

UNIVERSIDADE ABERTA



UNIVERSIDADE
AbERTA
www.uab.pt

Modelação Estatística no estudo da Depressão e impactos da Covid-19 em indivíduos com
50 ou mais anos

Sara Isabel Ribeiro Pires

Mestrado em Bioestatística e Biometria

2023

UNIVERSIDADE ABERTA



UNIVERSIDADE
AbERTA
www.uab.pt

Modelação Estatística no estudo da Depressão e impactos da Covid-19 em indivíduos com
50 ou mais anos

Sara Isabel Ribeiro Pires

Mestrado em Bioestatística e Biometria

Dissertação orientada pela

Professora Doutora Maria do Rosário Olaia Duarte Ramos, Departamento de Ciências e
Tecnologia, Universidade Aberta (orientadora)

e

Professora Doutora Ana Paula Fernandes, Departamento de Ciências e Tecnologia,
Universidade Aberta (coorientadora)

Abril de 2023

Agradecimentos

Primeiramente agradeço às minhas orientadoras por todo o apoio e compreensão ao longo deste percurso. Agradeço também a todos os professores de quem tive o prazer de ser aluna durante toda a minha caminhada de aprendizagem, nomeadamente aos professores do Mestrado em Bioestatística e Biometria.

À minha família pelo apoio incondicional, compreensão e sobretudo pela motivação nos momentos mais difíceis. Aos meus avós, que onde quer que estejam fiquem orgulhosos de mim.

Aos meus colegas de mestrado e licenciatura, nomeadamente à Joana e à Margarida, que em muito contribuíram para que a minha chegada até aqui.

Deixo ainda um especial agradecimento a todas as pessoas que, de alguma forma, me ajudaram na realização deste trabalho.

Resumo

Atualmente, como consequência da análise de vários indicadores demográficos, tais como a taxa de natalidade, a taxa de mortalidade e a taxa de prevalência de certas doenças, bem como do impacto do surgimento de novas doenças na população mundial, de que é exemplo a recente pandemia COVID19, torna-se urgente o investimento em estudos que possam contribuir para ajudar na tomada de decisões na área da Medicina, e das Políticas e Saúde Pública, entre outras. No âmbito desta problemática, foi criado a nível europeu o projeto SHARE, *Survey of Health, Aging and Retirement in Europe* que preserva um banco de dados multidisciplinar e transnacional de microdados sobre saúde, nível socioeconómico, entre outros fatores, obtidos através de um inquérito. Iniciado em 2004, conta com amostras dos indivíduos de 50 ou mais anos de em 28 países europeus e Israel.

Este trabalho teve como objetivo contribuir para novo conhecimento no que diz respeito à população portuguesa, uma vez que a temática do envelhecimento é premente, e os estudos realizados até à data mostram que há espaço para outras análises e perspetivas. Recorrendo a metodologias estatísticas em que se inclui a regressão logística à base de dados SHARE vaga 9 (Corona 2), procurou-se modelar a situação/sentimento de depressão reportada pelos inquiridos com base numa seleção de variáveis de saúde física, mental, comportamentais entre outras consideradas pertinentes.

Palavras-chave: Depressão, População idosa, Covid-19, Regressão Logística, Associações

Abstract

Currently, as a result of the analysis of various demographic indicators, such as the birth rate, death rate and the prevalence rate of certain diseases, as well as the impact of the emergence of new diseases on the world population, of which the pandemic is an example recent COVID19, it is urgent to invest in studies that can contribute to decision-making in the area of Medicine, and Public Health and Policies, among others. Within the scope of this issue, was created at European level the SHARE project, Survey of Health, Aging and Retirement in Europe, which maintains a multidisciplinary and transnational database of microdata on health, socioeconomic status, among other factors, obtained through a survey. Started in 2004, it features samples of the individuals aged 50 and over from 28 European countries and Israel.

This work aimed to contribute to new knowledge with regard to the Portuguese population, since the issue of aging is pressing societies, and those carried out to date show that there is room for other studies, other analyzes and perspectives. Using statistical methods, including logistic regression to the SHARE Wave 9 (Corona 2) database, we propose to model the situation/feeling of depression reported by respondents based on a selection of physical, mental and behavioral health variables among others considered relevant.

Keywords: Depression, Elderly Population, Covid-19, Logistic Regression, Associations

Índice

Lista de Figuras	v
Lista de Tabelas	vi
Lista de Abreviaturas, Siglas e Acrónimos	ix
1. Introdução	1
2. Enquadramento da questão de investigação	3
2.1 A patologia depressão	3
2.2 A pandemia COVID-19 e o seu impacto.....	4
2.3 População Portuguesa: descrição, visualização gráfica e esperança média de vida....	7
2.4 Idade real e idade sentida	11
2.5 Depressão em Portugal: relação com a idade e com a pandemia COVID-19	12
2.6 O Projeto SHARE: Survey of Health, Ageing and Retirement in Europe	16
3. Metodologias de Análise Estatística: Seleção e descrição	18
3.1 Testes de hipóteses paramétricos para amostras emparelhadas	18
3.2 Métodos Não Paramétricos	18
3.2.1 Testes de Ajustamento de Distribuições	19
3.2.2 Teste para Comparação entre amostras emparelhadas	21
3.2.3 Teste para comparação de duas ou mais amostras independentes	24
3.3 Análise de tabelas de contingência.....	26
3.3.1 Teste de Qui-quadrado de Independência	27
3.3.2 - Teste de sobre o coeficiente Kappa de Cohen	28
3.3.3 Medidas e Testes de Associação.....	29
4. Modelos Lineares Generalizados	31
4.2 Características do Modelo Linear Generalizado	33
4.3 Modelo de Regressão Logística	34
4.4 Método da máxima verosimilhança	36
4.5 Inferência em Modelos Lineares Generalizados	37
4.5.1 Teste de Hipóteses e Análise de Desempenho	38
4.5.2 Seleção de Modelos.....	40

4.5.3 Avaliação do Modelo	40
4.5.4 Capacidade Preditiva do Modelo.....	42
4.5.5 Interpretação dos Coeficientes.....	44
5. Metodologia	46
6. Aplicação a dados reais do Projeto Europeu SHARE.....	48
6.1 Descrição da Amostra	48
6.2 Classificação das Variáveis	50
6.3 Caracterização da Amostra e Relação entre as Variáveis	52
6.5 – Modelo de Regressão Logística.....	92
7. Considerações, conclusões e sugestões para trabalhos futuros.....	102
8. Bibliografia	105

Lista de Figuras

Figura 2. 1: Pirâmide Etária Portuguesa em 2021	7
Figura 2. 2: Estimativa e projeções da população residente em Portugal 1991 – 2080 (Fonte: INE).....	8
Figura 2. 3:Gráfico de Linhas da Esperança de vida à nascença em Portugal, por sexo (Fonte: Pordata)	10
Figura 2. 4:Diagrama de Barras com a Percentagem de pessoas que reportaram depressão crónica em 2019.....	13
Figura 2. 5: Diagrama de Linhas de Perturbações depressivas, Demência, Perturbações da ansiedade, em Portugal entre 2011 e 2016.....	15
Figura 4. 1: Representação do Modelo Logístico com parâmetros $\beta_0=20$ e $\beta_1=0.2$	35
Figura 6. 1: Boxplot das variáveis idade e idade que sente ter, por género	56
Figura 6. 2: Resíduos deviance do modelo	97
Figura 6. 3: Distância de Cook.....	97

Figura 6. 4: Curva ROC	98
------------------------------	----

Lista de Tabelas

Tabela 3.1: Tabela Classificativa do nível de concordância (Fonte: Julius Sim, 2005)	28
Tabela 4. 1: Matriz de Confusão	42
Tabela 4. 2:Tabela classificativa de AUC (Fonte: Hosmer e Lemeshow (2000)).....	44
Tabela 6. 1: Tabela de apresentação e classificação das variáveis em estudo	50
Tabela 6. 2: Quadro resumo de estatísticas descritivas das questões demográficas	52
Tabela 6. 3: Quadro resumo de estatísticas descritivas das variáveis "idade" e "idade que sente ter" por género	55
Tabela 6. 4: Tabela de Frequências das questões de saúde física e comportamentos de saúde.....	57
Tabela 6. 5: Tabela de Contingência das variáveis "Percepção do Estado de Saúde" e Género	58
Tabela 6. 6:Tabela de Contingência das variáveis "Tem fratura do quadril" e Género	59
Tabela 6. 7: Tabela de Contingência das variáveis "Tem diabetes ou açúcar elevado no sangue" e género	60
Tabela 6. 8: Tabela de Contingência das variáveis " Tem tensão alta ou hipertensão" e género.....	61
Tabela 6. 9: Tabela de Contingência das variáveis "Teve um enfarte agudo do miocárdio ou tem outro problema de coração" e Género	62
Tabela 6. 10: Coeficiente de Contingência de Pearson entre as variáveis "Teve um enfarte agudo do miocárdio ou tem outro problema de coração" e género	63
Tabela 6. 11: Tabela de Contingência das variáveis "Tem doença pulmonar crónica" e género.....	63
Tabela 6. 12: Coeficiente de Contingência de Pearson entre as variáveis "Tem doença pulmonar crónica" e género	64

Tabela 6. 13: Tabela de Contingência das variáveis "Tem cancro ou um tumor maligno" e Género	64
Tabela 6. 14: Tabela de Contingência das variáveis "Tem outra doença ou condição de saúde" e género	65
Tabela 6. 15: Tabela de Contingência das variáveis "Toma medicação regular" e género	66
Tabela 6. 16: Tabela de Contingência da variável "Saiu de casa nos 3 últimos meses" e Género	67
Tabela 6. 17: Tabela de Contingência da variável "Tomou Medicação Preventiva para a COVID-19" e Género	68
Tabela 6. 18: Tabela de Contingência das variáveis "Está vacinado contra a COVID-19" e Género	69
Tabela 6. 19: Tabela de contingência das variáveis "Pretende ser vacinado contra a COVID-19" e Género	70
Tabela 6. 20: Tabela de Frequências das questões de saúde mental	72
Tabela 6. 21: Tabela de Contingência das variáveis "No último mês sentiu-se nervoso, ansioso ou no limite" e género	72
Tabela 6. 22: Coeficiente de Contingência de Pearson entre as variáveis "No último mês sentiu-se nervoso, ansioso ou no limite " e género	73
Tabela 6. 23: Tabela de contingência das variáveis "No último mês esteve triste ou deprimido" e género	74
Tabela 6. 24: Coeficiente de Contingência de Pearson entre as variáveis "No último mês esteve triste ou deprimido " e género	75
Tabela 6. 25: Tabela de contingência das variáveis "Teve dificuldades em dormir recentemente" e Género	75
Tabela 6. 26: Coeficiente de Contingência de Pearson entre as variáveis "Teve dificuldades para dormir recentemente" e género	76
Tabela 6. 27: Tabela de contingência das variáveis "Sente-se Solitário" e Género	77
Tabela 6. 28: Coeficiente de Contingência de Pearson entre as variáveis "Sente-se solitário " e género	77
Tabela 6. 29: Tabela de Frequências de questões sobre a COVID-19	78

Tabela 6. 30: Tabela de Contingência das variáveis "Quantas vezes testou positivo para a COVID-19" e gênero.....	78
Tabela 6. 31: Tabela de Frequências de questões sobre a situação profissional.....	79
Tabela 6. 32: Tabela de Contingência entre as variáveis "Situação Atual de Emprego" e Gênero	80
Tabela 6. 33: Quadro de resultado do coeficiente de contingência de Pearson entre as variáveis "Situação atual de emprego" e gênero	81
Tabela 6. 34: Tabela de Frequências de questões sobre fatores sociais.....	81
Tabela 6. 35: Tabela de Contingência entre as variáveis "Durante o surto de COVID-19 utilizou pelo menos 1 vez a internet para fazer compras, procurar informação, entre outros" e gênero.....	82
Tabela 6. 36: Coeficiente de contingência de Pearson entre as variáveis "Durante o surto de COVID-19 utilizou pelo menos 1 vez a internet para fazer compras, procurar informação, entre outros" e gênero	83
Tabela 6. 37: Caracterização da Variável dependente com as questões demográficas	84
Tabela 6. 38: Caracterização da Variável dependente com as questões de saúde física e comportamentos de saúde.....	86
Tabela 6. 40: Caraterização da Variável Dependente com questões de saúde mental	89
Tabela 6. 41: Caracterização da variável dependente com questões sobre a COVID-19....	90
Tabela 6. 42: Caracterização da variável dependente com questões sobre a situação profissional.....	91
Tabela 6. 43: Caracterização da variável dependente com questões sobre fatores sociais	91
Tabela 6. 44: Sumário do modelo ajustado	94
Tabela 6. 45: Multicolinearidade	95
Tabela 6. 46: Matriz de Confusão do Modelo	98
Tabela 6. 47: Estimativas dos coeficientes e Odds Ratio.....	100

Lista de Abreviaturas, Siglas e Acrónimos

H₀ – Hipótese Nula

H₁ – Hipótese Alternativa

INE – Instituto Nacional de Estatística

MLG – Modelo Linear Generalizado

OMS – Organização Mundial de Saúde

SHARE - *Survey of Health, Ageing and Retirement in Europe*

SNS – Sistema Nacional de Saúde

SPSS - *Statistical Package for the Social Sciences*

1. Introdução

Com vista a explorar de modo mais profundo a aprendizagem dos Modelos Lineares Generalizados apresentados no decorrer da parte curricular do Mestrado em Bioestatística e Biometria da Universidade Aberta, bem como a sua aplicabilidade, foi considerado um problema de Saúde Pública que no contexto da pandemia mundial de COVID-19 atingiu o mundo de uma forma global expandindo-se a um ritmo muito rápido e inesperado. Deste modo, questionou-se a forma como esta pandemia veio impactar a saúde mental dos cidadãos portugueses, ou seja, qual o impacto no sentimento de depressão expresso pelos indivíduos em particular na população adulta, considerando a idade real e a idade sentida.

A base de dados que permitiu realizar este estudo foi obtida através do programa europeu SHARE - *Survey of Health, Ageing and Retirement in Europe*, que é um projeto multidisciplinar entre países europeus e que fornece dados relacionados com a saúde, estatuto socio-cultural, redes sociais e redes familiares.

Assim sendo, com recurso à base de dados recolhida, procede-se em primeiro lugar à sua análise exploratória, usando medidas de redução de dados e visualização gráfica. De seguida, visando o objetivo principal, são exploradas metodologias de modelação estatística que permitam ajustar um modelo para a variável binária representativa do estado de depressão sentido (com e sem) de acordo com as variáveis mais pertinentes para o efeito.

Uma vez verificado que os dados não cumprem o requisito da normalidade e devido ao facto da variável dependente ser dicotómica, surgem os Modelos Lineares Generalizados, tendo sido utilizado o Modelo de Regressão Logística.

De modo a alcançar os objetivos de investigação apresentados, esta dissertação foi estruturada em três pontos principais: apresentação de conteúdos estatísticos relevantes para o estudo, apresentação e explicação da metodologia adotada, apresentação dos resultados obtidos e por fim são apresentadas as principais conclusões deste estudo.

Primeiramente, na seção da apresentação de conteúdos estatísticos são abordados alguns conceitos de estatística não paramétrica e os Modelos Lineares Generalizados,

nomeadamente o Modelo de Regressão Logística. Este enquadramento realizado através da revisão de literatura permitiu aprofundar conhecimentos teóricos relevantes de modo a fundamentar as metodologias utilizadas neste estudo.

Na seção seguinte é feita uma apresentação e explicação da metodologia adotada, com o intuito de explicar detalhadamente o estudo realizado, isto é, são apresentadas as variáveis selecionadas, é feita a caracterização da amostra e são apresentados os procedimentos estatísticos utilizados.

Posteriormente são apresentados os resultados obtidos com a aplicação de métodos de Regressão Logística a uma base de dados proveniente do programa SHARE. A base de dados selecionada é constituída por 947 observações válidas relativas à população portuguesa. No questionário que serviu para obtenção desta base de dados são efetuadas questões que permitem obter variáveis de teor socioeconómico, de saúde física e mental, entre outras.

Por fim, são apresentadas as principais conclusões do estudo, assim como algumas limitações encontradas e sugestões para trabalhos futuros.

2. Enquadramento da questão de investigação

2.1 A patologia depressão

A depressão é uma doença grave que afeta negativamente a forma como o indivíduo se sente, pensa, e age - com falta de interesse por atividades, constituindo uma das principais causas de incapacidade em todo o mundo. Como tal, contribui para o aumento global de doenças e conseqüente sobrecarga nos Sistemas de Saúde. Esta patologia afeta todas as faixas etárias, mas parece ser mais evidente nas faixas etárias mais elevadas. Assim, uma população cuja demografia revela uma elevada taxa de idosos, incorre, conseqüentemente, num maior risco de vir a ter de lidar com problemas causados por distúrbios psiquiátricos geriátricos. Os idosos, em geral, enfrentam vários desafios conseqüentes de alterações físicas e psicológicas associadas ao processo natural de envelhecimento, pelo que, com a perspetiva de aumento da longevidade, se teme que a incidência de problemas de saúde mental venha a aumentar nas populações mais idosa e nos adultos em geral.

A depressão revela-se um preponderante fator de risco para incapacidade e mortalidade (Yaka et. al,2014). Nas faixas etárias mais elevadas esse risco é agravado, pelo facto de muitas vezes a depressão não ser corretamente diagnosticada como tal, confundindo-se os seus efeitos com os do envelhecimento natural, o que pode vir a interferir com a qualidade de vida do idoso. Talvez também seja de algum modo graças a isso, que na literatura se encontram vários estudos que mostram para este assunto taxas de prevalência diferentes e até inconsistentes.

Os fatores de risco para a depressão incluem alterações hormonais, situações de grande stress, eventos adversos – a nível familiar, profissional, social ou outros, dificuldade em dormir, competências desajustadas, excesso de atividades a executar, fortes exigências no cumprimento de prazos. Um fator de risco muito importante foi evidenciado na conseqüência das medidas de confinamento recentemente impostas pelo surgimento da pandemia COVID19, nomeadamente, o isolamento. Este fator impactou negativamente sobretudo na população idosa, constatando-se um agravamento substancial do número de doentes idosos com a patologia da depressão (INSA, 2020).

As manifestações de sintomas de depressão mais usuais incluem reações negativas, irritabilidade permanente e agressividade, o que muitas vezes, e não sendo devidamente interpretado como doença, pode conduzir à rejeição social e ao recurso a outras medidas menos desejáveis.

Com este trabalho de investigação pretende-se, como objetivo principal, a construção de um modelo fidedigno que tem como variável dependente a ocorrência de depressão, como percepção do indivíduo, e como variáveis explicativas algumas variáveis correspondentes a fatores pertinentes associados usualmente à idade mais avançada, como a existência de algumas doenças, e situações sociais.

2.2 A pandemia COVID-19 e o seu impacto

Em janeiro de 2020, em Wuhan (China), foi detetado um novo coronavírus, denotado por SARS-CoV-2 e responsável pela doença Covid-19. Rapidamente a OMS decretou Emergência de Saúde Pública de Âmbito Internacional e posteriormente classificou esta doença como pandemia, devido a sua rápida propagação e a elevada taxa de mortalidade associada, especialmente em pessoas com comorbilidades. Por se tratar de uma infeção respiratória a sua transmissão ocorre principalmente através de aerossóis e gotículas de secreções respiratórias. Em pessoas infetadas, os sintomas apresentados são os de uma infeção respiratória, isto é, tosse, febre, dificuldade respiratória, dores musculares e cansaço. Em situações de carácter mais grave, o quadro clínico poderá agravar-se conduzindo a uma pneumonia com insuficiência respiratória, falência de órgãos e eventual morte.

Atendendo à escalada da doença a nível mundial e rápida propagação da doença entre pessoas, foi necessário, em vários países, recorrer a confinamentos e a outras medidas sanitárias, como uso de máscara e desinfeção das mãos com álcool gel, na tentativa de evitar a evolução deste severo problema de saúde. A ocorrência da pandemia COVID-19 deu origem a uma crise a todos os níveis, não apenas na saúde, mas também a nível social e económico (Silva et. al, 2021).

No mundo atual e à medida que os países foram sendo afetados pelo COVID-19, foi aumentando o problema de isolamento na população mais idosa. Na tentativa de proteger

os maiores de 70 anos, e de procurar aliviar os sistemas de saúde demasiado sobrecarregados, assistiu-se à imposição de regras de confinamento, recolher obrigatório e isolamento social para mitigar a propagação da síndrome respiratória aguda grave pelo coronavírus (SARS-CoV-2). No entanto sabe-se que o isolamento social entre idosos constitui um grave problema de saúde pública, devido ao seu maior risco de problemas de saúde cardiovasculares, autoimunes, neuro-cognitivos e sobretudo de saúde mental. Santini et. al (2020) demonstraram que a desconexão social coloca os idosos em maior risco de depressão e ansiedade. O isolamento afeta desproporcionalmente os indivíduos idosos cujo único contacto social é fora de casa, em comunidades, centros e locais de culto. Aqueles que não têm família próxima ou amigos e que contam com o apoio de serviços voluntários ou assistência social podem ser colocados sob um risco adicional, tal como aqueles que já se encontram sozinhos, isolados ou excluídos.

É neste contexto que as tecnologias online podem e têm sido aproveitadas para fornecer suporte social, redes e um sentimento de pertença, embora possa haver disparidades no acesso ou na literacia em recursos digitais. As intervenções podem envolver simplesmente o contacto telefónico mais frequente com familiares próximos, amigos, organizações voluntárias, profissionais de saúde ou projetos de extensão comunitária, oferecendo suporte de pares durante os períodos de isolamento forçado.

Atzendorf & Gruber (2021) reforçam a ideia de que as medidas de controlo epidémico que visam introduzir o distanciamento social, ajudam a desacelerar a propagação da pandemia de COVID-19. Contudo, as suas consequências em termos de bem-estar mental podem ser negativas, principalmente para os idosos. No artigo mencionado e que se intitula “Depression and Loneliness of older adults in Europe and Israel after the first wave of covid-19”, os autores utilizaram dados da pesquisa “SHARE Corona”, constituída por aposentados entrevistados com 60 anos ou mais, de 25 países europeus e Israel, combinados com dados macro do “Oxford COVID-19 Government Response Tracker” que permite incluir macro indicadores no nível do país, ou seja, o número de mortes por cada 100000 indivíduos e o número de dias com medidas rigorosas de controle da epidemia, para além de diversas características individuais. Os resultados obtidos com este estudo mostram que ambos os

macro indicadores são influentes para aumentar os sentimentos de tristeza/depressão, mas que os fatores individuais são cruciais para explicar o aumento dos sentimentos de solidão após o primeiro confinamento. Modelos com termos de interação revelaram que os macro indicadores incluídos têm consequências negativas para o bem-estar, particularmente para os participantes mais velhos da pesquisa. Além disso, os resultados revelaram que especialmente aqueles que vivem sozinhos tiveram um risco mais elevado do aumento da solidão no tempo e conseqüentemente de depressão, após a primeira onda de COVID-19.

Mistry, et al., (2021) referem-se também à depressão como um dos transtornos mentais mais comuns, e revelam a sua preocupação com o aumento desta doença em todo o mundo provocado pela pandemia da doença de coronavírus 2019 (COVID-19), especialmente entre a população mais velha. O objetivo do seu estudo consistiu na medição dos sintomas depressivos prevalentes e seus associados entre adultos mais velhos, face à pandemia de COVID-19 em Bangladesh. A amostra transversal contou com 1.032 idosos de Bangladesh, com idade maior ou igual a 60 anos, os quais responderam a entrevistas telefônicas em outubro de 2020. Foi usado um questionário semiestruturado que permitiu recolher dados sobre características sociodemográficas e de estilo de vida dos participantes, o histórico de condições médicas pré-existentes e informações relacionadas à COVID-19. Os sintomas depressivos foram medidos usando a conhecida Escala de Depressão Geriátrica de 15 itens (GDS-15). Os autores recorreram ao modelo de regressão logística binária para identificar os principais fatores associados aos sintomas depressivos.

Zhao et al. (2023) apresentam um estudo realizado na China incidente na população idosa, que teve como objetivo examinar o papel mediador do isolamento social entre a mobilidade física e a função cognitiva e se existem diferenças quanto ao gênero nos efeitos mediadores. Com esse estudo, os autores concluem que, assim que haja condições para tal, reverter as medidas de isolamento social deve ser encarado como intervenção prioritária visando a prevenção do declínio cognitivo na população idosa e a promoção do envelhecimento com a qualidade de vida desejada.

2.3 População Portuguesa: descrição, visualização gráfica e esperança média de vida

Conhecer a demografia de um país revela-se, cada vez mais, um fator de elevada importância, uma vez que um bom conhecimento das idades e géneros dos indivíduos que habitam no país permite que sejam tomadas decisões e medidas, de forma atempada de modo a evitar constrangimentos em serviços e que o tipo de serviços prestados sejam adequados ao tipo de população residente.

Segundo dados estatísticos divulgados pelo INE (Instituto Nacional de Estatística), resultantes dos censos 2021, a população portuguesa é de aproximadamente 10 345 habitantes. De modo a compreender melhor o tipo de população que atualmente reside no nosso país, é apresentada a pirâmide etária para o ano mencionado.

