-	-
Em relacionamento	-0.39	0.2	-1.9
Faixa Etária			
9 anos ou menos*	-	-	-
10 a 19	0.83	0.35	2.35
20 a 29	0.22	0.36	0.6
30 a 39	0.3	0.4	0.79
40 a 49	0.28	0.42	0.67
50 a 59	1.05	0.48	2.16
60 anos ou mais	1.13	0.49	2.31
Nível de Educação			
Nenhum *	-	-	-
Alfabetização	-0.8	0.44	-1.81
Primário	0.75	0.36	2.07
Secundário	0.71	0.37	1.89
Superior	1.09	0.49	2.23
Total de fontes	0.08	0.04	1.99

Tabela 5.0.2: Estimação do modelo sobre comportamentos sócio-culturais durante a pandemia da COVID19

* : Categoria de referência

ANEXO B

COMPORTAMENTOS SÓCIO-CULTURAIS

O principal objectivo deste questionário é estudar os comportamentos sócio-culturais e padrões de movimento da população de modo a ajudar os órgãos competentes na tomada de decisões locais sobre as medidas da mitigação e contenção da propagação da pandemia da COVID-19 em Moçambique.

N.B.: Este inquérito é anónimo e todos os dados recolhidos são totalmente confidenciais.

ID do formulário _ _ _ _ _ _ _	
1. Consome bebidas alcóolicas?	<input type="checkbox"/> Não <input type="checkbox"/> Sim
2. Se sim. Onde é que consome com mais frequência?	<input type="checkbox"/> Casa <input type="checkbox"/> Bar <input type="checkbox"/> Barraca <input type="checkbox"/> Quintais <input type="checkbox"/> Outro
3. Se não consome em casa, quais razões podem estar associadas?	<input type="checkbox"/> Convite dos próximos <input type="checkbox"/> Simples Vontade
4. Você está preocupado com o impacto do coronavírus em você pessoalmente?	<input type="checkbox"/> Muito preocupado <input type="checkbox"/> Um pouco preocupado <input type="checkbox"/> Não tão preocupado <input type="checkbox"/> Nem um pouco preocupado
5. Com que frequência consumia o álcool em 2021?	<input type="checkbox"/> Todos os dias <input type="checkbox"/> Uma ou duas vezes por semana <input type="checkbox"/> Apenas finais de semana
6. Já fez o teste de COVID-19?	<input type="checkbox"/> Não <input type="checkbox"/> Sim
7. Se sim, qual foi a motivação?	<input type="checkbox"/> Sintomas <input type="checkbox"/> Viagem <input type="checkbox"/> Outros Serviços
8. Testou positivo a COVID-19?	<input type="checkbox"/> Não <input type="checkbox"/> Sim
9. Se sim, que cuidados tomou?	<input type="checkbox"/> Quarentena <input type="checkbox"/> Máscara <input type="checkbox"/> Distanciamento <input type="checkbox"/> Outros
10. Já faltou a algum local de frequência habitual por causa da covid-19?	<input type="checkbox"/> Não <input type="checkbox"/> Sim
11. Se sim, qual foi o seu principal motivo para não frequentar o tal local?	<input type="checkbox"/> Sintomas da covid <input type="checkbox"/> Medo de contaminar os outros <input type="checkbox"/> Medo de ser contaminado <input type="checkbox"/> Estado de calamidade
12. Por quanto tempo esteve ausente do local?	<input type="checkbox"/> 1-3 dias <input type="checkbox"/> 4-6 dias <input type="checkbox"/> Uma semana ou mais

Figura 5.0.2: Comportamentos Sócio-culturais - Página 1

ANEXO C

MÍDIA

O principal objectivo deste questionário é estudar os comportamentos sócio-culturais e padrões de movimento da população de modo a ajudar os órgãos competentes na tomada de decisões locais sobre as medidas da mitigação e contenção da propagação da pandemia da COVID-19 em Moçambique.

N.B.: Este inquérito é anónimo e todos os dados recolhidos são totalmente confidenciais.

