

UNIVERSIDADE ABERTA



**Análise estatística: Perfil e factores de risco da mortalidade causada por
doenças infecciosas em Maputo**

Fernando André Fernando Manjate

Mestrado em Estatística, Matemática e Computação, Ramo de Estatística
Computacional

2025

UNIVERSIDADE ABERTA



**Análise estatística: Perfil e factores de risco da mortalidade causada por
doenças infecciosas em Maputo**

Fernando André Fernando Manjate

Mestrado em Estatística, Matemática e Computação, Ramo de Estatística
Computacional

Dissertação orientada pela Professora Doutora Catarina S. Nunes

Fevereiro, 2025

Dedicatória

Dedico à minha mãe Marta Chiponze, ao meu pai André Manjate, as minhas irmãs Edna Manjate e Janice Januário. À eles não só dedico a dissertação, dedico a minha vida.

Dedico ainda à todos meus familiares e aos amigos que se tornaram familiares, por acreditam em mim e sempre me apoiarem.

Agradecimentos

À Deus, por estar sempre comigo e por mim.

À Professora Doutora Catarina S. Nunes, pela paciência, compreensão, simplicidade e a competência demonstrada na orientação desde a fase plano até a efectivação da dissertação.

Ao Dr. Emerson Guilande e à Direcção Científica do Hospital Central de Maputo pela disponibilização de dados.

A todos docentes da parte Curricular do Mestrado por terem me permitido ganhar algum aprendizado. Sem dúvida, foram imprescindíveis para a minha formação.

Aos meus docentes de Licenciatura em especial à Professora Doutora Dulcília Guezemane, ao Professor Doutor Natércio Mucavel (meu orientador da monografia de Licenciatura), ao Msc. Manuel Castigo e ao Professor Doutor Gimo Maundera por sempre me incentivarem a continuar com os estudos.

Aos meus familiares em especial mãe, pai, irmãos, prima Neyma por sempre me apoiarem nesta formação.

Aos meus colegas do Mestrado, em especial ao José Nhantigo, Armindo Saia, Wilson Paulo e Paulo Assunção que apesar de não os conhecer pessoalmente, virtualmente permitiram-me ganhar algum aprendizado através dos nossos debates.

Aos meus amigos em especial Milton Cumbi, Samuel Langa, Dário Joalinho e Felisberto Cumbane pelo apoio.

Aos meus colegas de trabalho, em especial à *Seal's team* (Assucenio Macamo, Helder Martins, Ermeltigio Matchane, Obet Muchave e Tesoura João) e à *team Faculdade* (Juliana Sitole e Ângela Cuamba), meus amigos que, em meio a rotina de estudos e do trabalho, garantiram bons momentos de confraternização e foram importantes para desanuviar.

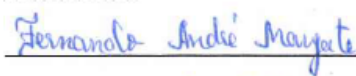
Por fim, agradeço a todos que directa e indirectamente contribuíram para a formação no nível de Mestrado e em particular para a realização da presente dissertação.

DECLARAÇÃO DE INTEGRIDADE

Declaro ter actuado com integridade na elaboração da presente dissertação. Confirmando que em todo trabalho conducente à sua elaboração não recorri à prática de plágio ou a qualquer outra forma de falsificação de resultados.

Mais declaro que tomei conhecimento integral do Regulamento Disciplinar da Universidade Aberta, publicado no Diário da República, 2ª série, n.º 215, de 6 de Novembro de 2013.

Universidade Aberta, 11 de Fevereiro de 2025

A handwritten signature in blue ink, reading "Fernando André Fernando Manjate", is written over a horizontal line.

Fernando André Fernando Manjate

Resumo

As doenças infecciosas com destaque para malária, tuberculose e HIV/SIDA constituem um grande problema de saúde pública para Moçambique, pelos custos empreendidos e pelo nível de mortalidade. Neste sentido, com base nestas doenças e nos dados de 2020 a 2022, a presente dissertação tem como objectivo central compreender o perfil e os factores de risco da mortalidade causada por doenças infecciosas no Hospital Central de Maputo. Para tal, a metodologia baseou-se numa análise estatística recorrendo essencialmente a técnicas da Estatística Descritiva e a Regressão Logística e tendo por base uma amostra de 687 doentes dos quais 364 tiveram alta e 323 foram óbitos. O perfil da amostra é marcado por uma taxa de mortalidade de 24% sendo a malária a maior causa (40%), género feminino (51,7%), idades entre 25 e 49 anos (52,9%), casados (44,7%), consumidores de álcool (68%, com 15 ou mais anos de idade), não fumadores (70%, com 15 ou mais anos de idade), ser proveniente do Grande Maputo (43,1%), não sofrer de nenhuma doença crónica (63,9%) e os que nunca tiveram um internamento anteriormente (59,1%). Quanto aos factores de risco comparou-se os critérios *Enter* (em duas fases), *Forward LR* e *Backward LR*, todos apresentam boas capacidades preditivas e de discriminação, confiabilidade substancial e ajustamentos adequados; os últimos dois apresentam resultados iguais e destacam-se como os de melhor desempenho pela simplicidade e menor razão de verosimilhança, tendo-se identificado os seguintes factores: ser casado, ser proveniente do centro e norte, ser fumante e sofrer duma doença crónica associada.

Palavras-chave: doenças infecciosas, mortalidade, regressão logística, perfil da mortalidade, factores de risco.

Abstract

Infectious diseases, especially malaria, tuberculosis and HIV/AIDS, are a major public health problem for Mozambique, due to the costs incurred and the level of mortality. In this sense, based on these diseases and data from 2020 to 2022, the main objective of this dissertation is to understand the profile and risk factors of mortality caused by infectious diseases at the Maputo Central Hospital. To this end, the methodology was based on a statistical analysis using essentially Descriptive Statistics and Logistic Regression techniques, and on a sample of 687 patients, of which 364 were discharged and 323 were deaths. The profile of the sample is marked by a mortality rate of 24% with malaria being the main cause (40%), female gender (51.7%), aged between 25 and 49 years (52.9%), married (44.7%), alcohol consumers (68%, aged 15 or over), non-smokers (70%, aged 15 or over), being from Greater Maputo (43.1%), not suffering from any chronic disease (63.9%) and those who have never been hospitalized before (59.1%). As for the risk factors, the *Enter* criteria (in two phases), *Forward LR* and *Backward LR* were compared, all of which have good predictive and discrimination capabilities, substantial reliability and adequate adjustments, and the last two present equal results and stand out as the best performers due to their simplicity and lower likelihood ratio. The following factors were identified: being married, coming from the center and north, being a smoker and suffering from an associated chronic disease.

Keywords: infectious diseases, mortality, logistic regression, mortality profile, risk factors.

Índice

1. Introdução.....	2
1.1. Objectivos.....	3
1.2. Estrutura da dissertação.....	4
2. Doenças infecciosas, mortalidade e estudos relacionados.....	6
2.1. Conceitos ligados a doenças infecciosas e a mortalidade	6
2.2. Estudos relacionados	8
3. Caracterização teórica dos métodos estatísticos.....	12
3.1. Algumas técnicas da Estatística Descritiva.....	12
3.1.1. Frequências	12
3.1.2. Gráficos.....	12
3.1.3. Medidas de tendência central.....	13
3.1.4. Medidas de dispersão	14
3.2. Modelos de regressão.....	15
3.3. Modelos Lineares Generalizados	19
3.3.1. Componentes dum modelo linear generalizado.....	20
3.4. Regressão logística.....	23
3.4.1. Regressão logística simples ou univariada	24
3.4.2. Transformação logit.....	25
3.4.3. Razão de chances (<i>OddsRatio</i>)	26
3.4.4. Regressão logística múltipla	27
3.4.5. Pressupostos da regressão logística	27
3.4.6. Ajustamento do modelo: O método de máxima verosimilhança.....	27
3.4.6.1. Matriz de informação	29
3.4.6.2. Intervalos de confiança.....	30
3.4.7. Teste de rácio de verosimilhança.....	31
3.4.8. Teste de ajustamento do modelo.....	32
3.4.9. Teste à significância dos coeficientes do modelo.....	33
3.4.10. Pseudo- <i>R</i> ²	34
3.4.11. Selecção de variáveis independentes	35
3.4.12. Análise dos resíduos	36
3.4.13. Classificação da previsão do modelo.....	38
3.4.13.1. Sensibilidade e especificidade do modelo.....	38
3.4.13.2. Curva de ROC.....	39
3.4.14. Kappa de confiabilidade de Cohen	39
4. Concepções metodológicas.....	43
4.1. Tipo de pesquisa.....	43
4.2. Procedimentos	44
4.3. Amostragem	45
5. Aplicação da análise estatística á base de dados	48

5.1.	Análise descritiva dos dados	48
5.1.1.	Análise univariada	48
5.1.2.	Análise bivariada	51
5.2.	Perfil da mortalidade causada por doenças infecciosas	56
5.3.	Factores de risco da mortalidade causada por doenças infecciosas: Aplicação da regressão logística	65
5.3.1.	Análise da multicolinearidade.....	65
5.3.2.	Modelo sem variáveis independentes	67
5.3.3.	Critério de selecção Enter	68
5.3.3.1.	Critério Enter reduzido.....	68
5.3.3.2.	Critério de selecção Enter Completo.....	76
5.3.4.	Critério de selecção <i>Forward LR</i>	85
5.3.5.	Critério <i>Backward LR</i>	94
5.4.	Discussão dos resultados	94
6.	Conclusões e considerações finais.....	100
	Referências bibliográficas.....	104

Índice de tabelas

Tabela 3.1: Distribuições da família exponencial.....	22
Tabela 3.2: Funções de ligação.....	22
Tabela 3.3: Nível de concordância do Kappa de Cohen.....	41
Tabela 4.1: Dados gerais sobre Altas no HCM no triênio 2020-2023.....	46
Tabela 4.2: Dados gerais sobre Óbitos no HCM no triênio 2020-2023.....	46
Tabela 5.1: Resumo da análise bivariada.....	55
Tabela 5.2: Taxa de mortalidade.....	57
Tabela 5.3: Cruzamento entre gênero e grupo etário dos óbitos.....	60
Tabela 5.4: Cruzamento entre gênero e estado civil dos óbitos.....	61
Tabela 5.5: Cruzamento entre grupo etário e estado civil dos óbitos.....	61
Tabela 5.6: Análise de multicolinearidade.....	65
Tabela 5.7: Resumo dos casos processados.....	65
Tabela 5.8: Codificação das variáveis independentes.....	66
Tabela 5.9: Classificação no modelo sem variáveis independentes.....	67
Tabela 5.10: Teste de rácio de verossimilhança utilizando o critério <i>Enter</i> reduzido (Omnibus).....	68
Tabela 5.11: Pseudos- R^2 do Critério <i>Enter</i> reduzido.....	69
Tabela 5.12 Teste de Hosmer e Lemeshow utilizando o critério <i>Enter</i> reduzido.....	69
Tabela 5.13: Classificação do modelo <i>Enter</i> reduzido.....	69
Tabela 5.14: Área da Curva ROC.....	70
Tabela 5.15: Confiabilidade utilizando o critério <i>Enter</i> Reduzido.....	71
Tabela 5.16: Significância das variáveis independentes pelo Critério <i>Enter</i> Reduzido.....	71
Tabela 5.17: Teste de rácio de verossimilhança pelo Critério <i>Enter</i> completo.....	76
Tabela 5.18: Pseudos- R^2 do Critério <i>Enter</i> Completo.....	77
Tabela 5.19: Teste de Hosmer e Lemeshow pelo Critério <i>Enter</i> Completo.....	77
Tabela 5.20: Classificação pelo Critério <i>Enter</i> Completo.....	77
Tabela 5.21: Área da Curva ROC do Critério <i>Enter</i> Completo.....	78
Tabela 5.22: Confiabilidade do Critério <i>Enter</i> Completo	79
Tabela 5.23: Significância das variáveis independentes pelo Critério <i>Enter</i> Completo.....	80
Tabela 5.24: Teste de rácio de verossimilhança (Omnibus) do Critério <i>Forward</i> LR.....	85
Tabela 5.25: Pseudos- R^2 do Critério <i>Forward</i> LR.....	86

Tabela 5.26: Teste de Hosmer e Lemshow do Critério <i>Forward LR</i>	86
Tabela 5.27: Classificação do Critério <i>Forward LR</i>	87
Tabela 5.28: Área da Curva ROC do Critério <i>Forward LR</i>	88
Tabela 5.29: Confiabilidade do critério <i>Forward LR</i>	89
Tabela 5.30: Significância das variáveis independentes pelo Critério <i>Forward LR</i>	89
Tabela 5.31: Resumo dos elementos de análise dos critérios da regressão logística	96
Tabela 5.32: Área da curva ROC da validação externa.....	97
Tabela 5.33: Previsão de modelo de eleição na amostra externa.....	98

Índice de gráficos

Figura 3.1: Esquema de um modelo de regressão.....	17
Fig.5.1: Mortalidade ou sobrevivência.....	48
Fig. 5.2: Mortalidade por doença	57
Fig. 5.3: Gráfico da Mortalidade por ano.....	58
Fig. 5.4: Gráfico da Mortalidade por doenças infecciosas no triênio 2020-2022.....	58
Fig. 5.5: Gráfico do género dos óbitos.....	59
Fig. 5.6: Gráfico do grupo etário dos óbitos.....	59
Fig. 5.7: Gráfico do estado civil dos óbitos.....	60
Fig.5.8: Gráfico dos óbitos consumidores ou não de álcool com 15 ou mais anos de idade.....	62
Fig.5.9: Gráfico dos óbitos fumantes ou não de álcool com 15 ou mais anos de idade.....	63
Fig.5.10: Proveniência dos óbitos.....	63
Fig.5.11: Gráfico da existência de doenças crónicas associadas.....	64
Fig.5.12: Histograma e boxplot de número de internamentos anteriores.....	64
Fig.5.13: Curva de ROC do Critério <i>Enter</i> Completo.....	70
Fig.5.14: Resíduos estudentizados e estandardizados pelo Critério <i>Enter</i> reduzido.....	75
Fig.5.15: Distância de Cook pelo Critério <i>Enter</i> reduzido.....	76
Fig.5.16: Curva ROC do Critério <i>Enter</i> Completo.....	78
Fig.5.17: Resíduos estudentizados e estandardizados.....	84
Fig.5.18: Diferença de Cook do critério <i>Enter</i> Completo.....	84
Fig.5.19: Curva ROC do Critério <i>Forward LR</i>	88
Fig.5.20: Resíduos estudentizados e estandardizados.....	92
Fig.5.21: Distância de Cook do critério <i>Forward LR</i>	93
Fig.5.22: Curva ROC da validação externa	97

Lista de abreviaturas

HCM- Hospital Central do Maputo

INS- Instituto Nacional de Saúde.

MCDI- Mortalidade causada por doenças infecciosas

MLG- Modelos Lineares Generalizados

ODS- Objectivos do Desenvolvimento Sustentável

OMS- Organização Mundial da Saúde

ONU- Organização das Nações Unidas

ROC- Característica de Operação do Receptor

TM- Taxa de mortalidade

VIF- Factor de inflação da variância

CAPÍTULO I: INTRODUÇÃO

1. Introdução

As doenças infecciosas são aquelas que surgem como resultado duma penetração e desenvolvimento ou multiplicação de um agente infeccioso (como vírus, bactérias, protozoários ou fungos) no organismo (Santos, 2023).

De acordo com a Agência de Informação de Moçambique (AIM, 2023), as doenças infecciosas continuam a ser um grande problema de saúde pública no continente africano.

Neste grupo de doenças, destacam-se a malária, a tuberculose e o HIV/SIDA que colectivamente representam 80% das doenças infecciosas com peso em África, causando, anualmente, a morte de mais de seis milhões de pessoas e tendo um custo económico superior a 800 mil milhões de dólares norte-americanos (o maior bolo que as nações investem em saúde pública) (AIM, 2023).

Em Moçambique, com base nos dados de 2012 a 2021, morrem aproximadamente 90 em cada 1000 pessoas por ano (Dados Mundiais, 2022) e apesar de Moçambique estar actualmente em fase de transição epidemiológica, em que há um aumento do peso das mortes causadas por doenças crónicas não transmissíveis, as doenças infecciosas continuam a ser uma importante causa de morte no país sobretudo em indivíduos na faixa etária compreendida entre os 15 e 49 anos de idade (Instituto Nacional de Saúde Moçambique [INS], 2023; Camacho *et al.*, 2022).

Verifica-se através do estudo de Camacho *et al.* (2022), feito com base nos dados de 2018 a 2021, que em Moçambique a malária, a tuberculose e o HIV/SIDA encontram-se entre as principais causas de morte em ambos os géneros e em quase todas as faixas etárias (com excepção da faixa dos 0-4 anos em que tuberculose não faz parte das principais causas e os maiores de 65 anos em que a malária não faz parte das principais causas).

Na mesma senda, a Organização das Nações Unidas (ONU, 2016) pontua a necessidade de se dirigir atenções para o grupo de doenças em referência, tendo entre os 17 Objectivos de Desenvolvimento Sustentável (ODS) o objectivo de saúde de qualidade no qual refere que “Até 2030, acabar com as epidemias de SIDA, tuberculose e doenças tropicais negligenciadas” (ONU, 2016, p.7).

Evidentemente, as doenças infecciosas com enfoque para a malária, a tuberculose e o HIV/SIDA representam um grande problema de saúde pública para Moçambique em particular e para o

continente africano em geral pelos custos e pelo nível mortalidade que estas causam, sendo assim importante realizar pesquisas em volta deste grupo de doenças para melhor entendimento e formulação de estratégias de modo a minimizar os seus impactos.

Várias pesquisas (algumas apresentadas no próximo capítulo) têm recorrido a métodos estatísticos para compreender aspectos epidemiológicos inclusive os relacionados com a mortalidade.

O presente estudo centra-se em compreender o perfil e os factores de risco da mortalidade causada por doenças infecciosas (MCDI) com enfoque para a malária, tuberculose e HIV/SIDA dos doentes do Hospital Central do Maputo (HCM) (a maior unidade sanitária de Moçambique) do ano 2020 a 2022 tendo em conta as variáveis: género, idade, estado civil, nível de distância da residência do internado ao hospital, existência de doença crónica associada, consumo de álcool, ser ou não fumante, número de internamentos anteriores.

O perfil em alusão poderá ser determinado através de recursos da estatística descritiva e os factores de risco da mortalidade poderão ser identificados através de uma técnica da análise estatística multivariada: a regressão logística.

Deste modo, acredita-se que conhecendo o perfil e os factores de risco da mortalidade causadas por doenças infecciosas poderá identificar-se as características dos grupos prioritários de doentes, o que criará melhoramento das políticas de trabalho e provavelmente reduzirá o nível de mortalidade em causa.

1.1.Objectivos

1.1.1. Objectivo geral:

- Compreender o perfil e os factores de risco da mortalidade causada por doenças infecciosas dos doentes no HCM.

1.1.2. Objectivos específicos:

- Apresentar uma análise descritiva dos dados recolhidos;
- Descrever o perfil da mortalidade causada por doenças infecciosas dos doentes no HCM através de técnicas da estatística descritiva (frequências, gráficos, tabelas, etc.);
- Determinar a taxa de mortalidade intra-hospitalar do HCM;

- Identificar os factores de risco da mortalidade causada por doenças infecciosas dos doentes no HCM aplicando a regressão logística.

1.2.Estrutura da dissertação

A presente dissertação encontra-se dividida em seis capítulos nomeadamente a introdução (capítulo I), as doenças infecciosas, mortalidade e estudos relacionados (capítulo II), a caracterização teórica dos métodos estatísticos (capítulo III), as concepções metodológicas (capítulo IV), a aplicação da análise estatística à base de dados (capítulo V) e as conclusões (capítulo VI)

O capítulo I, que é o presente capítulo, apresenta a dissertação trazendo a contextualização de aspectos gerais, os objectivos bem como estrutura do trabalho.

O capítulo II faz um enquadramento teórico de aspectos em volta das doenças infecciosas e mortalidade bem como apresenta estudos anteriores relacionados com o tema da presente dissertação.

O capítulo III apresenta os fundamentos teóricos que norteiam os métodos estatísticos usados na presente dissertação.

O capítulo IV indica os métodos usados para alcançar os objectivos do trabalho.

O capítulo V descreve os dados, mostra a aplicação dos métodos estatísticos usados e apresenta os resultados procurando responder a cada objectivo específico.

O capítulo VI, em forma de considerações finais e com base numa síntese do capítulo V, busca trazer as constatações obtidas no trabalho.

CAPÍTULO II:
DOENÇAS INFECCIOSAS, MORTALIDADE E
ESTUDOS RELACIONADOS

2. Doenças infecciosas, mortalidade e estudos relacionados

Visto que a presente dissertação trata sobre a mortalidade de doenças infecciosas, é importante realizar uma revisão bibliográfica em relação a estes elementos. Este capítulo aborda conceitos ligados às doenças infecciosas com um enfoque para a malária, a tuberculose e o HIV/SIDA; conceitos ligados à mortalidade bem como apresenta estudos anteriores (baseados na análise estatística).

2.1. Conceitos ligados a doenças infecciosas e a mortalidade

➤ Doença

Para a OMS, a doença é “conjunto de sinais e sintomas específicos que afectam um ser vivo alterando o seu estado normal de saúde” (Silva, 2017:2).

➤ Infecção

A infecção é “invasão de tecidos corporais de um organismo hospedeiro por parte de organismos capazes de causar doenças, a multiplicação destes organismos e a reacção destes tecidos no hospedeiro “(Silva, 2017:2).

➤ Malária

A malária é uma doença febril aguda transmitida ao homem pelo sangue, geralmente através da picada da fêmea do mosquito *Anopheles*, infectada pelo microrganismo *plasmodium* (Biblioteca Virtual da Saúde [BVS], 2009).

Os sintomas mais comuns desta doença são: calafrios, febres altas, dores de cabeça e musculares, aumento de batimentos cardíacos, aumento de baços e por vezes, delírios (BVS, 2009).

➤ HIV/SIDA

HIV é a sigla em inglês para o Vírus da Imunodeficiência Humana e SIDA significa Síndrome da Imunodeficiência Adquirida (BVS, 2016). SIDA é uma doença transmitida pelo HIV, caracterizada pelo enfraquecimento do sistema de defesa do corpo e pelo aparecimento de doenças oportunistas.

O HIV é transmitido por meio de relações sexuais desprotegidas com pessoa seropositiva, pelo sangue ou de mãe seropositiva (sem tratamento) para o filho durante a gestação, parto ou amamentação (BVS, 2016).

Os primeiros sintomas são muito parecidos com uma gripe como febre e mal-estar e por isso na maioria dos casos passa despercebida (BVS, 2016).

Ainda não há cura para esta doença, contudo há vários avanços científicos que possibilitam qualidade de vida à pessoa que tenha o vírus (BVS, 2016).

➤ **Tuberculose**

A tuberculose é uma doença infecciosa e transmissível que ataca principalmente os pulmões e é causada pelo *Mycobacterium tuberculosis* ou bacilo de Koch (BVS, 2020).

A transmissão desta doença é directa de pessoa para pessoa através da saliva sendo a aglomeração de pessoas o factor principal de transmissão (BVS, 2020). Esta pode decorrer através da fala, espirro ou tosse em que o agente infeccioso passa para o outro contaminando-o (BVS, 2020).

➤ **Morte**

A morte é “desaparecimento permanente de todas as evidências de vida a qualquer momento após o nascimento vivo (cessação pós-natal das funções vitais sem capacidade de ressuscitação)” (Camacho *et al.*, 2018:5).

➤ **Causa da morte**

As causas da morte são "todas aquelas doenças, condições mórbidas ou lesões que resultaram ou contribuíram para a morte ou as circunstâncias do acidente ou da violência que produziram tais lesões" (Camacho *et al.*, 2018:5).

➤ **Causa básica de morte**

A causa da morte a ser usada para fins de tabulação estatística primária foi designada como a causa básica da morte. A causa básica de morte é definida como:

“(a) a doença ou lesão que iniciou o conjunto de eventos que levaram directamente à morte, ou (b) as circunstâncias do acidente ou da violência que produziu a lesão fatal”(Camacho *et al.*, 2018:5)

➤ **Taxa de mortalidade bruta**

A taxa de mortalidade bruta é “o número de mortes em relação ao tamanho dessa população durante um determinado período, geralmente um ano. É expressa como o número de mortes por 1.000 habitantes por ano” (Camacho *et al.*, 2018:5).

➤ **Taxa mortalidade intra-hospitalar (TMI)**

A taxa de mortalidade intra-hospitalar:

É o número de mortes ocorridas dentro da Unidade Sanitária num determinado período específico em relação ao número total de altas na mesma unidade sanitária e período. Este é geralmente expressa em percentagem. (Camacho *et al.*, 2018).

2.2. Estudos relacionados

Como já foi referenciado anteriormente, o presente estudo versa essencialmente sobre o perfil e os factores de risco da mortalidade causada por doenças infecciosas e relacionados com esta temática existem variados estudos.

Os perfis epidemiológicos em geral e os ligados à mortalidade em particular constituem uma preocupação no âmbito científico e são geralmente determinados com base em técnicas da Estatística Descritiva. Neste âmbito, a seguir apresentam-se estudos em que são usados métodos estatísticos tendo foco nos que retratam a mortalidade e mais especificamente sobre as doenças que compõem o grupo das infecciosas.

Silva *et al.* (2023) traçaram o perfil epidemiológico dos internados com malária no sudeste do Pará (Brasil) tendo em conta 72 casos, através das frequências absoluta e relativa percentual e tabela tendo notado que a maior parte dos casos são do sexo masculino (76,38%), têm raça parda (54.16%) e estão na faixa etária dos 20 aos 39 anos (51,38%).

Souza (2017) apresenta o perfil da mortalidade das doenças infecciosas e parasitárias no Maranhão (Brasil) no período de 2003 a 2014 recorrendo a histogramas, frequências absolutas e relativas entre outras técnicas estatísticas descritivas, tendo notado que haviam 11585 óbitos por doença, uma taxa média de óbitos de 15,13/100000 habitantes (manifestando-se uma tendência crescente ao longo dos anos), as doenças com maior taxa de mortalidade foram HIV/SIDA, tuberculose e doenças infecciosas intestinais sendo que a primeira afectou maioritariamente a população entre 15 e 59 anos (50,9%) e a última afectou maioritariamente os menores de 5 anos (77,44%).

Macuvele *et al.* (2018), com recurso a frequências absolutas e relativas e a gráficos ilustrativos, avaliam o perfil clínico dos doentes que evoluíram para obtido no Centro de Saúde Mariem Ngoabyem Xai-Xai (Moçambique) entre Março de 2013 e Março de 2015, tendo notado que durante o período de análise houve 54 óbitos por tuberculose como doença básica onde 55,6% foram do sexo masculino e 44,4% do sexo feminino; a forma clínica da tuberculose mais frequente foi a pulmonar, 87% tinham serologia positiva para o HIV e 80% morreram em fase intensiva do tratamento.

Nhachungue *et al.* (2022) recorreram as frequências absolutas e relativas percentuais para estudar o perfil da mortalidade na região Centro de Moçambique tendo em conta os anos de 2019 e 2020, em que foram notificados 3030 óbitos dentre os quais 37,4% eram crianças menores de 5 anos. Notou-se, neste estudo, que dos 353 casos de morte neonatal, as causas principais foram prematuridade (41,4%) e sepses (25,2%); dos 780 casos de morte dos menores de 5 anos, as principais causas foram diarreias, malária e pneumonia com 29,5%, 14,6% e 7,7% respectivamente; na faixa etária dos 5 aos 14 anos, dos 158 óbitos, as principais causas foram traumas, doenças diarreicas e malária com 20,9%, 15,8% e 14,6% respectivamente; na faixa etária dos 15 aos 49, dos 820 óbitos, tem-se como principais causas o HIV/SIDA e trauma com 29% e 14,9% respectivamente e para os maiores de 50 anos (910 casos obtidos) têm as neoplasias e o HIV como as principais causas de morte com 17% e 16,8% respectivamente. Este estudo conclui que as doenças infecciosas, que são na sua maioria evitáveis, constituem a principal causa de morte.