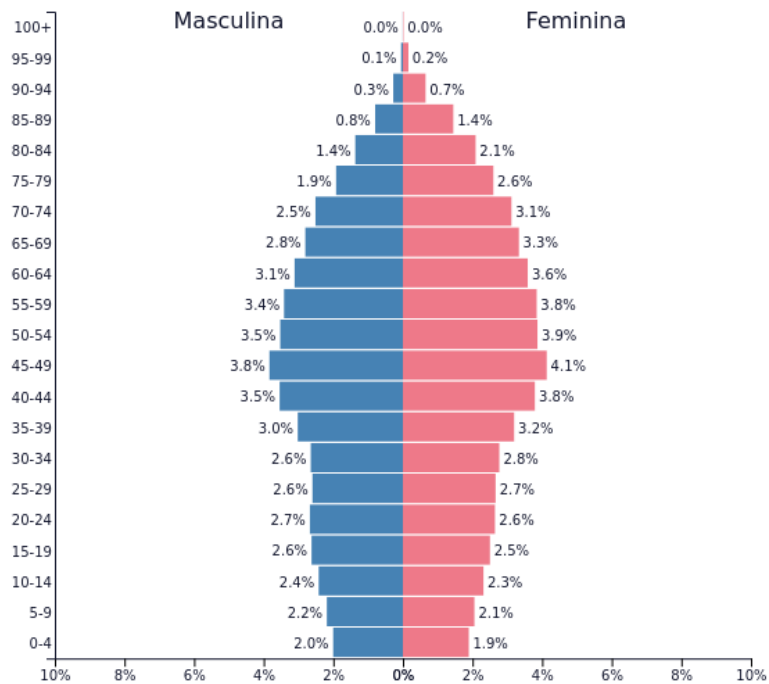


Figura 2. 1: Pirâmide Etária Portuguesa em 2021

Na figura 2.1, verifica-se a prevalência de uma população com idade superior a 45 anos o que revela o envelhecimento da população. Paralelamente, verifica-se que dificilmente ocorrerá nos próximos anos uma renovação da população, uma vez que as classes etárias mais baixas apresentam um reduzido número de cidadãos, o que evidencia que a longo prazo deverá ocorrer uma redução da população residente em Portugal. No entanto, com a crise gerada pela guerra Rússia-Ucrânia, cada vez se assiste mais à deslocação de

refugiados para o nosso País, principalmente jovens, o que poderá a médio-longo prazo vir a contrariar esta tendência.

Na figura 2.2 encontram-se representados os possíveis cenários para a evolução da população residente em Portugal até ao ano de 2080.

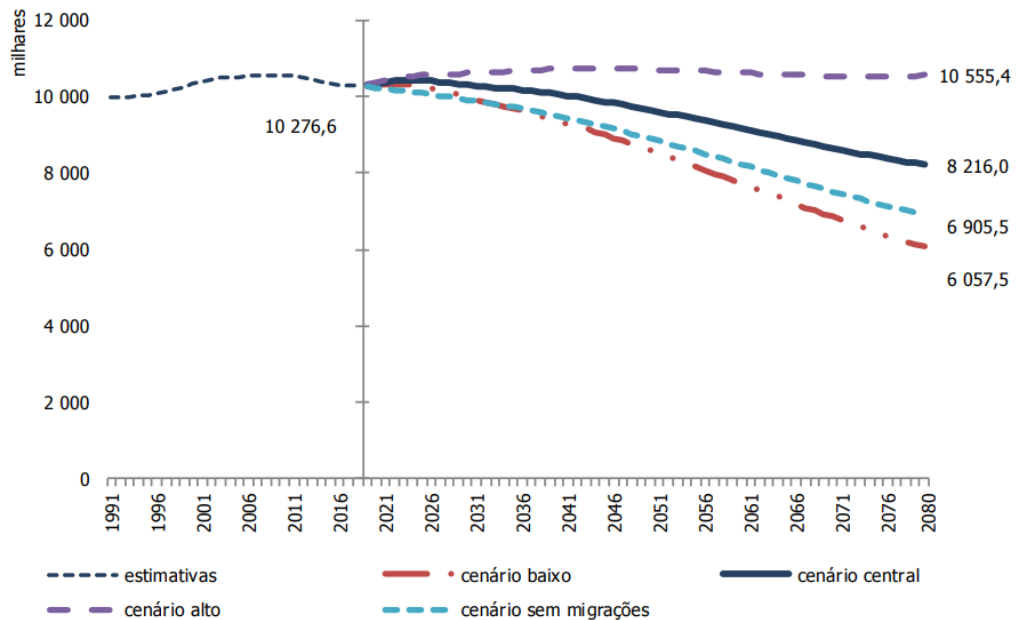


Figura 2. 2: Estimativa e projeções da população residente em Portugal 1991 – 2080 (Fonte: INE)

De acordo com as Projeções de População Residente em Portugal realizadas pelo INE a 30 de março de 2020, representadas na figura 2.2, baseadas na conjugação de elementos como fecundidade, mortalidade e migrações é possível estabelecer diferentes cenários de variação para a população portuguesa. Os cenários possíveis são:

- I. Cenário Baixo: neste cenário são consideradas as hipóteses mais pessimistas para os indicadores de fecundidade, mortalidade e migrações;
- II. Cenário Central: neste cenário são consideradas as hipóteses de evolução central para os indicadores de fecundidade, mortalidade e migrações;
- III. Cenário Alto: neste cenário são consideradas as hipóteses de evolução mais otimista para os indicadores de fecundidade, mortalidade e migrações;

IV. Cenário Sem Migrações: neste cenário é baseado no cenário central, mas sem considerar as migrações.

As projeções da população são realizadas com base nas estimativas anuais provisórias da população residente. O método de cálculo utilizado é dado por:

$$P_t = P_{t-1} + NV_t - Ob_t + I_t - E_t \quad (1)$$

onde,

P_t representa a população obtida para o ano t;

P_{t-1} representa a população no ano t-1;

NV_t representa o número de nados vivos estimado no ano t;

Ob_t representa o número de óbitos estimado no ano t;

I_t representa o número de imigrante estimado no ano t;

E_t representa o número de emigrantes estimado no ano t;

Prevê-se que no cenário central a população residente tenderá a diminuir, tomando valores próximos os 8,2 milhões de habitantes em 2080. Caso seja este o cenário correto, estima-se que já em 2042 a população portuguesa seja inferior a 10 milhões. Segundo o cenário baixo, também apresentado na figura anterior, o decréscimo populacional poderá ser ainda mais acentuado. Tal facto é justificado pelos reduzidos níveis de fecundidade e saldos migratórios. Contrariamente ao esperado através dos restantes cenários, pelo cenário alto é esperado um aumento populacional, possível através do aumento acentuado dos níveis de fecundidade e de um elevado e positivo saldo migratório.

O conceito de esperança média de vida à nascença é o número de anos, em média, que uma pessoa à nascença pode esperar viver, mantendo-se as taxas de mortalidade por idades observadas no momento de referência. Segundo dados do INE, no triénio 2019-2021

a esperança de vida à nascença situa-se, em Portugal nos 77,67 anos para os indivíduos do sexo masculino e em 83,37 anos para os cidadãos do sexo feminino. Tendo sido verificado um aumento de 1,90 e 1,48 anos, para os indivíduos do sexo masculino e feminino, respetivamente, face aos valores de 2008-2010.

Com base nos dados fornecidos pela plataforma PORDATA, foi possível gerar o gráfico apresentado em seguida na Figura 2.3.

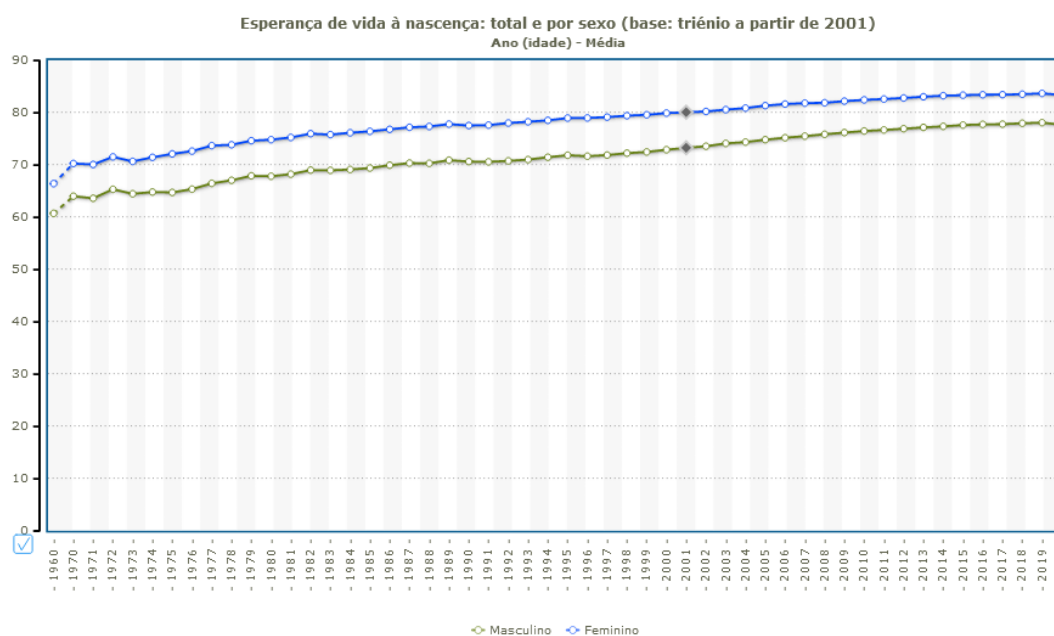


Figura 2. 3:Gráfico de Linhas da Esperança de vida à nascença em Portugal, por sexo (Fonte: Pordata)

Através do gráfico apresentado na figura 2.3, onde se encontra representada a esperança média de vida por género em Portugal, é notório que a esperança média de vida é bastante superior nas mulheres.

É importante frisar que, embora se verifique um aumento da esperança média de vida face ao primeiro ano considerado no gráfico, verifica-se um ligeiro decréscimo face ao ano anterior (2019). Segundo o INE esta ligeira diminuição deve-se ao elevado número de óbitos que se registaram no decorrer da pandemia de COVID-19, o que conduziu a uma redução da esperança média de vida na maioria do território nacional.

Em Portugal, o acompanhamento epidemiológico da COVID-19 tem sido assegurado pelo Instituto Ricardo Jorge, através do Departamento de Epidemiologia em colaboração com a Direção Geral de Saúde, Faculdade de Ciências da Universidade de Lisboa e a Universidade de Trás-os-Montes e Alto Douro. Estas entidades disponibilizaram relatórios de informação aos decisores do Ministério da Saúde, responsáveis pela gestão e planeamento de combate à pandemia. Durante o período de incidência mais preocupante da pandemia, parte da informação produzida por estas entidades foi divulgada semanalmente, através de um relatório sobre a curva epidémica, ao qual são adicionados os parâmetros de transmissibilidade como complemento informativo

2.4 Idade real e idade sentida

É sabido que a idade real de um indivíduo é determinada pela diferença entre a data atual e a data de nascimento. O conceito de idade sentida, também designada, segundo alguns autores, como idade percebida ou subjetiva, apresenta uma designação mais geral e alargada. Segundo Dieht et al. (2015), a idade sentida reflete a disparidade entre a idade real do indivíduo e a idade que este sente ter. Por outro lado, a idade ideal, é segundo Ward (2010), a idade que o indivíduo deseja ter comparativamente à sua idade real. Contudo, esta variável é muito pouco estudada, uma vez que não apresenta tanta pertinência, devendo ser dada prioridade, quando possível, ao estudo da variável idade sentida (Keyes & Westerhof, 2012). Dado que existe pouca informação fornecida por estas variáveis, são vários os estudos dedicados à procura de uma alternativa que permita responder, de forma mais fundamentada à passagem do tempo, de modo a compreender e quantificar a perceção do indivíduo sobre o seu próprio envelhecimento, Diehl et al.(2015).

O facto de um indivíduo poder refletir sobre o seu próprio envelhecimento, traduz-se numa grande mais-valia, visto que permite observar a relação entre a capacidade de reflexão e o funcionamento físico, psicológico e social, Westerhof & Wurn (2015). Ward (2010) justifica a existência da diferença entre a idade sentida e a idade real, em situações para as quais o indivíduo não se identifica com a sua idade real.

Existem vários estudos a nível internacional na área da idade sentida, embora seja escasso o número de estudos publicados que evidenciem a relação entre a idade sentida e doenças

diversas, entre as quais a depressão. De modo a prevenir a patologia de depressão na população idosa, Fiske, et al. (2009) afirmam ser muito importante refletir precisamente acerca desta relação entre a idade sentida e a depressão.

Em Keyes & Westerhof (2012) é realizado um estudo onde se verifica a existência de uma relação residual entre a idade sentida e a depressão, contudo as correlações encontradas são distintas quando se considera a idade sentida ou a idade real. Quanto à primeira, apresenta uma relação com o bem-estar mental e com o episódio de depressão major, já com a segunda verifica-se uma relação única com o bem-estar mental.

2.5 Depressão em Portugal: relação com a idade e com a pandemia COVID-19

A Depressão em Portugal afeta cerca 10% dos cidadãos e é, segundo a Ordem dos Psicólogos, um problema de Saúde Pública que afeta toda a população de forma direta ou indireta. A Organização Mundial de Saúde, descreve este flagelo como um dos problemas de saúde mais frequentes, referindo também que considera a depressão como a principal causa de incapacidade no mundo.

Segundo estudo realizado pelo PORDATA em 2017, sobre como adoecem os portugueses, a depressão era a segunda patologia que mais afetava a população portuguesa, atingindo cerca de 87 milhares de indivíduos.

Em 2019, antes da pandemia COVID-19, segundo dados do Eurostat, Portugal era o segundo país europeu com maior percentagem de cidadãos a sofrer de depressão, cerca de 12%, uma percentagem bastante superior à percentagem média registada na União Europeia. É importante referir que são as mulheres portuguesas aquelas que mais afirmam sofrerem de depressão crónica (16,4%), seguidas das eslovas (16%). (Figura 2.4)

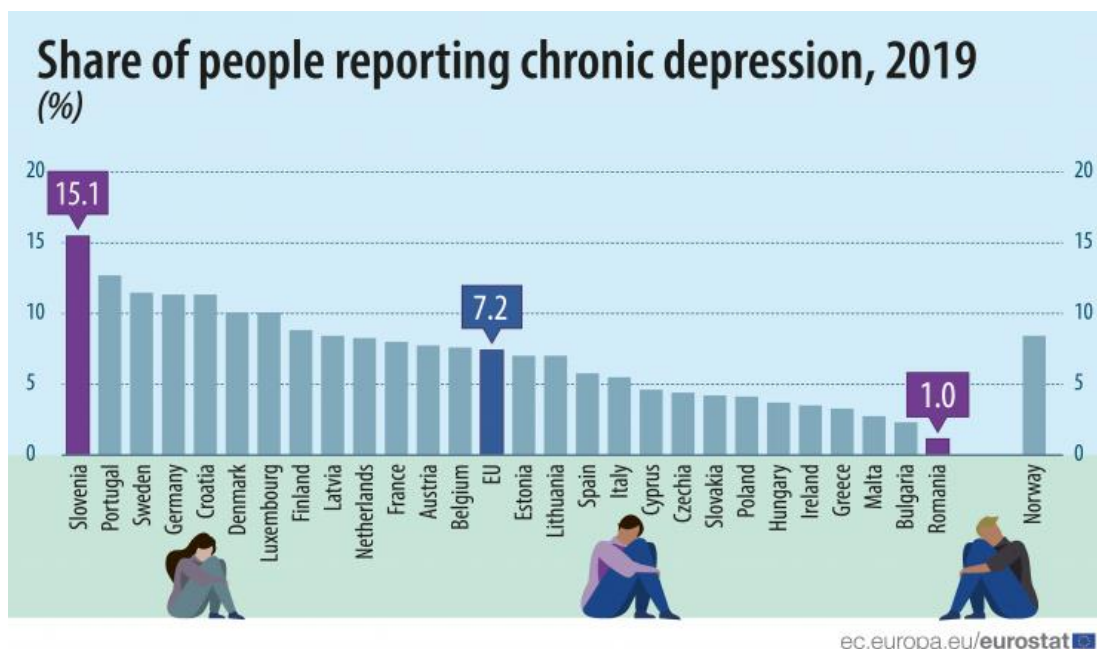


Figura 2. 4: Diagrama de Barras com a Percentagem de pessoas que reportaram depressão crónica em 2019

Durante a pandemia de COVID-19, à semelhança do que se verificou com as demais patologias, também os cuidados de saúde mental acabaram por ser condicionados, o que conduziu a um aumento de doentes com patologia de depressão ou stress pós-traumático. Segundo um estudo realizado no âmbito europeu, Portugal: Perfil de Saúde do País 2021, a pandemia causou um elevado sofrimento psicológico na população, sendo que um inquérito realizado a nível nacional, entre maio e julho de 2020, revelou que 27% da população apresentava sintomas de ansiedade moderada a grave e cerca de 26% apresentava sintomas compatíveis com stress pós-traumático. É nos cidadãos do sexo feminino com idades compreendidas entre os 18 e os 29 anos, nos desempregados e nos indivíduos que auferem baixos rendimentos, que as taxas de sofrimento psicológico tendem a ser classificadas de moderado a grave.

Segundo investigação no âmbito do projeto de doutoramento em Saúde Pública de Ana Aguiar, investigadora do ISPUP, a pandemia afetou a saúde mental das populações nos diferentes subníveis, atingido também a população com idades entre os 19 e os 39 anos, bem como os indivíduos com maior nível de literacia. Assim, é essencial identificar os

cidadãos mais propensos a desenvolver este tipo de patologias nos diferentes estratos populacionais. É igualmente necessário informar e desmitificar o campo da saúde mental.

Uma notícia avançada pela Agência Lusa em abril de 2022, refere que a pandemia causou um agravamento de questões decorrentes de ansiedade, depressão e dependência de álcool, sendo estimado pela Comissão de Acompanhamento do Plano Nacional de Saúde 2021-2030 que as necessidades de tratamentos na área da saúde sofreram um aumento decorrente destas questões. Os *stakeholders* consideram que estas necessidades poderão vir a ser decorrentes de determinantes demográficos, como a cobertura universal dos cuidados de água, desemprego, emprego precário, isolamento social, ou de determinantes comportamentais, como o uso em excesso de *internet*.

O trabalho conduzido pelo SNS e pelo Instituto Ricardo Jorge sobre o Estudo da Saúde Mental em Tempos de Pandemia, refere que cerca de 25% dos inquiridos apresentam sintomas moderados a graves de ansiedade, depressão e stress pós-traumático, resultado que vai de encontro a outros estudos realizados a nível mundial. Esse estudo teve como principal objetivo avaliar o impacto da pandemia na saúde mental e no bem-estar da população em geral e dos profissionais de saúde em particular. De acordo com os resultados obtidos, na população em geral são as mulheres e os jovens adultos aqueles que apresentam mais sintomas de depressão e ansiedade. Já no que diz respeito aos profissionais de saúde, 42% daqueles que se encontram em contacto direto com pacientes COVID19 apresentam sintomatologia de ansiedade moderada a grave, sendo que, ainda neste grupo, 43% apresentam elevados níveis de burnout. O mesmo estudo afirma que os indivíduos que conseguiram, durante o tempo de pandemia, manter passatempos, hobbies e a sua rotina diária apresentam um nível de ansiedade mais baixo, por isso, essas tarefas de ocupação são consideradas como possível fator de proteção à ocorrência de depressão.

É importante mencionar que segundo dados da OMS das principais causas de suicídio constam as patologias do foro psiquiátrico. Sabe-se que as mortes por suicídio chegam as 800 mil por ano em todo o mundo. Para a Liliana Paixão, médica psiquiatra "A depressão mata". Esta especialista refere que a depressão não tratada pode levar à negligência dos próprios cuidados, como por exemplo de higiene e de alimentação, ao ponto de a vida ficar

em risco. Pode ainda levar à negligência dos cuidados de dependentes, como por exemplo os filhos, cujo crescimento ficará afetado a todos os níveis pela doença de um dos pais (ou dos dois). (noticia Sapo *Life Style*)

Segundo dados do INE, referentes a 2019, a taxa de suicídio em Portugal é de 9,7 por cada 100 mil habitantes, ou seja, quase 10 cidadãos por cada 100 mil colocam fim à sua própria vida. Esta percentagem é cerca de quatro vezes mais elevada no sexo masculino (15,5) do que no sexo feminino (4,4). Na população idosa os resultados são bastante preocupantes uma vez que, num cidadão de idade superior a 65 anos, se regista uma taxa de suicídio de 32,6 no sexo masculino e de 8,1 no sexo feminino. Segundo dados do Pordata, em 2019 existiam em Portugal cerca de 2 262 mil pessoas com mais de 65 anos, sendo que se registaram cerca de 900 suicídios por ano, dos quais 700 foram cometidos por homens.

De acordo com a Direção Geral de Saúde, em Portugal as perturbações mentais e do comportamento apresentam uma forte influência no total de anos de vida saudável, sendo que estas perturbações representam 20,55% do total de anos vividos com incapacidade, ver figura 5.

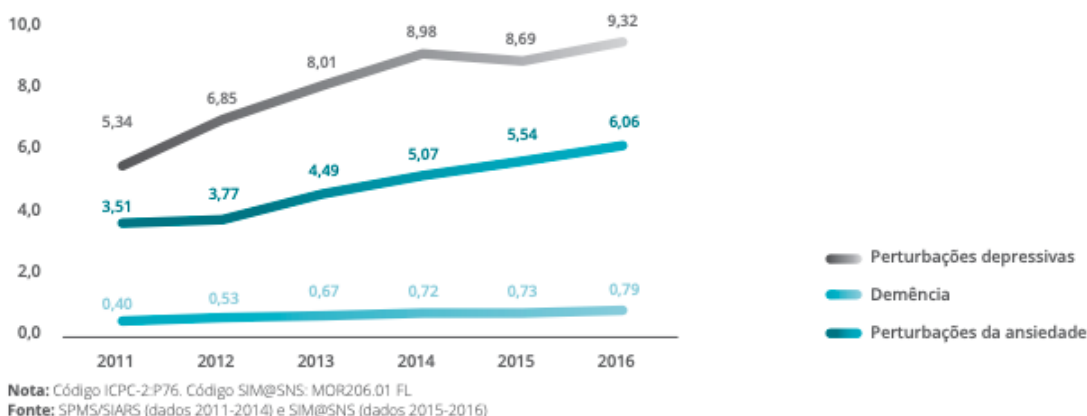


Figura 2. 5: Diagrama de Linhas de Perturbações depressivas, Demência, Perturbações da ansiedade, em Portugal entre 2011 e 2016

2.6 O Projeto SHARE: Survey of Health, Ageing and Retirement in Europe

No âmbito desta dissertação recorreremos aos dados do projeto SHARE para aplicação a dados reais, respeitantes ao inquérito realizado em Portugal no ano de 2021, numa fase em que a pandemia COVID estava instalada a nível mundial. O SHARE é um projeto europeu cujo objetivo é dar resposta ao apelo feito pela Comissão Europeia e que visa promover a pesquisa e estudar o impacto das políticas de saúde, sociais, económicas e ambientais ao longo da vida de cidadãos europeus. Desde 2004 até à atualidade foram inquiridos 140 000 cidadãos com mais de 50 anos em 28 países europeus e em Israel. Em Portugal este projeto é liderado pelo Instituto de Ciências Sociais da Universidade do Minho, sendo cofinanciado pela Fundação Calouste Gulbenkian e pela Fundação para a Ciência e Tecnologia. A nível europeu a coordenação é feita pelo *Munich Center for the Economics of Aging, Max Planck Institute for Social Law and Social Policy*. Este projeto está em harmonia com o *U.S. Health and Retirement Study* e o *English Longitudinal Study of Ageing*, sendo considerado um modelo de referência para várias pesquisas sobre envelhecimento em todo o mundo.

No ano de 2008 o projeto SHARE foi selecionado para integrar o *European Strategy Forum on Research Infrastructures*. Em 2011, e devido à sua importância a nível europeu, o projeto SHARE adquiriu um novo estatuto legal, formando o primeiro *European Research Infrastructure Consortium*, ao qual Portugal pertence como membro observador.

Segundo informação que consta do site oficial, a importância científica do SHARE prende-se particularmente com o facto de se tratar de um projeto longitudinal, que capta o carácter dinâmico do processo de envelhecimento. A sua dimensão multidisciplinar fornece uma perspetiva global deste processo enquanto procedimentos e práticas rigorosos, ao nível da conceção e implementação do projeto, assegurando a harmonização dos resultados.

Os inquéritos utilizados no projeto SHARE encontram-se disponíveis, em <http://www.share-project.org/home0.html> e são implementados com base no *Computer Assisted Personal Interview*. O inquérito que será alvo de análise nesta dissertação, e que se encontra em anexo neste documento, aborda o tema da saúde mental durante a

pandemia de COVID-19 e foi realizado em Portugal entre 5 de junho de 2021 e 4 de agosto de 2021.

3. Metodologias de Análise Estatística: Seleção e descrição

3.1 Testes de hipóteses paramétricos para amostras emparelhadas

De modo a testar diferença entre dois valores médios em amostras não independentes de igual dimensão é necessário recorrer às amostras emparelhadas. Considera-se D_1, \dots, D_n a amostra aleatória da diferença entre as amostras, de dimensão n .

As hipóteses a testar são:

$$H_0: \mu_D = k \quad Vs. \quad H_1: \mu_D \neq k$$

A hipótese alternativa (H_1), pode ser também unilateral direita, tendo-se $H_1: \mu_D > k$, ou unilateral esquerda, tendo-se $H_1: \mu_D < k$. A estatística de teste a considerar é dada por:

$$T = \frac{(\bar{D} - \mu_D)\sqrt{n}}{\widehat{\sigma}_D} \cap t_{n-1} \text{ sob } H_0 \text{ verdadeira} \quad (2)$$

Considerando o teste bilateral, a hipótese nula (H_0) deve ser rejeitada caso o valor observado da estatística de teste pertença a região crítica, isto é, caso $|t| < -t_{n-1, 1-\alpha/2}$.