ID do formulário _ _ _ _ _ _					
1.	Você considera que a mídia educa sobre os procedimentos a serem seguidos durante uma epidemia?				
	<input type="checkbox"/> Nunca	<input type="checkbox"/> Quase nunca	<input type="checkbox"/> Às vezes	<input type="checkbox"/> Quase sempre	<input type="checkbox"/> Sempre
2.	Considera que a comunicação social divulga comportamentos preventivos para ajudar no controlo da COVID-19?				
	<input type="checkbox"/> Nunca	<input type="checkbox"/> Quase nunca	<input type="checkbox"/> Às vezes	<input type="checkbox"/> Quase sempre	<input type="checkbox"/> Sempre
3.	Considera que os meios de comunicação social conscientizam sobre o COVID-19?				
	<input type="checkbox"/> Nunca	<input type="checkbox"/> Quase nunca	<input type="checkbox"/> Às vezes	<input type="checkbox"/> Quase sempre	<input type="checkbox"/> Sempre
4.	Você considera que a mídia educa sobre como cuidar de uma pessoa com COVID-19?				
	<input type="checkbox"/> Nunca	<input type="checkbox"/> Quase nunca	<input type="checkbox"/> Às vezes	<input type="checkbox"/> Quase sempre	<input type="checkbox"/> Sempre
5.	Você confia nas informações da mídia?				
	<input type="checkbox"/> Nunca	<input type="checkbox"/> Quase nunca	<input type="checkbox"/> Às vezes	<input type="checkbox"/> Quase sempre	<input type="checkbox"/> Sempre
6.	Você considera que a mídia está aumentando o medo, a ansiedade e a confusão?				
	<input type="checkbox"/> Nunca	<input type="checkbox"/> Quase nunca	<input type="checkbox"/> Às vezes	<input type="checkbox"/> Quase sempre	<input type="checkbox"/> Sempre
7.	Você recebe informações falsas nas redes sociais que possam representar risco à saúde da nossa população em relação aos sintomas e tratamento da COVID-19?				
	<input type="checkbox"/> Nunca	<input type="checkbox"/> Quase nunca	<input type="checkbox"/> Às vezes	<input type="checkbox"/> Quase sempre	<input type="checkbox"/> Sempre
8.	Com que frequência recebe informação relativa a Covid-19 dos seguintes meios/fontes?				
	Internet¹	<input type="checkbox"/> Quase nunca	<input type="checkbox"/> Às vezes	<input type="checkbox"/> Quase sempre	
	Mídia Social²	<input type="checkbox"/> Quase nunca	<input type="checkbox"/> Às vezes	<input type="checkbox"/> Quase sempre	
	Mídia tradicional³	<input type="checkbox"/> Quase nunca	<input type="checkbox"/> Às vezes	<input type="checkbox"/> Quase sempre	
	Amigos	<input type="checkbox"/> Quase nunca	<input type="checkbox"/> Às vezes	<input type="checkbox"/> Quase sempre	
	Colegas	<input type="checkbox"/> Quase nunca	<input type="checkbox"/> Às vezes	<input type="checkbox"/> Quase sempre	
	Familiares	<input type="checkbox"/> Quase nunca	<input type="checkbox"/> Às vezes	<input type="checkbox"/> Quase sempre	
	Agentes de saúde	<input type="checkbox"/> Quase nunca	<input type="checkbox"/> Às vezes	<input type="checkbox"/> Quase sempre	
	Órgãos oficiais⁴	<input type="checkbox"/> Quase nunca	<input type="checkbox"/> Às vezes	<input type="checkbox"/> Quase sempre	