Viera *et al.* (2018), recorrendo às frequências absolutas e relativas percentuais e à mediana, apresentam o perfil clínico dos doentes com HIV positivos internados nas Enfermarias de Medicina do Hospital Central da Beira (Moçambique) com desfecho clínico de óbito de Janeiro a Junho de 2015 tendo notado que os sintomas mais frequentes são a astenia (67%), perda de peso (51%), anorexia (49%) e febre (44%); as doenças que levaram a internamento foram anemia (45%), SIDA (42%), brancopneumonia (32%) e tuberculose (31%); os dias de internamentos variaram de 0 a 90 dias tendo uma mediana de 5 dias e as doenças de que contribuíram para menor sobrevivência foram a tuberculose tumores e AVC.

Uma outra preocupação no âmbito científico intrinsecamente ligada ao perfil epidemiológico da mortalidade são os factores de riscos e/ou preditores da mortalidade. Dentre tantas técnicas estatísticas, a mais usada nos estudos desta temática é a regressão logística apresentando um poder classificatório apreciável (maior que 80% nos estudos consultados). Neste sentido, a seguir apresentam-se alguns dos estudos com esta abordagem.

Santos (2019) pesquisou os determinantes da mortalidade por tuberculose e TB-HIV no sul do Brasil utilizando regressão logística identificado os seguintes determinantes: baixas rendas, superlotação das casas e consumo de álcool.

Uetela *et al.* (2018) realizou um estudo com recurso à um modelo de regressão logística ajustado para identificar os determinantes da infecção por HIV em meninas adolescentes e mulheres

jovens entre os 15 e 24 anos em Moçambique, tendo sido estes os seguintes: elevado número de parceiros, ser divorciada ou viúva e baixo nível de escolaridade.

Lopes *et al.* (2023) estudou os factores associados à mortalidade em unidades de terapia intensiv no Piauí (Brasil) com recurso à análise univariada através da taxa de mortalidade e testes Qui-quadrado e a análise multivariada através da regressão logística binária tendo detectado os seguintes factores: ser do sexo masculino, ter raça amarela, estar na fase idosa e estar nos internamentos de urgência ou prolongadas.

Gonsalves (2013) realizou um estudo que visava determinar os preditores da mortalidade em pessoas com síndrome coronária aguda recorrendo a regressão logística e algumas técnicas univariadas, concluindo que o sexo, a idade, o índice de massa corporal, a frequência cardíaca, pressão arterial, classe de Killip e classificação da função ventricular esquerda (VE) são preditores da morte.

CAPÍTULO III:
CARACTERIZAÇÃO TEÓRICA DOS
MÉTODOS ESTATÍSTICOS.

3. Caracterização teórica dos métodos estatísticos

A análise estatística constitui um elemento imprescindível no presente trabalho e por isso, neste capítulo, apresentam-se aspectos teóricos em torno dos métodos estatísticos a serem aplicados neste trabalho, desde os conceitos, os gráficos entre outros.

Inicia-se com algumas técnicas da Estatística Descritiva; segue-se para uma visão geral sobre os modelos de regressão e os modelos lineares generalizados como aspectos que antecedem e buscam elucidar o aspecto seguinte que é a metodologia da regressão logística.

3.1. Algumas técnicas da Estatística Descritiva

3.1.1. Frequências

As frequências correspondem ao número de repetições de uma observação, ou seja, quantas vezes determinado fenómeno acontece.

Frequências simples ou absolutas (f_i) são os valores que realmente representam o número de dados de cada observação (Côrrea, 2013). A soma das frequências simples é igual ao número total dos dados.

Frequências relativas (fr) são os valores das razões entre as frequências simples e o número total de dados (idem).

Multiplicando por 100% a frequência relativa, obtém-se a frequência relativa percentual.

3.1.2. Gráficos

Os gráficos possibilitam uma rápida visualização e compreensão do fenómeno bem como a representação de uma distribuição de frequências e séries estatísticas. Os gráficos devem apresentar simplicidade, clareza da informação que se deseja transmitir e tradução exacta da realidade dos factos (Assumpção *et al.*, 2018).

3.1.2.1. Gráfico de barras ou colunas

Esta forma gráfica refere-se à representação de uma série por meio de rectângulos, dispostos verticalmente (em colunas) ou horizontalmente (em barras) (Assumpção *et al.*, 2018).

3.1.2.2. Histograma

Esta representação apresenta-se em colunas ou em barras de um conjunto previamente tabulado e dividido em classes sejam elas uniformes ou não em que a base do rectângulo corresponde a uma classe e a altura corresponde à frequência com que o valor da classe ocorre no conjunto de dados (Assumpção *et al.*, 2018).

3.1.2.3. Gráficos em sectores ou circulares

Este gráfico é construído com base em um círculo e é utilizado sempre que se deseja ressaltar a participação do dado no total, em que o total é representado pelo círculo que por sua vez divide-se em sectores (Assumpção *et al.*, 2018).

Dependendo da quantidade de sectores que se pretende dividir, este gráfico pode se tornar confuso, por isso é relevante analisar os casos em que se pretende aplicar esta forma de representação gráfica dos dados.

3.1.3. Medidas de tendência central

3.1.3.1. Média aritmética

A média aritmética é a medida descritiva mais importante (Correia, 2013).

A média aritmética refere-se ao quociente da divisão da soma dos valores da variável pelo número deles (Correia, 2013; “Cálculo de”, 2013). Os mesmos autores garantem que as fórmulas para determinar esta medida são:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \text{ ou } \bar{X} = \frac{\sum_{i=1}^n X_i \times f_i}{n} \quad (3.1.1)$$

Em que X_i é o valor observado, f_i é a frequência absoluta do valor observado e n é o número total de observações.

Constituem vantagens da média aritmética: ser a medida mais conhecida e de maior emprego, ser facilmente calculável e depender de todos os valores do conjunto de dados (“Cálculo de”, 2013).

Por outro lado, a desvantagem principal da mesma é o facto de uniformizar valores de um conjunto de dados que nem sempre pode representá-los, isto é, a média é fortemente influenciada por valores extremos (muito grandes ou muito pequenos) (“Cálculo de”, 2013).

3.1.3.2. Mediana

O valor da mediana (Me) encontra-se no centro da série estatística organizada, de tal forma que o número de elementos situados antes desse valor (mediana) é igual ao número de elementos que se encontram após esse mesmo valor (mediana) (Correia, 2013).

Caso o tamanho da amostra seja par, a mediana corresponde à média aritmética dos dois valores que dividem o conjunto de dados em duas partes iguais.

3.1.3.3. Moda

A moda (Mo) é o valor que ocorre com maior frequência em um conjunto de dados (Correia, 2013; “Cálculo de”, 2013), isto é, é o valor de maior frequência absoluta.

Esta medida tem como vantagem principal o facto de não ser influenciada pelos extremos e apresenta as desvantagens de o seu valor não depender de todos valores do conjunto de dados, nem sempre existe e nem sempre é único o valor modal.

3.1.4. Medidas de dispersão

As medidas de dispersão são medidas estatísticas utilizadas para avaliar o grau de variabilidade de um conjunto de dados (Correia, 2013; “Cálculo de”, 2013). Servem ainda para aferir a representatividade da média.

3.1.4.1. Variância

A variância leva em consideração os valores extremos e os valores intermediários, isto é, expressa melhor os resultados obtidos (Correia, 2013). A variância relaciona os desvios em torno da média, ou, mais especificamente, é a média aritmética dos quadrados dos desvios (Correia, 2013).

De acordo com Correia (2013) e “Cálculo de” (2013) tem-se a seguinte fórmula para determinar a variância:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} \quad (3.1.2)$$

A fórmula acima é destinada a dados amostrais, quando se trata de dados populacionais, o de denominador passa a ser n ao invés de $n - 1$. O estimador S^2 é um estimador centrado para σ^2 (variância populacional).

3.1.4.2. Desvio-padrão

O desvio-padrão é a medida mais usada na comparação de diferenças entre conjuntos de dados, por ter grande precisão (Correia, 2013). O desvio padrão determina a dispersão dos valores em relação à média e é calculado por meio da raiz quadrada da variância.

O desvio-padrão é dado por:

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n - 1}} \quad (3.1.3)$$

Evidentemente, a fórmula acima é destinada a dados amostrais, quando tratar-se de dados populacionais altera-se o denominador do mesmo modo que na variância.

3.1.4.3. Coeficiente de variação

O coeficiente de variação (CV) é uma medida relativa de dispersão, útil para a comparação em termos relativos do grau de concentração (Correia, 2013; “Cálculo de”, 2013). O coeficiente de variação é a relação entre o desvio-padrão e a média (Correia, 2013; “Cálculo de”, 2013).

Esta medida é dada por:

$$CV = \frac{S}{\bar{X}} \times 100\% \quad (3.1.4)$$

De acordo com Correia (2013), diz-se que:

- A distribuição possui baixa dispersão se $CV \leq 15\%$;
- A distribuição possui média dispersão se $15\% < CV < 30\%$;
- A distribuição possui alta dispersão $CV \geq 30\%$.

3.2. Modelos de regressão

Uma das vias para a descrição e conhecimento de aspectos dos variados campos da actividade humana desde as ciências biológicas, exactas e sociais são os modelos matemáticos.

Os modelos matemáticos são representações de um sistema real objectivando a compreensão, descrição, apresentação do comportamento e a identificação das características mais essenciais duma determinada realidade (Sondré, 2007; Dutra & Franco, 2020). Trata-se de “um conjunto de equações que representam de uma forma quantitativa, as hipóteses que foram usadas na construção do modelo as quais se apoiam sobre o sistema real” (Sondré, 2007:4)

Nesta senda, tem-se os modelos estatísticos, que são casos particulares de modelos matemáticos. Os modelos estatísticos não só permitem a observação sobre como as variáveis se relacionam como também são capazes de evidenciar principais resultados e tendências de acordo com os dados em estudo (Cunha, 2023) através de ajustes com expressões matemáticas auxiliando assim numa previsão e conseqüentemente na tomada de decisão.

De acordo com Cunha (2023), tem-se as seguintes fases da modelação estatística são:

- Recolha de dados;
- Análise descritiva e exploratória, em que se apresenta as principais características dos dados, medidas de resumo (por exemplo medidas de tendência central, medidas de

dispersão, etc.), análise do conjunto de variáveis a fim de resumir as informações mais básicas, formular hipóteses iniciais e obter detalhes mais precisos sobre a os dados;

- Manipulação da base de dados, que consiste em extrair apenas informações necessárias;
- Diagnostico, em que se observam as correlações entre as variáveis e as tendências dos elementos de interesse, é nesta fase que o modelo inicial passa por testes até chegar ao modelo final.

Frequentemente, pelos estudos consultados, tem-se usado a análise de regressão no âmbito científico dos diversos campos de conhecimento.

O termo regressão surge nos trabalhos de Sir Francis Galton (1885) que se preocupava com a relação entre as alturas dos pais e filhos, tendo descoberto que os pais altos tendem a ter filhos altos e os pais baixos tendem a ter filhos baixos (Almeida, 2014; Moura, 2018) entretanto, a altura tendia a se descolar ou “regredir” para altura média da população como um todo (Almeida, 2014 Pontifícia Universidade do Rio de Janeiro [PUC-R], s.d). Nas palavras de Galton, isso chama-se “regressão à mediocridade” (Alunos do Mestrado em Administração [ADM], 2016)

No entanto, a interpretação actual da regressão é diferente. Actualmente, o conceito de regressão encontra-se ligado á determinação do comportamento de uma variável através da dependência desta em relação a uma ou mais variáveis.

A análise de regressão é uma técnica estatística que consiste na construção de uma equação matemática para investigar a relação entre uma variável resposta e uma ou mais variáveis independentes (Amaral *et al.*, 2009; Chein, 2019).

Notoriamente, a regressão é um modelo estatístico que se baseia numa expressão matemática em que uma variável resposta (dependente) é compreendida através de uma ou mais variáveis independentes visando fazer previsões e assim tomar decisões de acordo com as previsões feitas.

Neste sentido, de acordo com Cerqueira (2016) a relação estabelecida na regressão pode ser analisada como um processo em que X_j são as variáveis de entrada ou regressoras (inputs) e o valor de Y a é variável de saída ou resposta (output).

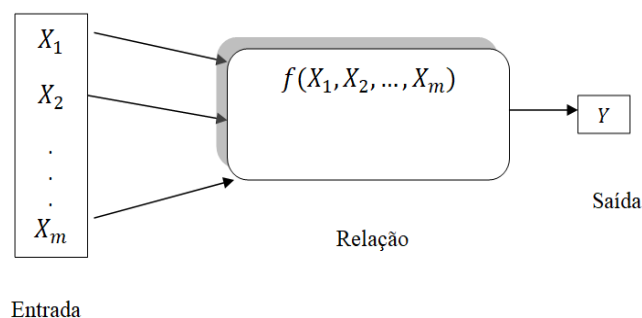


Figura 3.1: Esquema de um modelo de regressão. Fonte: Cerqueira (2016:13), Adaptada

X_1, X_2, \dots, X_m chamadas de variáveis de entrada, a relação $f(X_1, X_2, \dots, X_m)$ corresponde a equação do modelo através da qual se obtém o valor da variável de saída Y .

Tendo acima se tratado sobre o conceito de análise de regressão e de alguns elementos à volta do mesmo, importa apresentar de forma destacada os objectivos da análise de regressão. Nesta ordem de ideias, de acordo com Almeida (2014), a utilização da análise de regressão pode ter os seguintes objectivos:

- **Predição**, visto que com o modelo se deseja que uma parte considerável da variabilidade de Y seja explicada pelas variáveis X_j (com $j = 1, 2, \dots, m$) então pode obter-se (ou prever-se) Y através das variáveis correspondentes.
- **Seleção de variáveis**, geralmente não se conhece as variáveis X_j que contribuem significativamente para a variação de Y , no entanto, conduz-se uma pesquisa com um grande número de variáveis e a análise de regressão auxilia no sentido de eliminar aquelas variáveis cuja sua contribuição estatística não seja importante.
- **Estimação de parâmetros**, dado o modelo e conjunto de dados (amostra) referente às variáveis resposta e preditoras, estimar parâmetros, ou ainda ajustar o modelo aos dados, significa obter valores (estimativas) para os parâmetros, tendo-se por base a estrutura do modelo e os dados observados.
- **Inferências**, para além de estimar parâmetros, o ajuste de um modelo de regressão também busca realizar inferências sobre eles tais como testes de hipóteses e intervalos de confiança.

Os modelos de regressão podem ser classificados de acordo com o número de variáveis independentes, sendo chamada de regressão simples quando há apenas uma variável independente e quando há mais de uma variável independente é considerada regressão múltipla (Cerqueira, 2016).

Por outro lado, os modelos de regressão podem ser classificados de acordo com o tipo de função que o define, podendo ser de diferentes tipos como linear, polinomial, etc. (Cerqueira, 2016).

Há diversos métodos usados para a estimação de um modelo de regressão, sendo um dos mais usados o método de mínimos quadrados.

De acordo com Barreto *et al.* (2016: 111-112), o método de mínimos quadrados “(...) fundamenta-se em minimizar o erro quadrático médio das medidas, ou seja, consiste em encontrar estimadores para os parâmetros de forma que a soma dos quadrados dos desvios entre os valores estimados pelo modelo e os valores observados seja a menor possível.”

Ao tratar dos modelos de regressão, os de regressão linear constituem um ponto de partida.

Na regressão linear simples (onde apenas temos uma variável independente X , admite-se que existem n pares de valores de duas variáveis x_i e y_i com $i = 1, 2, \dots, n$) sendo Y uma função linear de X cujo modelo estatístico é dado por:

$$Y = \beta_0 + \beta_1 X + \epsilon \quad (3.2.1)$$

Na regressão linear múltipla assume-se que há uma relação linear entre a variável resposta e as m variáveis explicativas (Barreto *et al.*, 2016:111). Num modelo de regressão linear múltipla assume-se que existe uma variável dependente Y e m variáveis independentes X_j ($m > 1$). De acordo com Rodrigues (2012) e Almeida (2014), o modelo de regressão linear múltipla tem a forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \epsilon$$

Em que β_0, \dots, β_m são os parâmetros que se pretende estimar sendo que β_0 é o valor estimado de Y quando todas as variáveis explicativas são nulas e β_1, \dots, β_m medem a influência que X_j tem sobre Y ; β_j são constantes no modelo e ϵ é variável residual que representa a variação de Y não explicada pelo modelo (isto é, que deriva de outros factores).

O modelo de regressão a ser desenvolvido neste trabalho é o de regressão logística. Estes modelos distinguem-se da regressão linear pelo facto de a variável dependente ser nominal. Esta diferença reflecte-se tanto na escolha do modelo como também nas hipóteses a considerar.

No entanto, desde que se tenha em conta esta diferença, a regressão logística segue os mesmos princípios que a regressão linear, daí que, a breve descrição sobre a regressão linear acima feita, tal como foi referido, constitui um ponto de partida.

3.3. Modelos Lineares Generalizados

O modelo de regressão linear desenvolvido por Legendre e Gauss no início do século XIX foi a principal técnica de modelação estatística até meados do século XX apesar de vários modelos não lineares ou não normais terem sido desenvolvidos para fazer face a situações que não eram adequadamente explicadas pelo modelo linear normal antes dessa época (Turkman & Silva, 2000).

São exemplos disso, tal como referem McCullagh e Nelder (1989) e Lindsey (1997) citados por Turkman e Silva (2000): o modelo complementar log-log para ensaios de diluição, os modelos probit, logit para proporções, os modelos log-lineares para dados de contagens, entre outros.

Ainda segundo Turkman e Silva (2020), os exemplos acima referenciados são modelos que apresentam uma estrutura de regressão linear e possuem em comum o facto da variável resposta seguir uma distribuição dentro de uma família de distribuições com propriedades específicas: a família exponencial.

Os modelos lineares generalizados (MLG) introduzidos por Nelder e Wedderburn (1972) correspondem a uma teoria de unificação destes e mais modelos do ponto de vista conceptual e teórico da modelação estatística até então desenvolvida (Turkman & Silva, 2000; Cordeiro & Demétrio, 2008).

Os MLG são uma generalização dos modelos de regressão linear clássicos com principal característica de haver possibilidade lidar com variáveis resposta que não seguem a distribuição normal desde que esta siga uma distribuição que faça parte da família exponencial.

Alguns dos modelos que fazem parte dos MLG's são: modelo de regressão linear, modelo de análise de variância e co-variância, modelo de regressão logística, modelo de regressão de Poisson, modelo *probit* para os estudos de proporções entre outros (Turkman& Silva, 2000).

Notoriamente os MLG's incluem o modelo de regressão linear clássico e por isso os MLG's são uma extensão dos modelos de regressão linear. Esta extensão permite não só casos em que a variável resposta seja contínua como também casos em que esta seja categórica.

Nesta ordem de ideias, o conceito de MLG não procura trazer novos modelos de regressão, mas apenas enquadrar modelos num único conjunto desde que estes satisfaçam as características do grupo.

3.3.1. Componentes dum modelo linear generalizado

A perspectiva de McCullagh e Nelder (1989) define os modelos lineares generalizados através de três componentes fundamentais:

a) Componente aleatória

A variável resposta que se pretende modelar é aleatória, da qual se recolhem m variáveis independentemente distribuídas com média μ cuja distribuição de probabilidades faz parte da família das exponenciais.

b) Componente sistemática ou estrutural

Consiste numa combinação linear de variáveis predictoras havendo m variáveis predictoras e n observações, tal que:

$$\eta = X\beta = \beta_0 + \beta_1 X_1(i) + \dots + \beta_m X_m(i) \quad (3.3.1)$$

Para $i = 1, \dots, n$.

Os preditores podem ser variáveis numéricas ou indicadores associados a factores.

c) Função de ligação

A função de ligação (g) é uma função diferencial e monótona que associa as componentes aleatória e sistemática de tal forma que:

$$g(\mu_i) = \beta_0 + \beta_1 X_1(i) + \dots + \beta_m X_m(i) \quad (3.3.2)$$

A função g é invertível. Um MLG modela o valor esperado (μ) duma variável resposta (Y) com distribuição na família exponencial através da expressão:

$$\mu_i = g^{-1}(\beta_0 + \beta_1 X_1(i) + \dots + \beta_m X_m(i)) \quad (3.3.3)$$

3.3.2. Família exponencial

O conceito de família exponencial foi introduzido na Estatística pelo estatístico britânico Fisher (Vaz, 2020). Pelos aspectos já referenciados anteriormente, a família exponencial encontra-se directamente ligada aos MLG's, sendo, por isso, importante apresentar a sua estrutura.

De acordo com Turkman e Silva (2020) diz-se que uma variável aleatória Y tem distribuição pertencente a família da exponencial se a função densidade de probabilidade pode ser escrita na forma:

$$f(y|\theta, \varphi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\varphi)} + c(y, \varphi) \right\} \quad (3.3.4)$$

Onde θ e φ são parâmetros escalares; $a(\cdot)$, $b(\cdot)$ e $c(\cdot)$ são funções reais conhecidas e y são dados observados (variável de interesse).

Nesta definição, θ corresponde a forma canónica do parâmetro de localização (parâmetro natural ou função de ligação canónica) e φ corresponde ao parâmetro de dispersão associado á variância (Tukman & Silva, 2000; Lima, s.d).

Admite-se que quando o parâmetro φ é desconhecido pode ou não fazer parte da família exponencial, tal como bi-paramétrica como é geralmente definida e que a função $b(\cdot)$ é diferenciável tendo um suporte da distribuição que não depende dos parâmetros (Tukman & Silva, 2000).

Baseando-se em Turkman e Silva (2000) e Lima (s.d), a seguir tem-se as propriedades da família exponencial.

- O valor médio da distribuição é dado por:

$$\mu = E(Y) = b'(\theta) \quad (3.3.5)$$

- A variância da distribuição é dada por:

$$var(Y) = a(\varphi)b''(\theta) \quad (3.3.6)$$

De acordo com Turkman e Silva (2000), nota-se que em muitos casos, a função $a(\varphi)$ toma a forma:

$$a(\varphi) = \frac{\varphi}{\omega} \quad (3.3.7)$$

Onde ω é uma constante conhecida, obtendo-se a variância de Y como produto do parâmetro de dispersão por uma função apenas do valor médio.

Nestes casos, a função da família exponencial já definida toma a forma:

$$f(y|\theta, \phi) = \exp \left\{ \frac{\phi}{\omega} [y\theta - b(\theta)] + c(y, \phi, \omega) \right\} \quad (3.3.8)$$

Na tabela 3.1 a seguir apresenta-se as principais distribuições da família exponencial que a variável resposta pode seguir e o tipo de dados correspondentes.

Tabela 3.1: Distribuições da família exponencial

Distribuição	Tipo de dados
Normal	Contínuos
Poisson	Contagens
Bernoulli	Dicotómicas
Binomial	Proporções
Gama	Contínuos assimétricos
Normal inversa	Continua assimétrica

Fonte: Vaz (2020:14). Adaptado

Inúmeras funções podem ser consideradas funções de ligação, no entanto, a decisão de escolha está relacionada com o tipo de resposta que se pretende (Vaz, 2020). Na tabela 3.2 que se segue indicam-se as relações entre as principais distribuições mais comuns da componente aleatória e a função de ligação comumente usada.

Tabela 3.2: Funções de ligação

Distribuição	Função de ligação		Função de variância
	Designação	Notação	
Normal	Identidade	μ	1
Poisson	Logit	$\log\left(\frac{\mu}{1-\mu}\right)$	$\mu \frac{1-\mu}{n}$
Binomial	Logarítmica	$\log(\mu)$	μ
Gama	Recíproca	1	μ^2
Normal inversa	Quadrática inversa	$\frac{1}{\mu^2}$	μ^3

Fonte: Vaz (2020:14). Adaptado.

Importa referir que tal como os MLG's englobam a regressão linear, a família das exponenciais engloba a distribuição normal sendo que nesta distribuição a função ligação é a identidade.

3.4. Regressão logística

Há muitas situações sociais que pelos seus dados são qualitativas ou discretas ao invés de contínuas ou quantitativas (Barrios, 2021). Nestes casos, a regressão linear não é plausível. Neste âmbito, quando se aplica a regressão linear para variáveis dependentes dicotómicas (0 ou 1), de acordo com Campel (2000, citado por Barrios, 2021), registam-se alguns problemas que precisam de ser ultrapassados:

- Não linearidade, com uma variável dependente dicotómica, a representação gráfica da regressão linear toma a forma “S” e assim, a variação da covariável junto os limites 0 e 1 (correspondentes a cada categoria) faz com que a variável resposta varie de forma residual, enquanto no meio a mesma variação da covariável tem um efeito maior na variação da variável resposta;
- Aditividade, quando uma covariável empurra a variável resposta para próximo de um dos limites, os efeitos das outras covariáveis passam a ser irrelevantes na variável resposta;
- Distribuição não normal, com uma variável resposta dicotómica, apenas dois resíduos existem para todo o valor da covariável;
- Heterocedasticidade, a variância não é constante, isto é, junto dos limites os resíduos dos valores da covariável são pequenos e vão aumentando quando os valores da covariável tendem para o meio.

Uma forma razoável de ultrapassar estes problemas é a transformação logarítmica da regressão linear nos moldes descritos na secção sobre os MLG. Esta transformação tem o nome de Regressão Logística.

A Regressão Logística é uma técnica estatística usada nos casos em que a variável resposta é qualitativa nominal dicotómica pretendendo verificar a relação desta com uma série de variáveis explicativas sejam elas numéricas e/ou categóricas. Trata-se de um modelo linear generalizado ligado ao modelo *logit*.

A regressão logística compõe um grupo maior chamado de regressão categorial. A regressão categorial é destinada a casos em que a variável resposta é qualitativa e assume classes discretas e mutuamente exclusivas (Marôco, 2011), sendo, por isso, a regressão logística um caso particular da regressão categorial.

Existe uma extensão da regressão logística: a regressão multinomial, que se refere aos casos em que a variável resposta é qualitativa nominal politémica; no entanto a regressão logística é mais comum (Marôco, 2011). A regressão multinomial é também um caso particular da regressão categorial.

Hosmer e Lemshow (2000) indicam vantagens da regressão logística afirmando que este modelo é flexível e detém uma interpretação poderosa e objectiva.

Com o modelo de regressão logística, pode-se alcançar os seguintes objectivos (Gonzalez, 2018):

- Modelar a probabilidade de um evento decorrer de acordo com os valores das variáveis independentes (sejam estas quantitativas ou qualitativas);
- Estimar a probabilidade de um evento ocorrer contra a probabilidade do mesmo não ocorrer para uma variável seleccionada aleatoriamente;
- Prever o efeito do conjunto de variáveis sobre a variável resposta;
- Classificar as observações, estimando a probabilidade de uma observação estar numa determinada categoria.

A variável resposta Y da regressão logística é binária e por isso segue a distribuição de Bernoulli (Gonzalez, 2018).

$$Y = \begin{cases} 1, & \text{se ocorrer sucesso} \\ 0, & \text{se ocorrer fracasso} \end{cases}$$

A probabilidade p de sucesso varia de 0 a 1 ($0 \leq p \leq 1$) e a probabilidade de fracasso é $q = 1 - p$.