Para proceder aplicação deste teste de hipóteses devem ser verificados os seguintes pressupostos:

1. As observações das duas amostras emparelhadas devem ser independentes entre si, isto é, as observações de uma amostra não devem influenciar as observações da outra;
2. A diferença entre o par de observações deve seguir assintoticamente uma distribuição normal
3. Deve existir homogeneidade de variâncias, isto é a diferença entre o par de observações deve apresentar variâncias iguais para todas as diferenças

3.2 Métodos Não Paramétricos

Os métodos não paramétricos são mecanismos de inferência estatística, sendo a sua designação devida ao facto de estes não terem como objetivo testar ou estimar parâmetros de uma dada distribuição, mas sim averiguar o ajustamento de determinadas funções aos dados, analisar a independência ou comparar uma ou mais distribuições. No estudo e

descrição da abordagem a estes testes foi tomado como referência o livro editado pela Universidade Aberta, de Oliveira, T. (2004).

Se por um lado os métodos paramétricos são utilizados em dados quantitativos exigindo o cumprimento dos pressupostos de normalidade, independência e homocedasticidade subjacente aos dados, o uso dos métodos não paramétricos não se encontra condicionado pela validação prévia do tipo de distribuição populacional. Posto isto, sempre que o pressuposto de normalidade é violado e é necessário proceder à comparação de médias de amostras quer independentes quer dependentes, os métodos não paramétricos apresentam uma especial importância. Contudo, apesar de os seus pressupostos serem mais flexíveis, os métodos não paramétricos apresentam a desvantagem de não apresentar uma potência de teste tão elevada, fazendo com que não seja detetado um número elevado de diferenças entre as amostras em estudo. Em Oliveira, T. (2004) consta que, quando comparados os métodos não paramétricos com os métodos paramétricos verifica-se que os primeiros apresentam as seguintes vantagens:

- I. Possibilidade de utilização em amostras de pequena dimensão;
- II. A não necessidade de verificação de pressupostos de normalidade subjacente à população de onde provém os dados da amostra em estudo;
- III. Alguns testes de hipóteses podem ser utilizados com dados qualitativos, como por exemplo, dados qualitativos ordinais;
- IV. Facilidade de aplicação.

São de seguida descritos os testes usados no âmbito da aplicação a dados reais nesta dissertação.

3.2.1 Testes de Ajustamento de Distribuições

Os testes de ajustamento permitem averiguar se uma determinada distribuição teórica se ajusta convenientemente aos dados em análise, permitindo assim concluir se a amostra é proveniente de uma população que segue essa determinada distribuição. Genericamente podem ser estabelecidas as seguintes hipóteses a testar:

$$H_0: A \text{ função densidade de probabilidade de } X \text{ é } f_0$$

Vs.

H_1 : A função densidade de probabilidade de X não é f_0

Seguidamente serão abordados os testes de ajustamento mais utilizados: Teste de ajustamento *Kolmogorov-Smirnov* e Teste de ajustamento de *Shapiro-Wilk*.

O teste de ajustamento de *Kolmogorov-Smirnov* (K-S), introduzido por Kolmogorov e Smirnov em 1933, é uma estatística não paramétrica que permite averiguar a igualdade de distribuições de probabilidade contínuas, como é o caso da distribuição Normal. As hipóteses a testar com a utilização deste teste são:

H_0 : X segue uma distribuição F_0 *Vs.* H_1 : X não segue uma distribuição F_0

Através do teste de K-S é possível avaliar a diferença, em módulo, das distâncias entre a função de distribuição empírica em se pretende testar e a função de distribuição da amostra, matematicamente tem-se: $|F_n(x) - F_0(x)|$. A estatística de teste associada é dada por:

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)| \quad (3)$$

É de salientar que este teste de ajustamento foi desenvolvido admitindo que na hipótese nula (H_0) é especificada a distribuição que se pretende verificar, em situações em que tal não se verifique é necessário recorrer a métodos de estimação de modo a estimar os parâmetros da distribuição. Contudo, do ponto de vista prático, este teste é utilizado em situações em que os parâmetros da distribuição são desconhecidos.

Em 1967, através de estudos desenvolvidos por *Lilliefors*, foram feitas correções e desenvolvidas novas tabelas para o caso de se tratar de uma distribuição normal de parâmetros desconhecidos, e anos mais tarde, em 1969 para o caso de se tratar de uma distribuição exponencial.

Um teste de ajustamento alternativo ao KS é o teste de *Shapiro-Wilk*, publicado em 1965 por Samuel Sanford Shapiro e Martin Wilk. Este teste de ajustamento permite verificar se

uma amostra, composta por variáveis aleatórias independentes e identicamente distribuídas (i.i.d), é proveniente de uma população normalmente distribuída. Posto isto as hipóteses a testar são:

$H_0: X$ segue uma distribuição Normal Vs. $H_1: X$ não segue uma distribuição Normal

A estatística de teste associada, segundo Shapiro & Wilk (1965) é dada pela seguinte expressão:

$$W = \frac{(\sum_{j=1}^n a_j X_j)^2}{\sum_{j=1}^n (X_j - \bar{X})^2} \quad (4)$$

onde,

\bar{X} : representa a média amostral;

a_j obtém-se através do cálculo do vetor $(a_1, \dots, a_j) = \frac{m^T V^{-1}}{C}$, com $C = \|V^{-1}m\|$, onde m é o vetor de valores esperados das variáveis aleatórias i.i.d. e V a respetiva matriz de covariância.

Estudos desenvolvidos por Razli et al. (2011) com recurso à estimação de Monte Carlo permite concluir que o teste de *Shapiro-Wilk* apresenta uma maior potência, para um nível de significância α , quando comparado com vários testes de normalidade, como é o caso do teste de *Kolmogorov-Smirnov*.

3.2.2 Teste para Comparação entre amostras emparelhadas

O estudo de amostras emparelhadas é bastante frequente quando se pretende averiguar se a exposição a duas situações distintas do mesmo individuo conduz a diferenças estatisticamente significativas. Por não existir independência entre as variáveis neste tipo de estudos recorre-se a testes para amostras emparelhadas, dos quais se destacam o Teste dos Sinais, o Teste de Wilcoxon e o Teste de Friedman.

O teste dos sinais surge no século XVII, sendo o teste não paramétrico mais antigo de que há conhecimento. Este teste permite proceder à comparação de duas amostras

correlacionadas, sendo utilizado quando se pretende averiguar se uma das variáveis (X, Y) é superior à outra.

Contrariamente ao que se verifica no teste paramétrico análogo, teste paramétrico para a comparação de médias, neste teste apenas é considerada a diferença dos valores observados de cada par (X, Y) . Define-se a diferença como $D_i = X_i - Y_i$. A sua fácil aplicação faz com que este teste não paramétrico apresente uma desvantagem que se prende com a significativa perda de informação inerente às diferenças (D_i) uma vez que apenas é considerado o seu sinal e não a sua amplitude.

A sua aplicação prática passa pelo seguinte processo, para cada um dos pares aleatórios (X, Y) é determinada a diferença (D_i) e posteriormente é atribuído um sinal a cada um destes valores através das seguintes normas:

- I. Se $X_i > Y_i$ verifica-se que X_i é preferido a Y_i , então tem-se que $D_i = X_i - Y_i > 0$, sendo por isso atribuído o sinal "+";
- II. Se $X_i < Y_i$ verifica-se que X_i é preterido a Y_i , então tem-se que $D_i = X_i - Y_i < 0$, sendo por isso atribuído o sinal "-";
- III. Se $X_i = Y_i$ verifica-se uma situação de empate, então tem-se que $D_i = X_i - Y_i = 0$, sendo por isso atribuído o valor 0;

Sob a hipótese nula verdadeira as variáveis X_i e Y_i são oriundas da mesma distribuição, sendo que existe igual probabilidade de obter uma diferença positiva e negativa, tendo-se: $H_0: P(X < Y) = P(X > Y) = 0.5$.

Posto isto considera-se T a variável aleatória que representa o número de sinais ocorridos um menor número de vezes e t_0 o respetivo valor observado. Posto isto é possível considerar, para amostras de dimensão inferior a 20, que a estatística de teste é dada por *Binomial* $(n; 0.5)$, onde n representa o número de casos não empatados.

Caso se esteja perante uma amostra de grandes dimensões, deve ser utilizada a aproximação da distribuição Binomial à Normal, obtendo-se a seguinte estatística de teste.

$$Z = \frac{T - \frac{n}{2}}{\frac{\sqrt{n}}{2}} \tilde{\cap} N(0,1) \quad (5)$$

Considerando o teste bilateral, a hipótese nula deve ser rejeitada, caso $|Z| \leq -z_{1-\alpha/2}$.

O teste de Wilcoxon é uma alternativa não paramétrica ao teste t para a diferença de médias para amostras emparelhadas, fazendo inferência sobre a localização das distribuições de X e Y. Este teste apresenta a vantagem de não conduzir a tanta perda de informação como o teste dos Sinais. O teste de Wilcoxon tem em consideração a ordem atribuída a cada observação bem como a amplitude da diferença entre observações, contrariamente ao teste dos Sinais.

Segundo Reis et.al (2018) as hipóteses subjacentes a este teste são as seguintes:

$$H_0: E[X] = E[Y] \text{ Vs. } H_1: E[X] \neq E[Y],$$

sendo que a hipótese alternativa também poderá ser unilateral à direita ($H_1: E[X] > E[Y]$) ou unilateral à esquerda ($H_1: E[X] < E[Y]$).

Sendo a estatística de teste é dada por:

$$T = \frac{\sum_{i=1}^n R_i}{\sqrt{\sum_{i=1}^n R_i^2}} \cap N(0,1), \text{ sob } H_0 \text{ verdadeira} \quad (6)$$

onde,

$$R_i = \frac{D_i}{|D_i|} R'_i, \quad (7)$$

$$D_i = (Y_i - X_i) \quad (8)$$

Para a aplicação deste teste deve-se primeiramente proceder ao cálculo das diferenças $D_i = (Y_i - X_i)$. Seguidamente deve ser ordenado de forma crescente o valor absoluto das diferenças D_i atribuindo-se a cada uma a respetiva ordem, em situação de empate, a cada uma das observações empatadas deve ser atribuída a ordem média das que lhe seriam atribuídas em caso de não empate.

Posto isto, e para um teste bilateral, rejeita-se a hipótese nula ao nível de significância α para $T > z_{1-\frac{\alpha}{2}} \cup T < -z_{1-\frac{\alpha}{2}}$. Caso estejamos perante um teste unilateral direito, rejeita-se a hipótese nula ao nível de significância α para $T > z_{1-\alpha}$. Em situações em que o teste de hipóteses aplicado seja unilateral à esquerda, rejeita-se a hipótese nula ao nível de significância α para $T < z_{\alpha}$.

O teste de Friedman representa uma extensão do teste dos Wilcoxon para k amostras.

As hipóteses a testar são as seguintes:

H_0 : Todas as k populações apresentam a mesma distribuição, em média

Vs.

H_1 : A distribuição de pelo menos um das k populações é distinta, em média

Considera-se que $R(X_{ij})$ é a ordem atribuído à observação X_{ij} e que R_i é a soma das ordens das observações da j-ésima coluna.

A estatística de teste é dada por:

$$\chi_r^2 = \frac{12}{nk(k+1)} \sum_{i=1}^k (R_j)^2 - 3n(k+1) \cap \chi_{k-1}^2 \text{ sob } H_0 \text{ verdadeira} \quad (9)$$

A hipótese nula H_0 deve ser rejeitada, ao nível de significância α , se $\chi_r^2 > \chi_{k-1, 1-\alpha}^2$.

3.2.3 Teste para comparação de duas ou mais amostras independentes

Foi desenvolvido inicialmente em 1945 por F. Wilcoxon um teste que permitia comparar medidas de tendência central de duas amostras de igual dimensão. Posteriormente, em 1947, H.B. Mann e D.R. Whitney generalizaram a técnica de modo a permitir a aplicação do teste em amostras de diferentes dimensões, dando origem ao Teste de Mann-Whitney-Wilcoxon. Este teste permite a comparação de duas amostras independentes de modo a verificar se ambas são provenientes da mesma população, sendo a alternativa não paramétrica para o teste t para amostras independentes. Contrariamente ao que sucede

com o teste t, em que é testada a igualdade de médias, no teste de Mann-Whitney-Wilcoxon, segundo Reis et.al (2018), é testada a igualdade de medianas.

Considera-se que X e Y são amostras aleatórias independentes de dimensão n_1 e n_2 , respetivamente, onde $n_1 < n_2$ e $n = n_1 + n_2$. Admite-se ainda que as observações podem ser traduzidas numa ordem ou escala.

As hipóteses a testar são:

H_0 : A localização de X não diferente da localização de Y

Vs.

H_1 : A localização de X difere da localização de Y

Para amostras de grande dimensão, isto é para amostras de dimensão superior a 10 observações, é possível recorrer a aproximação à Distribuição Normal, utilizando-se por isso a seguinte estatística de teste:

$$Z = \frac{T - \mu_T}{\sigma_T} = \frac{T - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n + 1)}{12}}} \cap N(0,1), \text{ sob } H_0 \text{ verdadeira} \quad (10)$$

Posto isto, se a probabilidade associada ao valor de Z observado for inferior ao nível de significância, isto é, $P(Z \leq -|z|)$, rejeita-se a hipótese nula H_0 .

O Teste de Kruskal Wallis é considerado como sendo a alternativa não paramétrica à ANOVA, isto é, à Análise de Variância Simples que permite a comparação de k amostras independentes. O teste de Kruskal-Wallis é então utilizado quando se pretende apurar se as k amostras independentes são provenientes da mesma população. Consideram-se assim as seguintes hipóteses a testar:

H_0 : A localização das k populações é semelhante

Vs.

H_1 : A localização de pelo menos um das k populações é distinta

Considera-se que $R(X_{ij})$ é a ordem atribuída à observação X_{ij} , que R_i é a soma das ordens das observações da j -ésima amostra, isto é, $R_i = \sum_{j=1}^{n_i} R(X_{ij})$, e por último que o número total de observações é dado por $n = \sum_{i=1}^k n_i$.

A estatística de teste é dada por:

$$T = \frac{1}{S^2} \left(\sum_{i=1}^k \frac{R_i^2}{n_i} - \frac{n(n+1)^2}{4} \right) \cap \chi_{k-1}^2, \text{ sob } H_0 \text{ verdadeira} \quad (11)$$

onde,

$$S^2 = \frac{1}{n-1} \left(\sum_{i=1}^k \sum_{j=1}^{n_i} R(X_{ij})^2 - \frac{n(n+1)^2}{4} \right) \quad (12)$$

Em caso de não se verificar a existência de empates, a estatística de teste a utilizar é:

$$T = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n+1) \cap \chi_{k-1}^2, \text{ sob } H_0 \text{ verdadeira} \quad (13)$$

A hipótese nula H_0 deve ser rejeitada, ao nível de significância α , caso o valor observado da estatística de teste seja superior ao valor do quantil apropriado, isto é, se $T > \chi_{k-1, 1-\alpha}^2$.

3.3 Análise de tabelas de contingência

Por tabela de contingência, entende-se uma tabela de frequências onde é apresentado um conjunto de variáveis classificadas segundo um conjunto de características de acordo com classificações exaustivas e mutuamente exclusivas, ou seja, existem categorias suficientes nas variáveis de modo que todos os elementos da população sejam passíveis de classificação, de modo que cada elemento da população pertença a uma, e só uma categoria.

Em situações em que os dados em estudo são qualitativos ou quantitativos discretos e se pretende investigar se existe alguma relação entre as variáveis, é feita a construção de uma

tabela de contingência. A dimensão da tabela de contingência irá sempre variar consoante o número de variáveis que se pretende estudar, sendo possível obter tabelas bidimensionais (tabela com duas variáveis), tabelas tridimensionais (tabelas com três variáveis) ou tabelas multidimensionais (tabela com mais de três variáveis)

3.3.1 Teste de Qui-quadrado de Independência

De modo a verificar se existe independência entre variáveis qualitativas é aplicado um teste de Qui-quadrado de independência cujas hipóteses a testar são:

H_0 : As variáveis são independentes Vs. H_1 : As variáveis não são independentes

A estatística de teste a aplicar é dada por:

$$\chi^2 = \sum_{i=1}^L \sum_{j=1}^C \frac{(N_{ij} - e_{ij})^2}{e_{ij}} \tilde{\sim} \chi^2_{(L-1)(C-1)}, \text{ sob } H_0 \text{ verdadeira} \quad (14)$$

onde,

L: representa o número de linhas da tabela de contingência;

C: representa o número de colunas da tabela de contingência;

N_{ij} : representa os valores observados;

e_{ij} : representa os valores esperados.

Por observação da estatística de teste apresentada acima, verifica-se que existe uma dependência do valor da diferença ($N_{ij} - \hat{e}_{ij}$), ou seja, este valor será menor em caso de independência das variáveis em estudo.

O teste de Qui-quadrado exige que sejam verificados dois pressupostos:

1. No máximo apenas podem existir 20% das células com frequência esperada inferior a 5;
2. Nenhuma célula da tabela de contingência pode ter frequência esperada inferior ou igual a 1. (em particular para N de pequenas dimensões)

Posto isto, a região crítica é definida com sendo unilateral à direita, ou seja, a hipótese nula é rejeitada ao nível de significância α , para valores da estatística de teste superiores ou iguais ao quantil de probabilidade α de uma distribuição de Qui-Quadrado. Matematicamente tem-se: $\chi^2 > \chi_{(L-1)(C-1); \alpha}$.

Considerando o *valor-p* definido por $P(\chi_{LC}^2 > k | H_0 \text{ verdadeira})$, onde k é o valor observado da estatística de teste, a hipótese nula é rejeitada ao nível de significância α para valores de *p-value* $< \alpha$.

3.3.2 - Teste de sobre o coeficiente Kappa de Cohen

A medida de concordância *Kappa* de Cohen permite quantificar a concordância entre duas classificações independentes, provenientes de duas variáveis categóricas com categorias idênticas entre si. O seu cálculo é feito através da seguinte expressão:

$$k = \frac{p_o - p_e}{1 - p_e} \quad (15)$$

sendo, $p_o = \frac{O_D}{n}$, onde O_D é a soma das frequências observados na diagonal da tabela; e $p_e = \frac{E_D}{n}$, onde E_D é a soma das frequências esperadas na diagonal da tabela.

O nível de concordância pode ser qualificado da seguinte forma:

Tabela 3.1: Tabela Classificativa do nível de concordância (Fonte: Julius Sim, 2005)

Valor de K	Concordância
< 0	Má
$0 \leq k < 0,2$	Razoável
$0,2 \leq k < 0,45$	Moderada
$0,45 \leq k < 0,75$	Considerável
$0,75 \leq k < 1$	Quase perfeita
$k = 1$	Perfeita

De modo a verificar a significância da concordância entre as variáveis é possível realizar o teste sobre o coeficiente Kappa de Cohen, cujo as hipóteses a testar são:

$$H_0: k = 0 \quad Vs. \quad H_1: k \neq 0$$

A estatística de teste associada é dada por:

$$Z = \frac{k}{s(k)} \cap N(0,1) \text{ sob } H_0 \text{ verdadeira} \quad (16)$$

onde, k é a estimativa do coeficiente de concordância e

$$s(k) = \sqrt{\frac{p_0(1 - p_0)}{n(1 - p_e)^2}} \quad (17)$$

A hipótese nula deverá ser rejeitada se o valor observado da estatística de teste for superior ao quantil crítico $1 - \alpha/2$ da distribuição Gaussiana.

3.3.3 Medidas e Testes de Associação

Dado que o teste de Qui-quadrado para a independência apenas permite concluir sobre a possível existência de relação entre variáveis, é necessário recorrer às medidas de associação, para que caso exista relação, seja possível quantificá-la.

Estas medidas podem assumir valores entre 0 e 1, sendo que quando mais próximo da unidade maior será a associação existente entre as variáveis.

Um dos coeficientes que permite quantificar a associação existente entre duas variáveis é o Coeficiente de Contingência de Pearson. Este coeficiente pode ser utilizado quando uma ou ambas as variáveis são classificadas como variáveis qualitativas nominais.

O grau de associação entre as variáveis pode ser quantificado através do cálculo da seguinte expressão:

$$C = \sqrt{\frac{X^2}{n + X^2}} \quad (18)$$

onde,

X^2 representa o valor observado da estatística de teste de Qui-quadrado para a independência.

Este coeficiente apresenta a particularidade de nunca assumir o valor 1, mesmo em caso de associação total, o seu valor varia em:

$$0 \leq C \leq \sqrt{\frac{q-1}{q}} < 1, \text{ onde } q = \min(L, C) \quad (19)$$

4. Modelos Lineares Generalizados

O modelo de regressão linear foi proposto por Francis Galton e Karl Pearson no final do século XIX, e até meados do século XX a Análise de Regressão Linear foi tida como a técnica de modelação estatística mais usada, assentando nos pressupostos de normalidade, independência e homocedasticidade dos erros. A principal finalidade desta metodologia consiste em examinar a relação entre uma variável de resposta e uma ou mais variáveis explicativas, visando encontrar uma função capaz de explicar um determinado fenómeno em estudo. Esta metodologia permite compreender as causas das variações de um fenómeno e prever seu comportamento, com base nas variáveis explicativas utilizadas, no espaço de domínio destas. Para apenas uma variável explicativa tem-se o modelo de regressão linear simples. Para mais do que uma variável explicativa, tem-se o modelo de regressão linear múltipla.

Por forma a ultrapassar as limitações exigidas no que respeita aos referidos pressupostos, e como extensão dos modelos anteriores, Nelder e Wedderburn (1972) introduziram os Modelos Lineares Generalizados (MLG), agregando sob esta designação vários modelos estatísticos já existentes, tais como o modelo complementar log-log para ensaios de diluição (Fisher, 1922), os modelos para proporções do tipo probit - ver (Bliss, 1934) e logit - ver (Berkson, 1944) e (Rasch, 1960).

Para além da distribuição normal os MLG permitem outras distribuições para os erros, e possibilitam uma função de ligação relacionando a média da variável resposta à combinação linear das variáveis explicativas. Os modelos lineares generalizados constituem uma importante extensão dos modelos de regressão linear clássicos permitindo a análise de dados numa perspectiva de aplicação mais abrangente, e que não atendem obrigatoriamente aos pressupostos de normalidade e homogeneidade de variância. Permitem por exemplo modelar variáveis contínuas, na forma de contagem, binárias e categóricas. No entanto, com estes modelos continua a ser exigido assumir o pressuposto de independência dos erros, o que impede por exemplo a modelação de bases de dados espaciais, com medidas repetidas ou do tipo multinível, para os quais têm vindo a ser desenvolvidos modelos mistos e outros.

Todos os modelos criados no decorrer do século XX apresentavam uma estrutura de regressão linear tendo como ponto comum o facto de que a variável resposta seguir uma distribuição encontrada na família da distribuição exponencial, tendo por isso características e propriedades bastante peculiares.

É já em 1972, que pelas mãos de Nelder e Wedderburn, que surge o conceito de Modelos Lineares Generalizados, que corresponde a uma sinopse de todos os modelos existentes até então, tendo deste modo, surgido com o intuito de reunir, tanto do ponto de vista teórico como conceptual, toda a teoria de modelação estatística existente à data.

Segundo Turkman & Silva (2000), os Modelos Lineares Generalizados são limitados por manterem a estrutura de linearidade, isto porque, as distribuições se restringem à família exponencial e por exigirem a independência das respostas.

4.1 Família Exponencial

Uma distribuição diz-se ser da família exponencial se a sua função de densidade de probabilidade poder ser escrita na seguinte forma:

$$f(y|\theta, \phi) = e^{\frac{y\phi - b(\theta)}{a(\phi)} + c(y, \phi)} \quad (20)$$

Onde, θ e ϕ são parâmetros escalares e $a(\cdot)$, $b(\cdot)$ e $c(\cdot, \cdot)$ são funções reais conhecidas. O parâmetro θ considera-se um parâmetro de localização e ϕ um parâmetro de dispersão, que genericamente é conhecido. Sabe-se que a função $b(\cdot)$ é uma função diferenciável e o suporte da distribuição não depende dos parâmetros.

A família exponencial assume uma especial importância nos modelos lineares generalizados, pelo facto de nestes ser possível a utilização de dados discretos ou contínuos, dados que possuem assimetrias e dados que são restritos a um intervalo do conjunto dos reais.

Veja-se o seguinte exemplo:

Seja Y uma variável aleatória de resposta, que segue uma distribuição Normal, $Y \sim N(\mu, \sigma^2)$. A sua função densidade de probabilidade é dada por:

$$\begin{aligned} f(y; \mu, \sigma) &= \frac{1}{\sqrt{2\pi\sigma^2}} \times \exp\left\{\frac{-(y - \mu)^2}{2\sigma^2}\right\} \\ &= \exp\left\{\frac{y \times \mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{1}{2} \times \ln(2\pi\sigma^2) - \frac{y^2}{2\sigma^2}\right\} \end{aligned} \quad (21)$$

Obtém-se assim uma equação do tipo da anteriormente apresentada, com:

$$\theta = \mu; \quad b(\theta) = \frac{\mu^2}{2}; \quad a(\phi) = \sigma^2; \quad c(y, \phi) = \frac{1}{2} \times \ln(2\pi\sigma^2) - \frac{y^2}{2\sigma^2}$$

Seguidamente serão apresentados modelos contínuos e discretos cuja distribuição pertence à família exponencial.

4.2 Características do Modelo Linear Generalizado

Os modelos lineares generalizados (MLG) são caracterizados pela seguinte estrutura: componente aleatória, componente sistemática e a função de ligação.

1. Componente aleatória

Dado um certo vetor de covariáveis X_i , as variáveis Y_i são condicionalmente independentes cuja distribuição é da família exponencial, seguindo a forma apresentada anteriormente. Posto isto, sabe-se que o valor médio é dado por: $E(Y_i|x_i) = \mu_i = b'(\theta_i)$ com $i = 1, \dots, n$

2. Componente sistemática

Seja n_i , um preditor linear, definido como uma combinação linear de variáveis explicativas, matematicamente $n_i = X_i^T \beta$.

Este valor associa-se a cada valor da variável resposta Y_i . O vetor X_i^T é um vetor linha de dimensão $p \times 1$, X_{ik} representa a k -ésima covariável para i -ésimo indivíduo e β é um vetor linha de dimensão $p \times 1$ constituído por parâmetros desconhecidos.