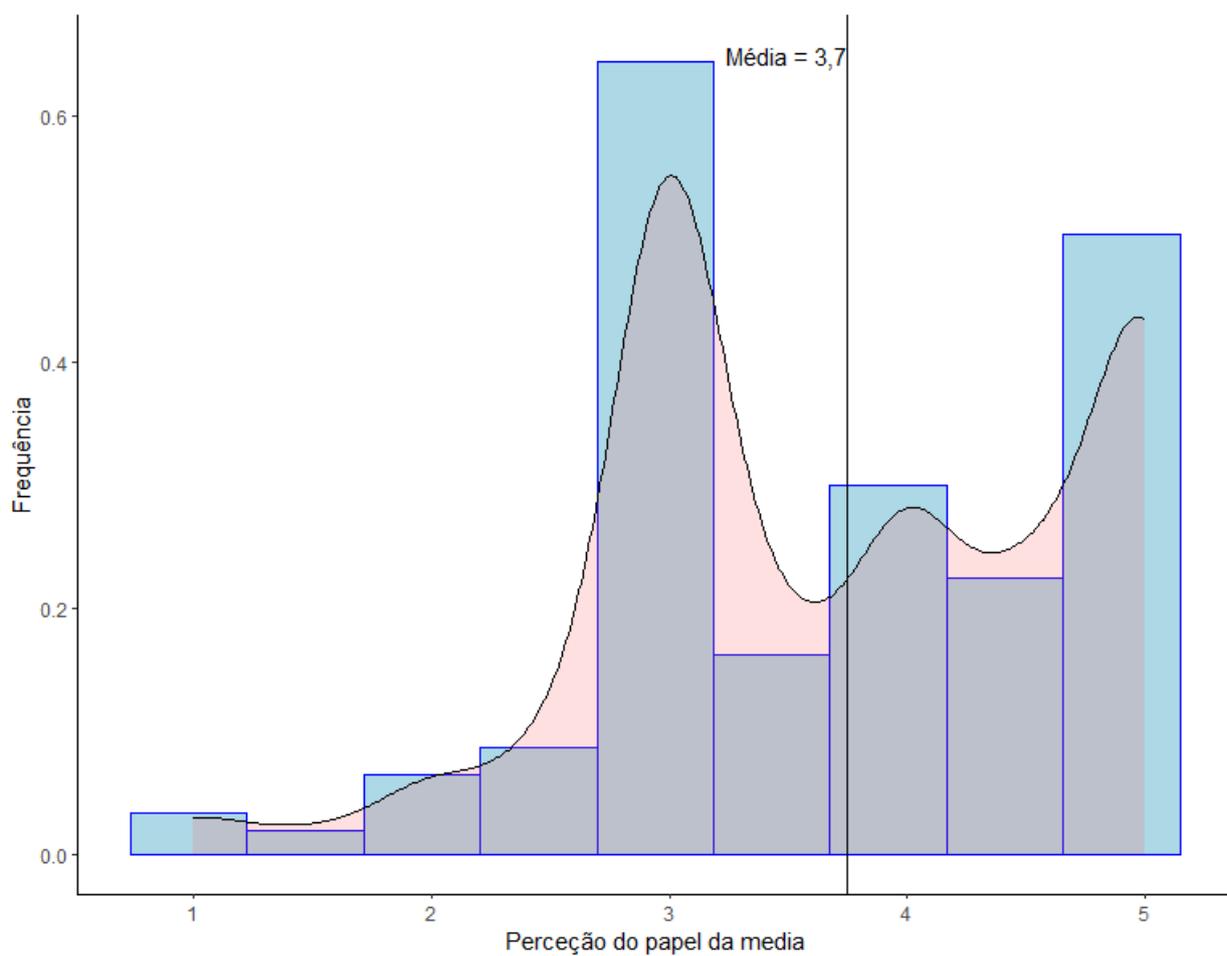
¹ Blogs, Notícias da internet

² Facebook, Twitter, Whatsapp, Instagram