Na regressão logística é feita a estimação de probabilidade desconhecida p dado uma combinação linear de variáveis independentes (Gonzalez, 2018).

3.4.1. Regressão logística simples ou univariada

Na regressão uma quantidade essencial é o valor médio condicional (o valor médio da variável resposta dado o valor da variável independente) que é expressa como $E[Y/X]$ em que Y representa a variável resposta e X a única variável explicativa. Neste sentido, tem-se $E[Y/X = x]$ que é o valor médio condicional dado $X = x$.

No modelo de regressão linear simples admite-se que o valor médio condicional pode ser expresso na seguinte forma:

$$E[Y/X = x] = \beta_0 + \beta_1 x \quad (3.4.1)$$

Neste caso, de modo a simplificar a notação pode-se considerar $\pi(x) = E[Y/X = x]$. De acordo com Cabral (2013), na regressão logística assume-se que:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (3.4.2)$$

3.4.2. Transformação logit

Um dos aspectos fundamentais da regressão logística é a transformação que a caracteriza: a função logit. De acordo com Cabral (2013), a função logit é dada por:

$$g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) \quad (3.4.3)$$

Substituindo $\pi(x)$ da equação (3.4.3) tem-se:

$$g(x) = \ln\left(\frac{\frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}}{1 - \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}}\right) \leftrightarrow g(x) = \ln\left(\frac{\frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}}{\frac{1}{1 + e^{\beta_0 + \beta_1 x}}}\right) \quad (3.4.4)$$

Simplificando a expressão (3.4.4), tem-se:

$$g(x) = \ln(e^{\beta_0 + \beta_1 x}) = \beta_0 + \beta_1 x \quad (3.4.5)$$

Da equação anterior, tem-se o modelo da regressão logística simples:

$$g(x) = \beta_0 + \beta_1 x \quad (3.4.6)$$

Em muitas literaturas é comum considerar-se $g(x) = \text{logit}(\pi)$. Neste sentido, também pode-se considerar como equação da regressão logística simples:

$$\text{logit}(\pi) = \beta_0 + \beta_1 x \quad (3.4.7)$$

Esta transformação, que culmina com o modelo de regressão logística, é denominada de transformação *logit* de $\pi(x)$ e a razão $\frac{\pi(x)}{1 - \pi(x)}$ é denominada *Odds Ratio*.

3.4.3. Razão de chances (*Odds Ratio*)

A razão de chances é “a probabilidade de que um evento ocorra dividido pela probabilidade de que ele não ocorra, ou seja é o quociente entre a quantidade de eventos e a quantidade livre de eventos” (Cerqueira, 2016:28). Trata-se de uma medida que compara a possibilidade de dois eventos sendo especificamente a razão entre a chance de ocorrer um evento em um grupo e a chance do mesmo ocorrer em outro grupo (Moura, 2018).

Evidentemente na *Odds Ratio* (OR) tem-se o quociente entre a *odds* do evento de interesse ($Y=1$) ocorrer nos indivíduos com $x=1$ e a *odds* desse evento ocorrer em indivíduos com $x=0$ (Vaz, 2020). A *odds* do evento de interesse ocorrer em $x=1$ e em $x=0$ são definidas por $\frac{\pi(1)}{1-\pi(1)}$ e $\frac{\pi(0)}{1-\pi(0)}$ respectivamente.

$$odds = \frac{\text{probabilidade do evento ocorrer}}{\text{probabilidade do evento não ocorrer}} \quad (3.4.8)$$

De acordo com Vaz (2020), a probabilidade do evento de interesse ocorrer nas categorias já referenciadas pode ser dada pelas igualdades:

$$\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}} \pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}} \quad (3.4.9)$$

Por consequência, o valor da OR é dado por:

$$OR = \frac{\pi(1)[1 - \pi(0)]}{\pi(0)[1 - \pi(1)]} \quad (3.4.10)$$

Cerqueira (2016), Moura (2018) e Vaz (2020) apresentam considerações em relação aos valores que OR pode tomar:

- $OR > 1$ indica que o evento é mais provável no primeiro grupo;
- $OR < 1$ indica que o evento é menos provável no primeiro grupo;
- $OR = 1$ indica que o evento é igualmente provável em ambos grupos.

Para confirmar a significância dos valores do OR é importante ter em conta o intervalo de confiança e o valor-p.

3.4.4. Regressão logística múltipla

Apresentou-se, na secção 3.4.1, a regressão logística simples, que é o caso em que há apenas uma variável independente. Nesta secção importa tratar, de forma análoga, sobre o caso em que há duas ou mais variáveis independentes, o que se denomina regressão logística múltipla.

Neste caso assume-se:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m}} \quad (3.4.11)$$

E por consequência tem-se o modelo de regressão logística múltipla:

$$g(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m \quad (3.4.12)$$

Pelos mesmos motivos apresentados na secção 3.4.2 pode-se considerar como modelo de regressão logística múltipla:

$$\text{logit}(\pi) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m \quad (3.4.13)$$

3.4.5. Pressupostos da regressão logística

De acordo com Marôco (2011:811), os pressupostos do modelo de regressão logística são:

- Linearidade e aditividade: o modelo $\text{logit}(\pi)$ é aditivo e linear (mas π não);
- Proporcionalidade e aditividade: a proporcionalidade de x_i é proporcional ao seu respectivo factor β_i ;
- Constância e efeito: a contribuição de cada variável independente é constante e independente da contribuição das outras variáveis independentes;
- Os erros são independentes e apresentam distribuição binomial;
- Os preditores não são multicolineares (a semelhança da regressão linear).

A validação dos pressupostos do modelo pode ser feita graficamente pela análise dos resíduos e o diagnóstico da multicolinearidade pode ser feito através da tolerância (Marôco, 2011).

3.4.6. Ajustamento do modelo: O método de máxima verosimilhança

Visto que os erros da regressão logística não seguem uma distribuição normal e nem apresentam variância constante, através do método de mínimos quadrados (usado na regressão linear) não se obtém estimativas óptimas dos coeficientes do modelo, sendo neste caso usado um outro método: o método de máxima verosimilhança (Marôco, 2011).

Este método estima os coeficientes de regressão que maximizam a probabilidade de encontrar as realizações da variável dependente amostradas, isto é, que tem maior probabilidade de replicar o padrão das observações (Marôco, 2011; Gonzalez, 2018).

A seguir faz-se o desdobramento dos estimadores de máxima verosimilhança para a regressão logística univariada recorrendo à perspectiva de Queiroz (2004, citado por Carvalho, 2022).

A função de máxima verosimilhança (L) é dada por:

$$L(\beta) = \prod_{i=1}^n f\left(\frac{y_i}{x_i}\right) \quad (3.4.14)$$

Logaritmizando a função (3.4.14), tem-se:

$$LL(\beta) = \ln(L(\beta)) = \ln\left(\prod_{i=1}^n f\left(\frac{y_i}{x_i}\right)\right) \quad (3.4.15)$$

$$LL(\beta) = \sum_{i=1}^n [y_i \ln(P(x_i)) + (1 - y_i) \ln(1 - P(x_i))] \quad (3.4.16)$$

$$LL(\beta) = \sum_{i=1}^n \left[y_i \ln\left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}\right) + (1 - y_i) \ln\left(1 - \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}\right) \right] \quad (3.4.17)$$

$$LL(\beta) = \sum_{i=1}^n [y_i(\beta_0 + \beta_1 x_i) - \ln(1 + e^{\beta_0 + \beta_1 x_i})] \quad (3.4.18)$$

Derivando $LL(\beta)$ em relação β a fim de encontrar os coeficientes, chega-se às seguintes igualdades:

$$\frac{\partial LL(\beta)}{\partial \beta_0} = \sum_{i=1}^n \left(y_i x_i \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) \quad (3.4.19)$$

$$\frac{\partial LL(\beta)}{\partial \beta_1} = \sum_{i=1}^n \left(y_i x_i - \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) \quad (3.4.20)$$

Os estimadores obtidos através do método de máxima verosimilhança são centrados e com variância mínima, sendo, por isso, eficientes e produzem estimativas com distribuição normal assintótica (Marôco, 2011). Estes estimadores medem a taxa de variação do *logit* para uma unidade de variação na variável dependente Y e a sua variável independente X (Gonzalez, 2018).

Tendo em conta o modelo de regressão logística multivariada e os procedimentos seguidos anteriormente para a verosimilhança da regressão logística univariada, de acordo com Carvalho (2022) tem-se os seguintes resultados para o método de verosimilhança da regressão logística multivariada:

$$LL(\beta) = \sum_{i=1}^n [y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)] \quad (3.4.21)$$

$$\frac{\partial LL(\beta)}{\partial \beta_0} = \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\pi}_i = 0 \quad (3.4.22)$$

$$\frac{\partial LL(\beta)}{\partial \beta_j} = \sum_{i=1}^n x_{ij} y_i - \sum_{i=1}^n x_{ij} \hat{\pi}_i = 0 \quad (3.4.23)$$

Com $j = \{1, \dots, m\}$ sendo que no modelo existem $m + 1$ coeficientes (incluindo β_0).

De referir que $\hat{\pi}_i$ é o estimador de máxima verosimilhança de π_i .

3.4.6.1. Matriz de informação

Com aplicação do método iterativo dos mínimos quadrados ponderados pode obter-se as raízes das equações (3.4.22) e (3.4.23) que são estimativas dos parâmetros da modelo da regressão logística multivariada e através destes estimar os coeficientes de máxima verosimilhança (Vaz, 2020). Estes estimadores são obtidos a partir da matriz de covariância de segundas derivadas da função log verossimilhança (Vaz, 2020).

$$\frac{\partial^2 LL(\beta)}{\partial \beta_j^2} = - \sum_{i=1}^n x_{ij}^2 \pi_i (1 - \pi_i) \quad (3.4.24)$$

$$\frac{\partial^2 LL(\beta)}{\partial \beta_j \partial \beta_l} = - \sum_{i=1}^n x_{ij} x_{il} \pi_i (1 - \pi_i) \quad (3.4.25)$$

Onde $j, l = 0, 1, \dots, m$

Seja uma matriz quadrada de ordem $m+1$ de termos negativos das derivadas parciais das equações (3.4.24) e (3.4.25) é denominada matriz de informação de Fisher e denotada por $I(\beta)$ (Vaz, 2020).

De acordo com o mesmo autor, a matriz de informação de Fisher pode ser obtida por:

$$I(\beta) = \begin{bmatrix} -\frac{\partial^2 LL(\beta; x)}{\partial \beta_1^2} & \cdots & -\frac{\partial^2 LL(\beta; x)}{\partial \beta_1 \partial \beta_m} \\ \vdots & \ddots & \vdots \\ -\frac{\partial^2 LL(\beta; x)}{\partial \beta_m \partial \beta_1} & \cdots & -\frac{\partial^2 LL(\beta; x)}{\partial \beta_m^2} \end{bmatrix} \quad (3.4.26)$$

De acordo com Vaz (2021), os valores β_1, \dots, β_m são parâmetros cujos estimadores de máxima verossimilhança são obtidos como soluções das equações:

$$\frac{\partial LL(\beta_j; x)}{\partial \beta_j} \quad (3.4.27)$$

As variâncias e co-variâncias dos coeficientes estimados são obtidas a partir da inversa da matriz $I(\beta)$ (idem).

De referir que a matriz de informação de Fisher pode ainda ser estimada por $I(\beta) = X'VX$ (Cabral, 2013; Maroco, 2011) em que X é uma matriz de ordem $nx(m+1)$ que contém o vector 1 e os valores observados das variáveis independentes e V é uma matriz quadrada de ordem n cuja diagonal é composta pelos valores na forma $\hat{\pi}_i(1 - \hat{\pi}_i)$ (Cabral, 2013).

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ 1 & x_{31} & x_{32} & \cdots & x_{3m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad V = \begin{bmatrix} \hat{\pi}_1(1 - \hat{\pi}_1) & 0 & \cdots & 0 \\ 0 & \hat{\pi}_2(1 - \hat{\pi}_2) & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \hat{\pi}_n(1 - \hat{\pi}_n) \end{bmatrix} \quad (3.4.28)$$

3.4.6.2. Intervalos de confiança

De acordo com Marôco (2011), o intervalo de confiança dos coeficientes da regressão logística é dada por:

$$] \beta_i - z_{1-\frac{\alpha}{2}} \widehat{SE}(\beta_i); \beta_i + z_{1-\frac{\alpha}{2}} \widehat{SE}(\beta_i) [\quad (3.4.29)$$

Em que $z_{1-\frac{\alpha}{2}}$ é o percentil $1 - \frac{\alpha}{2}$ da distribuição normal reduzida com α nível de significância.

De forma equivalente para o rácio de chances, tem-se:

$$] e^{\beta_i - z_{1-\frac{\alpha}{2}} \widehat{SE}(\beta_i)} ; e^{\beta_i + z_{1-\frac{\alpha}{2}} \widehat{SE}(\beta_i)} [\quad (3.4.30)$$

Estes intervalos de confiança são baseados na normalidade assintótica do método de máxima verosimilhança que só é válida para amostras de grande dimensão (considera-se, geralmente, amostras em que o número de observações seja de 10 a 20 vezes maior que o número de coeficientes do modelo) (Marôco, 2011).

A medida $\widehat{SE}(\beta_i)$ corresponde ao desvio ou erro padrão de β_i obtido através da matriz de informação de Fisher, sendo o i -ésimo termo da inversa desta matriz o seu valor (Marôco, 2011).

3.4.7. Teste de rácio de verosimilhança

De acordo com Marôco (2011), este teste usa-se para testar a significância do modelo ajustado e tem como hipóteses:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_m = 0$$

$$H_1: \exists j, \beta_j \neq 0 \quad (j = 1, 2, \dots, m)$$

A hipótese nula refere que o modelo não é estatisticamente significativo e a hipótese alternativa refere que o modelo é estatisticamente significativo (Marôco, 2011). Se o modelo não for significativo, não é possível prever a probabilidade de “sucesso” com base nas variáveis independentes (Marôco, 2011).

A estatística de teste à significância do modelo compara a verosimilhança do modelo só com a constante (modelo nulo ou reduzido: $\text{logit}(\pi) = \beta_0$) com a verosimilhança do modelo com variáveis independentes na combinação linear (modelo completo: $\text{logit}(\pi) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m$) e tem-se em conta o indicador mediocridade do ajustamento dos modelos aos dados, $-2LL(\beta)$, em que quando maior for o seu valor, pior é ajustamento e se $-2LL(\beta) = 0$, o ajustamento é perfeito.

O cálculo da estatística de teste à significância do modelo (G^2), segundo Marôco (2011), é geralmente feito das formas que seguem.

➤ Para o modelo nulo:

$$X_0^2 = -2LL_0 \sim \chi_{n-1}^2 \quad (3.4.31)$$

➤ Para o modelo completo:

$$X_C^2 = -2LL_C \sim \chi_{n-1-p}^2 \quad (3.4.32)$$

O mesmo autor, afirma que a estatística deste teste é dada por:

$$G^2 = X_0^2 - X_C^2 = -2LL_0 - (-2LL_C) = -2 \ln \left[\frac{L_0}{L_C} \right] \sim \chi_p^2 \quad (3.4.33)$$

Onde LL_0 e LL_C representam o logaritmo da função de verosimilhança dos modelos completo e nulo, respectivamente. Rejeitando a hipótese nula, conclui-se que pelo menos um dos parâmetros é estatisticamente diferente de zero e consegue consequentemente prever o $\text{logist}(\pi)$.

3.4.8. Teste de ajustamento do modelo

De acordo com Marôco (2011), para este teste são formuladas as seguintes hipóteses:

H_0 : *O modelo ajusta – se aos dados*

H_1 : *O modelo não se ajusta aos dados*

Se os dados estão agrupados em J células, resultantes de cruzamento de preditores qualitativos, usa-se a estatística (Hosmer & Lemshow, 2000):

$$X_p^2 = \sum_{j=1}^J \left(\frac{y_j - n_j \hat{\pi}_j}{\sqrt{n_j \hat{\pi}_j (1 - \hat{\pi}_j)}} \right)^2 = \sum_{j=1}^J \frac{(O_j - E_j)^2}{E_j} \quad (3.4.34)$$

Onde y_j é o número de sucessos na célula j , $\hat{\pi}_j$ é a probabilidade estimada de célula j , n_j é o número de elementos da célula j , usando a notação tradicional do qui-quadrado, O_j e E_j são número de sucessos observados e esperados, respectivamente. Uma das alternativas propostas nestes casos (por exemplo por McCullage & Nelder, 1989) é a estatística Deviance:

$$D = 2 \sum_{j=1}^J \left[y_j \ln \left(\frac{y_j}{n_j \hat{\pi}_j} \right) + (n_j - y_j) \ln \left(\frac{n_j - y_j}{n_j (1 - \hat{\pi}_j)} \right) \right] \quad (3.4.35)$$

As duas estatísticas apresentam $J-m-1$ graus de liberdade. Entretanto, se o modelo incluir dados quantitativos e se o n for próximo de j , as duas estatísticas não podem ser usadas com dados individuais.

Nestes casos, o teste qui-quadrado pode ser aplicado através de uma tabela de contingências $2g$ e fazendo g grupos definidos pelos decis de sucesso (geralmente $g=10$), O_j é o número de “sucessos” definidas pelo decil e E_j em cada classe decílica é dada por $E_j = n_j \hat{\pi}_j$. E assim tem-se a estatística de teste:

$$X_{HL}^2 = \sum_{j=1}^g \frac{(O_j - E_j)^2}{E_j} \quad (3.4.36)$$

Para os casos em que a amostra é grande X_{HL}^2 tem distribuição assintótica (χ_{g-2}^2) .

Alternativamente, o teste ao ajustamento do modelo pode ser feito com a estatística X_C^2 decrita na secção anterior.

3.4.9. Teste à significância dos coeficientes do modelo

Tendo-se tido a conclusão de que o modelo ajustado é significativo, para identificar quais as variáveis que influenciam significativamente o modelo usualmente recorre-se ao teste de Wald (Marôco, 2011). Neste caso, tem-se as hipóteses para cada regressor:

$$H_0: \beta_j = 0$$

$$H_0: \beta_j \neq 0, j=1,2,\dots,m$$

A estatística de teste é:

$$T_{Wald} = \frac{\hat{\beta}_l}{SE(\hat{\beta}_l)} \sim N(0,1) \quad (3.4.37)$$

Onde $\hat{\beta}_l$ é estimador de β_j e $SE(\hat{\beta}_l)$ é o estimador do erro padrão de β_j .

Marôco (2011) aponta algumas fragilidades do teste de Wald: o facto de ser verdadeiramente válida em amostras de grande dimensão (pois a distribuição é assintótica), ter maiores probabilidades de erros do tipo I e II em relação ao teste t da regressão linear e ser pouco indicada quando as variáveis independentes são qualitativas.

Neste sentido, para contornar estas fragilidades, o mesmo autor aponta que para testar a significância das variáveis é aconselhável optar por um teste de rácio de verosimilhança que compare a verosimilhança de um modelo completo com a verosimilhança de modelos alinhados (modelos cujas variáveis independentes são subconjunto das variáveis independentes do modelo completo): o teste Omnibus.

Neste caso, o teste Omnibus tem as mesmas hipóteses que o teste de Wald porém apresenta uma estatística de teste G^2 calculada pela diferença da verosimilhança do modelo em que a variável que se pretende avaliar a significância é eliminada ($-2LL_{-x_j}$) e a verosimilhança do modelo completo ($-2LL_C$).

$$G^2 = -2LL_{-x_j} - (-2LL_C) \sim \chi^2_{(1)} \quad (3.4.38)$$

Os graus de liberdade de G^2 são dados pela diferença de variáveis independentes dos dois modelos.

3.4.10. Pseudo- R^2

Na regressão linear é usada a estatística R^2 para avaliar em termos percentuais o efeito do modelo. Na regressão logística, uma vez que a variável dependente é dicotómica e variância desta depende da probabilidade de ocorrência dos valores, não é possível calcular o R^2 .

O pseudo- R^2 é similar, porém não é uma estatística equivalente ao R^2 sendo necessário tomar algumas precauções na sua interpretação. Existem variadas formas apontadas na literatura para calcular o pseudo- R^2 , são geralmente destacadas duas o de Cox e Smell (1989) e o de Nagelkerke (1991) citados em Maroco (2011), pelo facto de serem usadas no meio computacional.

De acordo com Marôco (2011) e Vaz (2020), o pseudo- R^2 de Cox e Smell é dado por:

$$R_{CS}^2 = 1 - e^{\frac{2(LL_C - LL_0)}{n}} \quad (3.4.39)$$

Mesmo que o ajustamento seja perfeito, esta estatística não atinge o valor 1 e face a isto Nagelkerke propôs uma correcção (Marôco, 2011; Vaz, 2020) que resultou no seguinte coeficiente:

$$R_N^2 = \frac{R_{CS}^2}{1 - e^{-\frac{2LL_0}{n}}} \quad (3.4.40)$$

Apesar do R_N^2 variar no intervalo de $[0; 1]$, Shtatland *et al.* (2002 citados por Marôco, 2011) concluíram que pseudo- R^2 de McFadden é o melhor dentre os demais, sendo este dado por:

$$R_{MF}^2 = 1 - \frac{LL_C}{LL_0} \quad (3.4.41)$$

3.4.11. Selecção de variáveis independentes

Tendo o conjunto das variáveis independentes que se pretende incluir no modelo, é importante identificar o conjunto de variáveis com poder preditor, que melhor explicam a variabilidade e probabilidade de sucesso.

Antes de apresentar os métodos comumente usados para a regressão logística, a seguir apresentam-se os métodos de selecção de forma genérica nas asserções de Marôco (2011) e Moura (2018).

- *Enter*, em que todas as variáveis em bloco são inseridas em um único passo, não tem exclusão de variáveis insignificantes;
- *Forward*, o modelo inicia com apenas a constante (β_0), de seguida cada variável é adicionada individualmente, sendo a primeira a que adiciona a de maior poder explicativo no modelo e assim sucessivamente até que nenhuma das variáveis restantes aumente o poder de explicação do modelo;
- *Backward*, inicia com todas as variáveis e retira individualmente a variável que adiciona o menor poder de explicação ao modelo até que restem somente as variáveis que expliquem significativamente uma parcela da variável dependente;
- *Stepwise*, é um método híbrido que incorpora os dois métodos anteriores, inicia com apenas uma variável (como no *forward*) mas a significância de adição de uma nova variável é testada como no método *backward*.

Os métodos mais usuais são os *stepwise (forward e backward)* (Ferreira, 2012), estes métodos apresentam a vantagem de eliminar uma variável independente cuja importância no modelo é reduzida face a novas variáveis (Marôco, 2011).

Tanto os métodos *stepwise forward* como os *stepwise backward* usuais podem possuir testes de remoção baseados numa das três formas: razão de verosimilhança fundamentada nas estimativas paramétricas condicionais, na probabilidade da estatística de razão de verosimilhança fundamentada nas estimativas de probabilidade parcial máximas ou na probabilidade da estatística Wald, no entanto os métodos *stepwise forward* apresentam um teste de entrada com base na significância da estatística escore.

3.4.12. Análise dos resíduos

O diagnóstico do modelo baseado na análise dos resíduos é feito para a validação da qualidade do modelo estimado, avaliando a discrepância entre os valores observados e os valores esperados.

A análise de resíduos permite identificar artefactos (*outliers*) e casos influentes na estimação do modelo. As medidas usadas para a análise de resíduos na regressão logística são basicamente as mesmas da regressão linear: resíduos, *leverage*, distância de Cook e DfBetas; entretanto na regressão logística é necessário considerar que as variâncias das observações não são constantes como acontece na regressão linear e as fórmulas de cálculo têm de ser adaptadas a esse facto (Marôco, 2011).

De acordo com Marôco (2011), os resíduos ou erros são a diferença entre o número de sucessos observados (y_i) e o número de sucessos estimados (\hat{y}_i) na célula i :

$$e_i = y_i - \hat{y}_i = y_i - n_i \hat{\pi}_i^2 \quad (3.4.42)$$

Onde n_i é o número de observações da célula i e $\hat{\pi}_i$ é a probabilidade de sucesso na célula i .

Estes resíduos não estão estandardizados. Para obter os resíduos estandardizados (também chamados de resíduos de Pearson) deve-se dividir o resíduo anterior pela estimativa do desvio-padrão:

$$e'_i = \frac{e_i}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}} \quad (3.4.43)$$

Estes resíduos não são completamente estandardizados pois a sua influência depende das observações e da sua influência na estimação dos coeficientes do modelo (Marôco, 2011). Esta influência pode ser estimada pela *leverage* através da expressão:

$$h_i = n_i \hat{\pi}_i(x_i) [1 - \hat{\pi}_i(x_i)] x_i' (X' V X)^{-1} x_i' \quad (3.4.44)$$

Onde $x_i' = [1, x_{1i}, x_{2i}, x_{3i}, x_{mi}]$ é o vector com as observações i em todas as variáveis independentes.

Existem também os resíduos estudentizados, que são dados pela expressão:

$$r_i = \frac{e_i'}{\sqrt{1 - h_i}} \quad (3.4.45)$$

Os resíduos estudentizados r_j e a *leverage* h_i para além de permitir a identificação de *outliers* também são usados para avaliar a influência de uma observação no ajustamento do modelo.

No mesmo âmbito tem-se uma medida análoga a distância de Cook que indica a variação dos resíduos quando se retira uma observação i do ajustamento do modelo (Marôco, 2011). Esta distância tem como base o resíduo de Pearson e a *leverage*, podendo calcular-se através da fórmula:

$$DC_i = r_i^2 \frac{h_i}{1 - h_i} \quad (3.4.46)$$

Os DfBetas estimam a influência de cada observação na estimação de cada um dos coeficientes de regressão (Marôco, 2018). Na regressão logística, estes são dados por:

$$DfBeta_{ji} = \hat{\beta}_j - \hat{\beta}_{j(-i)} \quad (3.4.47)$$

Onde $\hat{\beta}_j$ é a estimativa do modelo de regressão, $\hat{\beta}_{j(-i)}$ é a estimativa do coeficiente ajustado sem a observação i .

Pode dar-se as considerações idênticas às da regressão linear em que os valores acima de $2 \sqrt{\frac{m+1}{n}}$ (em que $m + 1$ é o número de coeficientes do modelo e n é a dimensão da amostra) são observações influentes.

3.4.13. Classificação da previsão do modelo

De acordo com Marôco (2011), tendo o modelo de regressão logística, é possível estimar a probabilidade $\hat{\pi}_i$ de cada um dos sujeitos pertencer ao grupo de 1 (“sucesso”) em relação ao grupo 0 (“insucesso”) através de uma das expressões seguintes:

$$\hat{\pi}_i = \frac{e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n x_{ni}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n x_{ni}}} \quad (3.4.48)$$

$$\hat{\pi}_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n x_{ni})}} \quad (3.4.49)$$

Neste caso, considera-se pertencente ao grupo 1 se a probabilidade de $Y_i = 1$ for maior que 0.5 e pertencente ao grupo 0, caso contrário.