3. Função de Ligação

A função de ligação tem como função relacionar o valor esperado da variável resposta com o preditor linear.

4.3 Modelo de Regressão Logística

Neste tópico, será abordado o modelo de regressão logística. Este é considerado, segundo Agresti (2002) um dos modelos lineares mais importantes, quando o objetivo é modelar uma variável resposta classificada como variável categórica binária, isto é, quando a variável pode assumir, um de dois valores possíveis, dado um determinado conjunto de variáveis explicativas.

Este modelo de regressão distingue-se do modelo de regressão linear através da distribuição seguida pela variável resposta. Em regressão linear a variável resposta segue uma distribuição de Bernoulli.

Este tipo de modelo é bastante utilizado não só na modelação de problemáticas de saúde, mas também pelas instituições de crédito para análise e gestão do risco de crédito. Risco de crédito consiste na probabilidade da contraparte (neste caso, o cliente), de não efetuar o reembolso total ou parcial do financiamento obtido junto de uma instituição de crédito.

O modelo de regressão logística simples é utilizado caso seja apenas uma variável explicativa. Neste caso assume-se que a variável resposta tem uma distribuição de probabilidade binomial, de parâmetros (m_i, π_i) , onde, m_i é o número de ensaios realizados e π_i a probabilidade de sucesso.

Seja x_i um vetor de k variáveis para o i -ésimo indivíduo, tem-se a seguinte função massa de probabilidade:

$$f(y|\pi) = \pi^y(1 - \pi)^{1-y}, y = 0, 1 \quad (22)$$

Seja $\pi(x_i) = P(Y_i=1|X)$, genericamente, a resposta média ao modelo linear é dada pela função:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}, i = 1, \dots, n \quad (23)$$

A representação gráfica deste modelo, genericamente, é semelhante à letra “S”. Foi utilizado o *software Rprogram* para representar o modelo, foram considerados os seguintes valores para os parâmetros -20 e $0,2$ para β_0 e β_1 respetivamente.

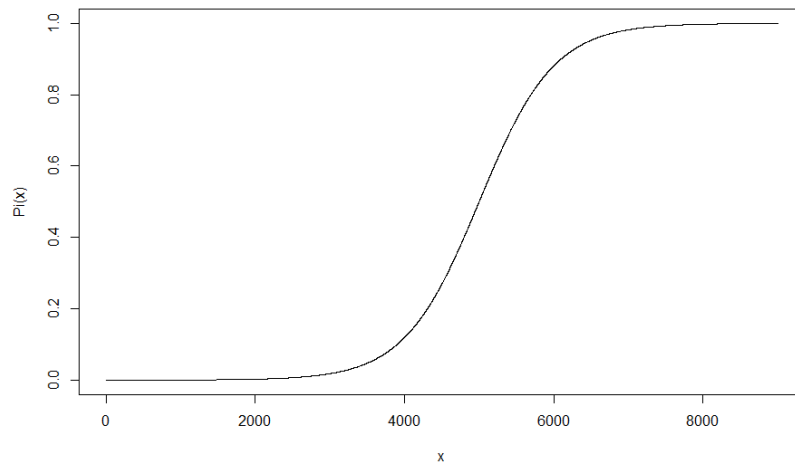


Figura 4. 1: Representação do Modelo Logístico com parâmetros $\beta_0=20$ e $\beta_1=0.2$

Para caracterizar a relação existente entre a variável resposta e a variável explicativa é calculado o logaritmo da razão entre a probabilidade de sucesso e a probabilidade de insucesso, isto é, a função *logit*.

$$\text{logit}[P(Y_i = 1|X)] = \ln \left[\frac{P(Y_i = 1|X)}{1 - P(Y_i = 1|X)} \right] \quad (24)$$

Ao coeficiente de regressão logística dá-se o nome de *Odds Ratio*, ou seja, a variação causada no valor do logaritmo por cada unidade aumentada em x . Este valor representa o risco da variável resposta tomar o valor 1 ($Y=1$) quando a variável explicativa toma o valor 1 ($X=1$) em relação a $X=0$.

O modelo de regressão logística múltiplo é análogo ao anteriormente apresentado, a diferença reside no facto de neste serem consideradas duas ou mais variáveis explicativas. Nesta situação a função logística é dada por:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}, i = 1, \dots, n \quad (25)$$

4.3.1 Inclusão de Interações no modelo

Segundo Pérez (2023), as interações e transformações em modelos de regressão logística permitem analisar relações mais complexas entre as variáveis explicativas e a variável dependente. As interações entre variáveis explicativas são necessárias quando o efeito de uma variável explicativa na variável dependente depende do valor de outra variável explicativa.

Os termos mais comuns e interpretáveis a serem adicionados são interações em pares e termos quadráticos. Contudo, podem ser considerados outros tipos de termos como por exemplo, interações de três termos e transformações cúbicas.

4.4 Método da máxima verossimilhança

De modo a estimar os parâmetros β_j , com $j = 1, \dots, k$, pode ser utilizado o método da máxima verossimilhança. A função de verossimilhança é dada por:

$$L(\beta) = \prod_{i=1}^n f(y_i | \theta_i, \phi) = e^{\sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{\phi} + \sum_{i=1}^n c(y_i, \phi)} \quad (26)$$

$$\ln(L(\beta)) = \sum_{i=1}^n \left(\frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi) \right) \quad (27)$$

Admitindo que se trata de observações independentes de uma amostra de dimensão n , considerando que se trata de regressão logística e atendendo à distribuição de Y , a função de máxima verossimilhança é:

$$L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \quad (28)$$

As estimativas para os coeficientes β_j são obtidas ao igualar as derivadas parciais da função de máxima verosimilhança a zero.

4.5 Inferência em Modelos Lineares Generalizados

Para permitir a aplicação válida da presente metodologia, importa assegurar determinadas características ao nível da qualidade do modelo e seu ajustamento aos dados, detetando incoerências/ discrepâncias. Desta forma, o processo de ajuste dos MLG pode ser dividido em três etapas: (1) formulação do modelo, (2) ajuste do modelo e (3) inferência, que, baseando-se na teoria assintótica de máxima verosimilhança, consiste na aplicação de testes de hipóteses relativos aos parâmetros do modelo e da qualidade do seu ajustamento aos dados.

De acordo com Cordeiro e Demétrio (2011), nesta etapa, estuda-se a precisão, bem como a interdependência das estimativas, testam-se os parâmetros do modelo, criam-se intervalos de confiança, analisam-se os respetivos resíduos e efetuam-se previsões.

A análise da precisão do modelo é efetuada com recurso da validação da maximização da previsão apresentada. A análise dos resíduos permite evidenciar a existência de tendências, possibilitando a avaliação da relação funcional entre média e variância e permitindo avaliar a necessidade de inclusão/ exclusão de variáveis explanatórias ao modelo.

Para avaliação do ajuste do modelo, normalmente adota-se o critério da razão da verosimilhança e a estatística de Pearson.

Segundo Cordeiro e Demétrio (2011), um modelo incorretamente ajustado aos dados pode apresentar uma ou mais das seguintes condições: (a) inclusão de um grande número de variáveis explanatórias ao modelo, muitas das quais redundantes e algumas explicando somente um pequeno percentual das observações; (b) formulação de um modelo bastante

pobre em variáveis explanatórias, que não revela nem reflete as características inerentes aos dados; (c) as observações são insuficientes para que falhas no modelo sejam detetadas.

4.5.1 Teste de Hipóteses e Análise de Desempenho

Os testes de hipóteses aplicados no âmbito dos MLG têm como objetivo responder a premissas relativas aos parâmetros β , sendo deduzidas de distribuições assintóticas de funções adequadas. As principais estatísticas de teste aplicadas são: (1) razão de verossimilhança, (2) Wald e (3) *Rao's Score*.

Para efetuar testes de hipóteses associadas a vários coeficientes, por norma, considera-se o teste da máxima verossimilhança. Pelo contrário, quando nos referimos a testes estatísticos com referência a um único coeficiente, o mais usual é a aplicação do teste de Wald.

A aplicação destes testes permite verificar a qualidade do ajustamento do modelo aos dados, por via da análise dos coeficientes associados a cada parâmetro são significativos, garantindo que o modelo apresenta uma consolidada discriminação das categorias da variável resposta.

Para verificar o poder de discriminação é verificado o desempenho do modelo quanto à sensibilidade e especificidade, usando as curvas ROC (*Receiver Operating Characteristic Curve*) e as tabelas de contingência.

O teste de Wald permite saber se determinada variável é significativa para o modelo ajustado. As hipóteses a testar são:

$$H_0: \beta_j = 0 \quad Vs. \quad H_1: \beta_j \neq 0 \quad \forall j = 1, \dots, k$$

A estatística de teste associada é dada por:

$$W = \frac{\hat{\beta}_j}{se_j} \cap N(0,1) \quad \text{sob } H_0 \text{ verdadeira} \quad (29)$$

onde,

$\hat{\beta}_j$ é o estimador de β_j

se_j é o erro padrão associado ao estimador de β_j

A hipótese nula deverá ser rejeitada para valores de *valor-p* inferiores ou iguais ao nível de significância estabelecido.

O teste de Hosmer e Lemeshow permite avaliar a qualidade de ajustamento do modelo aos dados. As hipóteses a testar são:

H_0 : O modelo é ajustado convenientemente aos dados

Vs.

H_1 : O modelo não é ajustado convenientemente aos dados

A estatística de teste a considerar é dada por:

$$C = \sum_{k=1}^g \frac{(O_k - e_k)}{e_k \left(1 - e_k/n'_k\right)} \tilde{\chi}^2_{(g-2)} \text{ sob } H_0 \text{ verdadeira} \quad (30)$$

onde,

n'_k representa o número de indivíduos no k-ésimo grupo

O_k representa o número de sucessos observado

e_k representa o número estimado de sucessos esperado

g representa o número total de grupos considerados

Segundo Hosmer et. al (2000), é possível considerar dois métodos distintos para a constituição dos grupos. No primeiro método, supondo a existência de 10 grupos tem-se $g = 10$, o primeiro grupo com $n'_1 = \frac{n}{10}$ indivíduos possuem os valores mais reduzidos de probabilidades ajustadas, por oposição o último grupo com $n'_{10} = \frac{n}{10}$ indivíduos possuem os valores mais elevados de probabilidades ajustadas. No segundo método, considerando

nomeante a existência de grupos, obtém-se ponto de corte definidos para valores $\frac{k}{10}$ com $k = 1, \dots, 9$, os grupos são constituídos por indivíduos com probabilidades ajustadas entre os pontos de corte.

Posto isto, a hipótese nula deverá ser rejeitada para valores de *valor-p* inferiores ou iguais ao nível de significância estabelecido.

4.5.2 Seleção de Modelos

Pretende-se que o modelo final seja constituído pelo menor número possível de co variáveis e que descreva de forma adequada a variável resposta.

O método *stepwise* é um método de seleção de co variáveis em três direções distintas: *backward*, *forward* e *both*. No método *stepwise backward* o processo inicia-se com o modelo completo, isto é, com todas as co variáveis do modelo, ao qual é retirado uma variável de cada vez. No método *stepwise forward*, o processo inicia-se com o modelo nulo, isto é, sem nenhuma co variável, ao qual vão sendo adicionadas à vez variáveis que são significativas na justificação do modelo. Por último, no método *stepwise both*, é uma conjugação dos dois métodos anteriormente descritos.

4.5.3 Avaliação do Modelo

O critério de infirmação de *Akaike* (*AIC*) é um dos métodos mais utilizado para a seleção de modelos, permitindo também fazer a comparação entre modelos distintos. A fórmula de cálculo deste critério é dada por:

$$AIC = -2[\ln(L) - k] \quad (31)$$

onde,

k representa o número de parâmetros incluídos no modelo

L representa o valor da verosimilhança para estimativa do modelo

Quando menor for o valor de *AIC*, menor será a quantidade de informação perdida com o ajustamento do modelo, o que conseqüentemente se traduz num melhor ajustamento.

A análise de resíduos é uma componente importante da avaliação do modelo. Através desta análise pretende-se não só avaliar a qualidade do modelo ajustado, como também auxiliar na identificação de observações ajustadas de forma inadequada. A diferença entre y_i e o seu valor estimado pelo modelo (\hat{y}_i) é denotado por resíduo (r_i).

O método mais utilizado para proceder à análise de resíduos é o resíduo da *deviance*, o seu cálculo é feito através da seguinte expressão:

$$r_D = \text{sinal}(y_i - \hat{y}_i) \sqrt{-2[y_i \ln(\hat{y}_i + (1 - y_i)) \ln(1 - \hat{y}_i)]} \quad (32)$$

O sinal da expressão acima apresentado será positivo se $y_i \geq \hat{y}_i$, caso contrário a expressão tomará um sinal negativo.

Estatisticamente, considera-se um *outlier* uma observação que se destaca das restantes através das co variáveis. Por vezes um *outlier* não é considerado uma observação influente, isto porque a sua eliminação não causa alterações em algumas estimativas dos coeficientes, o que aconteceria caso fosse uma observação influente.

De modo a avaliar se uma observação é influente quando retirada do modelo, é determinada a distância de Cook. O seu cálculo é feito através da seguinte expressão:

$$D_i = \frac{r_i^2 h_{ii}}{p(1 - h_{ii})}, i = 1, \dots, n \quad (33)$$

onde,

h_{ii} representa a distância entre a i -ésima observação em relação às restantes observações, r_i representa o resíduo da i -ésima observação e p representa uma constante.

São vários os autores que consideram que uma observação i é influente se $D_i > \frac{4}{n}$ (Kenneth A. Bollen, 1995). Contudo neste estudo, será utilizada a definição de Douglas C. Montgomery (2012), que considera que uma observação é influente se $D_i > 1$.

4.5.4 Capacidade Preditiva do Modelo

De modo a quantificar a capacidade discriminativa do modelo entre os indivíduos que verificam o acontecimento de interesse e os que não verificam, é estabelecido um ponto de corte uma vez que se verifica que os valores preditos do modelo estão contidos no intervalo $[0,1]$. Este ponto de corte permite classificar e quantificar o número de previsões positivas ou negativas, ou seja, quando o valor-predito é superior ao ponto de corte e consequentemente a variável resposta assume o valor 1, e por oposição, quando o valor-predito é inferior ao ponto de corte e consequentemente a variável resposta assume o valor 0. Posto isto é possível construir a seguinte matriz de confusão.

Tabela 4. 1: Matriz de Confusão

Valores Preditos	Valores Observados	
	1	0
1	Verdadeiros Positivos	Falsos Positivos
0	Falsos Negativos	Verdadeiros Negativos

O mais conveniente de obter é um número reduzido de falsos positivos e negativos. Posteriormente é possível determinar a sensibilidade, especificidade, precisão e eficiência do modelo.

- I. **Sensibilidade:** a sensibilidade ou taxa de verdadeiros positivos, representa a probabilidade de uma observação classificada como positiva ser efetivamente positiva. O seu cálculo é feito através da seguinte expressão:

$$S = \frac{\# \text{ verdadeiros positivos}}{\# \text{ verdadeiros positivos} + \# \text{ falsos negativos}} \quad (34)$$

- II. **Especificidade:** a especificidade ou taxa de verdadeiros negativos, representa a probabilidade de uma observação classificada como negativa ser efetivamente negativa. O seu cálculo é feito através da seguinte expressão:

$$E = \frac{\# \text{ verdadeiros negativos}}{\# \text{ verdadeiros negativos} + \# \text{ falsos positivos}} \quad (35)$$

- III. **Precisão**: proporção de predições corretas. Quando os conjuntos de dados em estudo são desequilibrados, esta medida pode levar a conclusões erradas quanto ao desempenho do modelo.

$$ACC = \frac{\# \text{ verdadeiros positivos} + \# \text{ verdadeiros negativos}}{\# \text{ dimensão da amostra}} \quad (36)$$

- IV. **Eficiência**: a eficiência corresponde ao resultado da média aritmética entre os valores da sensibilidade e especificidade.

$$EFF = \frac{\text{Sensibilidade} + \text{Especificidade}}{2} \quad (37)$$

Outro dos métodos que permite determinar a capacidade predita do modelo é a curva ROC. A curva ROC permite ilustrar a relação entre a percentagem de verdadeiros positivos e falsos positivos. Paralelamente também é possível avaliar a variação da sensibilidade e especificidade para todos os possíveis pontos de corte, o seu valor ótimo corresponde ao valor de corte onde se regista o valor máximo de sensibilidade e especificidade. Considerado a AUC (*Area Under the Curve*) é possível retirar conclusões relativas à performance do modelo em estudo.

Tabela 4. 2:Tabela classificativa de AUC (Fonte: Hosmer e Lemeshow (2000))

Área Abaixo da Curva	Poder de discriminação
$AUC = 0,5$	Sem discriminação
$0,5 \leq AUC < 0,7$	Discriminação Fraca
$0,7 \leq AUC < 0,8$	Discriminação Razoável
$0,8 \leq AUC < 0,9$	Discriminação Boa
$AUC \geq 0,9$	Discriminação Ótima

4.5.5 Interpretação dos Coeficientes

A *Odds Ratio* (OR) ou razão de chances, é uma medida de associação que permite interpretar, em termos relativos, os valores estimados dos coeficientes associados ao modelo. A *odd ratio* representa a razão entre a chance de sucesso entre os expostos e a chance de sucesso entre os elementos não expostos a uma determinada característica.

I. Variáveis Independentes Dicotômicas

Considerando a variável independente como sendo uma variável dicotômica que assume o valor 1 caso se verifique a presença de um determinado atributo e o valor 0 em caso contrário.

Admitindo que π_0 representa a probabilidade de obter sucesso na ausência do atributo e que π_1 representa a probabilidade de obter sucesso na presença do atributo. A *odds ratio* é calculada através da seguinte expressão:

$$OR = \frac{\pi_1 / 1 - \pi_1}{\pi_0 / 1 - \pi_0} \quad (38)$$

Dado que se trata da OR de uma variável dicotômica, tem-se $OR = e^{\beta_1}$. O intervalo de confiança de $100 \times (1 - \alpha)\%$ é dado pela seguinte expressão:

$$\left[e^{\hat{\beta}_j - z_{1-\alpha/2} \times s(\hat{\beta}_j)}; e^{\hat{\beta}_j + z_{1-\alpha/2} \times s(\hat{\beta}_j)} \right] \quad (39)$$

II. Variáveis Independentes Categóricas

Considerando a variável independente como sendo uma variável categórica com duas ou mais categorias, serão atribuídos $k - 1$ valores às k categorias, sendo que a categoria de referência assume o valor 1. A *odds ratio* é calculada de forma análoga à situação anteriormente apresentada.

Dada uma categoria k da variável i , tem-se $OR = e^{\beta_{ik}}$. O intervalo de confiança de $100 \times (1 - \alpha)\%$ é dado pela seguinte expressão:

$$\left[e^{\hat{\beta}_{ik} - z_{1-\alpha/2} \times s(\hat{\beta}_{ik})}; e^{\hat{\beta}_{ik} + z_{1-\alpha/2} \times s(\hat{\beta}_{ik})} \right] \quad (40)$$

III. Variáveis Independentes Contínuas

Considerando a variável independente como sendo uma variável contínua, assume-se que a sua função *logit* é uma variável linear.

Posto isto, tem-se $OR = e^{\{\beta_k \times c\}}$, onde c representa a variação de c unidades da variável x_k através da diferença entre *logits*. O intervalo de confiança de $100 \times (1 - \alpha)\%$ é dado pela seguinte expressão:

$$\left[e^{c \times \hat{\beta}_k - z_{1-\alpha/2} \times c \times s(\hat{\beta}_k)}; e^{c \times \hat{\beta}_k + z_{1-\alpha/2} \times c \times s(\hat{\beta}_k)} \right] \quad (41)$$

5. Metodologia

Primeiramente foi realizada uma breve análise exploratória das variáveis em estudo de modo a compreender melhor o comportamento das mesmas.

Embora a dimensão da amostra em estudo permita a utilização de metodologias paramétricas por aproximação no caso das variáveis numéricas, optou-se pelo uso de metodologias não paramétricas.

Posteriormente, dada a pertinência do tema e tendo em vista a posterior integração no modelo para a situação de depressão, será realizado um breve estudo de modo a perceber se existe uma correlação estatisticamente significativa entre a idade e a idade que sente ter.

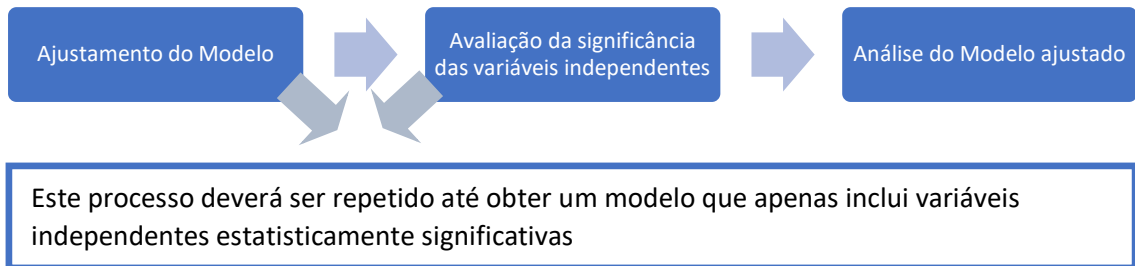
Em seguida, e com recurso a tabelas de contingência e testes estatísticos associados e ao cálculo de coeficientes de associação/contingência quando necessário, será verificada, para cada uma das variáveis que existem diferenças entre os géneros. Esta opção teve como objetivo avaliar a pertinência de efeitos de interação no modelo de predição da situação de depressão (sentida). Todos os pressupostos do teste de Qui-quadrado para a independência foram devidamente verificados.

Posteriormente dando continuidade à exploração das variáveis selecionadas como potenciais preditores do modelo, são realizadas comparações para identificar diferenças na variável dependente entre as categorias das variáveis independentes. Para tal serão realizados os testes de *Mann-Whitney* e *Kruskall-Wallis*, quando aplicável. Com base nos resultados obtidos com a realização dos testes de hipóteses referidos, serão selecionadas por fim as variáveis candidatas ao modelo.

Concluída a análise exploratória, inicia-se o ajustamento do modelo logístico aos dados e estudo. Inicialmente será ajustado um modelo com todas as variáveis independentes candidatas ao modelo, sendo posteriormente retiradas as variáveis que não se revelam significativas. A avaliação da significância das variáveis independentes é feita com recurso ao teste de *Wald*. Dada a permanência para o estudo será ainda ajustado modelos de

regressão logística com interações de modo a verificar se estes modelos permitem justificar de forma mais conveniente a variável dependente.

Por último, e após obter o modelo que melhor traduz a variável dependente será feita análise do modelo ajustado através de medidas de: análise de desempenho, avaliação do modelo e capacidade preditiva do modelo.



Para toda a inferência estatística foi estabelecido $\alpha = 0,05$ como valor crítico de significância, rejeitando-se a hipótese nula quando a probabilidade de erro tipo I foi inferior àquele valor (*valor-p* < 0,05).

Para aplicação dos métodos e obtenção dos resultados foi utilizado o *software* IBM SPSS versão 27.

6. Aplicação a dados reais do Projeto Europeu SHARE

6.1 Descrição da Amostra

Os dados que serão alvo de estudo nesta dissertação são provenientes de uma base de dados do projeto SHARE, já anteriormente apresentado neste trabalho.

A base de dados selecionada diz respeito à 9ª vaga do estudo realizado no verão de 2021, sendo por isso composta por dados recolhidos no contexto da pandemia COVID-19, numa fase de implementação mais ampla da doença. Esta base de dados é constituída por 309 variáveis de tipologia diversa correspondendo a 49253 observações, sendo que, do total da amostra, verifica-se que 1073 observações são relativas a inquéritos realizados à população portuguesa. Da subamostra da população portuguesa, apenas se registaram 947 observações válidas para este estudo. Em função do grau de importância das variáveis envolvidas, foram consideradas para análise neste trabalho as seguintes:

1. Questões demográficas:

Q1.1. Género (Feminino, Masculino)

Q1.2. Ano de nascimento

Q1.3. Idade que sente ter

Para efeitos de uniformização da escala, para comparação entre idade real e idade subjetiva (que sente ter), foi criada uma nova variável, denotada idade, que resulta da diferença entre 2021 (ano da realização do questionário) e o ano de nascimento do inquirido. Doravante, será utilizada esta nova variável.

2. Questões de saúde física e comportamentos de saúde:

Q2.1. Perceção do estado de saúde (Excelente, Muito Boa, Boa, Razoável, Má)

Q2.2. Tem fratura do quadril (Sim, Não)

Q2.3. Tem diabetes ou açúcar elevado no sangue (Sim, Não)

Q2.4. Tem tensão alta ou hipertensão (Sim, Não)

Q2.5. Teve um enfarte agudo do miocárdio ou tem outro problema de coração (Sim, Não)

Q2.6. Tem doença pulmonar crónica (Sim, Não)

Q2.7. Tem cancro ou um tumor maligno (Sim, Não)

Q2.8. Tem outra doença ou condição de saúde (Sim, Não)

Q2.9. Toma medicação regularmente (Sim, Não)

Q2.10. Saiu de casa nos últimos 3 meses (Sim, Não)

Q2.11. Tomou medicação preventiva para a COVID-19 (Sim, Não)

Q2.12. Está vacinado contra a COVID-19 (Sim, Não)

Q2.13. Se não está vacinado pretende ser vacinado contra COVID-19 (Não Aplicável; Sim, já tenho a vacinação agendada; Sim, quero ser vacinado; Não, não quero ser vacinado; Continuo indeciso)

3. Questões de saúde mental:

Q3.1 No último mês sentiu-se nervoso, ansioso ou no limite (Sim, Não)

Q3.2 No último mês esteve triste ou deprimido (Sim, Não)

Q3.3 Teve dificuldades em dormir recentemente (Problemas com o sono ou mudança recente no padrão, Sem problemas com o sono)

Q3.4 Sente-se solitário (Muitas Vezes, Algumas Vezes, Quase Nunca)

4. Questões sobre a COVID-19:

Q4.1. Quantas vezes testou positivo para a COVID-19 (Nunca, Apenas 1 vez, Entre 2 e 5 vezes, Entre 6 e 10 vezes)

5. Questões sobre situação profissional:

Q5.1. Situação atual de emprego (Reformado, Empregado ou Trabalhador Independente, Desempregado, Permanentemente doente ou incapacitado, Dona de Casa, Outro)

6. Questões sobre fatores sociais:

Q6.1. Durante o surto de COVID-19 utilizou pelo menos 1 vez a internet para fazer compras, procurar informação, entre outros (Sim, Não)

6.2 Classificação das Variáveis

Para melhor compreensão das variáveis em estudo, será apresentada de seguida na tabela 6.1 a classificação de cada uma das variáveis, bem como as opções de respostas dadas pelos inquiridos (seguindo listas pré-definidas).