³ Televisão, Jornais, Rádio

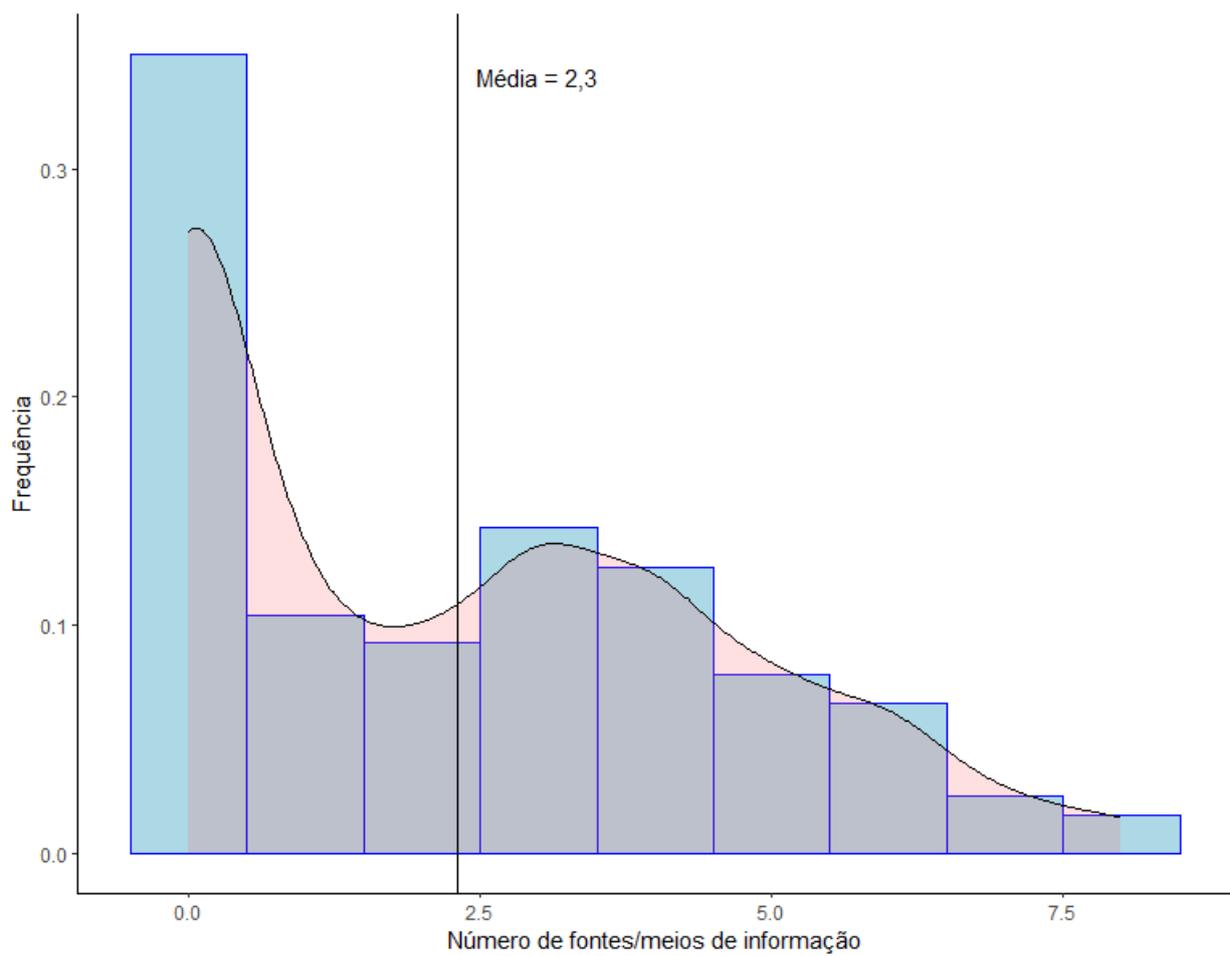
⁴ Órgãos governamentais (MISAU, INS, governo local, município, etc.)