É usual avaliar a qualidade de classificação do modelo comparando a percentagem global das classificações correctas obtidas com o modelo e as classificações correctas obtidas por acaso. Esta é calculada a partir do número de sujeitos observados em cada uma das 2 classes da variável dependente (C_j) pela expressão:

$$\text{Classificação correcta proporcional por caso (\%)} = 100 \times \sum_{j=1}^2 \left(\frac{C_j}{n} \right)^2 \quad (3.4.50)$$

3.4.13.1. Sensibilidade e especificidade do modelo

São medidas para a eficiência classificatória. A sensibilidade é a percentagem de classificações correctas no grupo 1 e a especificidade é a percentagem de classificação correcta no grupo 0 (Cerqueira, 2016; Carvalho, 2022). Formalmente, de acordo com Marôco (2011), tem-se:

$$\text{Sensibilidade} = P[\hat{Y} = 1 | Y = 1]$$

$$\text{Especificidade} = P[\hat{Y} = 0 | Y = 0]$$

Um modelo considera-se com boas capacidades preditivas quando possui sensibilidade e especificidade acima de 80%; com capacidades preditivas razoáveis, se estes valores estiverem entre 50% e 80% e com capacidades preditivas medíocres, se tiver uma percentagem abaixo de 50% (Marôco, 2011; Cerqueira, 2016).

3.4.13.2. Curva de ROC

No mesmo âmbito, tem-se também a Curva ROC (*Receiver Operating Characteristic*, que em português pode se traduzir como Característica de Operação do Receptor) que também avalia a capacidade do modelo discriminar entre sujeitos.

A área abaixo da curva de ROC varia de 0 a 1, em que se a área for de 0.5 significa que o modelo não discrimina e se estiver próxima de 1 maior é a capacidade de discriminação tendo em conta os grupos de “sucesso” (1) e insucesso (0) (Marôco, 2011; Cerqueira, 2016).

A avaliação da discriminação do modelo através da curva de ROC pode ser feita através de um gráfico simples e robusto que permite estudar a variação da sensibilidade e especificidade para diferentes pontos de quebra (Cabral, 2013). Considera-se um ponto de quebra C (geralmente usa-se 0.5) e comparar cada probabilidade com o valor de C.

A área (A) abaixo da curva de ROC fornece uma medida de discriminação, que segundo Marôco (2011) possui as seguintes linhas de orientação:

- $A=0.5$, o modelo não possui poder discriminativo;
- $0.5 < A < 0.7$, a discriminação do modelo é fraca;
- $0.7 \leq A < 0.8$, a discriminação do modelo é aceitável;
- $0.8 \leq A < 0.9$, a discriminação do modelo é boa;
- $A \geq 0.9$, a discriminação do modelo é excepcional.

3.4.14. Kappa de confiabilidade de Cohen

Em algumas bibliografias usa-se o termo concordância entre avaliadores como equivalente a confiabilidade entre avaliadores entretanto alguns autores como Matos (2014) e Alves *et al.*(2021) alertam que apesar da concordância influenciar na confiabilidade, estes termos não indicam o mesmo facto.

Os autores referidos esclarecem que a confiabilidade é uma medida de consistência entre avaliadores na ordenação ou posições relativas de avaliações de desempenho independentemente do valor absoluto da classificação de cada avaliador, enquanto a concordância mede o grau ou

frequência em que dois ou mais avaliadores que usam a mesma escala atribuírem exactamente a mesma classificação.

Para medir a confiabilidade entre dois avaliadores (ou modelos) geralmente é o usado o coeficiente Kappa de Cohen. O Kappa de Cohen representa a proporção de acordo entre avaliadores após ser retirada a proporção esperada de acordo devido o acaso (Santiago, 2016). Santiago (2016) afirma que este coeficiente é dado por:

$$K_{Cohen} = \frac{P_a - P_e}{1 - P_e} \quad (3.4.51)$$

Em que P_a é a proporção de concordância observada e P_e é a proporção esperada devido ao acaso.

A proporção de concordância observada é dada por sua vez:

$$P_a = \frac{1}{n} \sum_{k=1}^r n_{kk} = \sum_{k=1}^r p_{kk} \quad (3.4.52)$$

Onde n_{kk} indica o número de casos atribuídos à categoria i pelos dois avaliadores de um total de r categorias e n indica o número total de observações e assim tem-se $p_{kk} = \frac{n_{kk}}{n}$.

Tem-se também por sua vez a proporção esperada de acordo devido ao acaso dada pela soma dos produtos entre o total da linha pelo total da coluna:

$$P_e = \sum_{k=1}^r p_{k+} p_{+k} = \sum_{k=1}^r \frac{n_{k+}}{n} \times \frac{n_{+k}}{n} \quad (3.4.53)$$

Onde p_{k+} e p_{+k} indicam as respectivas frequências marginais.

O coeficiente de fiabilidade de Cohen é apenas aplicado para dados nominais ou ordinais perante dois avaliadores ou dois momentos (Cohen, 1960 citado por Matos, 2014).

Na regressão logística o Kappa de Cohen é um coeficiente de fiabilidade entre as classificações observadas e as obtidas pelo modelo.

A estatística em alusão assume valores entre -1 e 1 em que 1 significa confiabilidade perfeita, 0 significa que a concordância observada é exactamente a mesma que a concordância esperada devido ao acaso, os valores positivos indicam que a concordância observada é maior que esperada devido ao acaso (Cohen, 1960 citado por Matos, 2014).

Dentre as várias interpretações do Kappa de Cohen, a proposta por Landis e Koch (1977), apresentada na tabela 3.3, é mais abrangente.

Tabela 3.3: Nível de concordância do Kappa de Cohen.

Kappa	Nível de confiabilidade
< 0	Pobre
0 – 0.2	Fraca
0.21 – 0.4	Considerável
0.41 – 0.6	Moderado
0.61 – 0.8	Substancial
0.81 – 1	Quase perfeita

Fonte: Landis e Koch (1977). Adaptado

De acordo com Luz e Kufner (2008) é possível realizar um teste de hipóteses sobre o coeficiente Kappa através da seguinte estatística de teste:

$$\frac{K_{Cohen}}{\sqrt{\sigma_K^2}} \quad (3.4.54)$$

A estatística acima tem uma distribuição aproximadamente normal, sendo que a variância do Kappa é dada por:

$$\sigma_K^2 = \frac{P_a(1 - P_e)}{n(1 - P_e)^2} \quad (3.4.55)$$

As hipóteses do teste, de acordo com Paes e Silva (2012), são:

$$H_0: K = 0$$

$$H_1: K > 0$$

CAPÍTULO IV:
CONCEPÇÕES
METODOLÓGICAS

4. Concepções metodológicas

Neste capítulo mostra-se todos os passos a serem seguidos de modo a responder aos objectivos do trabalho. Apresenta-se, neste capítulo, o tipo de pesquisa sob três pontos de vista, os procedimentos metodológicos e amostragem.

4.1. Tipo de pesquisa

Pesquisar implica usar procedimentos científicos para achar as respostas ao problema proposto. Consequentemente, há que definir clara e concisamente o tipo de pesquisa a ser usada sob os vários pontos de vista.

A pesquisa quanto à natureza pode ser básica ou aplicada. No entanto, a presente pesquisa é aplicada pois pretende-se determinar os factores de risco da MCDI, com o intuito de a quem é de direito possa aplicar este conhecimento para minimizar o problema da mortalidade em causa.

Gil (2008) e Feitas e Prodanov (2013) afirmam que a pesquisa aplicada tem como objectivo gerar conhecimentos a serem aplicados de forma prática dirigidos à solução de problemas específicos, envolvendo, deste modo factos e interesses locais.

A pesquisa quanto à abordagem do problema pode ser: quantitativa, qualitativa ou mista. A presente pesquisa tem abordagem quantitativa visto que todos os procedimentos usados para alcançar os objectivos do trabalho são baseados em dados numéricos e em métodos estatísticos (técnicas descritivas e análise de regressão logística).

A pesquisa quantitativa é aquela que quantifica todas as informações para as analisar e classificar, assim ela recorre a recursos estatísticos como percentagem, moda, mediana, análise de regressão entre outros (Gil, 2008).

A pesquisa quanto aos objectivos pode ser descritiva, explicativa ou exploratória. A presente pesquisa engloba as facetas descritiva e explicativa, pois por um lado descreve as características dos óbitos causados pelas doenças infecciosas através do seu perfil e por outro lado procura encontrar os factores de risco que determinam a possibilidade de óbito.

A pesquisa descritiva visa descrever as características de determinada população ou fenómeno ou estabelecimento de relações entre variáveis (Gil, 2008).

A pesquisa explicativa tem por objectivo identificar os factores que determinam ou contribuem para a ocorrência de fenómenos; estas pesquisas aprofundam o conhecimento da realidade pois explicam o porquê das coisas (Gil, 2008).

4.2.Procedimentos

Pretendeu-se, neste estudo, recolher os dados dos doentes com doenças infecciosas, tanto os que tiveram alta como os que evoluíram para óbito dos anos de 2020 a 2022 do Hospital Central do Maputo, a maior unidade sanitária de Moçambique.

A pesquisa foi essencialmente documental. A pesquisa documental é aquela que se baseia em materiais que não receberam ainda um tratamento analítico ou que podem ser reelaborados de acordo com os objectivos da pesquisa (Gil, 2008; Freitas & Prodanov, 2013).

Neste sentido, recorreu-se aos documentos de registo do Hospital que apresentam dados sobre a malária, a tuberculose e o HIV/SIDA nos anos em causa e de seguida fez-se um tratamento estatístico.

Gil (2008:17) ao debruçar-se sobre o método estatístico afirma que “Este método fundamenta-se na aplicação da teoria estatística da probabilidade e constitui importante auxílio para a investigação em ciências” e que este “passa a se caracterizar por razoável grau de precisão, o que o torna bastante aceite por parte dos pesquisadores com preocupações de ordem quantitativa”.

Deste modo, precisou-se de dados sobre a mortalidade nas doenças infecciosas em causa para que se possa descrever o perfil da mortalidade através da Estatística Descritiva e precisou-se também de dados individuais de diferentes doentes, tanto os que evoluíram para óbito como os que tiveram alta para que através da regressão logística fosse possível determinar os factores de risco da mortalidade.

As características dos pacientes recolhidas foram género, idade, estado civil, nível de distância da residência do internado ao hospital, existência de doença crónica associada, consumo de álcool, ser ou não fumante, número de internamentos anteriores. Após recolhidas estas características, as mesmas foram introduzidas no *software*.

De salientar que para a obtenção dos dados, passou-se por um longo processo. Após ter dado entrada do pedido de dados no dia 17 de Janeiro de 2024, obteve-se uma autorização do Hospital que foi formalizada através duma carta de cobertura (em anexo I) no dia 30 de Maio de 2024 e uma autorização ética realizada pelo Comité Institucional de Bioética em Saúde (em anexo I) no dia 07 de Outubro de 2024.

A recolha de dados realizou-se entre os meses de Outubro e Novembro de 2024 e teve a duração de duas semanas.

A análise estatística foi feita utilizando o IBMSPSS® (Statistical Package for the Social Sciences) versão 27 como o *software* principal e o Excel como auxiliar. O SPSS é um programa flexível e de fácil utilização e bastante usado em variados estudos científicos em geral e em estudos relacionados com aspectos epidemiológicos em particular. O Excel serviu apenas para obtenção de algumas figuras (os gráficos que apresentam simultaneamente as cores azul e vermelha).

4.3.Amostragem

Pelos processos dos doentes disponibilizados pelo HCM (cerca de 1000) foi razoável uma amostra de 364 doentes que tiveram alta e 323 óbitos. No entanto, havia ainda interesse de nos restantes processos, fazer-se selecção de uma amostra para efeitos de validação externa. A amostra foi concebida por estratos e de forma proporcional.

Pela tabela 4.1 (criada através de dados do HCM, no anexo II), num total de 3518 altas, por diagnóstico tem-se 1251 para malária, 782 para tuberculose e 1485 para HIV correspondentes às percentagens 35,56%, 22,22% e 42,21% respectivamente. Sendo assim, dos 364 doentes que tiveram alta e fazem parte da amostra tem-se 129 com malária, 81 com tuberculose e 154 com HIV que correspondem às percentagens 35%, 44%, 22,25% e 42,31% respectivamente, que são percentagens próximas das anteriores.

Pela tabela 4.2 (criada através de dados do HCM, no anexo II), num total de 1110 óbitos, por diagnóstico tem-se 60 para malária, 61 para tuberculose e 989 para HIV correspondentes às percentagens 5,4%, 5,5% e 89,1% respectivamente. Sendo assim, dos 323 doentes que foram óbitos e fazem parte da amostra tem-se 17 com malária, 18 com tuberculose e 288 com HIV que

correspondem às percentagens 5,26%, 5,57% e 89,16% respectivamente, que são percentagens próximas às anteriores.

Tabela 4.1: Dados gerais sobre Altas no HCM no triénio 2020-2023

Diagnostico	2020	2021	2022	Total
Malária	401	387	463	1251
Tuberculose	251	277	254	782
HIV	724	460	301	1485
Total	1376	1124	1018	3518

Tabela 4.2: Dados gerais sobre Óbitos no HCM no triénio 2020-2023

Diagnostico	2020	2021	2022	Total
Malária	17	20	23	60
Tuberculose	22	17	22	61
HIV	425	257	307	989
Total	464	294	352	1110

Assim, para selecção dos processos dos doentes, separou-se os processos disponíveis em extractos (malária com alta, tuberculose com alta, HIV com alta, malária óbito, tuberculose óbito e HIV óbito).

Para cada extracto fazia-se o número de papelinhos correspondente ao número de processos disponíveis e marcava-se (com X) o número de papelinhos correspondente ao número de processos que deviam fazer parte da amostra, selava-se e espalhavam-se estes, de seguida colocavam-se nos processos e fizeram parte da amostra os processos com papelinhos marcados.

Nos mesmos moldes, para os processos que sobraram, seleccionou-se uma amostra de 181 doentes dos quais 96 tiveram alta e 85 foram óbitos para efeitos de validação externa. Dos que tiveram alta tem-se 34 com malária (35,4%), 21 com tuberculose (21,9%) e 41 com HIV/SIDA (42,7%) e dos que foram óbitos tem-se 5 com malária (5,8%), 7 com tuberculose (5,8%) e 72 com HIV/SIDA (88,3%).

**CAPÍTULO V:
APLICAÇÃO DA ANÁLISE
ESTATÍSTICA À BASE DE
DADOS**

5. Aplicação da análise estatística à base de dados

Neste capítulo aplica-se técnicas da Estatística Descritiva para se faça o desenho do perfil da MCDI e a Regressão Logística para identificação dos factores de risco da MCDI aos dados recolhidos no HCM. Esta aplicação é precedida de uma análise descritiva (univariada e bivariada) e sucedida pela discussão dos resultados.

Nos testes de hipóteses realizados adoptou-se um nível de significância de 5% ($\alpha = 0,05$).

5.1. Análise descritiva dos dados

Os dados apresentados indicam ser realísticos, visto que para as oito variáveis independentes existem valores sobre todos indivíduos da amostra e estes valores encontram-se dentro das possibilidades que cada variável possui de acordo com a sua natureza, não havendo assim evidências de ter havido alguma falha na concepção da base de dados.

Entretanto, existem dados em falta, visto que há alguns doentes com processos no HCM porém nos mesmos não existem dados relativos à alguma (s) variável (is) de interesse na presente pesquisa.

5.1.1. Análise univariada

Mortalidade ou sobrevivência

A mortalidade ou sobrevivência é a variável dependente ou resposta no presente estudo possuindo duas categorias: alta e óbito. Neste sentido, os doentes com alta são indicados com “0” e os que evoluíram para óbito são indicados por “1”.

São no total 687 doentes que compõem a amostra, dos quais 364 tiveram alta e 323 foram óbitos, o correspondente a 53% e 47% respectivamente (ver na tabela 3.1, em anexo III). Na Figura 5.1, através dum gráfico de barras visualiza-se os dados em alusão.

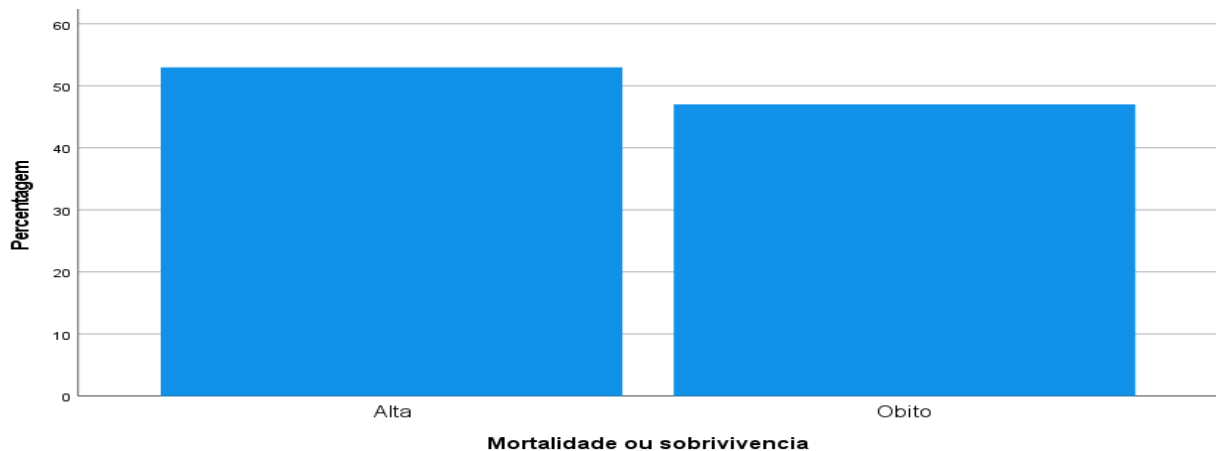


Fig.5.1: Mortalidade ou sobrevivência

Verifica-se que apesar do número das altas ser maior em relação ao número de óbitos, a diferença destes é ligeira sendo de 23 doentes, percentualmente correspondente a 6%.

Género

A variável género apresenta-se em duas categorias que são feminino e masculino, sendo que dos 687 doentes, 355 são do género feminino e 332 são do género masculino correspondentes á 51.7% e 48.3% respectivamente (ver na tabela 3.2, em anexo III). Assim, verifica-se que o género feminino representa a maioria dos doentes entretanto com uma ligeira diferença com o género masculino, sendo esta diferença de 23 doentes, o correspondente á apenas 3,2%.

Grupo etário

O grupo etário é uma variável que se divide em seis categorias, em que cada uma delas identifica um intervalo de idade, designadamente: menores que 1 ano, de 1 a 4 anos, de 5 a 14 anos, de 15 a 24 anos, de 25 a 49 anos e os de 50 ou mais anos de idade. Estes intervalos foram baseados nos estudos sobre Saúde que têm sido realizados pelo Instituto Nacional de Saúde (INS) em Moçambique, como por exemplo o Inquérito Nacional sobre HIV e SIDA em Moçambique do INS (2023).

Dos 687 doentes, têm-se 118 menores de 1 ano (17,2 %), 78 de 1 a 4 anos (11,4%), 85 de 5 a 14 anos (12,4%), 144 de 15 a 24 anos (21%), 215 de 25 a 49 anos (31,3%) e 47 de 50 ou mais anos de idade (ver na tabela 3.3, em anexo III).

Assim, nota-se que maior parte doentes do presente estudo encontram-se no grupo etário de 25 á 49 anos de idade seguido do grupo etário dos 15 a 24 anos. Estes dois grupos juntos perfazem 52,3%, isto é, de 25 a 49 anos encontram-se mais da metade dos doentes em estudo.

Estado civil

A variável estado civil encontra-se dividido em 4 categorias nomeadamente: solteiro (a), casado (a) ou vivendo maritalmente e viúvo (a). Dos doentes, 376 são solteiros, 235 são casados ou vivem maritalmente, 37 são divorciados e 21 são viúvos correspondentes á 54,7%, 34,2%, 5,4% e 3.1% respectivamente, perfazendo um total de 669 doentes correspondentes á 97,4% (ver na tabela 3.4, em anexo 1). Existem, no entanto, 18 doentes (2.6%) cujos processos não continham informação sobre o estado civil (ver na tabela 3.4, em anexo III).

Nota-se que mais da metade dos doentes são solteiros.

Proveniência

Com a variável proveniência, pretendia-se verificar a influência ou não do nível de distância da residência do doente com o Hospital, uma vez que o mesmo inclui no seu atendimento doentes que provenham de qualquer parte do país, inclusive transferidos dos demais Hospitais existentes no país.

Esta variável apresenta quatro categorias ordenadamente: Grande Maputo, Província de Maputo, região sul e fora da região sul. Grande Maputo é a categoria cujos doentes possuem residências mais próximas do Hospital. O Grande Maputo refere-se às cidades de Maputo, Matola e Boane e o distrito de Maracuene; a província de Maputo refere-se a todos os distritos da província do Maputo exceptuando o Grande Maputo, a região sul refere-se às províncias de Inhambane e Gaza (do sul do país) e fora da região sul refere-se às províncias das regiões centro e norte de Moçambique.

Dos doentes, 286 são provenientes do Grande Maputo, 190 da província do Maputo, 127 da região sul e 64 são de fora da região sul, correspondentes a 41,6%, 27,7%, 18,5% e 9,3% respectivamente, perfazendo um total de 667 doentes correspondentes a 97,1% (ver na tabela 3.5, em anexo III). Existem, no entanto, 20 doentes (2,9%) cujos processos não continham informação sobre a sua proveniência (ver na tabela 3.5, em anexo III).

Verifica-se que a maior parte dos doentes provém do Grande Maputo e que o número de doentes vai reduzindo à medida que a distância da residência ao Hospital aumenta.

Consumo de álcool

A variável consumo de álcool apresenta duas categorias: consumidores e não consumidores. Dos 687 doentes, 393 não consomem álcool e 275 consomem, o correspondente a 57,2% e 40% respectivamente perfazendo um total de 668 doentes (97,2%); existindo ainda 19 doentes (2,8%) cujos processos não continham informação sobre o consumo de álcool (ver na tabela 3.6, em anexo III).

Fumantes

A variável sobre fumantes apresenta duas categorias: fumantes e não fumantes. Dos 687 doentes, 587 não são fumantes e 72 são fumantes, o correspondente a 85,4% e 10,5% respectivamente perfazendo um total de 659 doentes (95,9%); existindo ainda 28 doentes (4,1%) cujos processos não continham informação sobre esta variável (ver na tabela 3.7, em anexo III).

Doença crónica associada

Com a variável doença crónica associada se pretendia saber se o doente para além da causa de internamento, sofria duma outra doença associada e que a mesma fosse crónica. Esta variável divide-se em duas categorias: os que sofriam de uma doença crónica e os que não sofriam duma doença crónica.

Dos 687 doentes, 548 não sofrem de doenças crónicas e 127 sofrem de pelo menos uma doença crónica, o correspondente á 79,8% e 18,5% respectivamente perfazendo um total 675 doentes (98,3%); existindo ainda 12 doentes (1,7%) cujos processos de internamento não continham informação sobre doenças associadas (ver na tabela 3.8, em anexo III).

Internamentos anteriores

Para 423 dos casos de estudo, que constituem a maioria (61,6%), tratava-se do primeiro internamento, isto é, nunca antes haviam tido um internamento; no entanto existem 167 casos que tiveram um internamento anterior, 72 casos que tiveram dois internamentos anteriores, 11 casos que tiveram três internamentos anteriores e apenas 3 casos que tiveram 4 internamentos anteriores correspondentes á 24,3%, 10,5%, 1,6% e 0,4% respectivamente, de referir que há 11 doentes (1,6%) cujos processos de internamento não continham informação sobre o número de internamentos anteriores (ver na tabela 3.9, em anexo III).

Visto que o número de internamentos anteriores é uma variável quantitativa, esta está passível de ser feita uma análise das medidas de tendência central, dispersão e assimetria.

O número de internamentos anteriores dos doentes que compõem o estudo varia de 0 á 4, com uma média de 0,53, um desvio padrão de 0,784 e um coeficiente de assimetria de 1,477 (ver na tabela 3.10, em anexo III). Assim, tendo em conta que a variável é discreta e pelo facto do coeficiente de assimetria ser maior que 1 a distribuição é assimétrica a direita e desta feita, confirma-se que uma grande parte dos doentes nunca antes teve um internamento (à esquerda da média 0,53 só existe "0").

5.1.2. Análise bivariada

Para testar a relação entre duas variáveis, recorre-se a alguns testes. Nos casos em que as variáveis são categorizadas usa-se o teste qui-quadrado de independência, entretanto, dependendo do caso, recorre-se ainda a outros testes auxiliares.

O teste de independência qui-quadrado é usado para verificar se existe independência entre duas variáveis categóricas medidas nas mesmas unidades experimentais (Correa, Queiroz & Trevisan, s.d; Aprender Estatística Fácil, 2024). O teste tem como hipótese nula: Não existe associação entre as duas variáveis (Correa, Queiroz & Trevisa, s.d; Aprender Estatística Fácil, 2024).

Através dos mesmos autores, pode-se indicar os seguintes pressupostos: os dados vêm de uma amostra aleatória; as categorias das variáveis são mutuamente exclusivas (cada indivíduo pertencendo a apenas uma categoria) e a frequência esperada em cada célula é pelo menos 5 em 80% das células.

Quanto ao primeiro pressuposto, este foi atendido visto que houve aleatoriedade na seleção da amostra, tal como se descreve na amostragem; nota-se em todas as variáveis que cada indivíduo pertence a apenas uma categoria, não sendo possível que um indivíduo pertença a duas ou mais categorias simultaneamente, isto é, as categorias das variáveis são mutuamente exclusivas atendendo assim o segundo pressuposto e verifica-se que em apenas uma célula é que se tem uma frequência menor que 5 (correspondente a 4 viúvos que tiveram alta) mas esta célula não representa mais de 20% das células no caso, atendendo assim ao terceiro pressuposto.

Para além do teste qui-quadrado, recorre-se também as significâncias dos coeficientes de contingências e do coeficiente Phi para as variáveis dicotômicas e recorre-se também a significância do coeficiente V de Cramer para casos em que a variável independente é nominal e apresenta mais de duas categorias. Estes dois coeficientes são considerados propícios para os casos referidos de acordo com Camacho (2013).

A seguir, testa-se a relação entre a variável resposta (mortalidade ou sobrevivência) e as variáveis independentes.

Mortalidade ou sobrevivência e género do internado

Dos 687 doentes, tiveram alta 188 do género masculino e 176 do género feminino e houve 167 óbitos do género masculino e 156 do género feminino (ver tabela 4.1.1, em anexo IV).

Pelo teste qui-quadrado, tem-se um valor-p de 0,989 (ver tabela 4.1.2, em anexo IV) que é maior que 0.05, sendo assim não se rejeita a hipótese de que não existe associação entre as duas variáveis, isto é, por este teste há independência entre a MCDI e o género do internado. A mesma constatação é confirmada pela significância dos coeficientes de contingência e Phi, que apresentam os mesmos resultados (ver tabela 4.1.3, em anexo IV).

Mortalidade ou sobrevivência e grupo etário do internado

Dos 687 doentes, tiveram alta 27 com menos de 1 ano de idade, 63 com idades entre 1 a 4 anos, 73 com idades entre 5 a 14 anos, 124 com idades entre 15 a 24 anos, 44 com idades entre 25 a 49 anos e 73 com 50 ou mais anos de idade e foram óbitos 91 com menos de 1 ano idade, 15 com idades entre 1 a 4 anos, 12 com idades entre 5 a 14 anos, 20 com idades entre 15 a 24 anos, 171 com idades entre 25 a 49 anos e 14 com 50 ou mais anos de idade (ver tabela 4.2.1, em anexo IV).

Pelo teste qui-quadrado, tem-se um valor- $p < 0.001$ (ver tabela 4.2.2, em anexo IV) que é muito menor que 0.05, sendo assim, rejeita-se a hipótese de que não existe associação entre as duas variáveis, isto é, por este teste há associação entre a MCDI e o grupo etário do internado. A mesma constatação é confirmada pela significância do coeficiente V de Cramer, que apresenta os mesmos resultados (ver tabela 4.2.3, em anexo IV).