Tabela 6. 1: Tabela de apresentação e classificação das variáveis em estudo

Questões demográficas		
Variável	Codificação*	Classificação
Género	1: Masculino 2: Feminino	Variável qualitativa nominal
Idade	Não Aplicável	Variável quantitativa continua
Idade que sente ter	Não Aplicável	Variável quantitativa continua
Questões de saúde física e comportamentos de saúde		
Perceção do Estado de Saúde	1: Excelente 2: Muito Boa 3: Boa 4: Razoável 5: Má	Variável qualitativa ordinal
Tem fratura do quadril	1: Sim 5: Não	Variável qualitativa nominal
Tem diabetes ou açúcar elevado no sangue	1: Sim 5: Não	Variável qualitativa nominal
Tem tensão alta ou hipertensão	1: Sim 5: Não	Variável qualitativa nominal
Teve um enfarte agudo do miocárdio ou tem outro problema de coração	1: Sim 5: Não	Variável qualitativa nominal
Tem doença pulmonar crónica	1: Sim 5: Não	Variável qualitativa nominal
Tem cancro ou um tumor maligno	1: Sim 5: Não	Variável qualitativa nominal
Tem outra doença ou condição de saúde	1: Sim 5: Não	Variável qualitativa nominal
Toma Medicação Regular	1: Sim 5: Não	Variável qualitativa nominal
Saiu de casa nos últimos 3 meses	1: Sim 5: Não	Variável qualitativa nominal
Tomou Medicação Preventiva para a COVID-19	1: Sim 5: Não	Variável qualitativa nominal

Está vacinado contra a COVID-19	1: Sim 5: Não	Variável qualitativa nominal
Se não está vacinado pretende ser vacinado contra COVID-19	-9: Não Aplicável 1: Sim, já tenho a vacinação agendada 2: Sim, quero ser vacinado 3: Não, não quero ser vacinado 4: Continuo indeciso	Variável qualitativa nominal
Questões de saúde mental		
No último mês sentiu-se nervoso, ansioso ou no limite	1: Sim 5: Não	Variável qualitativa nominal
No último mês esteve triste ou deprimido	1: Sim 5: Não	Variável qualitativa nominal
Teve dificuldades em dormir recentemente	1: Problemas com o sono ou mudança recente no padrão 2: Sem problemas com o sono	Variável qualitativa nominal
Sente-se solitário	1: Muitas Vezes 2: Algumas Vezes 3: Quase nunca ou nunca	Variável qualitativa ordinal
Questões sobre a COVID-19		
Quantas vezes testou positivo para a COVID-19	1: Nunca 2: Apenas 1 vez 3: Entre 2 e 5 vezes 4: Entre 6 e 10 vezes 5: Mais de 10 vezes	Variável qualitativa ordinal
Questões sobre situação profissional		
Situação atual de emprego	1: Reformado 2: Empregado ou Trabalhador Independente 3: Desempregado 4: Permanentemente doente ou incapacitado 5: Dona de Casa 97: Outro	Variável qualitativa nominal
Questões sobre fatores sociais		
Durante o surto de COVID-19 utilizou pelo menos 1 vez a internet para fazer compras, procurar informação, entre outros	1: Sim 5: Não	Variável qualitativa nominal

*Foi adotada a codificação original constante na base de dados

6.3 Caracterização da Amostra e Relações entre as Variáveis

Para uma análise adequada da base de dados em estudo e para melhor compreensão das variáveis selecionadas, recorreu-se na primeira fase à Estatística Descritiva, nomeadamente medidas de localização e de dispersão para as variáveis quantitativas, e análise de frequências para as variáveis qualitativas. Exploraram-se ainda alguns cruzamentos entre variáveis, para inspecionar eventuais associações.

Tabela 6. 2: Quadro resumo de estatísticas descritivas das questões demográficas

Questões demográficas		
Variável	Codificação (no SPSS)	Frequências
Género	1: Masculino	406 (42,9%)
	2: Feminino	541 (57,1%)
		Estatísticas
Idade	Não Aplicável	Mínimo: 45 Máximo: 98 Média: 71,73 Mediana: 71 Desvio Padrão: 8,05
Idade que sente ter	Não Aplicável	Mínimo: 8 Máximo: 100 Média: 67,12 Mediana: 68 Desvio Padrão: 11,95

Com base nos dados que constam da tabela 6.2 anteriormente apresentada é possível obter as seguintes conclusões:

Género: a amostra em estudo é constituída por um maior número de indivíduos do sexo feminino, o que demonstra coerência com a realidade da população portuguesa. Uma vez que são as mulheres aquelas que apresentam uma maior esperança média de vida, e na realidade existem em maior número na sociedade, a amostra revela-se representativa no que respeita a esta variável. Verificou-se contudo que a percentagem de respondentes de cada género é um pouco desequilibrada em relação às percentagens reais do grupo etário em estudo.

Idade: relativamente à idade dos indivíduos em estudo, verifica-se que está compreendida entre os 45 e os 98 anos e que em média o indivíduo tem 71 anos, o que mais uma vez se revela ser representativo da população portuguesa, que, tal como verificado anteriormente apresenta uma pirâmide etária com predominância de idosos (população com mais de 60 anos).

Idade que sente ter: no que respeita a esta variável verifica-se que existe uma grande amplitude amostral, uma vez que existem indivíduos que se sentem bastante jovens, com apenas 8 anos de idade indicados, e outros que se sentem bastante idosos, com 100 anos de idade. Verifica-se que em média os cidadãos inquiridos sentem ter, aproximadamente 67 anos.

Dada a diferença entre a idade real média e a idade média que o indivíduo sente ter, foi feito um estudo comparativo entre as duas variáveis, de modo a verificar se as diferenças registadas são estatisticamente significativas. Para tal, de modo a averiguar se existem diferenças significativas entre a idade dos indivíduos e a idade que sentem ter, e uma vez que não se verifica o pressuposto de normalidade das observações, será aplicado o teste não paramétrico para amostras emparelhadas, Teste de *Wilcoxon*, onde as hipóteses a testar são:

*H₀: A distribuição da idade sentida pelos indivíduos é idêntica
à distribuição da idade real*

Vs.

*H₁: A distribuição da idade sentida pelos indivíduos é diferente da
distribuição da idade real*

Com a aplicação do teste de hipóteses referido, e atendendo ao resultado obtido para o valor-*p* (valor-*p* < 0,001), rejeita-se H₀, ou seja, existe evidência estatística para afirmar que a idade sentida pelos indivíduos é diferente da sua idade real. Numa análise mais fina, o

resultado indica que a idade sentida apresenta, tipicamente, valores inferiores à idade real dos indivíduos, também em média.

As diferenças que se verificaram na base de dados em estudo, são também verificadas no contexto de outras análises já realizadas. Tal como é abordado em Rubin, D et. al (2006), que, com base numa amostra de 1470 adultos dinamarqueses, com idades compreendidas entre os 20 e os 97 anos, conclui que os indivíduos mais velhos tendem a sentir-se 20% mais jovens.

Idade que sente ter Vs. Idade real por género

De modo a averiguar se a diferença verificada anteriormente acontece também em cada um dos géneros de forma individual, será em seguida estudada a relação entre as variáveis: idade que sente ter e idade real, por género.

Apresentam-se na tabela 6.3, algumas medidas de localização e dispersão das variáveis mencionadas, por género.

Tabela 6. 3: Quadro resumo de estatísticas descritivas das variáveis "idade" e "idade que sente ter" por género

		Estatísticas
Género Masculino (n= 406)	Idade Real	Mínimo: 51 Máximo: 94 Média: 72,54 Mediana: 72 Desvio Padrão: 7,44
	Idade que sente ter	Mínimo: 20 Máximo: 95 Média: 67,95 Mediana: 69 Desvio Padrão: 11,69
Género Feminino (n= 541)	Idade Real	Mínimo: 45 Máximo: 98 Média: 71,12 Mediana: 71 Desvio Padrão: 8,43
	Idade que sente ter	Mínimo: 8 Máximo: 100 Média: 66,50 Mediana: 67 Desvio Padrão: 12,12

Por observação dos resultados da tabela 6.3, verifica-se que existe uma maior amplitude amostral na idade dos indivíduos do sexo feminino, tanto na idade real, como na idade que sente ter. Verifica-se também que, em média, os homens são cerca de 1 ano mais novos que as mulheres, contudo, em termos médios, sentem ter uma idade superior (67,95) aquela que as mulheres em média sentem ter (66,50).

Através do *boxplot* apresentado na figura 6.1. verifica-se a existência de alguns valores *outlier* sendo que nos *boxplot* relativos à idade sentida pelos indivíduos, em ambos os géneros, é evidente a existência de *outliers* severos, tal facto pode evidenciar a existência de problemas de saúde mental dos inquiridos.

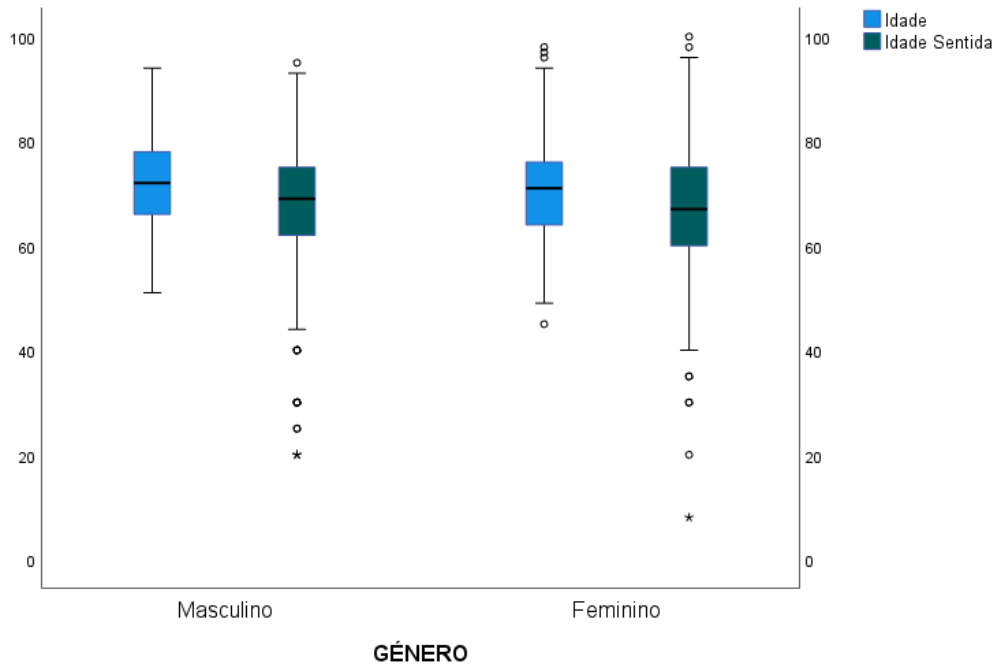


Figura 6. 1: Boxplot das variáveis idade e idade que sente ter, por gênero

Em seguida serão apresentadas as frequências obtidas nas respostas às questões classificadas anteriormente como sendo de saúde física e comportamentos de saúde.

Tabela 6. 4: Tabela de Frequências das questões de saúde física e comportamentos de saúde

Questões de saúde física e comportamentos de saúde		
Variável	Codificação	Frequências
Perceção do Estado de Saúde	1: Excelente	22 (2,3%)
	2: Muito Boa	50 (5,3%)
	3: Boa	272 (28,7%)
	4: Razoável	414 (43,7%)
	5: Má	189 (20%)
Tem fratura do quadril	1: Sim	48 (5,1%)
	5: Não	899 (94,9%)
Tem diabetes ou açúcar elevado no sangue	1: Sim	232 (24,5%)
	5: Não	715 (75,5%)
Tem tensão alta ou hipertensão	1: Sim	519 (54,8%)
	5: Não	428 (45,2%)
Teve um enfarte agudo do miocárdio ou tem outro problema de coração	1: Sim	161 (17%)
	5: Não	786 (83%)
Tem doença pulmonar crónica	1: Sim	104 (11%)
	5: Não	843 (89%)
Tem cancro ou um tumor maligno	1: Sim	88 (9,3%)
	5: Não	859 (90,7%)
Tem outra doença ou condição de saúde	1: Sim	375 (39,6%)
	5: Não	572 (60,4%)
Toma Medicação Regular	1: Sim	847 (99,4%)
	5: Não	100 (10,6%)
Saiu de casa nos últimos 3 meses	1: Sim	879 (92,8%)
	5: Não	68 (7,2%)
Tomou Medicação Preventiva para a COVID-19	1: Sim	47 (5%)
	5: Não	900 (95%)
Está vacinado contra a COVID-19	1: Sim	883 (93,2%)
	5: Não	64 (6,8%)
Se não está vacinado pretende ser vacinado contra COVID-19	-9: Não Aplicável*	883 (93,2%)
	1: Sim, já tenho a vacinação agendada	9 (1%)
	2: Sim, quero ser vacinado	36 (3,8%)
	3: Não, não quero ser vacinado	13 (1,4%)
	4: Continuo indeciso	6 (0,6%)

* Esta questão não é aplicável aos indivíduos que responderam de forma afirmativa à questão: Está vacinado contra a COVID-19?

Com base nos dados que constam da tabela anteriormente apresentada é possível obter as seguintes conclusões:

Percepção do Estado de Saúde: na amostra considerada existe uma prevalência de indivíduos que sentem ter um estado de saúde razoável ou bom (72,4%). De modo a verificar a independência entre esta variável e o género foi construída a tabela de contingência apresentada de seguida.

Tabela 6. 5: Tabela de Contingência das variáveis "Percepção do Estado de Saúde" e Género

		Género	
		Masculino	Feminino
Percepção do Estado de Saúde	Excelente	11 (2,71%)	11 (2,03%)
	Muito Boa	24 (5,91%)	26 (4,81%)
	Boa	131 (32,27%)	141 (26,06%)
	Razoável	171 (42,12%)	243 (44,92%)
	Má	69 (16,99%)	120 (22,18%)

Da observação da tabela acima (Tabela 6.5.) é de realçar que dos indivíduos do sexo feminino 44,92% consideram ter um estado de saúde razoável. No género masculino apenas 42.12% consideram que seu estado de saúde é razoável.

A fim de averiguar se existe independência entre as variáveis “percepção do estado de saúde” e “género”, será aplicado o teste de Qui-quadrado para a independência. Na aplicação deste teste consideram-se as seguintes hipóteses:

H_0 : A percepção do estado de saúde e o género são independentes

Vs.

H_1 : A percepção do estado de saúde e o género não são independentes

Através da aplicação do teste referido e atendendo ao resultado obtido para o *valor-p* (*valor-p* = 0.106 e $\chi^2_{\text{observado}} = 7.642$), não rejeitamos H_0 , logo não existe evidência estatística de associação significativa entre a percepção do estado de saúde e o género. Posto isto, não serão determinados os coeficientes de associação entre as variáveis.

Tem fratura do quadril: na amostra considerada apenas 48 indivíduos sofrem de fratura do quadril, sendo que 28 são mulheres. De modo a verificar a eventual associação entre esta variável e o género, foi construída a tabela de contingência e realizado novo teste, como apresentado em seguida.

Tabela 6. 6:Tabela de Contingência das variáveis "Tem fratura do quadril" e Género

		Género	
		Masculino	Feminino
Tem fratura do quadril	Sim	20 (4,93%)	28 (5,18%)
	Não	386 (95,07%)	513 (94,82%)

Para a aplicação do Teste de Qui-quadrado para a independência são consideradas as seguintes hipóteses a testar:

H_0 : *Ter fratura do quadril e género são independentes*

Vs.

H_1 : *Ter fratura do quadril e género não são independentes*

Com a aplicação do teste referido e atendendo ao resultado obtido para o *valor-p* (*valor-p* = 0,862 e $\chi^2_{\text{observado}} = 0.30$), não se rejeita H_0 , logo não existe evidência estatística de associação entre a ocorrência de fratura do quadril e o género.

Tratando-se de uma tabela de dimensão 2×2 , em alternativa é possível utilizar o teste exato de Fisher. Com a aplicação deste teste e atendendo ao valor de *valor-p* registado (*valor-p* = 0,882), obtém-se uma conclusão análoga.

Posto isto, não serão determinados os coeficientes de associação entre as variáveis.

Tem diabetes ou açúcar elevado no sangue: na amostra considerada existe uma prevalência de indivíduos que não apresentam a patologia de diabetes.

Tabela 6. 7: Tabela de Contingência das variáveis "Tem diabetes ou açúcar elevado no sangue" e género

		Género	
		Masculino	Feminino
Tem diabetes ou açúcar elevado no sangue	Sim	107 (26,35%)	125 (23,11%)
	Não	299 (73,65%)	416 (76,89%)

Por observação da tabela de contingência anterior verifica-se que 26,35% dos indivíduos do sexo masculino sofrem de diabetes, enquanto apenas 23,11% dos indivíduos do sexo feminino apresentam esta patologia (ou apenas açúcar elevado).

De modo a verificar a independência entre as duas variáveis, realizou-se o Teste de Qui-quadrado para a independência, sobre as seguintes hipóteses:

H_0 : *Ter diabetes ou açúcar elevado no sangue é independente do género*

Vs.

H_1 : *Ter diabetes ou açúcar elevado no sangue não é independente do género*

Com a aplicação do teste referido e atendendo ao resultado obtido para o *valor-p* (*valor-p* = 0,25 e $\chi^2_{observado} = 1.324$), não existe evidência estatística para afirmar que a patologia de diabetes não seja independente do sexo.

O teste exato de Fisher para as tabelas 2x2 conduz a uma conclusão análoga, atendendo ao valor de *valor-p* registado (*valor-p* = 0,253). Não se justifica assim o cálculo dos coeficientes de associação entre as variáveis.

Tem tensão alta ou hipertensão: na amostra considerada existe uma prevalência de indivíduos que apresentam a patologia de hipertensão.

Tabela 6. 8: Tabela de Contingência das variáveis " Tem tensão alta ou hipertensão" e género

		Género	
		Masculino	Feminino
Tem tensão alta ou hipertensão	Sim	231 (56,90%)	288 (53,23%)
	Não	175 (43,10%)	253 (46,77%)

Por observação da tabela de contingência anterior verifica-se que tanto nos indivíduos do género masculino como feminino existe uma prevalência da patologia de hipertensão.

De modo a verificar a independência entre esta variável e o género aplica-se novamente Teste de Qui-quadrado para a independência, sendo consideradas as seguintes hipóteses do teste:

H_0 : *Tem tensão alta ou hipertensão e género são independentes*

Vs.

H_1 : *Tem tensão alta ou hipertensão e género não são independentes*

Através da aplicação do teste referido, atendendo ao resultado obtido para o *valor-p* (*valor-p* = 0,262 e $\chi^2_{observado} = 1.256$), não existe evidência estatística de associação significativa entre as duas variáveis. Com o teste Exato de Fisher obtém-se uma conclusão análoga, como era esperado (*valor-p* = 0,291).

Teve um enfarte agudo do miocárdio (EAM) ou tem outro problema de coração: na amostra considerada existe uma prevalência de indivíduos que não apresentam problemas de coração.

Tabela 6. 9: Tabela de Contingência das variáveis "Teve um enfarte agudo do miocárdio ou tem outro problema de coração" e Género

		Género	
		Masculino	Feminino
Teve um enfarte agudo do miocárdio ou tem outro problema de coração	Sim	84 (20,69%)	77 (14,23%)
	Não	322 (79,31%)	464 (85,77%)

Por observação da tabela de contingência anterior verifica-se que, em ambos os géneros, não existe uma prevalência da doença.

Como anteriormente aplica-se o Teste de Qui-quadrado, neste caso para as seguintes hipóteses em teste:

H_0 : *Teve um enfarte agudo do miocárdio ou tem outro problema de coração e género são independentes*

Vs.

H_1 : *Teve um enfarte agudo do miocárdio ou tem outro problema de coração e género não são independentes*

Foi obtidos $valor-p = 0,009$ e $\chi^2_{observado} = 6.853$, rejeita-se a hipótese nula, logo existe evidência estatística para afirmar que existe uma associação entre o género e os tipos de patologia considerados. Utilizado o teste exato de Fisher obtém-se uma conclusão análoga, atendendo ao valor de $valor-p$ registado ($valor-p = 0,011$).

Dado que se verifica que as variáveis em estudo não são independentes faz sentido quantificar a relação existente entre elas. Para tal foi determinado o seguinte coeficiente de associação.

Tabela 6. 10: Coeficiente de Contingência de Pearson entre as variáveis "Teve um enfarte agudo do miocárdio ou tem outro problema de coração" e género

Coeficiente de Contingência de Pearson
0,085

Atendendo ao resultado obtido verifica-se uma fraca associação entre as variáveis.

Tem doença pulmonar crónica: na amostra considerada existe uma prevalência de indivíduos que não apresentam a patologia de doença pulmonar crónica.

Tabela 6. 11: Tabela de Contingência das variáveis "Tem doença pulmonar crónica" e género

		Género	
		Masculino	Feminino
Tem doença pulmonar crónica	Sim	54 (13,30%)	50 (9,24%)
	Não	352 (86,70%)	491 (90,76%)

Por observação da tabela de contingência anterior verifica-se que no género masculino 13.30% dos indivíduos apresentam doença pulmonar crónica, enquanto apenas 9,24% dos indivíduos do sexo feminino apresentam esta patologia.

Prosseguindo com o Teste de Qui-quadrado para a independência, para as seguintes hipóteses a testar:

H₀: Tem doença pulmonar crónica e género são independentes

Vs.

H_1 : Tem doença pulmonar crónica e género não são independentes

Atendendo ao *valor-p* obtido (*valor-p* = 0,048 e $\chi^2_{observado} = 3.908$), existe evidência estatística para afirmar que o facto do individuo sofrer de doença pulmonar obstrutiva crónica não é independente do género. A aplicação do teste exato de Fisher conduz a uma conclusão que contraria a anterior, atendendo ao valor de *valor-p* registado (*valor-p* = 0.058). Este último teste é o mais fidedigno para a situação de tabelas 2x2 (teste Exato), ainda assim optou-se por quantificar a associação existente entre as duas variáveis através do Coeficiente de Contingência de Pearson, cujo resultado é apresentado em seguida.

Tabela 6. 12: Coeficiente de Contingência de Pearson entre as variáveis "Tem doença pulmonar crónica" e género

Coeficiente de Contingência de Pearson
0,064

Atendendo ao resultado obtido verifica-se uma fraca associação entre as variáveis.

Tem cancro ou um tumor maligno: na amostra considerada existe uma prevalência de indivíduos que não apresentam a patologia de cancro ou tumor maligno.

Tabela 6. 13: Tabela de Contingência das variáveis "Tem cancro ou um tumor maligno" e Género

		Género	
		Masculino	Feminino
Tem cancro ou um tumor maligno	Sim	40 (9,85%)	48 (8,87%)
	Não	366 (90,15%)	493 (91,73%)

Por observação da tabela de contingência anterior verifica-se que 8,87% dos indivíduos do género feminino sofrem de cancro, enquanto 9,85% dos indivíduos do género masculino sofrem desta patologia.

De modo a verificar a independência entre as variáveis será aplicado um Teste de Qui-quadrado para a independência, sendo consideradas as seguintes hipóteses a testar:

H_0 : *Tem cancro ou um tumor maligno e género são independentes*

Vs.

H_1 : *Ter cancro ou um tumor maligno e género não são independentes*

Com a aplicação do teste de hipótese referido e atendendo ao resultado obtido para o valor-p ($\text{valor-p} = 0,607$ e $\chi^2_{\text{observado}} = 0.264$), não existe evidência estatística para afirmar que sofrer da patologia de cancro não é independente do sexo.

Tem outra doença ou condição de saúde: na amostra considerada existe uma prevalência de indivíduos que não apresentam outras patologias ou condições de saúde para além das anteriormente referidas.

Tabela 6. 14: Tabela de Contingência das variáveis "Tem outra doença ou condição de saúde" e género

		Género	
		Masculino	Feminino
Tem outra doença ou condição de saúde	Sim	147 (36,21%)	228 (42,14%)
	Não	259 (63,79%)	313 (57,86%)

Por observação da tabela de contingência anterior verifica-se que 42,14% dos indivíduos do sexo feminino sofrem de outras condições de saúde para além das referida. É aplicado o Teste de Qui-quadrado para a independência, sendo consideradas as seguintes hipóteses a testar:

H_0 : As variáveis tem outra doença ou condição de saúde e género são independentes

Vs.

H_1 : As variáveis tem outra doença ou condição de saúde e género não são independentes

Com a aplicação do teste referido e atendendo ao resultado obtido para o *valor-p* (*valor-p* = 0,064 e $\chi^2_{\text{observado}} = 3.418$), não existe evidência estatística para afirmar que a variável “Tem outra doença ou condição de saúde” não seja independente do sexo.

Dado que se trata de uma tabela de contingência de dimensão 2×2 é também passível de ser utilizado o teste exato de Fisher. Com aplicação deste teste obtém-se uma conclusão análoga, atendendo ao valor de *valor-p* registado (*valor-p* = 0,07).