Figura 5.0.3: Mídia



Autor: Inácio, Fernando Domingos

Figura 5.0.4: Distribuição da percepção do papel da mídia



Autor: Inácio, Fernando Domingos

Figura 5.0.5: Distribuição do número de fontes/meios de informação

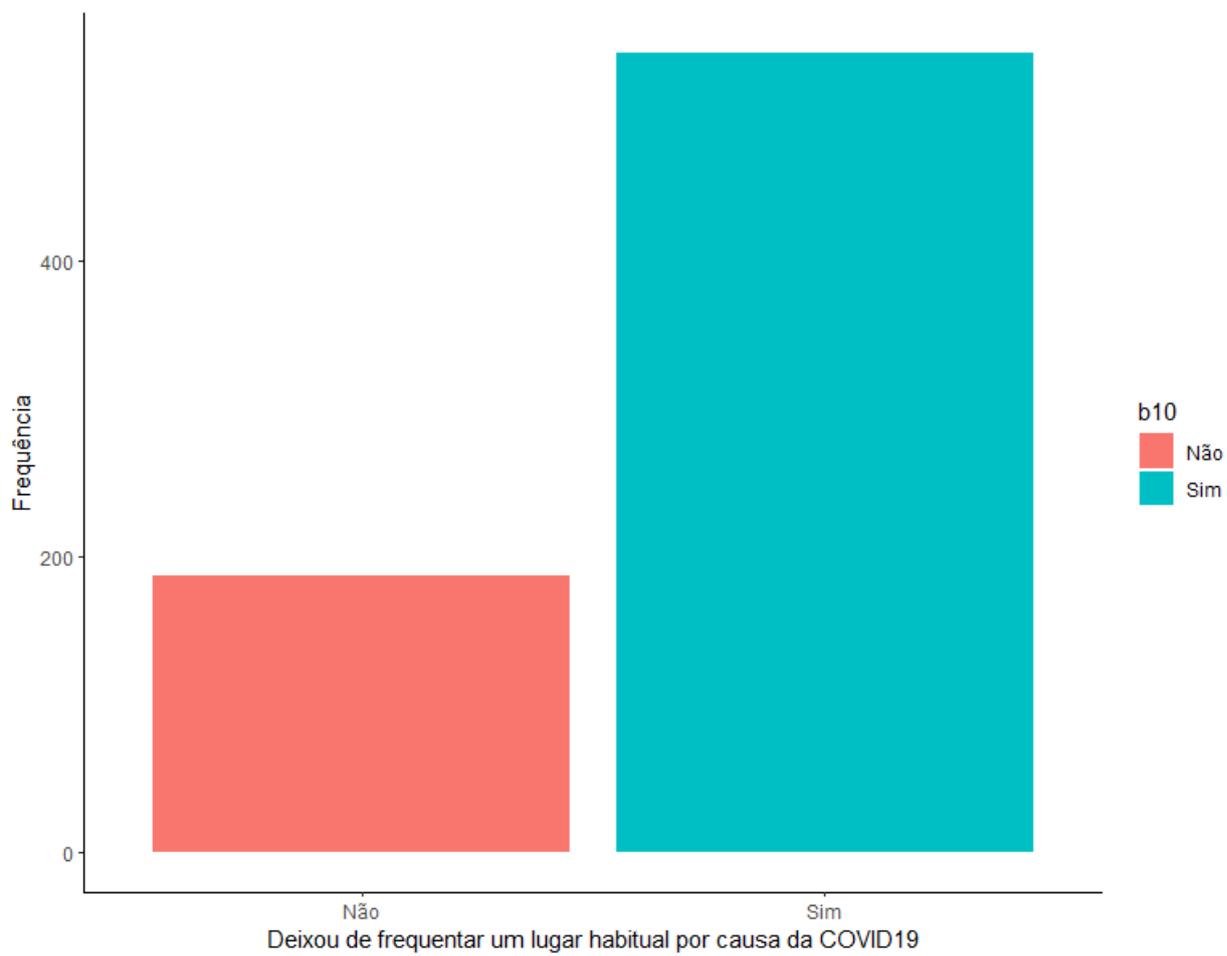