Mortalidade ou sobrevivência e estado civil do internado

Dos 669 doentes considerados, tiveram alta 247 solteiros, 92 casados ou que vivem maritalmente, 6 divorciados ou separados e 4 viúvos e foram óbitos 129 solteiros, 143 casados ou que vivem maritalmente, 31 divorciados e 17 viúvos (ver tabela 4.3.1, em anexo IV).

Pelo teste qui-quadrado, tem-se um valor- $p < 0.001$ (ver tabela 4.3.2, em anexo IV) que é muito menor que 0,05, sendo assim, rejeita-se a hipótese de que não existe associação entre as duas variáveis, isto é, por este teste há associação entre a MCDI e o estado civil do internado. A mesma constatação é confirmada pela significância do coeficiente V de Cramer, que apresenta os mesmos resultados (ver tabela 4.3.3, em anexo IV).

Mortalidade ou sobrevivência e Proveniência do internado

Dos 667 doentes considerados, tiveram alta 151 provenientes do Grande Maputo, 100 da província do Maputo, 78 da região sul e 25 são provenientes da zona centro ou norte do país e foram óbitos 135 provenientes do grande Maputo, 90 da província do Maputo, 49 da região sul e 39 são provenientes da zona centro ou norte do país (ver anexo 4.4.1, em anexo IV).

Pelo teste qui-quadrado, tem-se um valor- p de 0.035 (ver tabela 4.3.2, em anexo IV) que é menor que 0.05, sendo assim, rejeita-se a hipótese de que não existe associação entre as duas variáveis, isto é, por este teste há associação entre a MCDI e a proveniência do internado.

Mortalidade ou sobrevivência e consumo de álcool

Dos 668 doentes considerados tiveram altas 214 não consumidores de álcool e 140 consumidores e foram óbitos 179 não consumidores de álcool e 135 consumidores (ver tabela 4.5.1, em anexo IV).

Pelo teste qui-quadrado, tem-se um valor-p de 0.366 (ver tabela 4.5.2, em anexo IV) que é maior que 0.05, sendo assim não se rejeita a hipótese de que não existe associação entre as duas variáveis, isto é, por este teste há independência entre a mortalidade causada por doenças infecciosas e o consumo de álcool. A mesma constatação é confirmada pela significância dos coeficientes de contingência e Phi, que apresentam os mesmos resultados (ver tabela 4.5.3, em anexo IV).

Mortalidade ou sobrevivência e o ser ou não fumante

Dos 659 doentes considerados tiveram alta 342 não fumantes e 17 fumantes e foram óbitos 245 não fumantes e 55 fumantes (ver tabela 4.6.1, em anexo IV).

Pelo teste qui-quadrado, tem-se um valor-p < 0.001 (ver tabela 4.6.2, em anexo IV) que é muito menor que 0.05, sendo assim, rejeita-se a hipótese de que não existe associação entre as duas variáveis, isto é, por este teste há associação entre a MCDI e o facto de o doente ser ou não fumante. A mesma constatação é confirmada pela significância dos coeficientes de contingência e Phi, que apresentam os mesmos resultados (ver tabela 4.6.3, em anexo IV).

Mortalidade ou sobrevivência e doença crónica associada

Dos 675 doentes considerados, tiveram alta 329 que sofriam duma doença crónica associada e 30 que não sofriam e foram óbitos 219 que não sofriam duma doença crónica associada e 97 que não sofriam (ver anexo 4.7.1, em anexo IV).

Pelo teste qui-quadrado, tem-se um valor-p < 0.001 (ver tabela 4.7.2, em anexo IV) que é muito menor que 0.05, sendo assim, rejeita-se a hipótese de que não existe associação entre as duas variáveis, isto é, por este teste há associação entre a MCDI e existência de uma doença crónica associada. A mesma constatação é confirmada pela significância do coeficiente V de Cramer, que apresenta os mesmos resultados (ver tabela 4.7.3, em anexo IV).

Mortalidade ou sobrevivência e número de internamentos anteriores

De acordo com Camacho (2013) e Espírito-Santo e Daniel (2017), o coeficiente de correlação ponto bisserial é usado quando se pretende analisar a correlação entre um variável dicotômica e outra quantitativa. Por outro lado, Lira (2004), afirma o coeficiente Eta é adequada para casos em uma variável é nominal e a outra é quantitativa.

Deste modo, como a mortalidade ou sobrevivência é uma variável simultaneamente dicotômica e nominal e o número de internamentos é quantitativa, usa-se as medidas ponto bisserial e Eta para analisar as correlações das duas variáveis em alusão.

Pelo coeficiente de correlação ponto bisserial que é 0,036 e seu valor-p que é 0,349 (tabela 4.8.1, em anexo IV), visto que o coeficiente se encontra muito próximo a 0 e o valor-p é maior que 0.05, nota-se uma independência entre a mortalidade causada por doenças infecciosas e o número de internamentos anteriores. O valor da medida Eta que é de 0,082 (tabela 4.8.2, em anexo IV), consideravelmente próximo a 0 confirma a independência entre as duas variáveis.

A seguir tem-se a tabela 5.1 que faz resumo da análise bivariada acima detalhada.

Tabela 5.1: Resumo da análise bivariada

Variável testada com a mortalidade	Medidas de correlação usadas	Valor-p	Constatação
Gênero	Qui-quadrado, significâncias de Contingência e Phi	0.989	Há independência entre gênero e a mortalidade por doenças infecciosas
Grupo etário	Qui-quadrado e significância de V de Cramer	<0.001	Há dependência entre o grupo etário e a mortalidade por doenças infecciosas
Estado civil	Qui-quadrado e significância de V de Cramer	<0.001	Há dependência entre o estado civil e a mortalidade por doenças infecciosas
Proveniência	Qui-quadrado	0.035	Há dependência entre a proveniência e a mortalidade por doenças infecciosas

Consumo de álcool	Qui-quadrado, significâncias de Contingência e Phi	0.366	Há independência entre o consumo de álcool e mortalidade por doenças infecciosas
Ser ou não fumante	Qui-quadrado, significâncias de Contingência e Phi	<0.001	Há dependência entre ser ou não fumante a mortalidade por doenças infecciosas
Doença crónica associada	Qui-quadrado, significâncias de Contingência e Phi	<0.001	Há dependência entre a existência de uma doença crónica associada e a mortalidade por doenças infecciosas
Número de internamentos anteriores	Coeficiente e significância do ponto bisserial e Eta	0.036 comvalor- p=0.349 e 0.082	Há independência entre o número de internamentos e mortalidade por doenças infecciosas.

Desta forma, tem-se que a MCDI se encontra ligada ao grupo etário, estado civil, proveniência, ser fumante e existência de doença crónica associada e é independente do género, consumo de álcool e o número de internamentos anteriores.

5.2. Perfil da mortalidade causada por doenças infecciosas

5.2.1. Taxa de mortalidade

Através das tabelas 4.1 e 4.2 (da secção 4.3), resumem-se os dados da mortalidade na tabela 5.2 , encontrando e destacando uma estimativa das taxas de mortalidade (TM) por diagnóstico e nos anos em alusão.

De referir que se considera casos a soma das altas e dos óbitos e a taxa de mortalidade é determinada pela frequência relativa percentual dos óbitos em relação aos casos.

Tabela 5.2: Taxa de mortalidade

Diagnóstico	Ano 2020			Ano 2021			Ano 2022			Total		
	Casos	Óbitos	TM	Casos	Óbitos	TM	Casos	Óbitos	TM	Casos	Óbitos	TM
Malária	418	17	4.1%	407	20	4.9%	486	23	4.7%	1311	60	4.6%
Tuberculose	273	22	8.1%	294	17	5.8%	276	22	8%	843	61	7.2%
HIV	1149	425	34%	717	257	35.8%	608	307	50.5%	2474	989	40%
Total	1840	464	25.2%	1418	294	20.7%	1370	352	25.7%	4628	1110	24%

Tal como verifica-se na tabela anterior, tendo em conta o somatório nos três anos de estudo, houve um total de 1311 casos diagnosticados com malária dos quais 60 foram óbitos conferindo assim uma TM de 4,6%; houve um total de 843 casos com diagnóstico de tuberculose dos quais 61 foram óbitos o que confere uma TM de 7,2% e houve ainda 2474 casos de HIV dos quais 989 foram óbitos conferindo assim uma taxa de mortalidade de 40%. Na Figura 5.2 são visualizados os dados acima descritos.

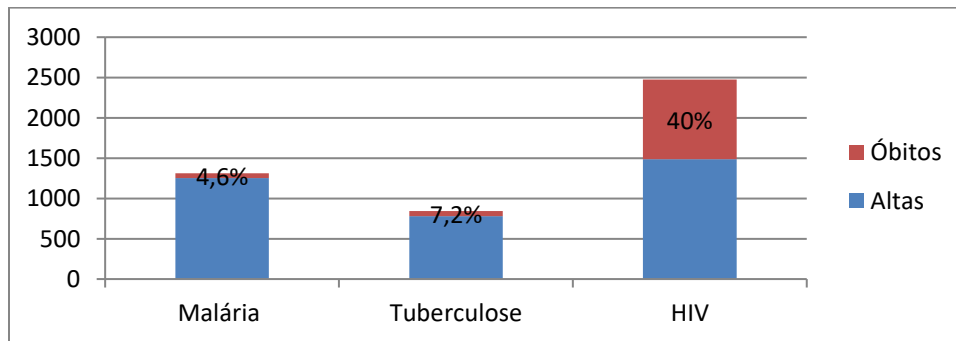


Fig. 5.2: Mortalidade por doença

Nota-se, deste modo, que o HIV para além de causar o maior número de óbitos também apresenta a maior taxa de mortalidade.

Tal como verifica-se na tabela 5.2, tendo em conta o somatório dos três diagnósticos em estudo, houve um total de 1840 casos diagnosticados as doenças infecciosas em 2020 dos quais 464 foram óbitos conferindo assim uma TM de 25,2%; houve um total de 1418 casos com doenças infecciosas em 2021 dos quais 294 foram óbitos o que confere uma TM de 20,7% e houve ainda em 2022, 1370 casos de doenças infecciosas dos quais 352 foram óbitos conferindo assim uma TM de 25,7%. Na Figura 5.3 são visualizados os dados acima descritos.

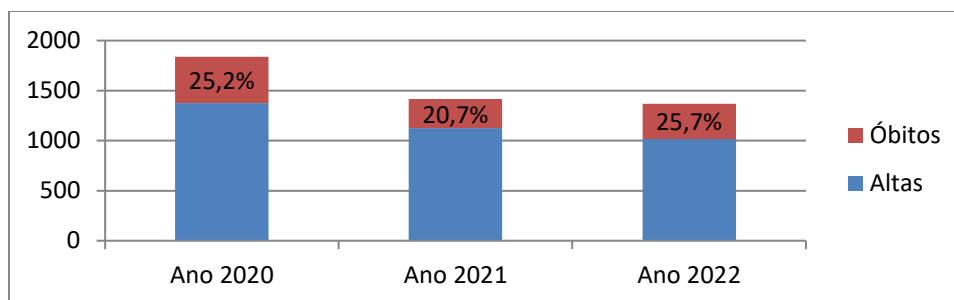


Fig. 5.3: Gráfico da Mortalidade por ano

Nota-se que o ano de 2022 teve o menor número de casos diagnosticados com as doenças infecciosas entretanto teve a maior TM, todavia, não existe uma diferença considerável entre as taxas.

Em geral, a partir dos dados das três doenças no triênio em alusão, diagnosticados 4628 casos dos quais 1110 foram óbitos tem-se uma taxa de 24% da mortalidade das doenças infecciosas.

Na Figura 5.4 visualizam-se os dados anteriormente descritos.

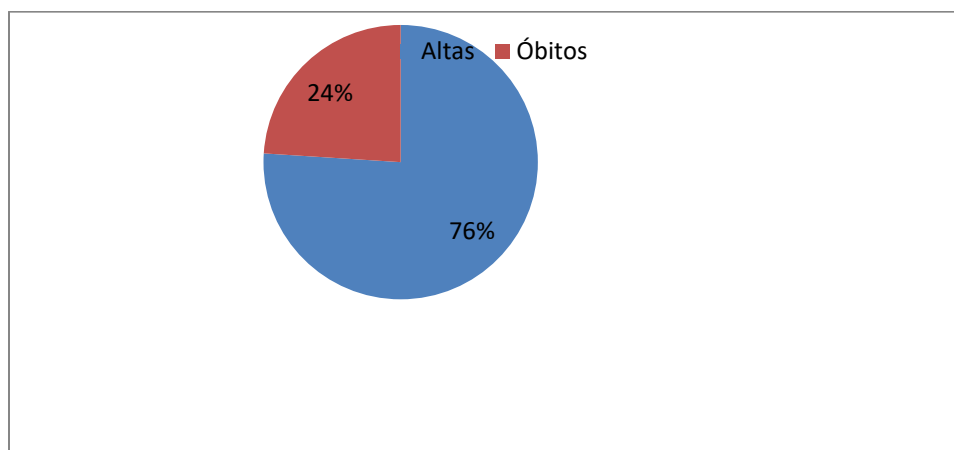


Fig. 5.4: Gráfico da Mortalidade por doenças infecciosas no triênio 2020-2022

Dando sequência à análise do perfil da mortalidade, separou-se da base de dados apenas os 323 casos que foram óbitos e nestes a seguir faz-se a descrição das oito variáveis.

5.2.2. Características biométricas dos óbitos

Pretende-se, nesta secção, tratar sobre as variáveis: género, grupo etário e estado civil.

Em relação ao género (Figura 5.5), pode se afirmar que dos 323 óbitos, 167 eram do género feminino e 156 do género masculino o que corresponde á 51,7% e 48,3% respectivamente (tabela 5.1, em anexo V).

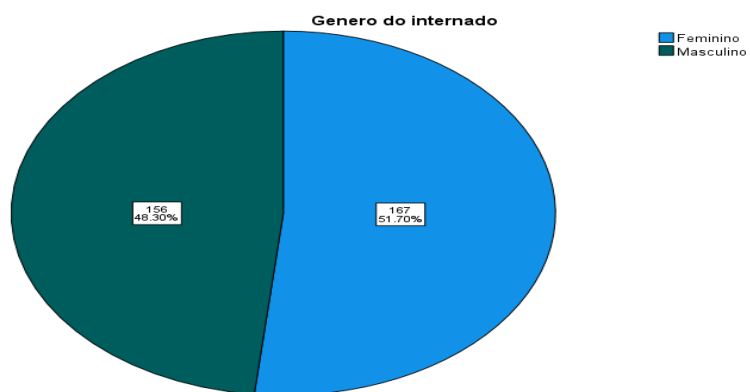


Fig. 5.5: Gráfico do género dos óbitos

Sendo assim, existe uma maioria de óbitos no género feminino, porém não se diferencia consideravelmente do género masculino.

Em relação ao grupo etário (Figura 5.6) pode se afirmar que os 323 óbitos, 91 tinha menos de um de idade, 15 possuíam idades entre 1 a 4 anos, 12 possuíam idades entre 5 a 14 anos, 20 possuíam idades entre 15 a 24 anos, 171 possuíam idades entre 25 a 49 anos e 14 possuíam 50 ou mais anos de idade (tabela 5.2, em anexo V).

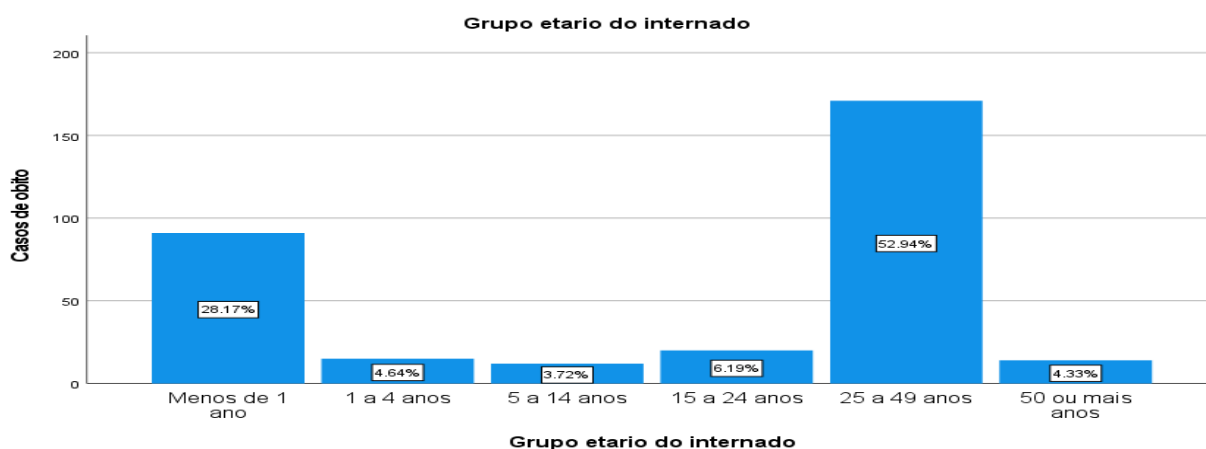


Fig. 5.6: Gráfico do grupo etário dos óbitos

Nesta óptica, destacam-se os óbitos que possuíam idades compreendidas entre 25 e 49 anos de idade seguido dos óbitos que tinham uma idade menor de 1 ano que percentualmente apresentam 52,9% e 28,2% respectivamente.

Em relação ao estado civil (Figura 5.7), pode se afirmar que dos 320 casos considerados, 129 eram solteiros, 143 eram casados ou viviam maritalmente, 31 divorciados ou separados e 17 viúvos (ver tabela 5.3, em anexo V).

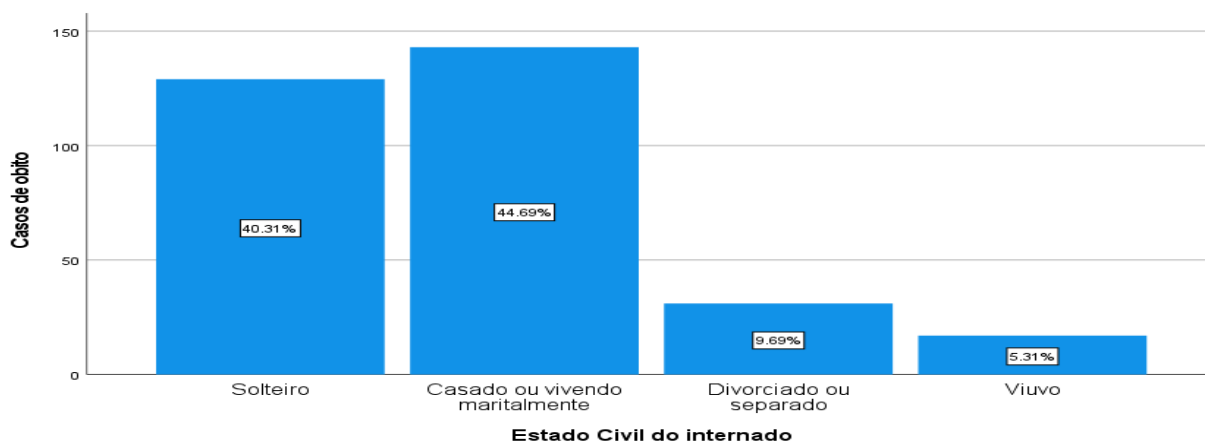


Fig. 5.7: Gráfico do estado civil dos óbitos

Neste caso, destacam-se os casados ou que viviam maritalmente seguidos dos solteiros apresentando 44,7% e 40,3% respectivamente

Cruzando o género e o grupo etário dos óbitos, é notório que o maior número de casos é o das mulheres com idades entre 25 e 49 anos (94 casos), tal como mostra a tabela 5.3.

Tabela 5.3: Cruzamento entre género e grupo etário dos óbitos.

		Grupo etário do internado						Total
		Menos de 1 ano	1 a 4 anos	5 a 14 anos	15 a 24 anos	25 a 49 anos	50 ou mais anos	
Género do internado	Feminino	41 (12,7%)	6 (1,8%)	9 (2,8%)	9 (2,8%)	94 (29,1%)	8 (2,5%)	167 (51,7%)
	Masculino	50 (15,5%)	9 (2,8%)	3 (0,9%)	11 (3,4%)	77 (23,8%)	6 (1,8%)	156 (48,3%)
Total		91 (28,2%)	15 (4,6%)	12 (3,7%)	20 (6,2%)	171 (52,9%)	14 (4,3%)	323 (100%)

Fonte: Output do SPSS (Adaptado)

Cruzando o género e estado civil do óbito é notório que o maior número é mulheres casadas ou vivendo maritalmente (74 casos) porém este número encontra-se próximo aos do mesmo estado civil no género masculino (69 casos) que sua vez se encontra próximo ao número de casos de solteiros homens e mulheres (67 e 62 casos respectivamente), tal como ilustra a tabela 5.4.

Tabela 5.4: Cruzamento entre género e estado civil dos óbitos.

		Estado Civil do internado				Total
		Solteiro	Casado ou vivendo maritalmente	Divorciado ou separado	Viúvo	
Género do internado	Feminino	62 (19,4%)	74 (23,1%)	20 (6,3%)	10 (3,1%)	166 (51,9%)
	Masculino	67 (20,9%)	69 (21,6%)	11 (3,4%)	7 (2,2%)	154 (48,1%)
Total		129 (40,3%)	143 (44,7%)	31 (9,7%)	17 (5,3%)	320 (100%)

Fonte: Output do SPSS (Adaptado)

Isto mostra que o perfil encontra-se ligado aos estados civis (solteiro e casado ou vivendo maritalmente) e não propriamente ao cruzamento das duas variáveis.

Cruzando o grupo etário e o estado civil, nota-se um maior número de óbitos em casados ou vivendo maritalmente com idades compreendidas entre os 25 e os 49 anos (127 casos), tal como ilustra a tabela 5.5 a seguir.

Tabela 5.5: Cruzamento entre grupo etário e estado civil dos óbitos.

		Estado Civil do internado				Total
		Solteiro	Casado ou vivendo maritalmente	Divorciado ou separado	Viúvo	
Grupo etário do internado	Menos de 1 ano	91 (28,4%)	0	0	0	91 (28,4%)
	1 a 4 anos	15 (4,7%)	0	0	0	15 (4,7%)
	5 a 14 anos	12 (3,8%)	0	0	0	12 (3,6%)
	15 a 24 anos	11 (3,4%)	6 (1,9%)	2 (0,6%)	1 (0,3%)	20 (6,3%)
	25 a 49 anos	0	127 (39,7%)	28 (8,8%)	13 (4,1%)	168 (52,5%)
	50 ou mais anos	0	10 (3,1%)	1 (0,3%)	3 (0,9%)	14 (4,4%)
Total		129 (40,3%)	143 (44,7%)	31 (9,7%)	17 (5,3%)	320 (100%)

Fonte: Output do SPSS (Adaptado)

5.2.3. Vícios dos óbitos

Nesta secção pretende-se tratar sobre duas variáveis: o consumo do álcool e o ser ou não fumante.

Em relação ao consumo do álcool, dos 314 casos considerados, tem-se 179 óbitos que não consumiam o álcool e 135 que consumiam, correspondentes a 57% e 43% respectivamente (tabela 5.4, em anexo V), sendo a maioria os que não consumiam álcool.

Entretanto, o número de óbitos que não consumiam álcool é consideravelmente influenciado por indivíduos que pela natureza da sua idade não consomem ou não deviam consumir álcool. Por via disto, através da tabela que faz o cruzamento entre o grupo etário e o consumo do álcool (tabela 5.5, em anexo V), tem-se 197 indivíduos com 15 ou mais anos de idade que evoluíram para óbito dos quais 63 não consumiam álcool e 134 consumiam, havendo, neste contexto, uma maioria dos que consumiam álcool (68%).

A Figura 5.8 mostra o consumo de óbitos para os possuíam 15 ou mais anos de idade

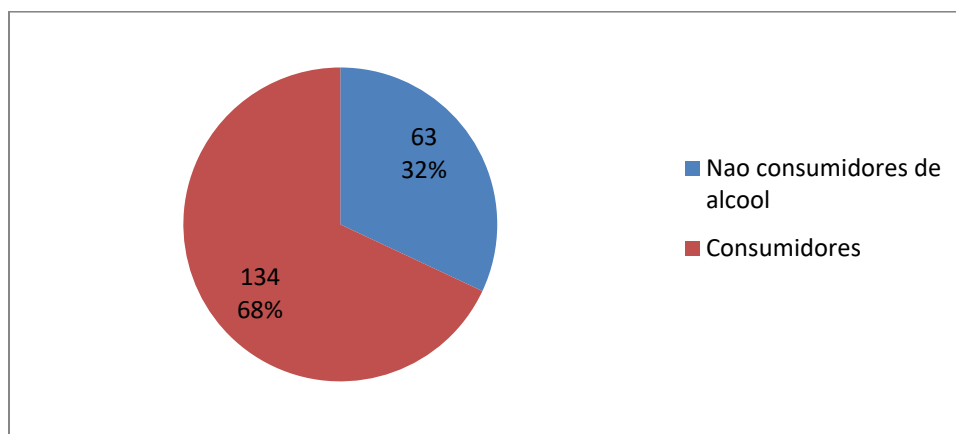


Fig.5.8: Gráfico dos óbitos consumidores ou não de álcool com 15 ou mais anos de idade

Em relação ao ser ou não fumante, dos 300 óbitos considerados, 245 não eram fumantes e 55 eram fumantes o correspondente a 81,7% e 18,3% respectivamente (tabela 5.6, em anexo V), sendo a maioria os não fumantes.

Pela mesma razão relatada acima sobre o consumo de álcool, faz-se o cruzamento entre o grupo etário e o ser o não fumante (tabela 5.7, em anexo V), tendo 184 casos que evoluíram para óbito com 15 ou mais anos idade dos quais 129 que não eram fumantes e 55 eram fumantes, havendo assim uma maioria de óbitos não fumantes (70%).

A Figura 5.9 mostra os fumantes ou não fumantes que evoluíram para óbito com 15 ou mais anos de idade

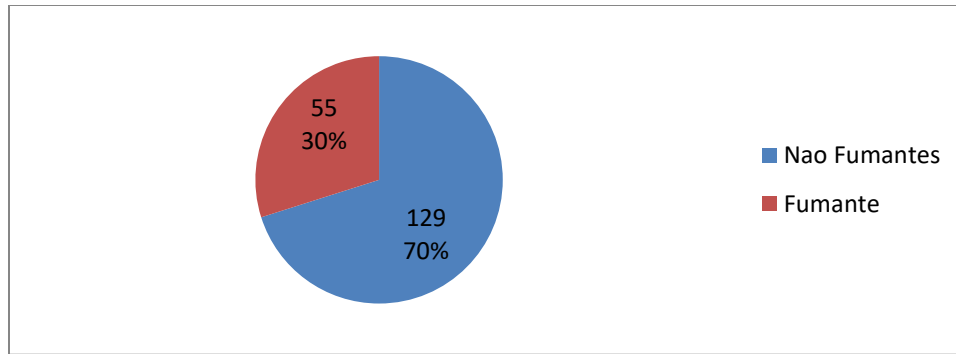


Fig.5.9: Gráfico dos óbitos fumantes ou não de álcool com 15 ou mais anos de idade

5.2.4. Outras características dos óbitos

Nesta secção, pretende-se tratar sobre as variáveis: proveniência, doenças crónicas associadas e internamentos anteriores.