Toma Medicação Regular: na amostra considerada existe uma prevalência de indivíduos que afirmam tomar medicação de forma regular.

Tabela 6. 15: Tabela de Contingência das variáveis "Toma medicação regular" e género

		Género	
		Masculino	Feminino
Toma Medicação Regular	Sim	361 (88,92%)	486 (89,83%)
	Não	45 (11,08%)	55 (10,17%)

Por observação da tabela de contingência anterior verifica-se que 89,83% dos indivíduos do sexo feminino toma medicação de forma regular. Para a aplicação do Teste de Qui-quadrado para a independência, são consideradas as seguintes hipóteses a testar:

H_0 : As variáveis toma medicação regular e género são independentes

Vs.

H_1 : As variáveis toma medicação regular e género não são independentes

Com a aplicação do teste referido e atendendo ao resultado obtido para o *valor-p* (*valor-p* = 0,649 e $\chi^2_{observado} = 0.207$), não existe evidência estatística para afirmar que a toma de medicação regular não seja independente do sexo.

Dado que se trata de uma tabela de contingência de dimensão 2×2 é também passível de ser utilizado o teste exato de Fisher. Com aplicação deste teste obtém-se uma conclusão análoga, atendendo ao valor de *valor-p* registado (*valor-p* = 0,67).

Saiu de casa nos últimos 3 meses: na amostra considerada existe uma prevalência de indivíduos que afirmam ter saído de casa nos últimos 3 meses.

Tabela 6. 16: Tabela de Contingência da variável "Saiu de casa nos 3 últimos meses" e Género

		Género	
		Masculino	Feminino
Saiu de casa nos últimos 3 meses	Sim	383 (94,33%)	496 (91,68%)
	Não	23 (5,67%)	45 (8,32%)

Por observação da tabela de contingência anterior verifica-se que 94,33% dos indivíduos do sexo masculino afirmam ter saído de casa nos últimos 3 meses.

De modo a verificar a independência entre esta variável e o género do indivíduo será aplicado um Teste de Qui-quadrado para a independência, sendo consideradas as seguintes hipóteses a testar:

H_0 : As variáveis saiu de casa nos últimos 3 meses e gênero são independentes

Vs.

H_1 : As variáveis saiu de casa nos últimos 3 meses e gênero não são independentes

Com a aplicação do teste referido e atendendo ao resultado obtido para o *valor-p* (*valor-p* = 0,118 e $\chi^2_{\text{observado}} = 2.449$), não existe evidência estatística para afirmar que a toma de medicação regular não seja independente do sexo.

Dado que se trata de uma tabela de contingência de dimensão 2×2 é também passível de ser utilizado o teste exato de Fisher. Com aplicação deste teste obtém-se uma conclusão análoga, atendendo ao valor de *valor-p* registado (*valor-p* = 0,128).

Tomou Medicação Preventiva para a COVID-19: na amostra considerada existe uma prevalência de indivíduos que afirmam não ter tomado medicação preventiva para a COVID-19.

Tabela 6. 17: Tabela de Contingência da variável "Tomou Medicação Preventiva para a COVID-19" e Género

		Género	
		Masculino	Feminino
Tomou Medicação Preventiva para a COVID-19	Sim	18 (4,43%)	29 (5,36%)
	Não	388 (95,57%)	512 (94,64%)

Por observação da tabela de contingência anterior verifica-se que apenas 5,36% dos indivíduos do sexo feminino tomaram medição preventiva para a COVID-19. De modo a verificar a independência das variáveis é aplicado o Teste de Qui-quadrado para a independência, sendo consideradas as seguintes hipóteses a testar:

H_0 : As variáveis tomou medicação preventiva para a COVID 19 e género são independentes

Vs.

H_1 : As variáveis tomou medicação preventiva para a COVID 19 e género não são independentes

Com a aplicação do teste referido e atendendo ao resultado obtido para o *valor-p* (*valor-p* = 0,516 e $\chi^2_{\text{observado}} = 0.423$), não existe evidência estatística para afirmar que o facto de o indivíduo tomar medicação preventiva para a COVID-19 não seja independente do seu género.

Dado que se trata de uma tabela de contingência de dimensão 2×2 é também passível de ser utilizado o teste exato de Fisher. Com a aplicação deste teste obtém-se uma conclusão análoga, atendendo ao valor de *valor-p* registado (*valor-p* = 0,549).

Está vacinado contra a COVID-19: na amostra considerada existe uma prevalência de indivíduos que afirmam estarem vacinados contra a COVID-19.

Tabela 6. 18: Tabela de Contingência das variáveis "Está vacinado contra a COVID-19" e Género

		Género	
		Masculino	Feminino
Está vacinado contra a COVID-19	Sim	382 (94,09%)	501 (92,61%)
	Não	24 (5,91%)	40 (7,39%)

Por observação da tabela de contingência anterior verifica-se que 94.09% dos indivíduos do sexo masculino se encontram vacinados contra a COVID-19.

De modo a verificar a independência entre esta variável e o género do indivíduo será aplicado um Teste de Qui-quadrado para a independência, sendo consideradas as seguintes hipóteses a testar:

H_0 : As variáveis está vacinado contra a COVID 19 e género são independentes

Vs.

H_1 : As variáveis está vacinado contra a COVID 19 e género não são independentes

Com a aplicação do teste referido e atendendo ao resultado obtido para o *valor-p* (*valor-p* = 0,368 e $\chi^2_{observado} = 0.809$), não existe evidência estatística para afirmar que o facto se encontrar vacinado contra a COVID-19 não seja independente do seu género.

Dado que se trata de uma tabela de contingência de dimensão 2×2 é também passível de ser utilizado o teste exato de Fisher. Com aplicação deste teste obtém-se uma conclusão análoga, atendendo ao valor de *valor-p* registado (*valor-p* = 0,433).

Se não está vacinado pretende ser vacinado contra a COVID-19: para análise desta variável serão apenas considerados os indivíduos que responderam negativamente à questão: Está vacinado contra a COVID-19?, sendo por isso considerado um total de 64 indivíduos.

Tabela 6. 19: Tabela de contingência das variáveis "Pretende ser vacinado contra a COVID-19" e Género

		Género	
		Masculino	Feminino
Pretende ser vacinado contra a COVID-19	Sim, já tenho a vacinação agendada	3 (0,74%)	6 (1,11%)
	Sim, quero ser vacinado	14 (3,47%)	22 (4,07%)
	Não, Não quero ser vacinado	6 (1,48%)	7 (1,29%)
	Continuo indeciso	1 (0,25%)	5 (0,92%)

Por observação da tabela de contingência anterior verifica-se que 4,07% dos indivíduos do sexo feminino apresentam interesse em ser vacinado contra a COVID-19.

Atendendo ao facto de não se verificarem os pressupostos necessários para a aplicação do teste de Qui-quadrado para a independência, visto que 62.5% das células da tabela de contingência apresenta uma frequência esperada inferior a 5, é necessário recorrer ao teste de Qui-quadrado por Simulação de Monte Carlo, de modo a verificar a independência entre esta variável e o género do indivíduo, sendo consideradas as seguintes hipóteses a testar:

H_0 : *Pretende ser vacinado contra a COVID 19 e género
são independentes*

Vs.

H_1 : *Pretende ser vacinado contra a COVID 19 e género
não são independentes*

Com a aplicação do teste referido e atendendo ao resultado obtido para o *valor-p* (*valor-p* = 0,695 e $\chi^2_{\text{observado}} = 2.362$), não existe evidência estatística de associação da intenção do indivíduo ser vacinado contra a COVID-19 e o género.

Em seguida serão apresentadas as frequências obtidas as respostas as questões classificadas anteriormente como sendo de saúde mental.

Tabela 6. 20: Tabela de Frequências das questões de saúde mental

Questões de saúde mental		
Variável	Codificação	Frequências
No último mês sentiu-se nervoso, ansioso ou no limite	1: Sim	497 (52,5%)
	5: Não	450 (47,5%)
No último mês esteve triste ou deprimido	1: Sim	416 (43,9%)
	5: Não	531 (56,1%)
Teve dificuldades em dormir recentemente	1: Problemas com o sono ou mudança recente no padrão	410 (43,3%)
	2: Sem problemas com o sono	537 (56,7%)
Sente-se solitário	1: Muitas Vezes	76 (8%)
	2: Algumas Vezes	222 (23,4%)
	3: Quase nunca ou nunca	649 (68,5%)

No último mês sentiu-se nervoso, ansioso ou no limite: verifica-se a prevalência se inquiridos que afirmam se terem sentido nervosos ansiosos ou no limite no último mês. De modo a verificar a independência entre esta variável e o género do indivíduo inquirido foi construída a tabela de contingência apresentada em seguida.

Tabela 6. 21: Tabela de Contingência das variáveis "No último mês sentiu-se nervoso, ansioso ou no limite" e género

		Género	
		Masculino	Feminino
No último mês sentiu-se nervoso, ansioso ou no limite	Sim	173 (42,61%)	324 (59,89%)
	Não	233 (57,39%)	217 (40,11%)

Da observação da tabela acima é importante mencionar que se verifica que 59,89% dos indivíduos do género feminino se sentiram nervosos, ansiosos ou no limite no último mês.

Em seguida, será aplicado o teste de Qui-quadrado para a independência. Para a aplicação deste teste consideram-se as seguintes hipóteses:

H_0 : As variáveis no último mês sentiu – se nervoso, ansioso ou no limite
e gênero são independentes

Vs.

H_1 : As variáveis no último mês sentiu – se nervoso, ansioso ou no limite
e gênero não são independentes

Com a aplicação do teste referido e atendendo ao resultado obtido para o *valor-p* (*valor-p* < 0,001 e $\chi^2_{observado} = 27.765$), existe evidência estatística para afirmar que a variável “no último mês sentiu-se nervoso, ansioso ou no limite” é não independente do sexo. Atendendo ao fato de se tratar de uma tabela de contingência de dimensão 2×2 , também seria possível aplicar o teste exato de Fisher, sendo obtido o mesmo valor para *valor-p*.

Atendendo o resultado obtido através do teste de Qui-quadrado, e dado que se verifica a existência de associação entre as variáveis é determinado o coeficiente de associação apresentado de seguida.

Tabela 6. 22: Coeficiente de Contingência de Pearson entre as variáveis "No último mês sentiu-se nervoso, ansioso ou no limite " e gênero

Coeficiente de Contingência de Pearson
0,169

Atendendo ao resultado obtido verifica-se uma fraca associação entre as variáveis.

No último mês esteve triste ou deprimido: verifica-se a prevalência de inquiridos que afirmam não se terem sentido tristes ou deprimidos no último mês. De modo a verificar a independência entre a futura variável dependente do modelo e o gênero do individuo inquirido foi construída a respetiva tabela de contingência

Tabela 6. 23: Tabela de contingência das variáveis "No último mês esteve triste ou deprimido" e gênero

		Gênero	
		Masculino	Feminino
No último mês esteve triste ou deprimido	Sim	125 (30,79%)	291 (53,79%)
	Não	281 (69,21%)	250 (46,21%)

Da observação da tabela acima é importante mencionar que 53,79% dos indivíduos do sexo feminino se sentiram tristes ou deprimidos no último mês, enquanto apenas 30,79% dos indivíduos do gênero masculino apresentaram esta sintomatologia.

Em seguida, será aplicado o teste de Qui-quadrado para a independência. Para a aplicação deste teste consideram-se as seguintes hipóteses:

H_0 : *As variáveis no último mês sentiu – se triste ou deprimido e gênero são independentes*

Vs.

H_1 : *As variáveis no último mês sentiu – se triste ou deprimido e gênero não são independentes*

Com a aplicação do teste referido e atendendo ao resultado obtido para o *valor-p* (*valor-p* < 0,001 e $\chi^2_{observado} = 49.818$), existe evidência estatística para afirmar que a variável no último mês sentiu-se triste ou deprimido é não independente do sexo. Atendendo ao fato de se tratar de uma tabela de contingência de dimensão 2×2 , também seria possível aplicar o teste exato de Fisher, sendo obtido o mesmo valor para *valor-p*.

Atendendo o resultado obtido através do teste de Qui-quadrado, e dado que se verifica a existência de associação foi determinado o seguinte coeficiente de associação.

Tabela 6. 24: Coeficiente de Contingência de Pearson entre as variáveis "No último mês esteve triste ou deprimido " e género

Coeficiente de Contingência de Pearson
0,224

Atendendo ao resultado obtido verifica-se uma fraca associação entre as variáveis.

Teve dificuldades em dormir recentemente: verifica-se a prevalência se inquiridos que afirmam não ter tido problemas com o sono. De modo a verificar a independência entre esta variável e o género do individuo inquirido foi construída a tabela de contingência apresentada em seguida.

Tabela 6. 25: Tabela de contingência das variáveis "Teve dificuldades em dormir recentemente" e Género

		Género	
		Masculino	Feminino
Teve dificuldades em dormir recentemente	Problemas com o sono ou mudança recente no padrão	135 (33,25%)	275 (50,83%)
	Sem problemas com o sono	271 (66,75%)	266 (49,17%)

Da observação da tabela acima é importante mencionar que se verifica que 50,83% dos indivíduos do género feminino apresentam dificuldades com o sono.

Em seguida, será aplicado o teste de Qui-quadrado para a independência. Para a aplicação deste teste consideram-se as seguintes hipóteses:

H_0 : As variáveis teve dificuldades em dormir recentemente e género são independentes

Vs.

*H₁: As variáveis teve dificuldades em dormir recentemente
e gênero não são independentes*

Com a aplicação do teste referido e atendendo ao resultado obtido para o *valor-p* (*valor-p* < 0,001 e $\chi^2_{observado} = 29.2$), existe evidência estatística para afirmar que a variável teve dificuldades em dormir recentemente não independente do sexo. Atendendo ao fato de se tratar de uma tabela de contingência de dimensão 2×2 , também seria possível aplicar o teste exato de Fisher, sendo obtido o mesmo valor para *valor-p*.

Atendendo ao resultado obtido através do teste de Qui-quadrado, e dado que se verifica a existência de associação entre as variáveis foi determinado o coeficiente de contingência de Pearson.

Tabela 6. 26: Coeficiente de Contingência de Pearson entre as variáveis "Teve dificuldades para dormir recentemente" e gênero

Coeficiente de Contingência de Pearson
0,173

Atendendo ao resultado obtido verifica-se uma fraca associação entre as variáveis.

Sente-se solitário: verifica-se a prevalência se inquiridos que afirmam quase nunca se terem sentido solitários. De modo a verificar a independência entre esta variável e o gênero do indivíduo inquirido foi construída a tabela de contingência apresentada em seguida.

Tabela 6. 27: Tabela de contingência das variáveis "Sente-se Solitário" e Género

		Género	
		Masculino	Feminino
Sente-se solitário	Muitas vezes	17 (4,19%)	59 (10,91%)
	Algumas vezes	69 (17%)	153 (28,28%)
	Quase nunca ou nunca	320 (78,81%)	329 (60,81%)

Da observação da tabela acima é importante mencionar que se verifica que apenas 10,91% dos indivíduos do sexo feminino se sentiu muitas vezes solitário.

Em seguida, será aplicado o teste de Qui-quadrado para a independência. Para a aplicação deste teste consideram-se as seguintes hipóteses:

H_0 : As variáveis sente – se solitário e género são independentes

Vs.

H_1 : As variáveis sente – se solitário e género não são independentes

Com a aplicação do teste de hipóteses e atendendo ao resultado obtido para o *valor-p* (*valor-p* < 0,001 e $\chi^2_{observado} = 36.618$), existe evidência estatística para afirmar que a variável sente-se solitário é não independente do sexo.

Atendendo ao resultado obtido através do teste de Qui-quadrado, e dado que se verifica a existência de associação entre as variáveis foi determinado o coeficiente de contingência de Pearson.

Tabela 6. 28: Coeficiente de Contingência de Pearson entre as variáveis "Sente-se solitário" e género

Coeficiente de Contingência de Pearson
0,193

Atendendo ao resultado obtido verifica-se uma fraca associação entre as variáveis.

Em seguida serão apresentadas as frequências obtidas as respostas as questões classificadas anteriormente como sendo sobre a COVID-19.

Tabela 6. 29: Tabela de Frequências de questões sobre a COVID-19

Questões sobre a COVID-19		
Variável	Codificação	Frequências
Quantas vezes testou positivo para a COVID-19	1: Nunca	570 (60,2%)
	2: Apenas 1 vez	197 (20,8%)
	3: Entre 2 e 5 vezes	150 (15,8%)
	4: Entre 6 e 10 vezes	19 (2%)
	5: Mais de 10 vezes	11 (1,2%)

Quantas vezes testou positivo para a Covid-19: por observação das frequências verifica-se a prevalência de indivíduos que nunca contraíram a patologia de COVID-19. De modo a verificar a independência entre esta variável e o género do individuo inquirido foi construída a tabela de contingência apresentada em seguida.

Tabela 6. 30: Tabela de Contingência das variáveis "Quantas vezes testou positivo para a COVID-19" e género

		Género	
		Masculino	Feminino
Quantas vezes testou positivo para a COVID-19	Nunca	235 (57,88%)	335 (61,92%)
	Apenas 1 vez	91 (22,41%)	106 (19,59%)
	Entre 2 e 5 vezes	66 (16,26%)	84 (15,53%)
	Entre 6 e 10 vezes	7 (1,72%)	12 (2,22%)
	Mais de 10 vezes	7 (1,72%)	4 (0,74%)

De modo a verificar a independência entre esta variável e o género do indivíduo será aplicado um Teste de Qui-quadrado para a independência, sendo consideradas as seguintes hipóteses a testar:

H_0 : *As variáveis quantas vezes testou positivo para a Covid19 e género são independentes*

Vs.

H_1 : *As variáveis quantas vezes testou positivo para a Covid19 e género não são independentes*

Com a aplicação do teste de hipótese referido e atendendo ao resultado obtido para o valor-p ($\text{valor-p} = 0,432$ e $\chi^2_{\text{observado}} = 3.812$), não existe evidência estatística para afirmar que o número de vezes que o indivíduo contraiu a patologia de COVID-19 não seja independente do seu género.

Em seguida é apresentada a caracterização das respostas as questões classificadas como sendo sobre a Situação Profissional dos inquiridos.

Tabela 6. 31: Tabela de Frequências de questões sobre a situação profissional

Questões sobre situação profissional		
Variável	Codificação	Frequências
Situação atual de emprego	1: Reformado	704 (74,3%)
	2: Empregado ou Trabalhador Independente	122 (12,9%)
	3: Desempregado	21 (2,2%)
	4: Permanentemente doente ou incapacitado	16 (1,7%)
	5: Dona de Casa	58 (6,1%)
	97: Outro	26 (2,7%)

Situação atual de emprego: por observação das frequências verifica-se a prevalência de indivíduos que se encontram reformados, o que já seria espectável tendo em conta a idade da maioria dos inquiridos. De modo a verificar a independência entre esta variável e o género do indivíduo inquirido foi construída a tabela de contingência apresentada em seguida.

Tabela 6. 32: Tabela de Contingência entre as variáveis "Situação Atual de Emprego" e Género

		Género	
		Masculino	Feminino
Situação Atual de Emprego	Reformado	342 (84,24%)	362 (66,91%)
	Empregado ou Trabalhador Independente	41 (10,10%)	81 (14,97%)
	Desempregado	7 (1,72%)	14 (2,59%)
	Permanentemente doente ou incapacitado	8 (1,97%)	8 (1,49%)
	Dona de Casa	0 (0%)	58 (10,72%)
	Outro	8 (1,97%)	18 (3,33%)

De modo a verificar a independência entre esta variável e o género do indivíduo será aplicado um Teste de Qui-quadrado para a independência, sendo consideradas as seguintes hipóteses a testar:

H_0 : As variáveis situação atual de emprego e género são independentes

Vs.

H_1 : As variáveis situação atual de emprego e género não são independentes

Com a aplicação do teste referido e atendendo ao resultado obtido para o *valor-p* (*valor-p* < 0,001 e $\chi^2_{observado} = 59.833$), existe evidência estatística para afirmar que a situação atual de emprego não é independente do seu género.

Atendendo ao resultado obtido através do teste de Qui-quadrado, e dado que se verifica a existência de associação entre as variáveis foi determinado o coeficiente de contingência de Pearson.

Tabela 6. 33: Quadro de resultado do coeficiente de contingência de Pearson entre as variáveis "Situação atual de emprego" e género

Coeficiente de Contingência de Pearson
0,244

Atendendo ao resultado obtido verifica-se uma fraca associação entre as variáveis.

Em seguida serão apresentadas as frequências obtidas as respostas as questões classificadas anteriormente como sendo sobre fatores sociais.

Tabela 6. 34: Tabela de Frequências de questões sobre fatores sociais

Questões sobre fatores sociais		
Variável	Codificação	Frequências
Durante o surto de COVID-19 utilizou pelo menos 1 vez a internet para fazer compras, procurar informação, entre outros	1: Sim	402 (42,4%)
	5: Não	545 (57,6%)

Durante o surto de Covid-19 utilizou pelo menos 1 vez a internet para fazer compras, procurar informações, entre outros: por observação das frequências acima apresentadas verifica-se a prevalência de inquiridos que não recorrem à internet com o intuito de fazer compras ou obter informações, durante o surto de COVID-19. De modo a verificar a independência entre esta variável e o género do indivíduo inquirido foi construída a tabela de contingência apresentada em seguida.

Tabela 6. 35: Tabela de Contingência entre as variáveis "Durante o surto de COVID-19 utilizou pelo menos 1 vez a internet para fazer compras, procurar informação, entre outros" e gênero

		Gênero	
		Masculino	Feminino
Durante o surto de COVID-19 utilizou pelo menos 1 vez a internet para fazer compras, procurar informação, entre outros	Sim	189 (46,55%)	213 (39,37%)
	Não	217 (53,45%)	328 (60,63%)

De modo a verificar a independência entre esta variável e o gênero do indivíduo será aplicado um Teste de Qui-quadrado para a independência, sendo consideradas as seguintes hipóteses a testar:

H_0 : Utilização de internet durante COVID – 19 (...) e gênero são independentes

Vs.

H_1 : Utilização de internet durante COVID

– 19 (...) e gênero não são independentes

Com a aplicação do teste referido e atendendo ao resultado obtido para o *valor-p* (*valor-p* = 0,027 e $\chi^2_{observado} = 4.895$), existe evidência estatística para afirmar que a situação atual de emprego não é independente do seu gênero. Dado que se trata de uma tabela de contingência 2×2 era igualmente válida a aplicação do teste Exato de Fisher, cujo resultado seria similar.

Atendendo ao resultado obtido através do teste de Qui-quadrado, e dado que se verifica a existência de associação entre as variáveis será determinado o coeficiente de contingência de Pearson.

Tabela 6. 36: Coeficiente de contingência de Pearson entre as variáveis "Durante o surto de COVID-19 utilizou pelo menos 1 vez a internet para fazer compras, procurar informação, entre outros" e género

Coeficiente de Contingência de Pearson
0,072

Atendendo ao resultado obtido verifica-se uma fraca associação entre as variáveis.

6.4 Variável Dependente e as Variáveis Independentes

Dado que o objetivo deste estudo é ajustar um modelo que permita determinar as variáveis que melhor justificam o facto do individuo se sentir deprimido ou triste, considera-se que a variável dependente é no último mês esteve triste ou deprimido (doravante denotada sintomatologia de depressão), sendo que as restantes variáveis são consideradas como variáveis independentes.

Em seguida será realizado um estudo comparativo no estado de tristeza ou depressão do cidadão inquirido. Para tal será utilizado o teste de hipóteses mais adequado à tipologia da variável em causa.

Os resultados obtidos para as questões demográficas encontram-se sumarizados na tabela abaixo.

Tabela 6. 37:Caracterização da Variável dependente com as questões demográficas

Questões demográficas VS Depressão(S=1/N=0)		
Variável		Valor-p
Género	Teste de Qui-quadrado	< 0,001
Idade	Teste de Mann-Whitney	0,006 Ordem média (grupo 1): 501.70 Ordem média (grupo 0): 452.30
Idade que sente ter		< 0,001 Ordem média (grupo 1): 513.70 Ordem média (grupo 0): 442.90

Género: de modo a verificar se existe associação entre a sintomatologia de depressão e o género é aplicado o teste de Qui-quadrado. Considerando $valor - p < 0.05$, rejeita-se a hipótese nula de independência entre as variáveis. Posto isto, é possível concluir que o sentimento de depressão depende do género. Com base nas estatísticas amostrais apresentadas no ponto anterior assim como nas tabelas de contingência, é possível constatar que a depressão é superior nos indivíduos do sexo feminino.

Idade: de modo a comparar os dois grupos definidos pela variável sintomatologia de depressão (S/N) relativamente à mediana da idade (real) foi aplicado o teste de hipóteses não paramétrico de Mann-Whitney. Considerando $valor - p < 0.05$, rejeita-se a hipótese nula de igualdade entre os grupos. Posto isto, é possível concluir que a depressão entre os dois grupos é significativamente diferente. Com base e estatísticas amostrais os cidadãos que afirmam terem-se sentido tristes ou deprimidos apresentam, em média, uma idade mais elevada.

Idade que sente ter: de modo a comparar os dois grupos definidos pela variável sintomatologia de depressão (S/N) relativamente à mediana da idade que sente ter foi aplicado o teste de hipóteses não paramétrico de Mann-Whitney. Considerando $valor - p < 0.05$, rejeita-se a hipótese nula de igualdade entre os grupos. Posto isto, é possível

concluir que a depressão entre os dois grupos é significativamente diferente. Com base em estatísticas amostrais os cidadãos que afirmam terem-se sentido tristes ou deprimidos, sentem ter uma idade mais elevada.

Relativamente aos resultados obtidos para as questões de saúde física e comportamentos de saúde encontram-se sumarizados na tabela abaixo.