Figura 5.0.6: Distribuição por cumprimento da medida de isolamento

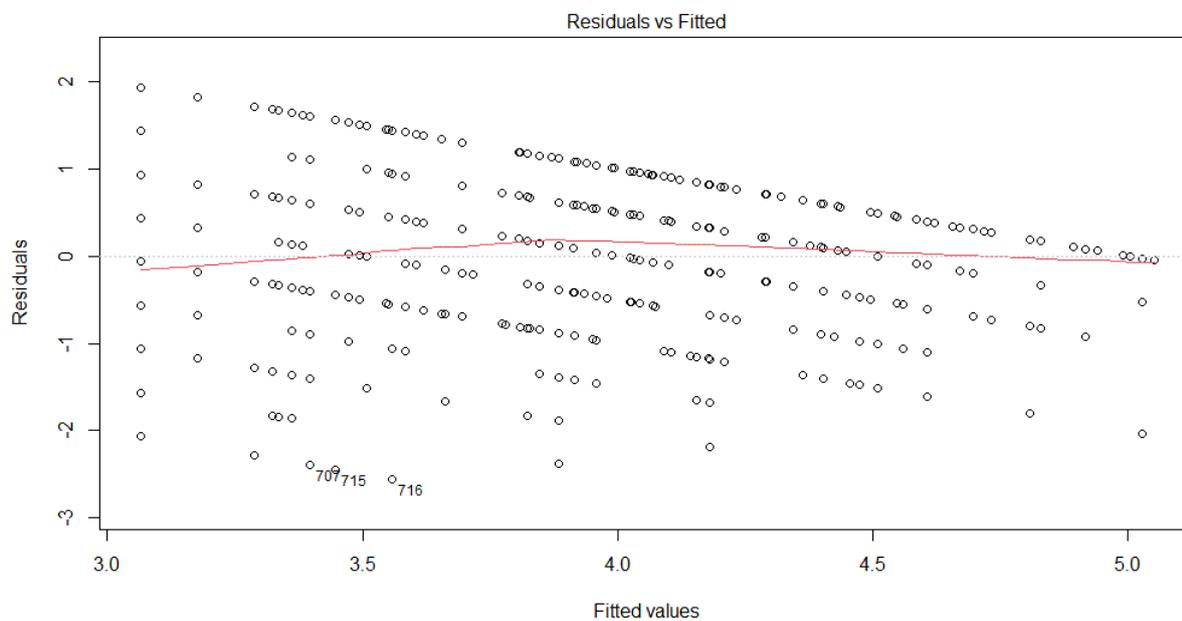


Figura 5.0.7: Resíduos padronizados - Modelo da percepção do papel da mídia

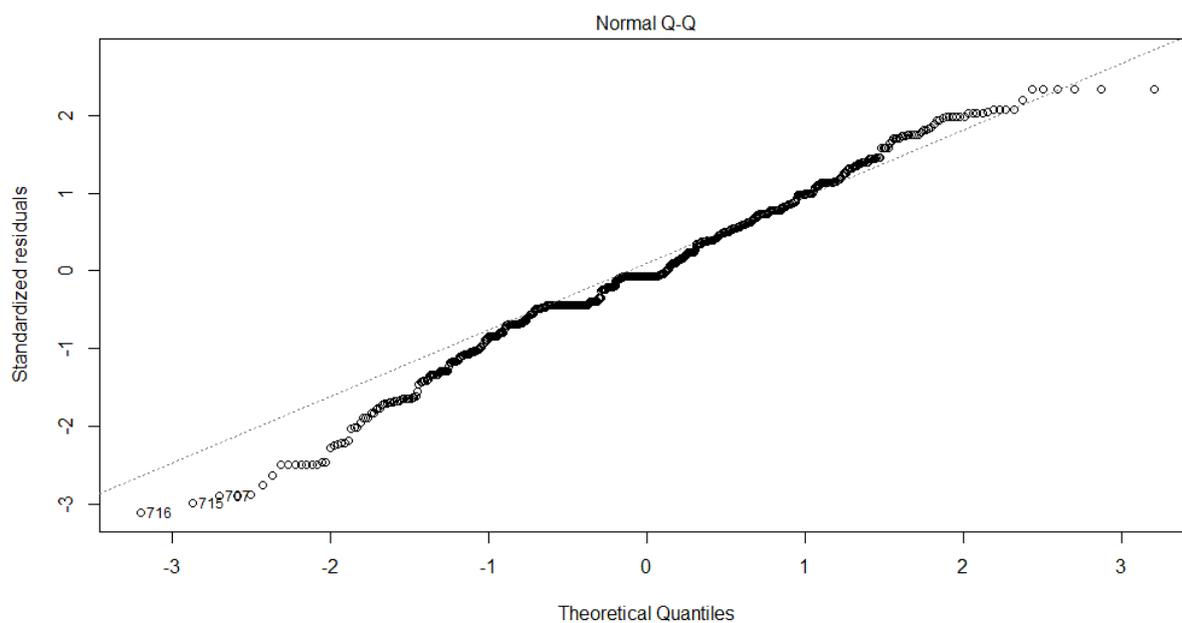


Figura 5.0.8: Normal QQ-Plot - Modelo da percepção do papel da mídia