Em relação a proveniência (Figura 5.10), dos 313 casos considerados, 135 são provenientes do Grande Maputo, 90 da província de Maputo, 49 da região sul e 39 são de fora da região sul correspondente a 43,1%, 28,8%, 15,7% e 12,5% respectivamente (tabela 5.8, em anexo V).

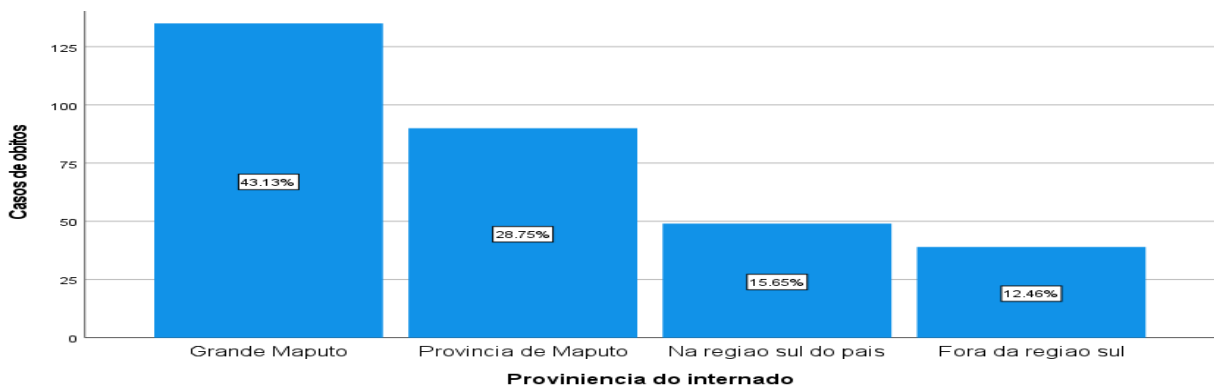


Fig.5.10: Proveniência dos óbitos

Destacam-se os óbitos provenientes do Grande Maputo seguido dos provenientes da província do Maputo.

Em relação a existência duma doença crónica associada, dos 316 casos considerados, 219 não sofriam de alguma doença crónica associada e 97 sofriam o que corresponde a 69,3% e 30,7% respectivamente (tabela 5.9, em anexo V), sendo a maioria os que não sofriam de doença crónica, tal como mostra a Figura 5.11.

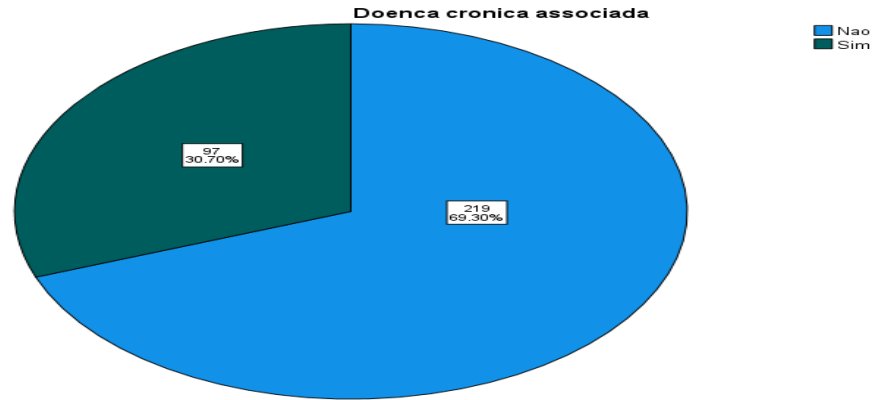


Fig.5.11: Gráfico da existência de doenças crónicas associadas

Em relação ao número de internamentos anteriores, baseado nas tabelas 5.10 e 5.11 em anexo V, verifica-se que o numero de internamentos anteriores varia de 0 á 4 e nota-se que uma média de 0.56 o que implica que em média os óbitos tiveram 0 a 1 internamento.

A moda e a mediana são ambas iguais a 0, o que denota uma maioria dos óbitos que nunca antes haviam tido um internamento; mais da metade dos óbitos nunca antes haviam sido internados (189 óbitos com 0 internamentos anteriores, correspondente à 59,4%).

Importa também referir que existiram casos consideráveis de um 1 internamento (89 casos, correspondente à 28%) e existe uma minoria de casos com 3 e 4 internamentos anteriores que apresentam 4 casos (1,3%) e 2 casos (0,6%) respectivamente. Estes últimos 6 casos são *outliers*.

Na Figura 5.12, através do histograma e do *boxplot*, visualiza-se os dados anteriormente descritos.

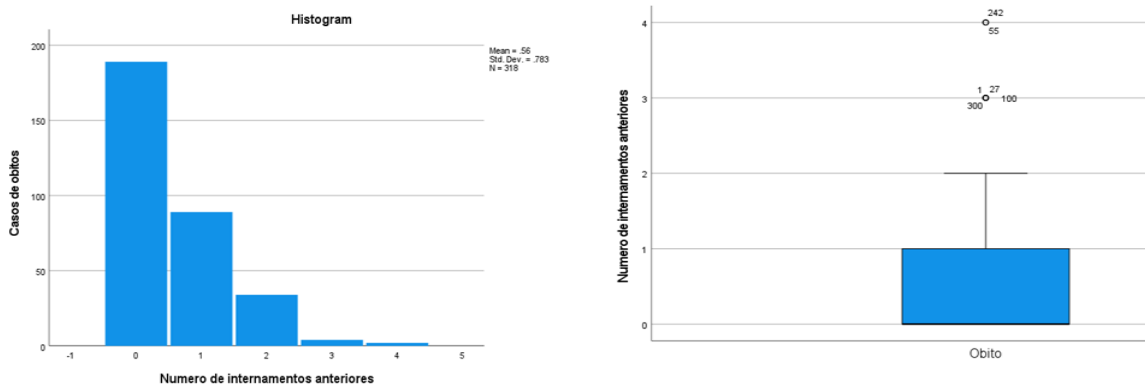


Fig.5.12: Histograma e boxplot de número de internamentos anteriores

5.3. Factores de risco da mortalidade causada por doenças infecciosas: Aplicação da regressão logística

5.3.1. Análise da multicolinearidade

Para que não exista multicolinearidade, os valores de tolerância devem ser maiores que 0.1 e os valores de VIF (factor de inflação da variância) devem ser menores que 10 (Hill & Hill, 2005 citado por Neves, 2017). A tabela 5.6 (estrato da tabela 6.1, em anexo VI) mostra os valores de tolerância e de VIF para cada variável.

Tabela 5.6: Análise de multicolinearidade

Variável	Estatística de colinearidade	
	Tolerância	VIF
Género do internado	0.982	1.018
Grupo etário do internado	0.387	2.586
Estado Civil do internado	0.510	1.962
Proveniência do internado	0.991	1.009
Consumo de álcool no internado	0.610	1.640
Fumante ou não fumante	0.894	1.119
Doença crónica associada	0.940	1.064
Número de internamentos anteriores	0.920	1.087

Na tabela 5.6 verifica-se que os valores de tolerância são maiores que 0.1 e os de VIF são menores que 10..

Por outro lado, Marôco (2011) afirma que correlações superiores a 0,9 podem indicar problemas de colinearidade. Pela tabela 6.2 (em anexo VI), todas as correlações encontram-se abaixo de 0,9.

Por conseguinte, dados estes dois aspectos, é razoável afirmar que se cumpre assim com o pressuposto de ausência de multicolinearidade

Para identificação dos factores de risco recorre-se a regressão logística. Esta regressão será feita através métodos de selecção de variáveis: Enter, *Forward LR* e *Backward LR*.

A tabela 5.7 que apresenta o resumo de casos processados.

Tabela 5.7: Resumo dos casos processados

UnweightedCases ^a		N	Percent
Selected Cases	Included in Analysis	608	88.5
	Missing Cases	79	11.5
	Total	687	100.0
Unselected Cases		0	.0
Total		687	100.0

a. If weight is in effect, see classification table for the total number of cases.

Fonte: Output do SPSS

Verifica-se que para análise da regressão logística se tem em conta 608 doentes que corresponde 88,5% dos 687 casos que compõem o estudo, neste contexto, não fazem parte desta análise 79 casos correspondente á 11,5%. A não inclusão destes 79 casos decorre devido a inexistência de informação de pelo menos uma variável independente das oito em estudo. Na tabela 5.8 apresenta-se a forma como se encontram codificadas as variáveis categóricas independentes.

Tabela 5.8: Codificação das variáveis independentes

		Frequency	Parameter coding				
			(1)	(2)	(3)	(4)	(5)
Grupo etário do internado	Menos de 1 ano	113	.000	.000	.000	.000	.000
	1 a 4 anos	70	1.000	.000	.000	.000	.000
	5 a 14 anos	81	.000	1.000	.000	.000	.000
	15 a 24 anos	129	.000	.000	1.000	.000	.000
	25 a 49 anos	175	.000	.000	.000	1.000	.000
	50 ou mais anos	40	.000	.000	.000	.000	1.000
Proveniência do internado	Grande Maputo	256	.000	.000	.000		
	Província de Maputo	170	1.000	.000	.000		
	Na região sul do país	119	.000	1.000	.000		
	Fora da região sul	63	.000	.000	1.000		
Estado Civil do internado	Solteiro	353	.000	.000	.000		
	Casado ou vivendo maritalmente	202	1.000	.000	.000		
	Divorciado ou separado	33	.000	1.000	.000		
	Viúvo	20	.000	.000	1.000		
Doença crónica associada	Não	490	.000				
	Sim	118	1.000				

Consumo de álcool no internado	Não	364	.000				
	Sim	244	1.000				
Fumante ou não fumante	Não	548	.000				
	Sim	60	1.000				
Género do internado	Feminino	316	.000				
	Masculino	292	1.000				

Fonte: Output do SPSS

Neste contexto, usou-se o contraste indicador (dado por *default* no SPSS) para a recodificação da variável e considera-se como referência a primeira categoria de cada variável.

5.3.2. Modelo sem variáveis independentes

Este modelo considera que todos indivíduos fazem parte da categoria de maior frequência em termos da variável resposta, tal como mostra a tabela 5.9.

Tabela 5.9: Classificação no modelo sem variáveis independentes

Observed		Predicted		
		Mortalidade ou sobrevivência		PercentageCorrect
		Alta	Óbito	
Mortalidade ou sobrevivência	Alta	329	0	100.0
	Óbito	279	0	.0
OverallPercentage				54.1

Fonte: Output do SPSS

Neste caso, a categoria de maior frequência é dos doentes que tiveram alta que é de 54.1%. Esta percentagem é a percentagem de acerto do modelo.

Naturalmente, há variáveis que não foram incluídas neste modelo, é importante saber se estas seriam úteis para predizer a mortalidade causada por doenças infecciosas. Neste contexto, a tabela 6.3 em anexo VI apresenta a significância das variáveis que independentes não inclusas no modelo e tem-se estatística geral comum valor- $p < 0.001$ e conseqüentemente menor que 0.05, então num nível de confiança de 95%, existe pelo menos uma variável independente que contribui significativamente no modelo.

Visto que foi estatisticamente comprovado que há variáveis independentes que contribuem a predição da mortalidade, a seguir usa-se os critérios de selecção já mencionados para identificar alguns factores de risco.

5.3.3. Critério de selecção *Enter*

O critério *Enter* inclui todas as variáveis independentes no modelo e por isso apresenta um modelo complexo, no entanto realizou-se o critério em duas fases: uma com apenas as variáveis correlacionadas a MCDI (tidas através da análise bivariada) denominada *Enter* reduzido e outra com todas as variáveis denominada *Enter* completo.

5.3.3.1. Critério *Enter* reduzido

O critério *Enter* reduzido contém apenas as 5 variáveis correlacionadas a MCDI e na sua análise são incluídas 624 observações correspondentes a 90,8% dos dados em estudo (ver tabela 7.1, em anexo VII). Este número encontra-se acima do estabelecido na tabela 5.7, visto que existem observações cujos dados em falta são das variáveis não correlacionadas (que foram excluídas neste critério), todavia para os restantes critérios são incluídas observações estabelecidas na tabela 5.7.

A tabela 5.10 apresenta o teste de rácio de verossimilhança nos passos (*step*), blocos (*block*) e no modelo final (*model*).

Tabela 5.10: Teste de rácio de verossimilhança utilizando o critério *Enter* reduzido (Omnibus)

		Chi-square	Df	Sig.
Step 1	Step	352.741	13	.000
	Block	352.741	13	.000
	Model	352.741	13	.000

Fonte: Output do SPSS

A hipótese nula deste teste diz que o modelo é não estatisticamente significativo e porque tem-se a estatística Qui-quadrado $\chi^2 = 352,741$ e um valor- $p < 0.001$ (consequentemente menor que 0.05) então rejeita-se a hipótese nula. Assim pode-se afirmar que o modelo com variáveis independentes através do critério *Enter* reduzido é significativo em relação ao modelo sem variáveis independentes, isto é, existe pelo menos uma variável independente significativa para prever a MCDI.

A tabela de resumo do modelo (tabela 7.2, em anexo VII) apresenta a estatística -2LL (-2Log *likelihood*) que é um meio para avaliar a qualidade de ajustamento.

O valor de -2LL é 508.296, entretanto não existe no SPSS uma forma directa para encontrar o p-valor sendo possível através duma transformação (ver imagem A, em anexo VII) com 594 graus

de liberdade resultante de 624-13-1 em que 624 é a amostra considerada na regressão logística e 13 são os graus de liberdade considerando as variáveis independentes, tendo-se assim obtido um valor-p de 0.995 (imagem A, em anexo VII).

Com este valor-p (> 0.05) não se rejeita a hipótese nula de que o modelo se ajusta aos dados, isto é, há evidências estatísticas de que o modelo completo do critério *Enter* reduzido se ajusta aos dados.

Em relação ao pseudo- R^2 , o SPSS fornece os valores Cox & Nell e Nalgekerke (ver tabela 7.2, em anexo VII) contudo para uma análise completa deste aspecto é necessário o Pseudo- R^2 de McFadden que de acordo com Marôco (2011) facilmente pode ser determinado do seguinte modo:

$$R_{MF}^2 = 1 - \frac{(-2LL)}{(-2LL) + \chi^2} = 1 - \frac{508,296}{508,296 + 352,741} = 0,41$$

Assim, tem-se a seguinte tabela dos pseudo- R^2 :

Tabela 5.11: Pseudos- R^2 do Critério *Enter* reduzido

Pseudo- R^2	Cox & Nell	Nalgekerke	McFadden
Valor	$R_{CS}^2 = 0,432$	$R_N^2 = 0,577$	$R_{MF}^2 = 0,41$

Os valores de pseudos- R^2 são intermediários (nem tão próximos de 1 e nem tão próximos de 0) o que sugere uma qualidade moderada do ajuste do modelo.

Na tabela 5.12 tem-se o teste de Hosmer e Lemshow apresentando a estatística $\chi_{HL}^2 = 12,393$ e um valor-p=0.134 (> 0.05) e por isso não se rejeita a hipótese nula de que os dados ajustam-se ao modelo, o que implica que os valores esperados estão próximos dos valores observados, isto é, há evidências estatísticas de que o modelo completo do critério *Enter* reduzido se ajusta aos dados.

Tabela 5.12 Teste de Hosmer e Lemeshow utilizando o critério *Enter* reduzido

Step	Chi-square	Df	Sig.
1	12.393	8	.134

Fonte: Output do SPSS

Na tabela 5.13 apresenta-se a classificação observada e a prevista do modelo ajustado.

Tabela 5.13: Classificação do modelo *Enter* reduzido

Observed		Predicted		
		Mortalidade ou sobrevivência		PercentageCorrect
		Alta	Óbito	
Mortalidade ou sobrevivência	Alta	271	66	80.4
	Óbito	50	237	82.6
OverallPercentage				81.4

Fonte: Output do SPSS

Existem 66 casos falsos positivos, isto é, são 66 doentes que tiveram alta mas o modelo considera que foram óbitos e existem 50 casos de falsos negativos, isto é, são 50 doentes que foram óbito mas o modelo considera que tiveram alta.

O modelo ajustado pelo critério *Enter* (reduzido) apresenta uma sensibilidade de 82.6%, esta é a percentagem dos óbitos correctamente classificados pelo modelo e uma especificidade de 80,4% que representa a percentagem dos doentes que tiveram alta correctamente classificados pelo modelo. Em geral, 81,4% dos doentes foram correctamente classificados.

Visto que a sensibilidade e a especificidade encontram-se acima de 80%, pelo exposto na secção 3.4.13.1, o modelo possui boas capacidades de previsão.

A tabela 5.14 apresenta os dados de análise da curva de ROC que são visualizados na Figura 5.13.

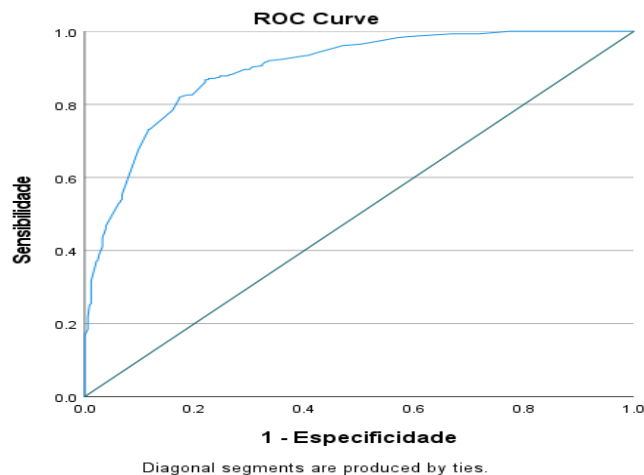


Fig.5.13: Curva de ROC do Critério Enter reduzido

Tabela 5.14: Área da Curva ROC

Test Result Variable(s): Predicted probability				
Area	Std. Error ^a	AsymptoticSig. ^b	Asymptotic 95% ConfidenceInterval	
			LowerBound	UpperBound

.896	.012	.000	.872	.920
------	------	------	------	------

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group.

Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Fonte: Output do SPSS

Tem-se uma área $A = 0,896$ que pelos critérios descritos na secção 3.4.13.2, pelo facto de estar no intervalo $0,8 \leq A < 0,9$ tem uma discriminação boa e apresenta um valor- $p < 0.001 (< 0.05)$ e por isso rejeita-se a hipótese nula de que $A = 0.5$ e aceita-se a hipótese de que $A > 0.5$. Assim, fica estatisticamente evidente que a área é significativamente superior ao valor de quebra (0.5) e apresenta uma boa capacidade de discriminar os doentes que foram óbitos dos doentes que tiveram alta.

Na tabela 5.15 tem-se os resultados da confiabilidade da classificação através do kappa de Cohen.

Tabela 5.15: Confiabilidade utilizando o critério *Enter Reduzido*

		Value	Asymptotic Standard Error ^a	Approximate T ^b	Approximate Sig nificance
Measure of Agreement	Kappa	.627	.031	15.692	.000
N of Valid Cases		624			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Fonte: Output do SPSS

Tem-se um valor- $p < 0.001$ do kappa de Cohen que é menor que 0.05 rejeitando-se assim a hipótese nula de que o kappa seja igual à 0 e aceitando a hipótese alternativa de que o kappa é maior que zero.

Nota-se ainda que o valor de kappa é 0,627, estando no intervalo de 0,61 a 0,8 e por isso, através da tabela 3.3, assume-se que existe uma confiabilidade substancial entre os valores observados e os resultados de mortalidade obtidos pelo modelo.

Já se notou que existe pelo menos uma variável independente que contribui para a previsão da MCDI, sendo necessário identificar de quais se tratam. Nesta perspectiva, a tabela 5.16 a seguir permite essa avaliação a partir do teste de Wald e da razão de chances.

Tabela 5.16: Significância das variáveis independentes pelo Critério Enter Reduzido

	B	S.E.	Wald	Df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Step 1 ^a			158.502	5	.000			
Grupo etário do internado								
Grupo etário do internado(1)	-2.990	.413	52.307	1	.000	.050	.022	.113
Grupo etário do internado(2)	-3.225	.412	61.288	1	.000	.040	.018	.089
Grupo etário do internado(3)	-3.942	.452	76.153	1	.000	.019	.008	.047
Grupo etário do internado(4)	-.810	.670	1.462	1	.227	.445	.120	1.654
Grupo etário do internado(5)	-3.304	.766	18.581	1	.000	.037	.008	.165
Estado Civil do internado			8.437	3	.038			
Estado Civil do internado(1)	.354	.605	.343	1	.558	1.425	.435	4.665
Estado Civil do internado(2)	1.755	.804	4.762	1	.029	5.781	1.196	27.949
Estado Civil do internado(3)	1.566	.892	3.084	1	.079	4.786	.834	27.468
Proveniência do internado			15.141	3	.002			
Proveniência do internado(1)	.400	.276	2.104	1	.147	1.492	.869	2.563
Proveniência do internado(2)	-.483	.313	2.387	1	.122	.617	.334	1.139
Proveniência do internado(3)	1.158	.401	8.331	1	.004	3.183	1.450	6.986
Fumante ou não fumante(1)	1.936	.453	18.291	1	.000	6.934	2.855	16.841
Doença crónica associada(1)	1.743	.307	32.240	1	.000	5.715	3.131	10.431

	Constant	.885	.254	12.106	1	.001	2.424		
--	----------	------	------	--------	---	------	-------	--	--

a. Variable(s) entered on step 1: Grupo etario do internado, Estado Civil do internado, Proveniência do internado, Fumante ou não fumante, Doença crónica associada.

Fonte: Output do SPSS

Tendo em conta as hipóteses apresentadas na secção 3.4.9, num nível de confiança de 95%, afirma-se que todas as variáveis incluídas no modelo contribuem significativamente para a mortalidade ou sobrevivência causada por doenças infecciosas nomeadamente: grupo etário ($\chi^2_{Wald} = 158.502; p < 0.001$), estado civil ($\chi^2_{Wald} = 8.437; p = 0.038$), proveniência ($\chi^2_{Wald} = 15.141; p = 0.002$), ser fumante ($\chi^2_{Wald} = 18.291; p < 0.001$), e sofrer de doenças crónicas associadas ($\chi^2_{Wald} = 32.240; p < 0.001$), pois possuem valores-p menores que 0.05.

Tendo em conta que em variáveis categóricas com mais de duas categorias tem-se estatísticas diferentes para cada categoria (diferente da categoria de referência), é de referir que:

- No grupo etário tem-se como categoria de referência os doentes com menos de 1 ano de idade e destacam-se os com 1 a 4 anos de idade ($\chi^2_{Wald} = 52.307; p < 0.001$), os com 5 aos 14 anos de idade ($\chi^2_{Wald} = 61.288; p < 0.001$), os com 15 aos 24 anos de idade ($\chi^2_{Wald} = 76.153; p < 0.001$) e os com 50 ou mais anos de idade ($\chi^2_{Wald} = 18.581; p < 0.001$);
- No estado civil tem-se como categoria de referência os solteiros e destacam-se os separados ou divorciados ($\chi^2_{Wald} = 4.762; p = 0.029$);
- Quanto á proveniência tem-se como categoria de referência os provenientes do Grande Maputo e destacam-se os provenientes de fora da região sul ($\chi^2_{Wald} = 8.331; p = 0.004$).

Importa salientar que a constante do modelo pelo critério Enter reduzido é significativa ($\chi^2_{Wald} = 8.885; p = 0.001$).

Para verificar o nível de contribuição das variáveis independentes na variável resposta, a seguir analisa-se a razão de chances (OR), que se encontra na coluna $EXP(B)$ da tabela, tendo em conta os intervalos de confiança num nível de 95%.

No grupo etário dos 1 a 4 anos tem-se $OR = 0.050$ e $IC =]0.022; 0.113[$ e assim tem-se a variação percentual da razão de chances de -95%, o que significa que a chance de um doente dos 1 a 4 anos ser óbito pelas doenças infecciosas é 95% menor que a dos doentes com menos de 1 ano de idade (categoria de referência).

No grupo etário dos 5 a 14 anos tem-se $OR = 0.04$ e $IC =]0.018; 0.089[$ e assim tem-se a variação percentual da razão de chances de -96%, o que significa que a chance de um doente dos 5 a 14 anos ser óbito é 96% menor que a dos doentes com menos de 1 ano de idade.

No grupo etário dos 15 a 24 anos tem-se $OR = 0.019$ e $IC =]0.008; 0.047[$ e assim tem-se a variação percentual da razão de chances de -98,1%, o que significa que a chance de um doente dos 15 a 24 anos ser óbito é 98,1% menor que a dos doentes com menos de 1 ano de idade.

No grupo etário dos 50 ou mais anos de idade tem-se $OR = 0.037$ e $IC =]0.008; 0.165[$ e assim tem-se a variação percentual da razão de chances de -96,3%, o que significa que a chance de um doente com 50 ou mais anos ser óbito é 96,3% menor que a dos doentes com menos de 1 ano de idade.

O estado civil separado ou divorciado apresenta $OR = 5.781$ e $IC =]1.196; 30.949[$, o que significa a chance de um separado ou divorciado ser óbito é mais que 5 vezes maior que a do solteiro (categoria de referência).

A proveniência de fora da zona sul apresenta $OR = 3.183$ e $IC =]1.45; 6.986[$, o que significa que os doentes que provem de fora da zona sul têm mais de 3 vezes maior chance de ser óbito em relação aos provenientes do Grande Maputo (categoria de referência).

O ser fumante apresenta $OR = 6.934$ e $IC =]2.885; 16.841[$, o que significa que a chance de um doente fumante ser óbito devido a doenças infecciosas é mais de 6 vezes maior que a do não fumante (categoria de referência)

A existência duma doença crónica associada apresenta $OR = 5.715$ e $IC =]3.131; 10.431[$, a chance de um doente sofrendo de pelo menos uma doença crónica ser óbito devido a doenças infecciosas é mais de 5 vezes maior que um doente que não sofre de nenhuma doença crónica.

Quando $OR=1$, as categorias não afectam a razão de chances. O estado civil divorciado e o ser fumante apresentam razões de chances (OR) com intervalos de confiança mais largos o que denota uma maior variabilidade na chance de ocorrer em relação as categorias de referencia. Contudo, tanto estas duas categorias como outras acima analisadas as OR, são significativas (valores-p < 0.05) e os intervalos de confiança não incluem $OR=1$.

Por outro lado, as outras categorias, para além de não serem significativas (valores-p > 0.05), os seus intervalos de confiança incluem $OR=1$.

O modelo de regressão logística através do critério *Enter* reduzido é:

$$\begin{aligned} \text{logit}(\hat{\pi}) = & 0,885 - 2,99GE(1) - 3,225GE(2) - 3,942GE(3) - 0,81GE(4) - 3,304GE(5) \\ & + 0,354EC(C) + 1,755EC(D) + 1,566EC(V) + 0,4P(M) - 0,483P(S) \\ & + 1,158P(FS) + 1,936F + 1,743DC. \end{aligned}$$

No modelo acima, GE representa o grupo etário em que 1, 2, 3, 4 e 5 representam os grupos de 1 a 4 anos, de 5 a 14 anos, de 15 a 24 anos, de 25 a 49 anos e os de 50 ou mais anos de idade respectivamente; EC representa o estado civil em que C, D e V representam os casados ou vivendo maritalmente, divorciados ou separados e viúvos respectivamente; P representa a proveniência em que M, S e FS representam província Maputo, região Sul e Fora da região Sul respectivamente; F representa os fumantes; DC representa a existência de doença crónica associada.

Pelo critério *Enter* reduzido tem-se como factores de risco da MCDI: ser separado ou divorciado, ser proveniente de fora da região sul, ser fumante e sofrer de uma doença crónica associada a uma doença infecciosa. Os grupos etários 1 a 4 anos, 5 a 14 anos, 15 a 24 anos e 50 ou mais anos de idade constituem factores protectores do risco de MCDI.