Tabela 6. 38: Caracterização da Variável dependente com as questões de saúde física e comportamentos de saúde

Questões de saúde física e comportamentos de saúde VS Sintomatologia de Depressão (S=1/N=0)		
Variável		Valor-p
Perceção do Estado de Saúde	Teste de Mann-Whitney	< 0,001 Ordem média (grupo 1): 565.45 Ordem média (grupo 0): 402.36
Se não está vacinado pretende ser vacinado contra COVID-19	Teste de Qui-quadrado	0,208
Tem fratura do quadril		0,008
Tem diabetes ou açúcar elevado no sangue		0,091
Tem tensão alta ou hipertensão		0,002
Teve um enfarte agudo do miocárdio ou tem outro problema de coração		0,005
Tem doença pulmonar obstrutiva crónica		0,266
Tem cancro ou um tumor maligno		0,035
Tem outra doença ou condição de saúde		< 0,001
Toma Medicação Regular		0,020
Saiu de casa nos últimos 3 meses		0,021
Tomou Medicação Preventiva para a COVID-19		0,425
Está vacinado contra a COVID-19		0,417

Perceção do estado de saúde: de modo a comparar as medianas das variáveis: sintomatologia de depressão e perceção do estado de saúde foi aplicado o teste de hipóteses não paramétrico de Mann-Whitney. Considerando *valor* – $p < 0,05$, rejeita-se a hipótese nula de igualdade entre os grupos. Atendendo ao resultado obtido serão realizadas as comparações múltiplas de modo a localizar as diferenças entre os grupos. Os resultados obtidos considerando a variável depressão como variável dependente encontram-se descritos na tabela seguinte:

Tem fratura do quadril: de modo a verificar se existe associação entre o sentimento de depressão e a fratura do quadril é aplicado o teste de Qui-quadrado. Considerando *valor* – $p < 0.05$, rejeita-se a hipótese nula de independência entre as variáveis. Posto isto, é possível concluir que o sentimento de depressão depende do facto do individuo apresentar ou não a fratura do quadril.

Tem diabetes ou açúcar elevado no sangue: de modo a verificar se existe associação entre o sentimento de depressão e a patologia de diabetes é aplicado o teste de Qui-quadrado. Considerando *valor* – $p > 0.05$, não se rejeita a hipótese nula de independência entre as variáveis. Conclui-se assim, que esta variável não é significativa para o modelo.

Tem tensão alta ou hipertensão: de modo a verificar se existe associação entre o sentimento de depressão e a patologia de hipertensão é aplicado o teste de Qui-quadrado. Considerando *valor* – $p < 0.05$, rejeita-se a hipótese nula de independência entre as variáveis. Posto isto, é possível concluir que existe uma associação entre o sentimento de depressão e a patologia.

Teve um enfarte agudo do miocárdio ou tem outro problema de coração: de modo a verificar se existe associação entre o sentimento de depressão e a patologia de enfarte agudo do miocárdio é aplicado o teste de Qui-quadrado. Considerando *valor* – $p < 0.05$, rejeita-se a hipótese nula de independência entre as variáveis. Posto isto, é possível concluir que existe uma associação entre o sentimento de depressão e esta patologia.

Tem doença pulmonar obstrutiva crónica: de modo a verificar se existe associação entre o sentimento de depressão e a doença pulmonar obstrutiva crónica é aplicado o teste de

Qui-quadrado. Considerando $valor - p > 0.05$, não se rejeita a hipótese nula de independência entre as variáveis. Conclui-se assim, que esta variável não é significativa para o modelo.

Tem cancro ou um tumor maligno: de modo a verificar se existe associação entre o sentimento de depressão e a patologia cancro é aplicado o teste de Qui-quadrado. Considerando $valor - p < 0.05$, rejeita-se a hipótese nula de independência entre as variáveis. Posto isto, é possível concluir que existe uma associação entre o sentimento de depressão e esta patologia.

Tem outra doença ou condição de saúde: de modo a verificar se existe associação entre o sentimento de depressão e a presença de outra doença ou condição de saúde é aplicado o teste de Qui-quadrado. Considerando $valor - p < 0.05$, rejeita-se a hipótese nula de independência entre as variáveis. Posto isto, é possível concluir que existe uma associação entre o sentimento de depressão e esta patologia.

Toma Medicação regular: de modo a verificar se existe associação entre o sentimento de depressão e a toma de medicação de forma regular é aplicado o teste de Qui-quadrado. Considerando $valor - p < 0.05$, rejeita-se a hipótese nula de independência entre as variáveis. Posto isto, é possível concluir que existe uma associação entre o sentimento de depressão e a toma regular de medicação.

Saiu de casa nos últimos 3 meses: de modo a verificar se existe associação entre o sentimento de depressão e o facto do indivíduo ter saído de casa nos últimos 3 meses é aplicado o teste de Qui-quadrado. Considerando $valor - p < 0.05$, rejeita-se a hipótese nula de independência entre as variáveis. Posto isto, é possível concluir que existe uma associação entre o sentimento de depressão e esta variável.

Tomou medicação preventiva para a COVID-19: de modo a verificar se existe associação entre o sentimento de depressão e toma de medicação preventiva para a COVID-19 é aplicado o teste de Qui-quadrado. Considerando $valor - p > 0.05$, não se rejeita a hipótese nula de independência entre as variáveis. Conclui-se assim, que esta variável não é significativa para o modelo.

Está vacinado contra a COVID-19: de modo a verificar se existe associação entre o sentimento de depressão e a variável “está vacinado contra a COVID-19” é aplicado o teste de Qui-quadrado. Considerando $valor - p > 0.05$, não se rejeita a hipótese nula de independência entre as variáveis. Conclui-se assim, que esta variável não é significativa para o modelo.

Se não está vacinado e pretende ser vacinado contra COVID-19: de modo a verificar se existe associação entre o sentimento de depressão e o facto do indivíduo que não se encontra vacinado contra a COVID-19 pretender ser vacinado é aplicado o teste de Qui-quadrado. Considerando $valor - p > 0.05$, não se rejeita a hipótese nula de independência entre as variáveis. Conclui-se assim, que esta variável não é significativa para o modelo.

Relativamente aos resultados obtidos para as questões de saúde mental encontram-se sumarizados na tabela abaixo.

Tabela 6. 39: Caraterização da Variável Dependente com questões de saúde mental

<u>Questões de saúde mental VS Sintomatologia de Depressão (S=1/N=0)</u>		
Variável		Valor-p
No último mês sentiu-se nervoso, ansioso ou no limite	Teste de Qui-quadrado	< 0,001
Teve dificuldades em dormir recentemente		< 0,001
Sente-se solitário		< 0,001

No último mês sentiu-se nervoso, ansioso ou no limite: de modo a verificar se existe associação entre o sentimento de depressão e o sentimento de nervosismo ou ansiedade no último mês é aplicado o teste de Qui-quadrado. Considerando $valor - p < 0.05$, rejeita-se a hipótese nula de independência entre as variáveis. Posto isto, é possível concluir que existe uma associação entre o sentimento de depressão e esta variável.

Teve dificuldades em dormir recentemente: de modo a verificar se existe associação entre o sentimento de depressão e a dificuldade para dormir recentemente é aplicado o teste de Qui-quadrado. Considerando *valor – p* < 0.05, rejeita-se a hipótese nula de independência entre as variáveis. Posto isto, é possível concluir que existe uma associação entre o sentimento de depressão e esta variável.

Sente-se solitário: de modo a verificar se existe associação entre o sentimento de depressão e o sentimento de solidão é aplicado o teste de Qui-quadrado. Considerando *valor – p* < 0.05, rejeita-se a hipótese nula de independência entre as variáveis. Posto isto, é possível concluir que existe uma associação entre o sentimento de depressão e esta variável.

Relativamente aos resultados obtidos para as questões sobre a COVID-19 encontram-se sumarizados na tabela abaixo.

Tabela 6. 40: Caracterização da variável dependente com questões sobre a COVID-19

<u>Questões sobre a COVID-19 VS Sintomatologia de Depressão (S=1/N=0)</u>		
Variável		<i>Valor-p</i>
Quantas vezes testou positivo para a COVID-19	Teste de Mann-Whitney	0,465

Quantas vezes testou positivo para a COVID-19: de modo a comparar as medianas das variáveis: sintomatologia de depressão e quantas vezes testou positivo para a COVID-19 foi aplicado o teste de hipóteses não paramétrico de Mann-Whitney. Considerando *valor – p* > 0.05, não se rejeita a hipótese nula de igualdade entre os grupos. Conclui-se assim, que esta variável não é significativa para o modelo.

Relativamente aos resultados obtidos para as questões sobre a situação profissional encontram-se sumarizados na tabela abaixo.

Tabela 6. 41: Caracterização da variável dependente com questões sobre a situação profissional

Questões sobre situação profissional VS Sintomatologia de Depressão (S=1/N=0)		
Variável		Valor-p
Situação atual de emprego	Teste de Qui-quadrado	0,06

Situação atual de emprego: de modo a verificar se existe associação entre o sentimento de depressão e situação atual de emprego é aplicado o teste de Qui-quadrado. Considerando $valor - p > 0.05$, não se rejeita a hipótese nula de independência entre as variáveis. Conclui-se assim, que esta variável não é significativa para o modelo.

Relativamente aos resultados obtidos para as questões sobre fatores sociais encontram-se sumarizados na tabela abaixo.

Tabela 6. 42: Caracterização da variável dependente com questões sobre fatores sociais

Questões sobre fatores sociais VS Sintomatologia de Depressão (S=1/N=0)		
Variável		Valor-p
Durante o surto de COVID-19 utilizou pelo menos 1 vez a internet para fazer compras, procurar informação, entre outros	Teste de Qui-quadrado	0,004

Durante o surto de COVID-19 utilizou pelo menos 1 vez a internet para fazer compras, procurar informação, entre outros: de modo a verificar se existe associação entre o sentimento de depressão e a utilização da internet para realizar compras ou procurar informação durante o surto de COVID-19 é aplicado o teste de Qui-quadrado. Considerando $valor - p < 0.05$, rejeita-se a hipótese nula de independência entre as

variáveis. Posto isto, é possível concluir que existe uma associação entre o sentimento de depressão e esta variável.

Com base na análise apresentada conclui-se que as variáveis independentes candidatas ao modelo são: género, idade, idade que sente ter, perceção do estado de saúde, tem fratura do quadril, tem tensão alta ou hipertensão, teve um enfarte agudo do miocárdio ou tem outro problema de coração, tem cancro ou um tumor maligno, tem outra doença ou condição de saúde, toma medicação regular, saiu de casa nos últimos 3 meses, no último mês sentiu-se nervoso ansioso ou no limite, teve dificuldades em dormir recentemente, sente-se solitário e durante o surto de COVID-19 utilizou pelo menos 1 vez a internet para fazer compras, procurar informação, entre outros.

6.5 – Modelo de Regressão Logística

Em seguida pretende-se ajustar o modelo que permita determinar se um indivíduo apresentará a sintomatologia depressiva. Para tal e com base nas conclusões anteriores, foram apenas consideradas as variáveis que se demonstraram ser candidatas ao modelo. Tendo-se o seguinte:

- I. **Variável dependente:** sintomatologia de depressão (sentiu-se triste ou deprimido) – variável dicotómica em que 1 representa presença de depressão (percecionada) e 0 representa a ausência de depressão (ou seja, responde não a sentir-se triste ou deprimido)

- II. **Variáveis independentes (candidatas ao modelo):** género, idade, idade que sente ter, perceção do estado de saúde, tem fratura do quadril, tem tensão alta ou hipertensão, teve um enfarte agudo do miocárdio ou tem outro problema de coração, tem cancro ou um tumor maligno, tem outra doença ou condição de saúde, toma medicação regular, saiu de casa nos últimos 3 meses, no último mês sentiu-se nervoso ansioso ou no limite, teve dificuldades em dormir recentemente, sente-se solitário e durante o surto de COVID-19 utilizou pelo menos 1 vez a internet para fazer compras, procurar informação, entre outros.

III. **Interações consideradas:** Tendo em consideração que se verificou a existência de algumas associações entre as variáveis, foi ainda explorados modelos de regressão logística com a componente de interação entre variáveis, nomeadamente: género *versus* idade; género *versus* percepção do estado de saúde e género *versus* sentiu-se nervoso ou no limite, contudo os resultados obtidos nos vários modelos considerando as interações referidas não revelaram ter uma boa capacidade preditiva, pelo que não foram considerados nem serão apresentados neste documento

Posto isto foi ajustado um modelo de regressão logística binária aos dados. Para este ajustamento foi utilizado o método de *Backward* com recurso ao teste de *Wald*, com vista a proceder à eliminação de variáveis que, devido *valor-p* do teste de Wald não se revelaram significativas para o modelo. Dado que um conjunto significativo das variáveis independentes são categóricas, foi considerada como categoria de referência a última categoria de cada uma das variáveis. No caso da variável dependente a categoria de referência considerada é a não presença de sintomatologia de depressão.

Com a utilização deste método de seleção de variáveis, obtiveram-se vários modelos, tendo sido apenas considerado o modelo que com base no teste de Hosmer e Lemeshow para a avaliação da qualidade do modelo e teste de Wald para a avaliação da significância das variáveis, apresentava apenas variáveis estatisticamente significativas e não violava a hipótese nula de que o modelo se ajusta bem aos dados.

Através da tabela abaixo é possível obter os valores estimados dos coeficientes que representam o efeito que cada variável tem na variável depressão (variável dependente), o desvio-padrão, os valores observados do teste de Wald e os respetivos *valores-p*.

Tabela 6. 43: Sumário do modelo ajustado

	Estimativa Coef.	Desvio Padrão	Teste de Wald	Valor-p
Constante	2,667	0,814	10,735	0,001
Idade	-0,23	0,011	4,796	0,029
Género: Masculino	0,708	0,172	16,968	< 0,001
Perceção do estado de Saúde: Excelente	0,781	0,587	1,773	0,183
Perceção do estado de Saúde: Muito Bom	2,093	0,521	14,146	< 0,001
Perceção do estado de Saúde: Bom	1,089	0,248	19,221	<0,001
Perceção do estado de Saúde: Fraco	0,429	0,221	3,749	0,053
Tem Fratura do Quadril: Sim	-0,965	0,385	6,288	0,012
No último mês sentiu-se nervoso ansioso ou no limite: Sim	-2,101	0,174	146,578	<0,001
Sente-se solitário: Muitas Vezes	-1,328	0,340	15,224	< 0,001
Sente-se solitário: Algumas Vezes	-1,274	0,200	40,703	< 0,001

Um dos pressupostos que deve ser verificado é a multicolinearidade entre as variáveis explicativas, visto que se pretende que estas não estejam correlacionadas entre si. Para tal, é necessário proceder ao cálculo do VIF que é uma medida que permite quantificar a existência de multicolinearidade. O VIF da variável k é calculado através da expressão:

$$VIF_k = \frac{1}{1 - R_k^2} \quad (48)$$

onde R_k^2 é o coeficiente de determinação obtido através da regressão da variável k nas restantes variáveis explicativas. É importante referir que este coeficiente apenas pode

assumir valores entre 0 e 1, sendo que quanto mais elevado for o seu valor melhor será o ajustamento do modelo aos dados.

Quando $VIF > 5$ considera-se que existe multicolinearidade elevada e valores de VIF próximos de 1 significa que o preditor não se encontra correlacionado com as demais variáveis. Os resultados obtidos encontram-se apresentados na tabela 6.44.

Tabela 6. 44: Multicolinearidade

Variáveis	VIF
Género	1,07
Idade	1,053
Perceção do estado de Saúde	1,131
Tem fratura do Quadril	1,010
No último mês sentiu-se nervoso, ansioso ou no limite	1,152
Sente-se solitário	1,111

Atendendo ao facto de que os valores obtidos para VIF são todos próximos de 1 tudo indica que não existe multicolinearidade entre as variáveis (deve ser acutelado o facto de estarmos a lidar com algumas variáveis dummy para representar as categorias de variáveis nominais).

Foi realizado o teste de Hosmer e Lemeshow de modo a avaliar de forma global o modelo, nomeadamente ao nível da sua eficiência em descrever a variável resposta. Foi obtido um $\text{valor-}p = 0,615$, pelo que não se rejeita a hipótese nula de que o modelo se ajusta bem aos dados. O modelo obtido é o seguinte:

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = 2.667 - 0.23Idade + 0.708Genero: Masculino$$

$$+ 0.781Perceção\ do\ Estado\ de\ Saúde: Excelente$$

$$+ 2.093Perceção\ do\ Estado\ de\ Saúde: Muito\ Bom$$

$$+ 1.089Perceção\ do\ Estado\ de\ Saúde: Bom$$

$$+ 0.429Perceção\ do\ Estado\ de\ Saúde: Fraco$$

$$- 0.965Tem\ Fratura\ do\ Quadril: Sim$$

$$- 2.101No\ último\ mês\ sentiu_{se} nervoso\ ansioso\ ou\ no\ limite: Sim$$

$$- 1.328Sente_{se} solitário: muitas\ vezes$$

$$- 1.274Sente_{se} solitário: algumas\ vezes$$

Posto isto, torna-se necessário efetuar a análise de resíduos. A análise de resíduos irá permitir avaliar a qualidade do ajustamento do modelo apresentado aos dados. De modo a realizar essa análise foi construído o gráfico (figura 6.1) no qual se encontram representados os resíduos normalizados. Este gráfico é bastante útil uma vez que permite identificar *outliers*. É necessário que os resíduos não se distribuam de forma padronizada em torno de $y = 0$ e que, pelo menos 95% destes resíduos sejam pertencentes ao intervalo $[-2,2]$. Por observação da figura 6.1, é possível verificar que os resíduos de encontram maioritariamente contidos no intervalo pretendido e que se distribuem em torno de zero. Esta representação gráfica pode ser também utilizada para a avaliação da qualidade de ajustamento do modelo. Posto isto, e atendendo à análise realizada constata-se que o gráfico não apresenta um padrão sistemático o que indica que o modelo poderá ser correto.

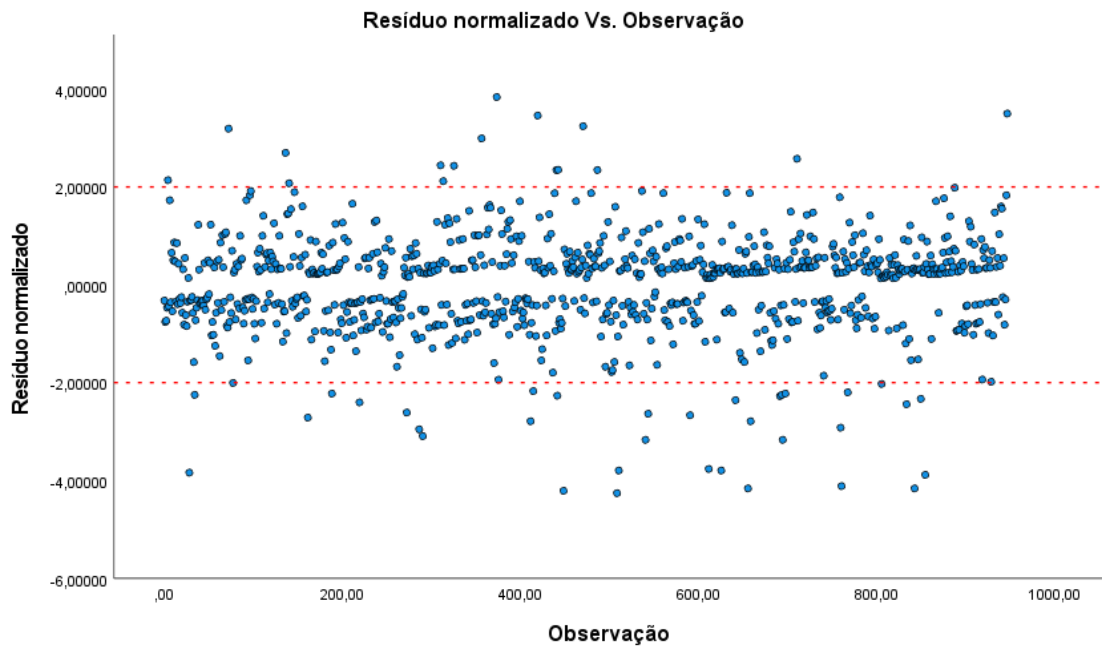


Figura 6. 2: Resíduos deviance do modelo

De modo a investigar a existência de observações influentes foi analisada a distância de Cook. O gráfico encontra-se abaixo, sendo possível verificar que todas as distâncias são inferiores a 1, pelo que considerando o critério apresentado no capítulo 4.5.3 é possível verificar que não existem observações influentes.

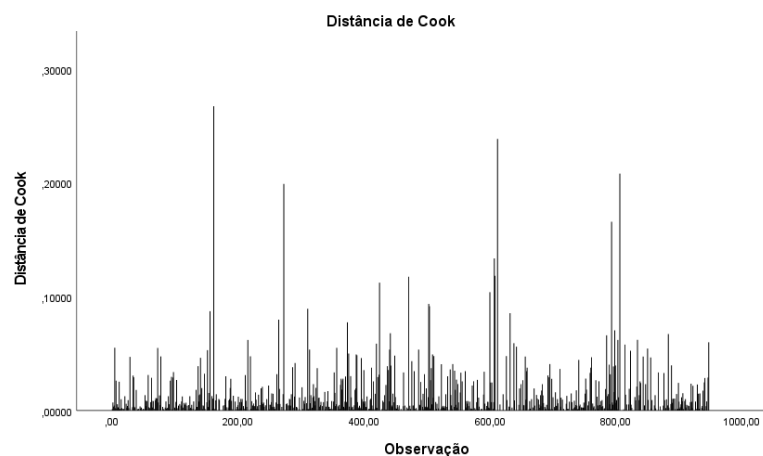


Figura 6. 3: Distância de Cook

Pela análise da curva ROC, na figura 6.4, pretende-se avaliar o ajustamento do modelo, nomeadamente na sua capacidade discriminatória. O modelo apresenta uma especificidade de 24% e uma sensibilidade de 80,4%. Posto isto, é possível através do modelo prever corretamente 24% dos valores referentes à sintomatologia de depressão. 80,4% das observações referentes à não presença de sintomatologia depressiva são bem classificados. O valor da AUC é de 78,2%, o que se traduz numa discriminação razoável.

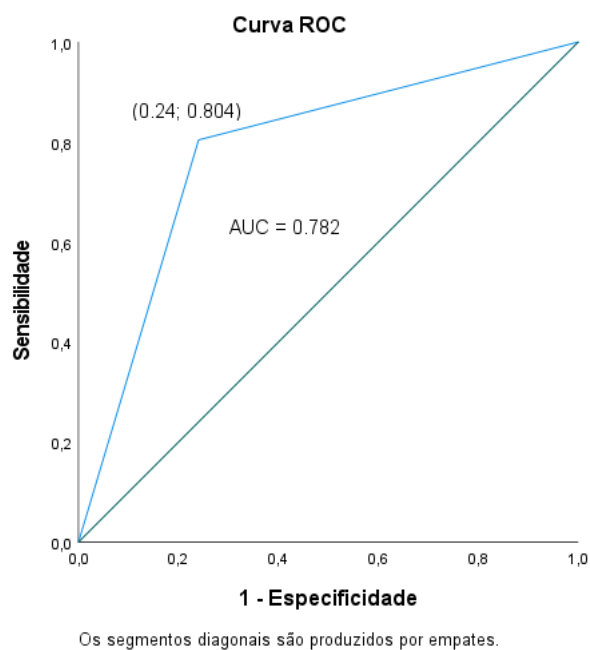


Figura 6. 4: Curva ROC

Tabela 6. 45: Matriz de Confusão do Modelo

		Valores Observados		
		Sim	Não	Total
Valores Previstos	Sim	316	104	420
	Não	100	427	527
	Total	416	531	947

A análise da matriz de confusão tem como objetivo testar o ajustamento do modelo de modo a verificar se o modelo classifica corretamente as observações, sendo possível calcular medidas que visam avaliar a qualidade do modelo. Obteve-se, assim através da tabela, os seguintes resultados para as medidas de avaliação da qualidade do modelo:

1. Sensibilidade: 59,96%
2. Especificidade: 80,41%
3. Precisão: 78,46%
4. Eficiência: 70,19%

Assim é possível concluir que o modelo ajustado contém apenas 59,96% de capacidade de prever corretamente as observações que são classificadas como presença de sintomas da patologia depressiva. Por oposição, o modelo tem a capacidade de prever corretamente 80,41% das classificações relativas à não presença de sintomas da doença. A proporção de previsões feitas corretamente é de 78,46%. A eficácia do modelo é de 70,19%, o que se pode traduzir numa capacidade de previsão razoável do modelo ajustado.

Considerando agora o estudo da Odds Ratio, neste caso em concreto, OR representa a razão entre a chance de apresentar Depressão (perceção/sintomas) e a chance de não apresentar sintomatologia da patologia de depressão, sendo que o acontecimento de interesse é não apresentar sintomatologia da patologia de depressão.

Tabela 6. 46: Estimativas dos coeficientes e Odds Ratio

	Estimativa	Odds Ratio	IC para Odds Ratio
Constante	2,667	14,403	
Idade	-0,23	0,977	(0,957; 0,998)
Género: Masculino	0,708	2,031	(1,45; 2,845)
Perceção do estado de Saúde: Excelente	0,781	2,184	(0,692; 6,895)
Perceção do estado de Saúde: Muito Bom	2,093	8,109	(2,922; 22,509)
Perceção do estado de Saúde: Bom	1,089	2,971	(1,826; 4,834)
Perceção do estado de Saúde: Fraco	0,429	1,535	(0,995; 2,369)
Tem Fratura do Quadril: Sim	-0,965	0,381	(0,179; 0,810)
No último mês sentiu-se nervoso ansioso ou no limite: Sim	-2,101	0,122	(0,087; 0,810)
Sente-se solitário: Muitas Vezes	-1,328	0,265	(0,136; 0,516)
Sente-se solitário: Algumas Vezes	-1,274	0,280	(0,189; 0,414)

Pela análise da tabela supra apresentada, é possível verificar que a chance de um individuo não apresentar a sintomatologia de depressão relativamente a outro que apresenta a sintomatologia depressiva, será devido a:

- Por cada aumento unitário na idade do individuo, a razão entre a chance do individuo não ter sintomas depressivos face a um individuo que tem diminui 23%.
- O risco de um individuo do género masculino não apresentar sintomatologia da patologia depressiva relativamente a um individuo que apresenta aumenta 70,8%.
- Como fator de proteção ao surgimento de sintomas da patologia surge a perceção do estado de saúde e o individuo ser do sexo masculino.