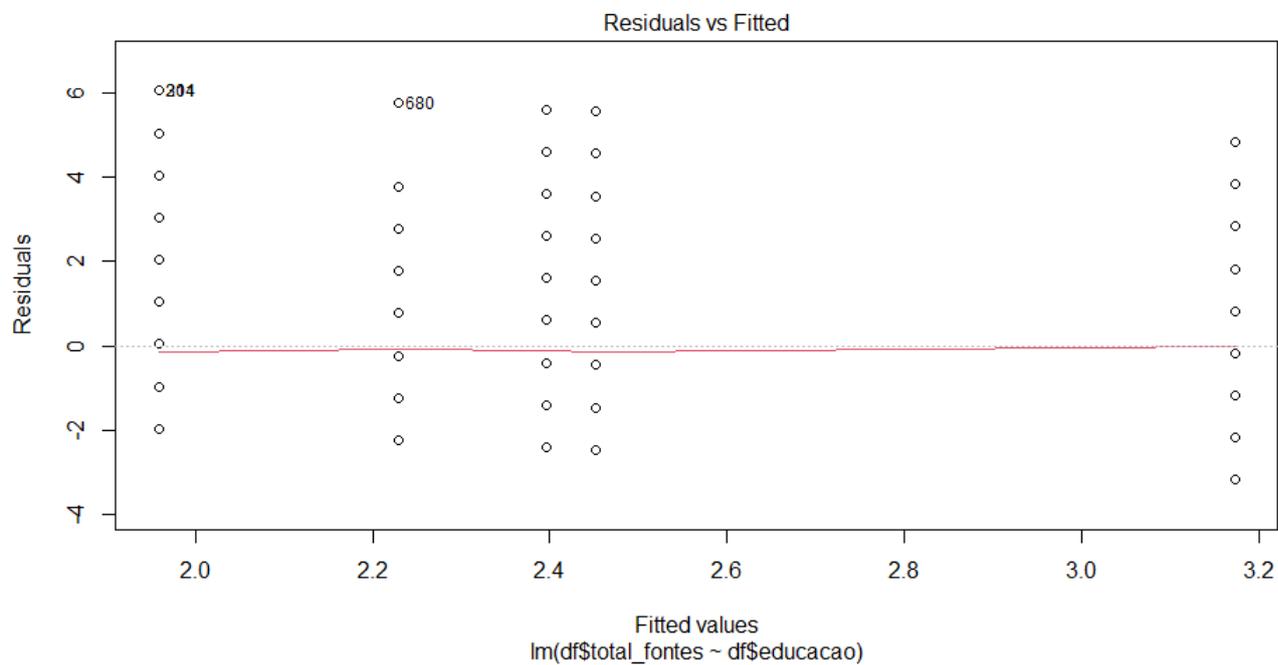


Figura 5.0.9: Resíduos padronizados - Modelo do número de fontes de informação

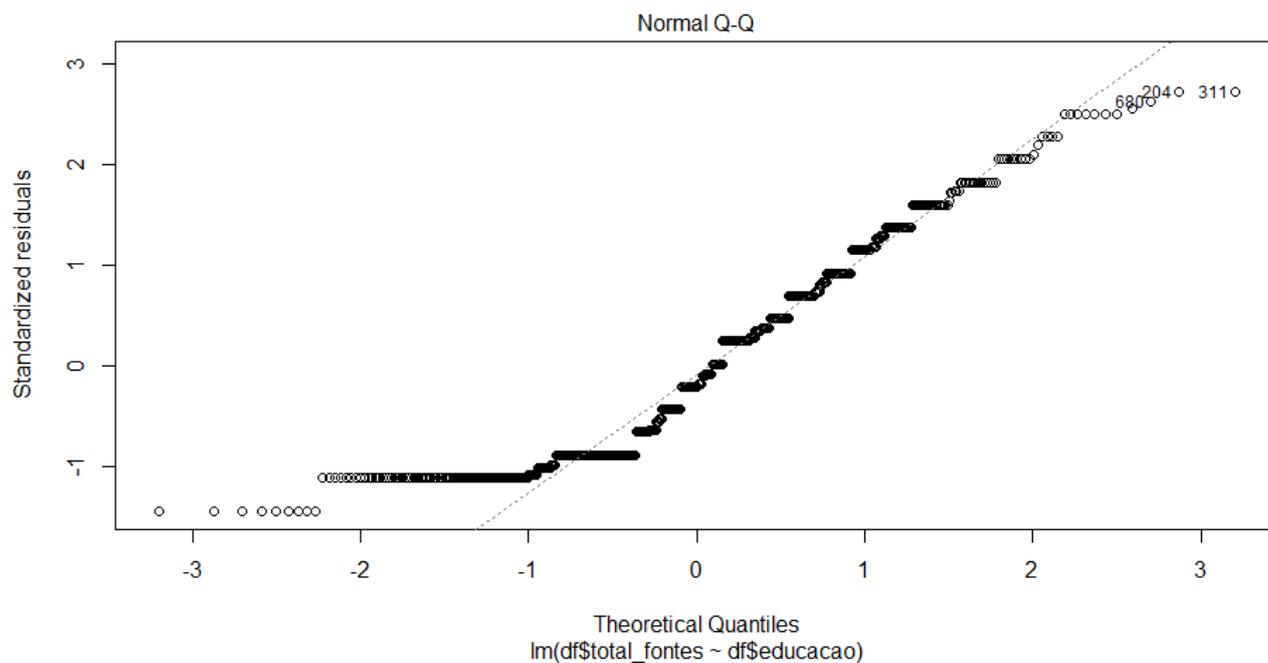


Figura 5.0.10: Normal Q-Q-Plot - Modelo do número de fontes de informação