Para a detecção de *outliers* apresenta-se a Figura 5.14 com o gráfico dos resíduos estudentizados e estandardizados.

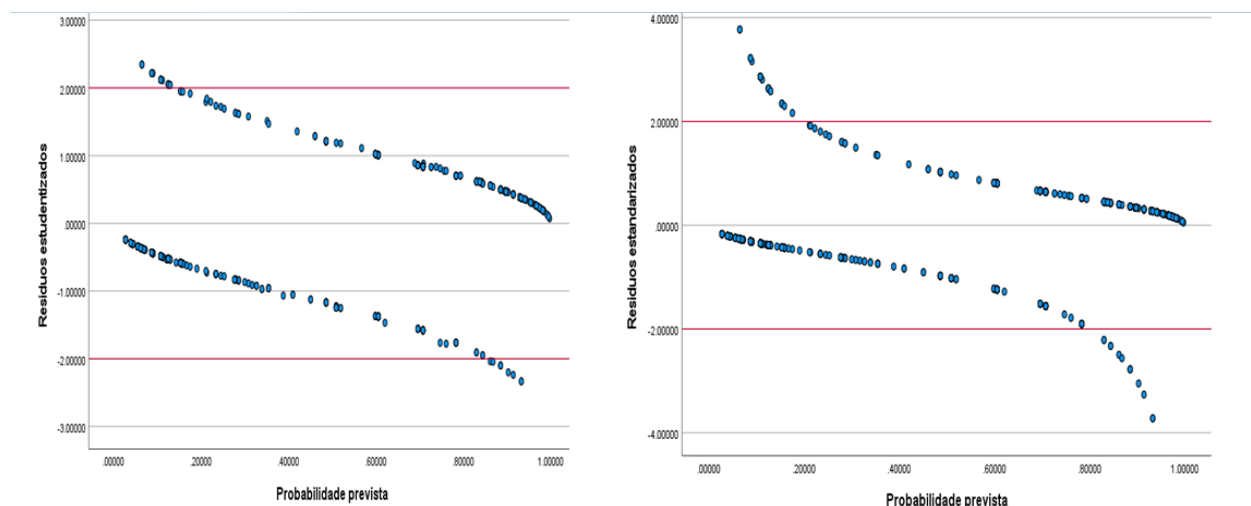


Fig.5.14: Resíduos estudentizados e estandardizados pelo Critério Enter reduzido

São candidatas a outliers as observações cujos módulos dos resíduos estandardizados ($zresid$) são maiores que 2. As mesmas observações apresentam módulos dos resíduos estudentizados ($sresid$) também maiores que 2.

As observações candidatas a *outliers* ainda podem ser verificadas na tabela 7.3 (em anexo VII). De referir que as observações que fazem parte da tabela são observações incorrectamente classificadas, isto é, são falsos negativos e falsos positivos.

Para avaliar a influência das observações no modelo recorre-se a distância de Cook, apresentada na Figura 5.15, do qual nota-se não existir nenhuma observação influente visto que todas as observações apresentam uma distância de Cook menor que 1.

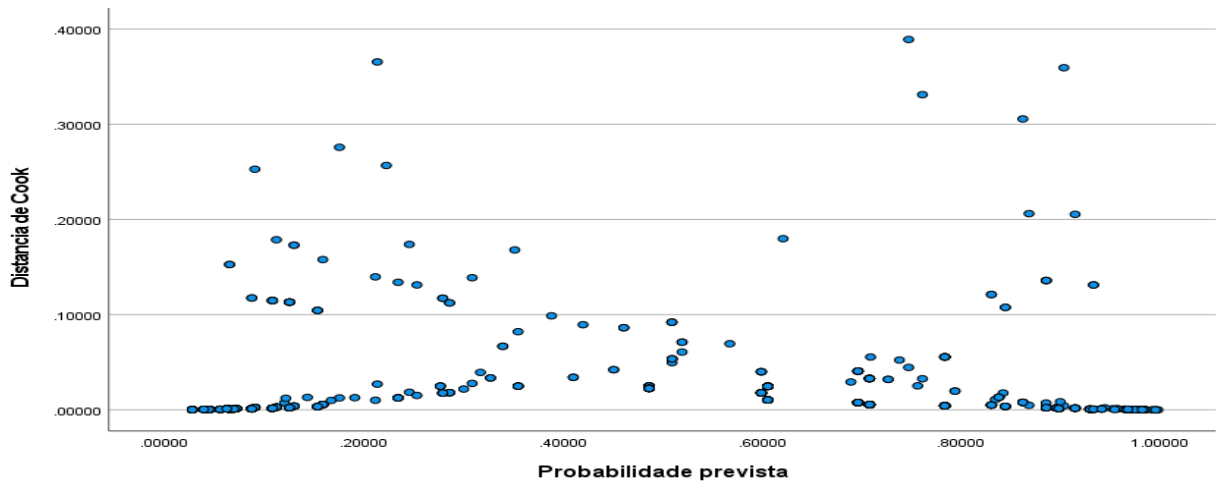


Fig.5.15: Distância de Cook pelo Critério *Enter* reduzido

Para avaliar a influência de cada observação em cada variável, tem-se em anexo os gráficos de *DfBetas*, tendo se verificado em todas as variáveis módulos de *DfBetas* menores que 1 e consequentemente menores que 2, o que mostra não existem observações influentes para nenhum dos coeficientes.

5.3.3.2. Critério de selecção *Enter* Completo

A tabela 5.17 apresenta o teste de rácio de verossimilhança nos passos (*step*), blocos (*block*) e no modelo final (*model*). O critério *Enter* inclui todas as variáveis independentes no modelo e por isso realiza a análise em apenas um passo, bloco e modelo e contém os mesmos valores.

Tabela 5.17: Teste de rácio de verossimilhança pelo Critério *Enter* completo

		Chi-square	Df	Sig.
Step 1	Step	346.518	16	.000
	Block	346.518	16	.000
	Model	346.518	16	.000

Fonte: Output do SPSS

Tem-se a estatística Qui-quadrado $\chi^2 = 346,518$ e um valor- $p < 0.001$ (< 0.05) então rejeita-se a hipótese nula de que o modelo não é significativo. Assim se pode afirmar que o modelo com variáveis independentes através do critério *Enter* completo é significativo em relação ao modelo sem variáveis independentes, isto é, existe pelo menos uma variável independente significativa para prever a MCDI.

A tabela resumo do modelo (tabela 8.1, em anexo VIII) apresenta a estatística -2LL (-2Log likelihood) cujo valor é 492.233 e de forma análoga ao modelo anterior faz-se o cálculo do valor-p (imagem B, em anexo VIII) com 591 graus de liberdade (resultante de 608-16-1 em que 608 é a amostra considerada na regressão logística e 16 são os graus de liberdade considerados nas variáveis independentes), tendo-se obtido um p-valor de 0.999

Com este valor-p (> 0.05) não se rejeita a hipótese nula de que o modelo se ajusta aos dados, isto é, há evidências estatísticas de que o modelo completo do critério *Enter* se ajusta aos dados.

De forma análoga ao modelo anterior tem-se pseudo- R^2 de McFadden:

$$R_{MF}^2 = 1 - \frac{(-2LL)}{(-2LL) + \chi^2} = 1 - \frac{492,233}{492,233 + 346,518} = 0,413$$

Assim, tem-se a tabela 5.18 com os pseudo- R^2 .

Tabela 5.18: Pseudo- R^2 do Critério *Enter* Completo

Pseudo- R^2	Cox & Nell	Nalgekerke	McFadden
Valor	$R_{CS}^2 = 0,434$	$R_N^2 = 0,581$	$R_{MF}^2 = 0,413$

Os valores de pseudo- R^2 são intermédios (nem tão próximos de 1 e nem tão próximos de 0) o que sugere uma qualidade moderada do ajuste do modelo.

Na tabela 5.19 tem-se o teste de Hosmer e Lemshow apresentando a estatística $\chi_{HL}^2 = 8.791$ e um p-valor=0.360 (> 0.05) e por isso não se rejeita a hipótese (nula) de que os dados ajustam-se ao modelo o que implica que os valores esperados estão próximos dos valores observados, isto é, há evidências estatísticas de que o modelo completo do critério *Enter* completo se ajusta aos dados.

Tabela 5.19: Teste de Hosmer e Lemshow pelo Critério *Enter* Completo

Step	Chi-square	Df	Sig.
1	8.791	8	.360

Fonte: Output do SPSS

Na tabela 5.20 apresenta-se a classificação observada e a prevista do modelo ajustado.

Tabela 5.20: Classificação pelo Critério *Enter* Completo

Observed	Predicted	
	Mortalidade ou sobrevivência	PercentageCorr

		Alta	Óbito	ect
Mortalidade ou sobrevivência	Alta	264	65	80.2
	Óbito	45	234	83.9
OverallPercentage				81.9

Fonte: Output do SPSS

Existem 65 casos falsos positivos, isto é, são 65 doentes que tiveram alta mas o modelo considera que foram óbitos e existem 45 casos de falsos negativos, isto é, são 45 doentes que foram óbito mas o modelo considera que tiveram alta.

O modelo ajustado pelo critério *Enter* completo apresenta uma sensibilidade de 83,9%, esta é a percentagem dos óbitos correctamente classificados pelo modelo e apresenta uma especificidade 80.2% que representa a percentagem dos doentes que tiveram alta correctamente classificados pelo modelo. Em geral, 81,9% dos doentes foram correctamente classificados.

Visto que a sensibilidade e a especificidade encontra-se acima de 80%, de acordo com os intervalos da secção 3.4.13.1, o modelo possui boas capacidades de previsão.

A tabela 5.21 apresenta-se os dados de análise da curva de ROC que são visualizados na Figura 5.16.

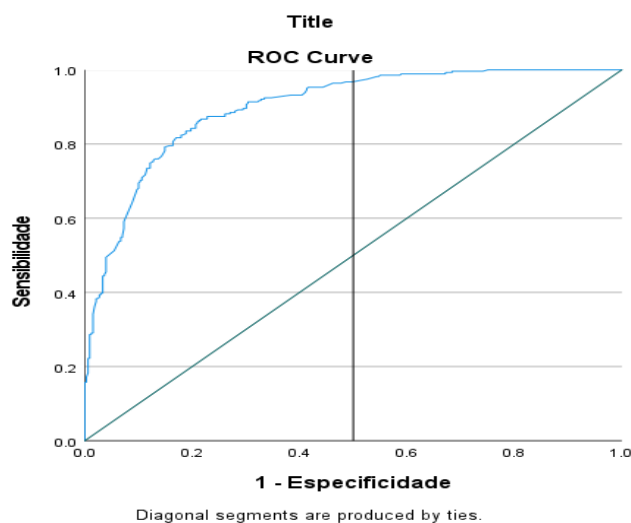


Fig.5.16: Curva ROC do Critério *Enter* Completo

Tabela 5.21: Área da Curva ROC do Critério *Enter* Completo

Test Result Variable(s): Predicted probability

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.898	.012	.000	.874	.922

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Fonte: Output de SPSS

Neste sentido, tem-se uma área $A = 0,898$ que pelos critérios descritos na secção 3.4.13.2, pelo facto de estar no intervalo $0,8 \leq A < 0,9$ tem uma discriminação boa e apresenta um valor- $p < 0.001$ (< 0.05) e por isso rejeita-se a hipótese nula de que $A = 0.5$ e aceita-se a hipótese de que $A > 0.5$. Assim, fica estatisticamente evidente que a área é significativamente superior ao valor de quebra (0.5) e apresenta uma boa capacidade de discriminar os doentes que foram óbitos dos doentes que tiveram alta.

Na tabela 5.22 tem-se os resultados de análise do coeficiente Kappa de confiabilidade.

Tabela 5.22: Confiabilidade do Critério Enter Completo

		Value	Asymptotic Standard Error ^a	Approximate T ^b	Approximate Significance
Measure of Agreement	Kappa	.638	.031	15.758	.000
N of Valid Cases		608			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Fonte: Output do SPSS

Tem-se um valor- $p < 0.001$ do kappa de Coehen que é menor que 0,05 rejeitando-se assim a hipótese nula de que o kappa seja igual à 0 e aceitando a hipótese alternativa de que o kappa é maior que zero.

Nota-se ainda que o valor de Kappa é 0,638, estando no intervalo de 0,61 a 0,8 e por isso, através da tabela 3.3, assume-se que existe uma confiabilidade substancial entre os valores observados e os resultados de mortalidade obtidos pelo modelo.

A tabela 5.23 permite avaliar a contribuição das variáveis independentes no modelo a partir do teste de Wald e da razão de chances.

Tabela 5.23: Significância das variáveis independentes pelo Critério *Enter* Completo

	B	S.E.	Wald	Df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Step 1 ^a								
Género do internado(1)	.000	.230	.000	1	1.000	1.000	.637	1.570
Grupo etário do internado			149.275	5	.000			
Grupo etário do internado(1)	-3.003	.415	52.275	1	.000	.050	.022	.112
Grupo etário do internado(2)	-3.244	.414	61.300	1	.000	.039	.017	.088
Grupo etário do internado(3)	-3.894	.511	58.055	1	.000	.020	.007	.055
Grupo etário do internado(4)	-.849	.715	1.408	1	.235	.428	.105	1.739
Grupo etário do internado(5)	-3.623	.846	18.356	1	.000	.027	.005	.140
Estado Civil do internado			8.660	3	.034			
Estado Civil do internado(1)	.406	.611	.442	1	.506	1.501	.453	4.969
Estado Civil do internado(2)	1.843	.810	5.178	1	.023	6.315	1.291	30.883
Estado Civil do internado(3)	1.630	.909	3.217	1	.073	5.104	.860	30.309
Proveniência do internado			14.123	3	.003			
Proveniência do internado(1)	.392	.282	1.933	1	.164	1.479	.852	2.569
Proveniência do internado(2)	-.461	.320	2.076	1	.150	.631	.337	1.181
Proveniência do internado(3)	1.145	.403	8.067	1	.005	3.143	1.426	6.926
Consumo de álcool no internado(1)	-.091	.332	.075	1	.784	.913	.476	1.751
Fumante ou não fumante(1)	1.936	.466	17.267	1	.000	6.932	2.781	17.278
Doença crónica associada(1)	1.749	.312	31.462	1	.000	5.749	3.120	10.593

Número de internamentos anteriores	.057	.156	.133	1	.716	1.058	.779	1.437
Constant	.889	.297	8.935	1	.003	2.432		

a. Variable(s) entered on step 1: Género do internado, Grupo etário do internado, Estado Civil do internado, Proveniência do internado, Consumo de álcool no internado, Fumante ou não fumante, Doença crónica associada, Número de internamentos anteriores.

Fonte: Output do SPSS

Através das hipóteses já apresentadas na secção 3.4.9, pelo critério *Enter* completo num nível de confiança de 95% pode-se afirmar que as variáveis que contribuem significativamente para a mortalidade ou sobrevivência por doenças infecciosas são: grupo etário ($\chi^2_{Wald} = 149.275; p < 0.001$), estado civil ($\chi^2_{Wald} = 8.660; p = 0.034$), proveniência ($\chi^2_{Wald} = 14.123; p = 0.003$), ser fumante ($\chi^2_{Wald} = 17.267; p < 0.001$), e sofrer de doenças crónicas associadas ($\chi^2_{Wald} = 31.462; p < 0.001$), pois possuem valores-p menores que 0.05.

Tendo em conta em variáveis categóricas com mais de duas categorias tem-se estatísticas diferentes para cada categoria (diferente da categoria de referência), é de referir que:

- No grupo etário tem-se como categoria de referência os doentes com menos de 1 ano de idade e destacam-se os com 1 a 4 anos de idade ($\chi^2_{Wald} = 52.275; p < 0.001$), os com 5 aos 14 anos de idade ($\chi^2_{Wald} = 61.3; p < 0.001$), os com 15 aos 24 anos de idade ($\chi^2_{Wald} = 58.055; p < 0.001$) e os com 50 ou mais anos de idade ($\chi^2_{Wald} = 18.356; p < 0.001$);
- No estado civil tem-se como categoria de referência os solteiros e destacam-se os separados ou divorciados ($\chi^2_{Wald} = 5.178; p = 0.023$);
- Quanto á proveniência tem-se como categoria de referência os provenientes do Grande Maputo destacam-se os provenientes de fora da região sul ($\chi^2_{Wald} = 8.067; p = 0.005$).

Num nível de confiança de 95% pelas mesmas hipóteses e pelo mesmo critério pode-se afirmar que as variáveis género ($\chi^2_{Wald} < 0.001; p > 0.999$), consumo de álcool ($\chi^2_{Wald} = 0.075; p = 0.784$) e o número de internamentos anteriores ($\chi^2_{Wald} = 0.133; p = 0.716$) são variáveis que não contribuem significativamente para a previsão da mortalidade causada por doenças infecciosas.

Importa salientar que a constante do modelo pelo critério *Enter* completo é significativa ($\chi^2_{Wald} = 8.935; p = 0.003$).

Para verificar o nível de contribuição das variáveis independentes na variável resposta, a seguir analisa-se a razão de chances (OR), que se encontra na coluna EXP(B) da tabela, tendo em conta os intervalos de confiança num nível de 95%.

No grupo etário dos 1 a 4 anos tem-se $OR = 0.050$ e $IC =]0.022; 0.112[$ e assim tem-se a variação percentual da razão de chances de -95%, o que significa que a chance de um doente dos 1 a 4 anos ser óbito pelas doenças infecciosas é 95% menor que a dos doentes com menos de 1 ano de idade (categoria de referência).

No grupo etário dos 5 a 14 anos tem-se $OR = 0.039$ e $IC =]0.017; 0.088[$ e assim tem-se a variação percentual da razão de chances de -96,1%, o que significa que a chance de um doente dos 5 a 14 anos ser óbito é 96,1% menor que a dos doentes com menos de 1 ano de idade.

No grupo etário dos 15 a 24 anos tem-se $OR = 0.020$ e $IC =]0.007; 0.055[$ e assim tem-se a variação percentual da razão de chances de -98%, o que significa que a chance de um doente dos 15 a 24 anos ser óbito é 98% menor que a dos doentes com menos de 1 ano de idade.

No grupo etário dos 50 ou mais anos de idade tem-se $OR = 0.027$ e $IC =]0.005; 0.140[$ e assim tem-se a variação percentual da razão de chances de -97,3%, o que significa que a chance de um doente com 50 ou mais anos ser óbito é 97.3% menor que a dos doentes com menos de 1 ano de idade.

O estado civil separado ou divorciado apresenta $OR = 6.315$ e $IC =]1.291; 30.883[$, o que significa a chance de um separado ou divorciado ser óbito é mais que 6 vezes maior que a do solteiro (categoria de referência).

A proveniência de fora da zona sul apresenta $OR = 3.143$ e $IC =]1.426; 6.926[$, o que significa que os doentes que provenham de fora da zona sul têm mais de 3 vezes maior chance de ser óbito em relação aos provenientes do Grande Maputo (categoria de referência).

O ser fumante apresenta $OR = 6.932$ e $IC =]2.781; 17.278[$, o que significa que a chance de um doente fumante ser óbito devido a doenças infecciosas é mais de 6 vezes maior que a do não fumante (categoria de referência)

A existência duma doença crónica associada apresenta $OR = 5,748$ e $IC =]3,120; 10,593[$, a chance de um doente sofrendo de pelo menos uma doença crónica ser óbito devido a doenças infecciosas é mais de 5 vezes maior que um doente que não sofre de nenhuma doença crónica.

O estado civil divorciado e o ser fumante apresentam razões de chances (OR) com intervalos de confiança mais largos o que denota uma maior variabilidade na chance de ocorrer em relação as categorias de referencia. Contudo, tanto estas duas categorias como outras acima analisadas as OR, são significativas (valores-p < 0.05) e os intervalos de confiança não incluem OR=1.

Importa referir que a variável género mostra ser irrelevante para previsão da MCPI através do critério *Enter* completo pois para além de possuir um valor-p muito acima de 0.05 (valor-p > 0.999) e OR=1 (que não afecta a razão de chances) possui também um coeficiente ínfimo e por isso esta variável é removida do modelo.

O modelo de regressão logística através do critério *Enter* completo é:

$$\begin{aligned} \text{logit}(\hat{\pi}) = & 0,889 - 3,003GE(1) - 3,344GE(2) - 3,894GE(3) - 0,849GE(4) - 3,623GE(5) \\ & + 0,406EC(C) + 1,843EC(D) + 1,63EC(V) + 0,392P(M) - 0,462P(S) \\ & + 1,145P(FS) - 0,91CA + 1,936F + 1,749DC + 0,057IA \end{aligned}$$

No modelo acima CA representa os consumidores de álcool e IA representa o número de internamentos anteriores.

Neste modelo não se inclui a variável género por ter coeficiente ínfimo.

Pelo critério *Enter* Completo tem-se como factores de risco da MCDI: ser separado ou divorciado, ser proveniente de fora da região sul, ser fumante e sofrer de uma doença crónica associada a uma doença infecciosa. Os grupos etários 1 a 4 anos, 5 a 14 anos, 15 a 24 anos e 50 ou mais anos de idade constituem factores protectores do risco de MCDI.

Para a detecção de *outliers* tem-se na Figura 5.17 o gráfico dos resíduos estudentizados e estandardizados.

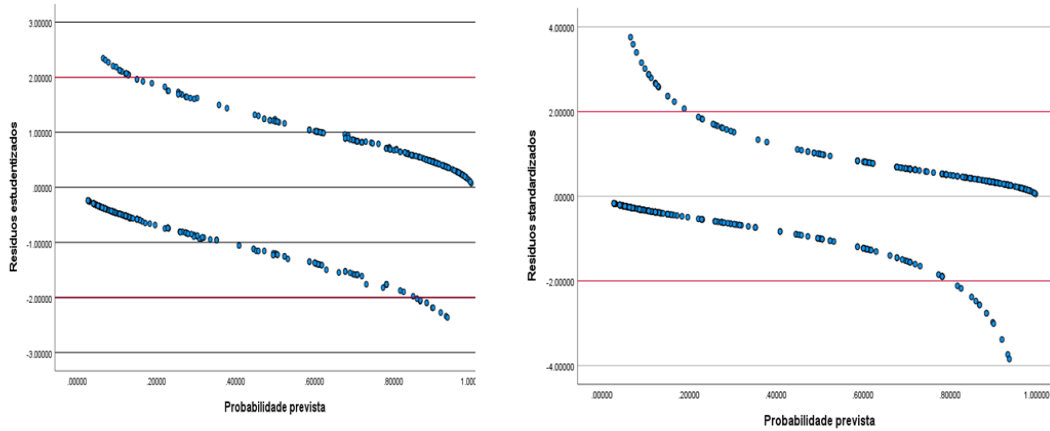


Fig.5.17: Resíduos estudentizados e estandardizados

São candidatas a *outliers* as observações que possuem os resíduos estandardizados ($zresid$) maiores que 2. As mesmas observações apresentam resíduos estudentizados ($sresid$) também maiores que 2.

As candidatas a *outliers* ainda podem ser verificadas na tabela 8.2 (em anexo VIII). De referir que as observações que fazem parte da tabela são observações incorrectamente classificadas isto é, são falsos negativos e falsos positivos.

Para avaliar a influência das observações no modelo recorre-se a distância de Cook apresentada na Figura 5.18, do qual nota-se não existir nenhuma observação influente visto que todas observações apresentam uma distância de Cook menor que 1.

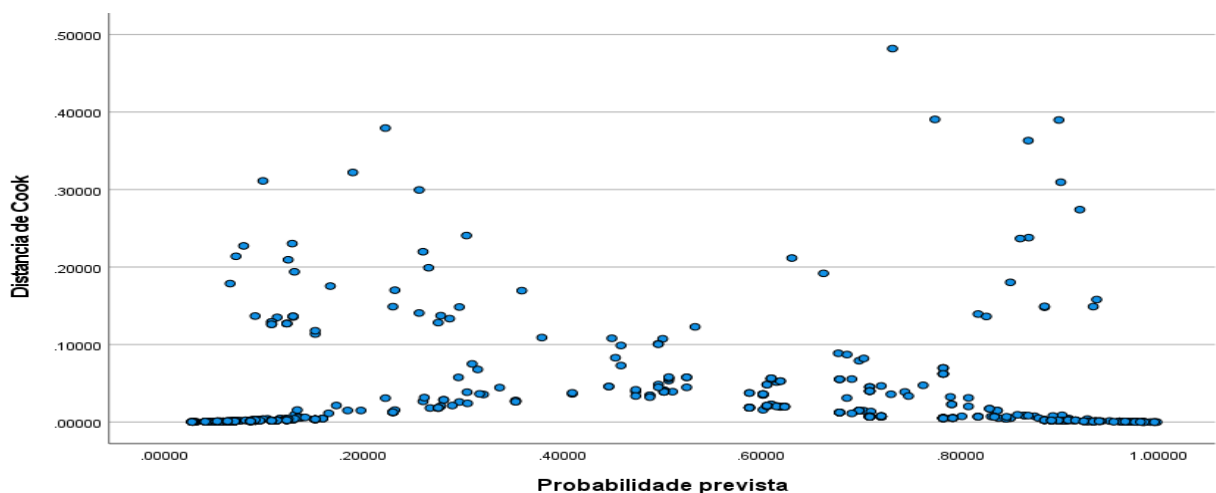


Fig.5.18: Diferença de Cook do critério *Enter* Completo

Para avaliar a influência de cada observação em cada variável, tem-se em anexo VIII os gráficos de *DfBet*as, tendo se verificado em todas variáveis módulos de *DfBet*as menores que 1 e conseqüente menores que 2, o que mostra não existem observações influentes para nenhum dos coeficientes.

5.3.4. Critério de selecção *Forward LR*

Trata-se do critério *forward* baseado no rácio de verossimilhança que descreve a sua análise em etapas, sendo a última etapa a que melhor descreve o modelo incluindo todas as variáveis significativas.

Neste caso, pelo critério tem-se o melhor modelo na quinta etapa, tal como mostra a tabela 5.24 do teste do rácio de verossimilhança a seguir, sendo assim a quinta etapa é a etapa de referência neste critério.

Tabela 5.24: Teste de rácio de verossimilhança (Omnibus) do Critério *Forward LR*

		Chi-square	Df	Sig.
Step 1	Step	261.895	5	.000
	Block	261.895	5	.000
	Model	261.895	5	.000
Step 2	Step	36.642	1	.000
	Block	298.537	6	.000
	Model	298.537	6	.000
Step 3	Step	24.724	1	.000
	Block	323.260	7	.000
	Model	323.260	7	.000
Step 4	Step	13.566	3	.004
	Block	336.826	10	.000
	Model	336.826	10	.000
Step 5	Step	9.486	3	.023
	Block	346.312	13	.000
	Model	346.312	13	.000

Fonte: Output do SPSS

Na quinta etapa tem-se $\chi^2 = 9,486$ para a etapa e $\chi^2 = 346,312$ para o bloco e para o modelo e nestes três aspectos apresenta p-valores menores que 0,05. Assim rejeita-se a hipótese nula de que o modelo não é estatisticamente significativo, sendo assim o modelo com variáveis independentes através do critério *Forward LR* é significativo em relação ao modelo sem variáveis independentes, isto é, existe pelo menos uma variável independente significativa para predizer a MCDI.