- Como fator de risco surge a fratura do quadril, ter-se sentido nervoso ansioso ou no limite no último mês ou o sentimento de solidão.

Relativamente aos intervalos de confiança obtidos para a Odds Ratio é de salientar que, para algumas das variáveis incluídas no modelo, nomeadamente, “Género”, “Perceção do estado de saúde” e “No último mês sentiu-se nervoso, ansioso ou no limite”, a amplitude do intervalo é grande o que se traduz numa fraca precisão do parâmetro estimado.

7. Considerações, conclusões e sugestões para trabalhos futuros

Com a investigação que se pretendia realizar nesta dissertação tinha como principal objeto estudar e fatores que permitam explicar a depressão, sobretudo em população idosa durante o período da COVID-19.

O facto de se tratarem de dados de variáveis qualitativas, dados de difícil comparação e de serem desconhecidos os parâmetros da distribuição esta investigação tornou-se desafiante para obter conclusões.

Ao explorar possíveis interações para auxiliar no ajustamento do modelo, encontraram-se associações significativas entre o género e outras variáveis explicativas, nomeadamente “Teve um enfarte agudo do miocárdio”, “Tem doença pulmonar crónica”, “Sentiu-se nervoso, ansioso ou no limite “, “No último mês esteve triste ou deprimido” (variável dependente do modelo), “Teve dificuldades em dormir recentemente”, “Sente-se solitário” e “Situação atual de trabalho”. Verificou-se, no entanto, que as interações exploradas não se revelaram significativas em nenhum caso em detrimento das variáveis individuais.

Para ajustar o modelo que permitisse responder ao principal objetivo deste estudo recorreu-se a modelos de regressão categorial, nomeadamente ao modelo de regressão logística. Os modelos de regressão categorial revelam ser uma ferramenta de especial importância na análise de variáveis qualitativas nominais. O facto de não fazer suposições sobre o comportamento probabilístico das variáveis independentes é uma vantagem. Nesta dissertação confirmou-se a utilidade deste tipo de modelos, para a modelagem de variáveis provenientes da área da saúde, nomeadamente psicológica.

Tendo em consideração o modelo determinado, as variáveis que melhor explicam a variável dependente, depressão, são: idade, género, perceção do estado de saúde, tem fratura do quadril e sente-se solitário. Esta última variável integra um índice global de Depressão abordado em outros estudos e vagas do SHARE, mas que esta base de dados relativa à Vaga 9 não permite calcular.

Nenhuma das variáveis relacionadas com a COVID-19 considerada neste estudo se revelou significativa para o modelo ajustado. Contudo, é notório que durante o período de

pandemia e como consequência dos consecutivos confinamentos e restrições impostas, os indivíduos, especialmente os mais idosos acabaram por ficar mais isolados, aumentando assim o sentimento de solidão. Tal facto é corroborado no estudo de Atzendorf et. al (2021) cujos resultados revelam que são especialmente aqueles que vivem sozinhos que tiveram um maior risco de aumentar a solidão após a primeira onda de COVID-19.

Segundo um estudo realizado por Haowei Wang (2022) sobre o luto durante a pandemia de COVID-19 e o seu impacto na saúde mental dos enlutados, foi possível concluir que existe uma forte associação entre o falecimento do cônjuge e problemas de saúde de mental, tanto antes como depois da pandemia. Com este estudo, também realizado com recurso a dados do programa SHARE, é possível concluir que os indivíduos cujos cônjuges faleceram de COVID-19 apresentam um risco mais elevado de depressão e solidão, contudo não se verifica a existência de problemas ou dificuldades em dormir. Embora o falecimento de uma familiar vítima de COVID-19 pareça relevante para o estudo e para a explicação do estado de saúde mental dos indivíduos, esta acabou por não se revelar significativa para o modelo ajustado, por não ser efetivamente ou por insuficiência de dados, pelo que não foi considerada.

No estudo conduzido por Wester (2022), foram analisadas as implicações da pandemia na saúde mental das populações idosas em países europeus, utilizando dados do SHARE. Os resultados demonstram que, embora o risco de depressão e distúrbios do sono tenha diminuído durante a pandemia, houve, pelo contrário, um aumento significativo do sentimento de solidão. Contudo, devido à ausência de dados prévios à COVID-19 em Portugal, não foi possível realizar uma análise comparativa desse período específico. No entanto, a importância desse estudo comparativo antes e depois da pandemia é relevante.

Uma limitação do modelo é a estimação de alguns coeficientes. Obtiveram-se alguns intervalos de confiança para os Odds Ratio que são pouco precisos, ou seja, maior incerteza na predição.

Um dado relevante seria fazer a comparação de dados obtidos numa fase inicial da pandemia com uma fase em que a doença já estava mais proliferada, contudo Portugal não participou no primeiro inquérito realizado.

Este estudo foi feito com recurso ao *software* IBM SPSS, mas realizá-lo em outro *software* como R poderia adicionar valor ao mesmo, pois permitiria comparar o modelo ajustado de ambos os *softwares*, explorar outros métodos de estimação não disponíveis no SPSS e daí extrair informação adicional que pudesse revelar-se pertinente. Esta análise já está em perspectiva para estudos que se prevê dar continuidade

Outros estudos que podem ser realizados futuramente é verificar se existe alguma relação entre o tipo de dieta alimentar do indivíduo e a depressão seria igualmente interessante.

A aplicação de outras metodologias estatísticas, tais como: outros modelos de regressão categorial ou técnicas de *machine learning* como as redes neuronais artificiais poderiam gerar resultados interessantes ao nível da capacidade preditiva possibilitando obter novas conclusões e desenvolvimentos relevantes para este estudo e para esta problemática.

8. Bibliografia

Abuladze, L., G. Opikova and K. Lang. (2020). Factors associated with incidence of depressiveness among the middle-aged and older Estonian population. *SAGE Open Medicine*: 1-12 DOI: 10.1177/2050312120974167.

Abu-Rayya, H.M. (2006). Depression and Social Involvement among Elders. *The Internet Journal of Health* 5(1).

Agência Lusa. (2022, 8 de abril). *Ansiedade, depressão e dependência do álcool maiores problemas decorrentes da pandemia*. Observador. <https://observador.pt/2022/04/08/ansiedade-depressao-e-dependencia-do-alcool-maiores-problemas-decorrentes-da-pandemia/>

Agência Lusa. (2022, 26 de setembro). *Homens vão viver menos 4,8 meses, mulheres menos 3,6. A pandemia mexeu com a esperança de vida à nascença*. CNN Portugal. <https://cnnportugal.iol.pt/esperanca-media-de-vida/portugal/homens-vaio-viver-menos-4-8-meses-mulheres-menos-3-6-a-pandemia-mexeu-com-a-esperanca-de-vida-a-nascenca/20220926/633181f80cf2f9a86eb77ff1>

Agresti, Alan (2002). *Categorical Data Analysis*. Wiley.

Atzendorf, J. and S. Gruber (2021). Depression and Loneliness of Older Adults in Europe and Israel after the First Wave of Covid-19. *European Journal of Ageing* DOI: 10.1007/s10433-021-00640-8.

Bashkin, O., R. Horne and I. Peytremann Bridevaux (2017). [Influence of Health Status on the Association Between Diabetes and Depression Among Adults in Europe: Findings From the SHARE International Survey](#). *Diabetes Spectrum* 30:3 DOI: 10.2337/ds16-0063.

Blazer DG, Hybels CF Origins of depression in later life. *Psychol Med*. 2005;35(9):1241–52.

Berkson, J. (1944), "Application of the Logistic Function to Bio-Assay," *Journal of the American Statistical Association*, 39, pp. 357-365.

Bliss, C.I. (1934). THE METHOD OF PROBITS. *Science*, 79 2037, 38-9 .

Bodner, E. and Y. S. Bergman. (2016). Loneliness and depressive symptoms among older adults: the moderating role of subjective life expectancy. *Psychiatry Research* 237: 78-82. DOI: 10.1016/j.psychres.2016.01.074.

Boeder, Jordan Desman. (2021). *The Development of Self-Perceptions of Aging: The Interplay Between Society and the Self Across the Lifespan*. CGU Theses & Dissertations, 200. https://scholarship.claremont.edu/cgu_etd/200. doi: 10.5642/cguetd/200

Bogdanova, B. and Z. Vladimirov (2021). [Cross-sectional differences in the Level of Depression since the COVID-19 Outbreak: A Study on SHARE Data](#). *Revista Inclusiones* 8: 22-43

Bogdanova, B. and B. Nedev (2021). [The Effect of Activities as Prevention Tool for Elderly People in Europe from Depression](#). *AIP Conference Proceedings* 2333(1) DOI: 10.1063/5.0042807.

Bollen, Kenneth A., Jackman, Robert W. (1985). Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases. *Sociological Methods Research* Vol 13, n.º 4, página 510–542. DOI:10.1177/0049124185013004004.

Börsch-Supan, A. (2022). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 9. COVID-19 Survey 2 Interview Date. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w9caintd.800

Brandão, M. P. and M. Fonseca-Cardoso (2020). [Associations Between Depressive Symptoms and Being Employed or Retired in Older Adults with Type 2 Diabetes](#). *Primary Care Diabetes* 14: 464-468 DOI: 10.1016/j.pcd.2020.01.006.

Brothers, A., & Diehl, M. (2017). Feasibility and efficacy of the AgingPlus Program: Changing views on aging to increase physical activity. *Journal of Aging and Physical Activity*, 25(3), 402-411. doi:10.1123/japa.2016-0039

Buber, I. and H. Engelhardt. (2011). The association between age and depressive symptoms among older men and women in Europe. Findings from SHARE. *Comparative Population Studies* 36(1): 103-125. DOI: 10.4232/10.CPoS-2011-02en.

Castelló, J. V. and C. Tubianosa (2020). Linking Mediterranean Diet and Lifestyle with Cardio Metabolic Disease and Depressive Symptoms: A Study on the Elderly in Europe.. *International Journal of Environmental Research and Public Health* 17(19): 7053. DOI: 10.3390/ijerph17197053.

Centro de Estudos de Comunicação e Sociedade. (s.d.). *Investigadora Alice Matos coordena a parte portuguesa do projeto SHARE-Survey of Health, Ageing and Retirement in Europe*. Centro de Estudos de Comunicação e Sociedade. <https://www.cecs.uminho.pt/investigadora-alice-matos-coordena-a-parte-portuguesa-do-projecto-share-survey-of-health-ageing-and-retirement-in-europe/>

Delgado, H. (2021, 24 de setembro). Esperança de vida à nascença em Portugal aumentou para 81,06 anos. DN. <https://www.dn.pt/sociedade/esperanca-de-vida-a-nascenca-em-portugal-aumentou-para-8106-anos--14155407.html>

Demétrio, C., & Cordeiro, G. (2011). Modelos Lineares Generalizados e Extensões. São Paulo: Departamento de Ciências Exatas, ESALQ, USP.

Feigl, P., & Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, 826-838. Doi.org/10.2307/2528247

Fiske, A., Wetherell, J. L., & Gatz, M. (2009). Depression in older adults. *Annual review of clinical psychology*, 5, 363–389. <https://doi.org/10.1146/annurev.clinpsy.032408.153621>

Fisher, RA (1922). Sobre os fundamentos matemáticos da estatística teórica. *Philosophical Transactions of the Royal Society of London. Série A*, contendo papéis de caráter matemático ou físico, 222 (594-604), 309-368. doi.org/10.1098/rsta.1922.0009

Flood M, Buckwalter KC.(2009). Recommendations for mental health care of older adults: Part 1—an overview of depression and anxiety. *J Gerontol Nurs.* 35(2):26–34.

Fox, J. (2009). *Applied Regression Analysis and Generalized Linear Models*. Sage Publication

Fundação Calouste Gulbenkian. (s.d.). *SHARE – Survey of Health, Ageing and Retirement in Europe*. Fundação Calouste Gulbenkian. <https://gulbenkian.pt/projects/share-survey-of-health-ageing-and-retirement-in-europe-2/>

Keyes, C., & Westerhof, G. (2012). Chronological and subjective age differences in flourishing mental health and major depressive episode. *Aging & Mental Health*, 16(1), 67-74. doi: 10.1080/13607863.2011.596811

Haowei Wang, PhD, Emily Smith-Greenaway, PhD, Shawn Bauldry, PhD, Rachel Margolis, PhD, Ashton M Verdery, PhD Mourning in a Pandemic: The Differential Impact of COVID-19 Widowhood on Mental Health, *The Journals of Gerontology: Series B*, Volume 77, Issue 12, December 2022, Pages 2306–2316, <https://doi.org/10.1093/geronb/gbac085>

Hosmer, D. W. e Lemeshow, S. (2000). *Applied Logistic Regression*, 2nd Edition. Wiley.

Instituto Nacional de Saúde Doutor Ricardo Jorge. (2020). *Saúde Mental em tempos de pandemia – SM – COVID-19*. Lisboa: INSA.

Joana Silva, Anna Bernard, Francisco Espiga e Madalena Gaspar, O impacto da Covid-19 na Pobreza e Desigualdade em Portugal, e o efeito mitigador das políticas de proteção. PROSPER (Católica-Lisbon's Center of Economics for Prosperity), <https://oobservatoriosocial.fundacaolacaixa.pt/-/o-impacto-da-covid-19-na-pobreza-e-desigualdade-em-portugal-e-o-efeito-mitigador-das-politicas-de-protecao>

Julius Sim , Chris C Wright, The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements, *Physical Therapy*, Volume 85, Issue 3, 1 March 2005, Pages 257–268, <https://doi.org/10.1093/ptj/85.3.257>

Klusmann, V., Sproesser, G., Wolff, J. K., & Renner, B. (2017). Positive Self-perceptions of Aging Promote Healthy Eating Behavior Across the Life Span via Social-Cognitive Processes. *The Journals of Gerontology: Series B*, 74(5), 735-744. doi:10.1093/geronb/gbx139

Leal, Margarida M. (1956). *Análise de Tabelas de Contingência*. Volume I. Tese de mestrado. Faculdade de Ciências da Universidade de Lisboa.

McCullagh, P. e Nelder, J. A. (1989). *Generalized Linear Models*, 2nd Edition. Chapman and Hall.

Mistry, S.K., Mehrab Ali, A.R.M., Hossain, M.B., Yadav, U.N., Ghimire, S., Rahman, M.A., Irfan, A.M., Huque, R. (2021). Exploring depressive symptoms and its associates among Bangladeshi older adults amid COVID-19 pandemic: findings from a cross-sectional study. *Social Psychiatry and Psychiatric Epidemiology*, 56:1487–1497
<https://doi.org/10.1007/s00127-021-02052-6>

Montgomery, Douglas C., Peck, Elizabeth A., Vining, G. Geoffrey (2012). *Introduction to Linear Regression Analysis*. (5a ed). Wiley. ISBN: 978-0-470-54281-1.

Murteira, B., Ribeiro, C. S., Silva, J. A., & Pimenta, C. (2001). Dados Estatísticos. Em B. Murteira, C. S. Ribeiro, J. A. Silva, & C. Pimenta, *Introdução à Estatística* (pp. 53-96). Amadora: McGraw-Hill.

Nelder, J. A., Wedderburn, W. M. (1972) *Generalized Linear Models*. *Wiley Journal of the Royal Statistical Society. Series A (General)*, Vol. 135 pp.370-38.

Noronha, N. d. (2022, 4 de outubro). *400 mil portugueses sofrem de depressão. Doença é muitas vezes escondida e em casos limite pode levar à morte*. SAPO Lifestyle.
<https://lifestyle.sapo.pt/saude/noticias-saude/artigos/400-mil-portugueses-sofrem-de-depressao-doenca-e-muitas-vezes-escondida-e-em-casos-limite-pode-levar-a-morte>

OCDE/Observatório Europeu dos Sistemas e Políticas de Saúde (2021), Portugal: Perfil de Saúde do País 2021, Estado da Saúde na UE, OCDE, Paris/Observatório Europeu dos Sistemas e Políticas de Saúde, Bruxelas.

Oliveira, Teresa (2004) *Estatística Aplicada*. Universidade Aberta. Lisboa.

Página oficial do Programa SHARE. Disponível em: <http://www.share-project.org.pt/> .

Peréz, F.L. (2023). Modelos de Regressão com resposta binária.
<http://leg.ufpr.br/~lucambio/ADC/ADC02.html#2.2.5>

Pestana, D. D., & Velosa, S. F. (2010). Análise Inicial dos Dados: Estatística Descritiva e Análise Exploratória. Em D. D. Pestana, & S. F. Velosa, *Introdução à Probabilidade e à Estatística* (pp. 81-190). Lisboa: Fundação Calouste Gulbenkian.

Pirâmides Populacionais do Mundo desde 1950 até 2100. (s.d.). PopulationPyramid.net.
<https://www.populationpyramid.net/pt/portugal/2021/>

PORDATA - Estatísticas, gráficos e indicadores. *Esperança de vida à nascença: total e por sexo (base: triénio a partir de 2001)*. (s.d.).
[https://www.pordata.pt/portugal/esperanca+de+vida+a+nascenca+total+e+por+sexo+\(base+trienio+a+partir+de+2001\)-418](https://www.pordata.pt/portugal/esperanca+de+vida+a+nascenca+total+e+por+sexo+(base+trienio+a+partir+de+2001)-418)

PORDATA - Estatísticas, gráficos e indicadores. *Como adoecem os portugueses?* (s.d.-a).
<https://www.pordata.pt/Publicacoes/Infografias/Como+adoecem+os+portugueses+-188>

Ramos, M. R. e Serranho, P. (s.d.). Bioestatística em SPSS – Notas de Apoio. Universidade Aberta, Universidade de Coimbra.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.

Razali, N. M., & Wah, Y. B. (2011). Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *Journal of Statistical Modeling and Analytics*, 2, 21-33.

Redação. (2021, 13 de setembro). *Portugal é o segundo país da UE com mais pessoas a sofrer de depressão crónica*. idealista/news. <https://www.idealista.pt/news/financas/economia/2021/09/10/48855-portugal-e-o-segundo-pais-da-ue-com-mais-pessoas-a-sofrer-de-depressao-cronica>

Redação. (2021b, 8 de outubro). *Saúde Mental em Portugal – é urgente falar sobre isso*. Revista Líder. <https://lidermagazine.sapo.pt/saude-mental-em-portugal-e-urgente-falar-sobre-isso/>

Reis, E., Melo, P., Andradade, R., Calapez, T. (2018) *Estatística Aplicada Volume 2* (6ª edição). Lisboa: Edições Sílabo

Rubin, D., & Berntsen, D. (2006). People over forty feel 20% younger than their age: Subjective age across the lifespan. *Psychon Bull Review*, 13(5), 776-780. doi: 10.3758/BF03193996

Santini, Z. I., Jose, P. E., York Cornwell, E., Koyanagi, A., Nielsen, L., Hinrichsen, C., Meilstrup, C., Madsen, K. R., & Koushede, V. (2020). Social disconnectedness, perceived isolation, and symptoms of depression and anxiety among older Americans (NSHAP): a longitudinal mediation analysis. *The Lancet. Public health*, 5(1), e62–e70. [https://doi.org/10.1016/S2468-2667\(19\)30230-0](https://doi.org/10.1016/S2468-2667(19)30230-0)

Seabra, D. (2022, 13 de janeiro). *Um em cada cinco portugueses sofreram sintomas de ansiedade e depressão durante a pandemia*. Notícias U.Porto. <https://noticias.up.pt/um-em-cada-cinco-portugueses-sofreram-de-ansiedade-e-depressao-durante-a-pandemia/>

Serviço Nacional de Saúde. (2021, 15 de janeiro). *Estudo Saúde Mental em Tempos de Pandemia (SM-COVID19): principais resultados*. Sistema Nacional de Saúde. <https://www.insa.min-saude.pt/estudo-saude-mental-em-tempos-de-pandemia-sm-covid19-principais-resultados/>

Serviço Nacional de Saúde. (s.d.-a). Covid-19: curva epidémica e parâmetros de transmissibilidade. Sistema Nacional de Saúde. <https://www.insa.min-saude.pt/category/areas-de-atuacao/epidemiologia/covid-19-curva-epidematica-e-parametros-de-transmissibilidade/>

Serviço Nacional de Saúde. (s.d.). *Novo coronavírus SARS-CoV-2 | Covid-19*. Consultado a 28/04/2022. Disponível em <https://www.insa.min-saude.pt/category/areas-de-atuacao/doencas-infeciosas/novo-coronavirus-sars-cov-2-covid-19/>

Sugiyama, K.(2019). *Generalized linear models*. Springer

Tiple P., Sharma S, Srivastava A. (2006). Psychiatric morbidity in geriatric people. *Indian J Psychiatry*. 48(2):88.

Turkman, M., & Silva, G. (2000). *Modelos Lineares Generalizados – da teoria à prática* -. Lisboa, SPE.

Yaka, E., Keskinoglu, P., Ucku, R., Yener, G. G., & Tunca, Z. (2014). Prevalence and risk factors of depression among community dwelling elderly. *Archives of Gerontology and Geriatrics*, 59(1), 150-154.

Ward, R. (2010). How old am I? Perceived age in middle and later life. *The International Journal of Aging and Human Development*, 71(3), 167-184. doi: 10.2190/AG.71.3.a

Wester, C. et al. (2022) "Longitudinal changes in mental health following the COVID-19 lockdown: Results from the Survey of Health, Ageing and Retirement in Europe", *Annals of Epidemiology*. doi: [10.1016/j.annepidem.2022.05.010](https://doi.org/10.1016/j.annepidem.2022.05.010). <https://www.sciencedirect.com/science/article/pii/S1047279722000898?via%3Dihub>

Zhao, Y., Shi, L., Jiang, Z. *et al*. The phenotype and prediction of long-term physical, mental and cognitive COVID-19 sequelae 20 months after recovery, a community-based cohort study in China. *Mol Psychiatry* (2023). <https://doi.org/10.1038/s41380-023-01951-1>

Anexo I

Inquérito Realizado

Os dados que foram alvo de estudo desta dissertação foram obtidos através de um inquérito realizado no âmbito do projeto SHARE. Em seguida é apresentado, na língua inglesa, o inquérito realizado. Devido a extensão do inquérito apenas são apresentadas as questões que foram consideradas para este estudo, o inquérito completo encontra-se disponível em: <http://www.share-project.org/home0.html>

CADN042_

IWER: Note sex of respondent (ask if unsure).

1. Male
2. Female

CADN003_

In which year were you born? [1900...2005]

CAS140_

Many people feel older or younger than they actually are. What age do you currently feel?

IWER: Write in age in years.

[0...1000] years old

CAPH003_

Would you say your health is excellent, very good, good, fair, or poor?

1. Excellent
2. Very good
3. Good
4. Fair

5. Poor

CAH004_

Do you have any of the following illnesses or health conditions? Please answer yes or no to each

category:

IWER: With this we mean that a doctor has told you that you have this condition, and that you are

either currently being treated for or bothered by this condition.

IWER: READ OUT.

CAH004_1 Hip fracture?

CAH004_2 Diabetes or high blood sugar?

CAH004_3 High blood pressure or hypertension?

CAH004_4 A heart attack including myocardial infarction or coronary thrombosis or any other heart problem including congestive heart failure?

CAH004_5 Chronic lung disease such as chronic bronchitis or emphysema?

CAH004_6 Cancer or malignant tumour, including leukaemia or lymphoma, but excluding minor skin cancers?

CAH006_

Do you regularly take prescription drugs?

1. Yes

5. No

CAH110_

During the last three months, did you ever leave your home?

1. Yes

5. No

CAH017_

Did you take any drugs or medicine as a prevention against Covid-19?

IWER: Drugs or medicine can include vitamins or common dietary supplements.

1. Yes

5. No

CAHC117_

Have you been vaccinated against Covid-19?

IWER: Code "yes" if R received at least one vaccination against Covid-19.

1. Yes

5. No

IF CAHC117_ = 5

CAHC118_

Do you want to get vaccinated against Covid-19?

IWER: Read out.

1. Yes, I already have a vaccination scheduled

2. Yes, I want to get vaccinated

3. No, I do not want to get vaccinated

4. I'm still undecided

ENDIF

CAH020_

In the last month, have you felt nervous, anxious, or on edge?

1. Yes

5. No

IF CAH020_ =1

CAMH002_

In the last month, have you been sad or depressed?

IWER: If participant asks for clarification, say 'by sad or depressed, we mean miserable, in low spirits,

or blue'.

1. Yes

5. No

CAMH007_

Have you had trouble sleeping recently?

IWER: DO NOT READ OUT.

1. Trouble with sleep or recent change in pattern

2. No trouble sleeping

CAMH037_

How much of the time do you feel lonely? Often, some of the time, or hardly ever or never?

1. Often
2. Some of the time
3. Hardly ever or never

CAEP005_CurrentJobSit

Which of the following best describes your current employment situation?

Code only one.

IWER: Only if R is in doubt then refer to the following: 1. Retired (retired from own work, including

semi-retired, partially retired, early retired, pre-retired). Retired refers to retired from own work only.

Recipients of survivor pensions who do not receive pensions from own work should not be coded as

retired. If they do not fit in categories 2 through 5, they should go into other.

IWER: Read out

1. Retired
2. Employed or self-employed (including working for family business)
3. Unemployed
4. Permanently sick or disabled
5. Homemaker

97. Other

CAIT104_

Since the outbreak of Corona, have you used the Internet, for e-mailing, searching for information,

making purchases, or for any other purpose at least once?

1. Yes

5. No