A tabela do resumo do modelo (tabela 9.1, em anexo IX) apresenta a estatística -2LL (-2Log *likelihood*) cujo valor é de 492.439 e de modo análogo aos critério anteriores tem-se o valor-p de 0.999 (> 0.05) e assim não se rejeita a hipótese nula de que o modelo se ajusta aos dados, isto é, há evidências estatísticas de que o modelo completo do critério *Forward LP* se ajusta aos dados. Analogamente aos critérios anteriores, calcula-se o pseudo- R^2 de McFadden:

$$R_{MF}^2 = 1 - \frac{(-2LL)}{(-2LL) + \chi^2} = 1 - \frac{492,439}{492,439 + 346,312} = 0,413$$

Na tabela do resumo do modelo já existem os outros dois pseudos- R^2 e assim tem-se a tabela 5.25.

Tabela 5.25: Pseudos- R^2 do Critério *Forward LR*

Pseudo- R^2	Cox & Nell	Nalgekerke	McFadden
Valor	$R_{CS}^2 = 0,434$	$R_N^2 = 0,580$	$R_{MF}^2 = 0,413$

Os valores dos pseudo- R^2 são intermédios o que sugere uma qualidade moderada do ajuste do modelo.

Na tabela 5.26 tem-se o teste de Hosmer e Lemshow apresentando a estatística $\chi_{HL}^2 = 12,044$ e um valor-p=0,149 (> 0.05) e por isso não se rejeita a hipótese nula de que os dados ajustam-se ao modelo o que implica que os valores esperados estão próximos dos valores observados, isto é, há evidências estatísticas de que o modelo completo do critério *Forward LR* se ajusta aos dados.

Tabela 5.26: Teste de Hosmer e Lemshow do Critério *Forward LR*

Step	Chi-square	Df	Sig.
1	.000	4	1.000
2	2.931	6	.818

3	.896	6	.989
4	6.672	8	.572
5	12.044	8	.149

Fonte: Output do SPSS

Na tabela 5.27 apresenta-se a classificação observada e a prevista do modelo ajustado.

Tabela 5.27: Classificação do Critério *Forward LR*

Observed		Predicted		
		Mortalidade ou sobrevivência		PercentageCorrect
		Alta	Óbito	
Mortalidade ou sobrevivência	Alta	268	61	81.5
	Óbito	52	227	81.4
OverallPercentage				81.4
Mortalidade ou sobrevivência	Alta	265	64	80.5
	Óbito	45	234	83.9
OverallPercentage				82.1
Mortalidade ou sobrevivência	Alta	266	63	80.9
	Óbito	43	236	84.6
OverallPercentage				82.6
Mortalidade ou sobrevivência	Alta	261	68	79.3
	Óbito	42	237	84.9
OverallPercentage				81.9
Mortalidade ou sobrevivência	Alta	267	62	81.2
	Óbito	50	229	82.1
OverallPercentage				81.6

Fonte: Output do SPSS

Existem 62 doentes que o modelo considera que tiveram alta mas foram óbitos (falsos positivos) e existem 50 doentes que o modelo considera que foram óbito mas tiveram alta (falsos negativos).

O modelo ajustado pelo critério *Forward LR* apresenta uma sensibilidade de 82,1%, esta é a percentagem dos óbitos correctamente classificados pelo modelo e apresenta uma especificidade 81,2% que representa a percentagem dos doentes que tiveram alta correctamente classificados pelo modelo. Em geral, 81,6% dos internados foram correctamente classificados.

Pelo exposto na secção 3.4.13.1, já que a sensibilidade e a especificidade são maiores que 80%, pode-se afirmar que o modelo apresenta boa capacidade preditiva.

A tabela 5.28 apresenta os resultados de análise da curva de ROC que são visualizados na figura 5.19.

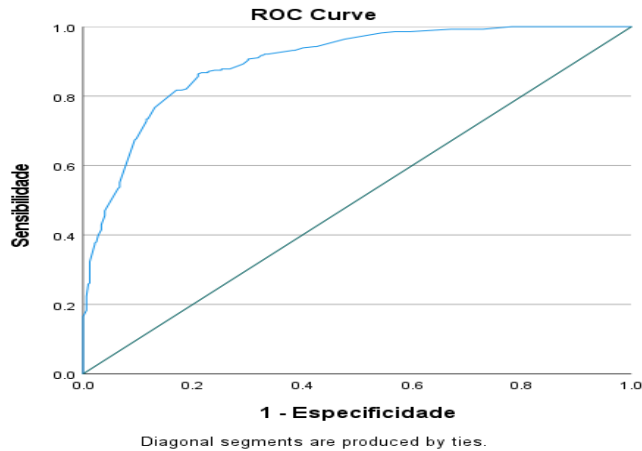


Fig.5.19: Curva ROC do Critério *Forward LR*

Tabela 5.28: Área da Curva ROC do Critério *Forward LR*

Test Result Variable(s): Predicted probability

Area	Std. Error ^a	AsymptoticSig. ^b	Asymptotic 95% Confidence Interval	
			LowerBound	UpperBound
.898	.012	.000	.874	.922

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group.

Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Fonte: Output do SPSS

Neste sentido, tem-se uma área $A = 0.898$ que encontra-se no intervalo $0,8 \leq A < 0,9$ e por isso pelos critérios descritos na secção 3.4.13.1 o modelo apresenta uma descrição boa e já que apresenta um valor- $p < 0,001 (< 0,05)$, rejeita-se a hipótese nula de que $A = 0,5$ e aceita-se a hipótese de que $A > 0,5$.

Assim, fica estatisticamente evidente que a área é significativamente superior ao valor de quebra (0.5) e apresenta uma boa capacidade de discriminar os doentes que foram óbitos dos doentes que tiveram alta.

Na tabela 5.29 tem-se os dados de análise do coeficiente Kappa de confiabilidade.

Tabela 5.29: Confiabilidade do critério *Forward LR*

		Value	Asymptotic Standard Error ^a	Approximate T ^b	Approximate Significance
Measure of Agreement	Kappa	.630	.032	15.553	.000
N of Valid Cases		608			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Fonte: Output do SPSS

Tem-se um valor-p < 0.001 do kappa de Coehen que é menor que 0,05 rejeitando-se assim a hipótese nula de que o kappa seja igual á 0 e aceitando a hipótese alternativa de que o kappa é maior que zero.

Nota-se ainda que o valor de Kappa é 0,630, estando no intervalo de 0,61 a 0,8 e por isso, através da tabela 3.3, assume-se que existe uma confiabilidade substancial entre os valores observados e os resultados de mortalidade obtidos pelo modelo.

A tabela 5.30 permite a avaliação das variáveis independentes que contribuem significativamente para o modelo a partir do teste de Wald e da razão de chances.

Tabela 5.30: Significância das variáveis independentes pelo Critério *Forward LR*

	B	Wald	Sig.	Exp(B)	IC (95%) de Exp(B)	
					Limite inferior	Limite superior
Grupo etário do internado		157.116	.000			
Grupo etário do internado(1)	-3.007	52.511	.000	.049	.022	.112
Grupo etário do internado(2)	-3.244	61.495	.000	.039	.017	.088
Grupo etário do internado(3)	-3.928	74.357	.000	.020	.008	.048
Grupo etário do internado(4)	-.864	1.631	.202	.421	.112	1.587
Grupo etário do internado(5)	-3.598	20.746	.000	.027	.006	.129
Estado Civil do internado		8.594	.035			
Estado Civil do internado(1)	.383	.394	.530	1.466	.444	4.841

Estado Civil do internado(2)	1.818	5.081	.024	6.158	1.268	29.911
Estado Civil do internado(3)	1.612	3.167	.075	5.012	.849	29.582
Proveniência do internado		14.037	.003			
Proveniência do internado(1)	.383	1.858	.173	1.466	.846	2.541
Proveniência do internado(2)	-.452	2.016	.156	.637	.341	1.187
Proveniência do internado(3)	1.141	8.050	.005	3.131	1.423	6.888
Fumante ou não fumante(1)	1.935	17.276	.000	6.923	2.780	17.239
Doença crónica associada(1)	1.752	31.753	.000	5.763	3.134	10.598
Constant	.906	12.187	.000	2.476		

A tabela acima é o extracto da tabela 9.2 em anexo IX, contendo os principais elementos de análise da quinta etapa.

O critério *Forward LR* apresenta apenas as variáveis que num nível de 95% de confiança contribuem significativamente para a mortalidade ou sobrevivência causada pelas doenças infecciosas, sendo elas: grupo etário ($\chi^2_{Wald} = 157.116; p < 0.001$), estado civil ($\chi^2_{Wald} = 8.594; p = 0.035$), proveniência ($\chi^2_{Wald} = 14.037; p = 0.003$), ser fumante ($\chi^2_{Wald} = 17.276; p < 0.001$), e sofrer de doenças crónicas associadas ($\chi^2_{Wald} = 31.753; p < 0.001$).

Para variáveis categóricas com mais de duas categorias, é de referir que:

- No grupo etário, que tem como referência os doentes com menos de 1 ano de idade, destacam-se os com 1 a 4 anos de idade ($\chi = 52.511; p < 0.001$), os com 5 aos 14 anos de idade ($\chi^2_{Wald} = 61.495; p < 0.001$), os com 15 aos 24 anos de idade ($\chi^2_{Wald} = 74.357; p < 0.001$) e os com 50 ou mais anos de idade ($\chi^2_{Wald} = 120.746; p < 0.001$);
- No estado civil, que se tem como categoria de referência os solteiros, destacam-se os separados ou divorciados ($\chi^2_{Wald} = 5.081; p = 0.024$);
- Quanto á proveniência, que se tem como categoria de referência os provenientes do Grande Maputo, destacam-se os provenientes de fora da região sul ($\chi^2_{Wald} = 8.050; p = 0.005$).

Para verificar o nível de contribuição das variáveis independentes na variável resposta, a seguir analisa-se a razão de chances (OR), que encontra-se na coluna Exp(B) da tabela, tendo em conta os intervalos de confiança num nível de 95%.

No grupo etário dos 1 a 4 anos tem-se $OR = 0.049$ e $IC =]0.022; 0.112[$ e assim tem-se a variação percentual da razão de chances de -95.1%, o que significa que a chance de um doente dos 1 a 4 anos ser óbito pelas doenças infecciosas é 95.1% menor que a dos doentes com menos de 1 ano de idade (categoria de referência).

No grupo etário dos 5 a 14 anos tem-se $OR = 0.039$ e $IC =]0.017; 0.088[$ e assim tem-se a variação percentual da razão de chances de -96,1%, o que significa que a chance de um doente dos 5 a 14 anos ser óbito é 96,1% menor que a dos doentes com menos de 1 ano de idade.

No grupo etário dos 15 a 24 anos tem-se $OR = 0.020$ e $IC =]0.008; 0.048[$ e assim tem-se a variação percentual da razão de chances de -98%, o que significa que a chance de um doente dos 15 a 24 anos ser óbito é 98% menor que a dos doentes com menos de 1 ano de idade.

No grupo etário dos 50 ou mais anos de idade tem-se $OR = 0.027$ e $IC =]0.006; 0.129[$ e assim tem-se a variação percentual da razão de chances de -97,3%, o que significa que a chance de um doente com 50 ou mais anos ser óbito é 97,3% menor que a dos doentes com menos de 1 ano de idade.

O estado civil separado ou divorciado apresenta $OR = 6.158$ e $IC =]1.268; 29.911[$, o que significa a chance de um separado ou divorciado ser óbito é mais que 6 vezes maior que a do solteiro (categoria de referência).

A proveniência de fora da zona sul apresenta $OR = 3.131$ e $IC =]1.423; 6.888[$, o que significa que os doentes que provenham de fora da zona sul têm mais de 3 vezes maior chance de ser óbito em relação aos provenientes do Grande Maputo (categoria de referência).

O ser fumante apresenta $OR = 6.923$ e $IC =]2.780; 17.239[$, o que significa que a chance de um doente fumante ser óbito devido a doenças infecciosas é mais de 6 vezes maior que a do não fumante (categoria de referência).

A existência duma doença crónica associada apresenta $OR = 5.763$ e $IC =]3.134; 10.598[$, a chance de um doente sofrendo de pelo menos uma doença crónica associada ser óbito devido a doenças infecciosas é mais de 5 vezes maior que um doente que não sofre de nenhuma doença crónica.

O estado civil divorciado e o ser fumante apresentam razões de chances (OR) com intervalos de confiança mais largos o que denota uma maior variabilidade na chance de ocorrer em relação às categorias de referência. Contudo, tanto estas duas categorias como outras acima analisadas as OR, são significativas (valores-p < 0.05) e os intervalos de confiança não incluem OR=1.

O modelo de regressão logística através do critério *Forward LR* é:

$$\begin{aligned} \text{logit}(\hat{\pi}) = & 0,906 - 3,007GE(1) - 3,244GE(2) - 3,928GE(3) - 0,864GE(4) - 3,598GE(5) \\ & + 0,383EC(C) + 1,818EC(D) + 1,612EC(V) + 0,383P(M) - 0,452P(S) \\ & + 1,141P(FS) + 1,935F + 1,752DC. \end{aligned}$$

Pelo critério *Forward LR*, tem-se como factores de risco da MCDI: ser separado ou divorciado, ser proveniente de fora da região sul, ser fumante e sofrer de uma doença crónica associada a uma doença infecciosa. Os grupos etários 1 a 4 anos, 5 a 14 anos, 15 a 24 anos e 50 ou mais anos de idade constituem factores protectores do risco de MCDI.

Para a detecção de *outliers* tem-se, a seguir, o gráfico dos resíduos estudentizados e estandardizados.

Todas as observações candidatas a outliers são incorrectamente classificadas (falsos positivos ou falsos negativos).

Na Figura 5.20 tem-se os gráficos dos resíduos estudentizados e estandardizados.

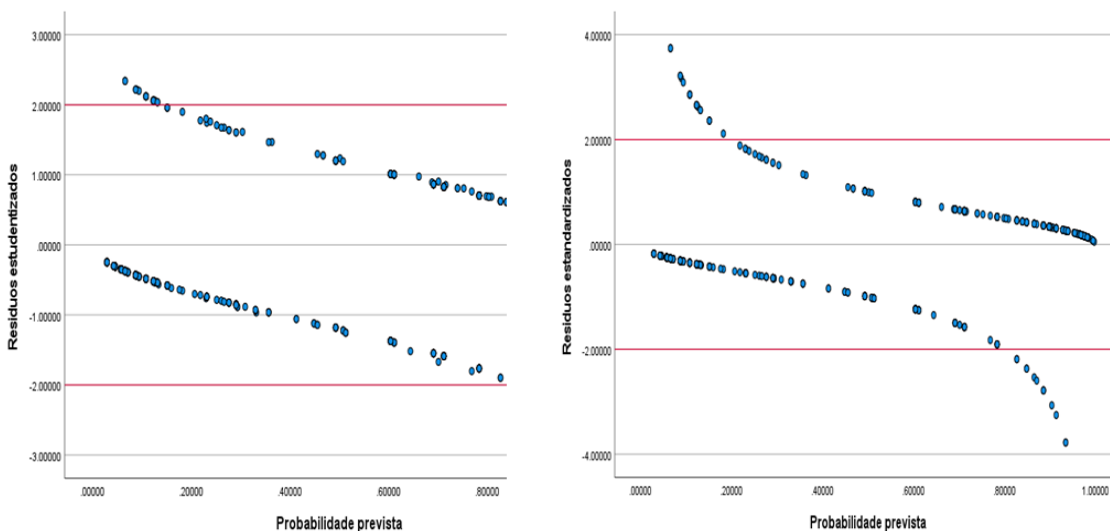


Fig.5.20: Resíduos estudentizados e estandardizados

São candidatas a outliers as observações que possuem os resíduos estandardizados ($zresid$) maiores que 2. As mesmas observações apresentam resíduos estudentizados ($sresid$) também maiores que 2.

As candidatas a *outliers* ainda podem ser verificadas na tabela 9.3 (em anexo IX). De referir que as observações que fazem parte da tabela são observações incorrectamente classificadas isto é, são falsos negativos e falsos positivos.

Para avaliar a influência das observações no modelo recorre-se à distância de Cook apresentada na Figura 5.21, do qual nota-se não existir nenhuma observação influente visto que todas observações apresentam uma distância de Cook menor que 1.

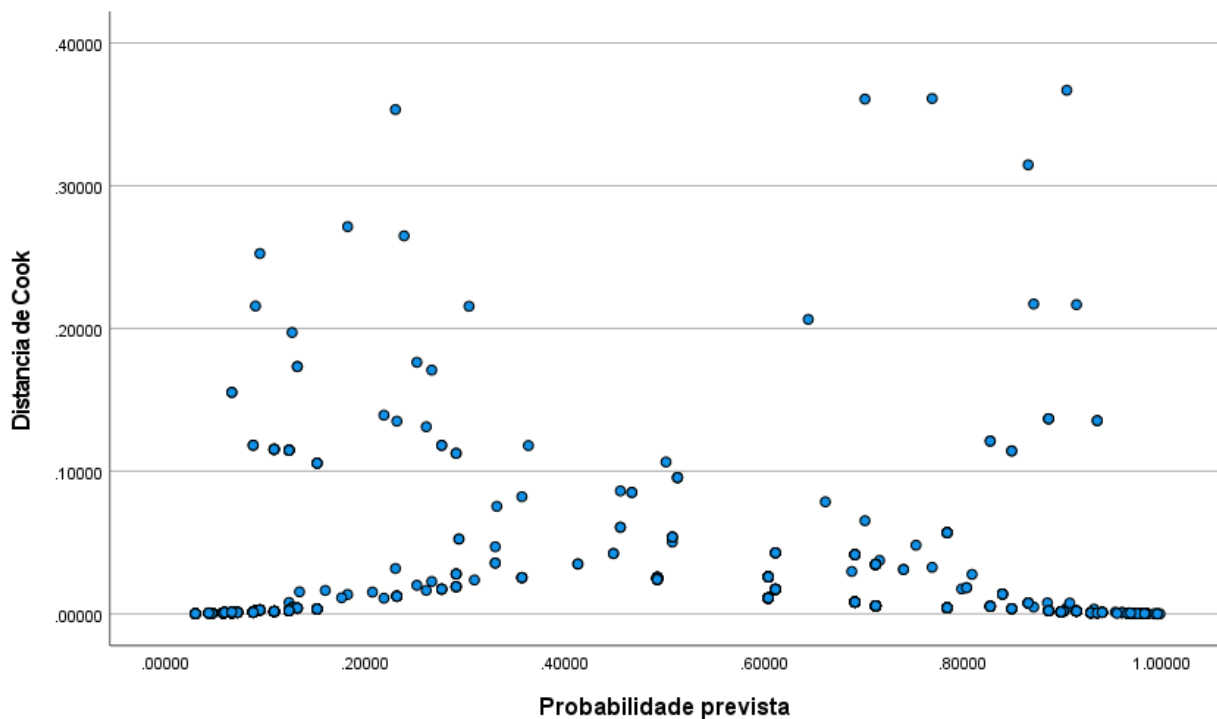


Fig.5.21: Distância de Cook do critério *Forward LR*

Para avaliar a influência de cada observação em cada variável, tem-se em anexo os gráficos de $DfBetas$, tendo se verificado em todas variáveis módulos de $DfBetas$ menores que 1 e consequentemente menores que 2, o que mostra não existem observações influentes para nenhum dos coeficientes.

5.3.5. Critério *Backward LR*

Este critério possui os mesmos resultados que os do critério *Forward LR* diferenciando-se apenas na etapa em que os critérios descrevem melhor o modelo, sendo a quinta etapa para o critério *Forward LR* e quarta etapa para o critério *Backward LR*.

Por conseguinte, a análise feita para o critério *Forward LR* é válida para o critério *Backward LR*. Os resultados do modelo utilizando o Critério *Backward LR* encontram-se em anexo X.

5.4. Discussão dos resultados

Com base no triénio analisado, o perfil da mortalidade causada por doenças infecciosas é marcado por uma TM de 24% sendo a malária a doença com a maior TM (40%) entre as doenças analisadas.

Este perfil é marcado ainda pelo género feminino (51,7%), idades entre 25 e 49 anos (52,9%), estado civil casado (44,7%), consumidores de álcool (57% em geral e 68% para internados com 15 ou mais anos de idade), não fumantes (81,7% em geral e 70% para internados com 15 ou mais anos de idade), ser proveniente do Grande Maputo (43,1%), não sofrer de nenhuma doença crónica (63,9%) e pelos que nunca tiveram um internamento anteriormente (59,1%).

Contudo, apesar de o perfil ser importante pois descreve as frequências de algumas características apresentadas pelos doentes que evoluem para óbitos, é imprescindível a determinação dos factores de risco da mortalidade, dado que o facto de uma característica ter maior frequência em óbitos não significa ser um factor de risco.

Por exemplo os não fumantes são a maioria dos óbitos porém ser fumante constitui um factor de risco da MCDI em relação aos não fumantes; ser proveniente do Grande Maputo é a categoria de proveniência de maior frequência entre os óbitos, porém os doentes provenientes das regiões centro e norte constituem um factor de risco em relação aos provenientes do Grande Maputo.

Anteriormente, na secção 5.3, foram utilizados os critérios *Enter*, *Forward LR* e *Backward LR* da regressão logística para a selecção das variáveis significativas na determinação de factores de risco da MCDI.

O critério *Enter* foi realizado em duas fases: *Enter* reduzido, em que se tem apenas as variáveis correlacionadas com a mortalidade (assinaladas através da análise bivariada) e *Enter* completo, que apresenta todas as variáveis em estudo.

O critério *Forward LR* e *Backward LR* apresentam resultados iguais, diferenciando-se apenas na etapa de referência (sendo quinta e quarta respectivamente).

O critério *Enter* quando aplicado às mesmas observações dos critérios *Forward LR* e *Backward LR* e nas variáveis correlacionadas a mortalidade (que são as mesmas consideradas como significativas nas últimas etapas dos critérios *Forward LR* e *Backward LR*) apresenta resultados iguais a estes dois critérios (ver anexo XI).

No entanto, o critério *Enter* reduzido, descrito na secção 5.3.3.1, apresenta as variáveis correlacionadas com a MCDI, em que as observações consideradas estão em número maior que as dos outros critérios (por razões anteriormente explicadas) e por isso apresenta uma diferença ligeira em relação aos critérios *Forward LR* e *Backward LR* nos seus resultados.

Estes factos mostram que independentemente do critério usado, para o presente estudo, os resultados são iguais desde que se considere as mesmas observações e as mesmas variáveis (correlacionadas com a mortalidade).

Neste sentido, visto que o critério *Enter* por sua natureza coloca todas as variáveis no modelo trazendo um modelo complexo em relação a outros critérios, mostra-se importante realizar este modelo em fases considerando primeiro um com apenas as variáveis correlacionadas obtidas através da análise bivariada.

Os critérios adoptados, em geral, apresentam estatísticas próximas e que denotam o mesmo significado.

Em relação aos *outliers* e observações influentes registou-se em todos os critérios observações candidatas a *outliers* (recorrendo a resíduos estudentizados e estandardizados) entretanto não houve nenhuma observação influente sob ponto de vista geral dos modelos (recorrendo à distância de Cook) e nem sob ponto de vista das variáveis (recorrendo a *dfBetas*).

Para indicação do critério que teve melhor desempenho, tem-se a seguir a tabela 5.31 que resume os principais elementos para cada critério. Os elementos de análise considerados são: razão de verossimilhança, sensibilidade, especificidade, curva ROC, Kappa de Cohen e número de variáveis.

Tabela 5.31: Resumo dos elementos de análise dos critérios da regressão logística

Elemento de análise	Critério <i>Enter</i> reduzido	Critério <i>Enter</i> Completo	Critérios <i>Forward LR</i> e <i>Backward LR</i>
Razão de verossimilhança	508,296	492,233	492,439
Sensibilidade	82,6%	83,9%	82,1%
Especificidade	80,4%	80,2%	81,2%
Classificação do modelo	81,4%	81,9%	81,6%
Curva ROC	0,896	0,898	0,898
Kappa de Cohen	0,627	0,638	0,630
Número de variáveis independentes	5	8	5

A sensibilidade, a especificidade, a classificação do modelo, a área da curva ROC e o kappa de Cohen apresentam valores muito próximos e quanto à interpretação encontram-se em intervalos com o mesmo significado.

Assim, os modelos em estudo apresentam boas capacidades preditivas e de discriminação dos doentes que foram óbitos e os que tiveram alta, e uma confiabilidade substancial.

Há diferenças consideráveis no valor de razão de verossimilhança (-2LL) e no número de variáveis. No critério *Enter* reduzido, apesar de ter-se um número reduzido de variáveis (5), tem-se o valor da -2LL consideravelmente maior que dos outros critérios e quanto maior o valor pior é o ajustamento.

O critério *Enter* completo possui um valor de -2LL menor que todos, entretanto possui uma diferença muito ligeira a dos critérios *Forward LR* e *Backward LR* e porque estes últimos

apresentam um modelo mais simples (5 variáveis independentes) então considera-se estes critérios os de melhor desempenho.

Neste contexto, elege-se como melhores critérios de selecção *Forward LR* e *Backward LR*, tendo como modelo:

$$\begin{aligned} \text{logit}(\hat{\pi}) = & 0,906 - 3,007GE(1) - 3,244GE(2) - 3,928GE(3) - 0,864GE(4) - 3,598GE(5) \\ & + 0,383EC(C) + 1,818EC(D) + 1,612EC(V) + 0,383P(M) - 0,452P(S) \\ & + 1,141P(FS) + 1,935F + 1,752DC. \end{aligned}$$

Com uma amostra externa de 181 doentes, aplicando o modelo acima tem-se uma área da curva de ROC de 0.875 e significativa (p-valor < 0.05) o que significa que o modelo realizou uma discriminação boa dos que tiveram alta dos que foram óbito.

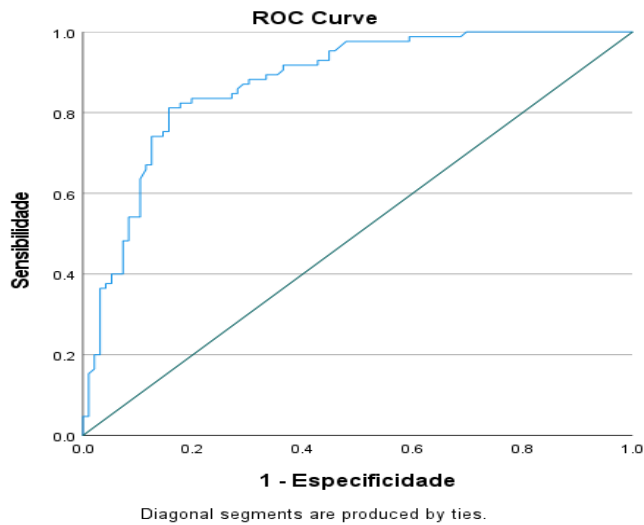


Fig.5.22: Curva ROC da validação externa

Tabela 5.32: Área da curva ROC da validação externa

TestResultVariable(s): Probabilidade prevista				
Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.875	.026	.000	.824	.925

The test result variable(s): Probabilidade prevista has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Fonte: Output do SPSS

Na tabela 5.33, nota-se que na amostra externa, dos 96 doentes que tiveram alta, 79 foram correctamente classificados pelo modelo eleito e dos 85 doentes que foram óbitos, 70 foram correctamente classificados. Assim, tem-se uma percentagem de acerto das altas de 82,3% e dos óbitos de 82,4%.

Em geral, dos 181 casos, 149 foram correctamente classificados, havendo assim um acerto de 82,3% .

Tabela 5.33: Previsão de modelo de eleição na amostra externa

		Mortalidade prevista na amostra externa		Total	Percentagem de acerto
		Alta	Obito		
Mortalidade na amostra externa	Alta	79	17	96	82,3%
	Obito	15	70	85	82,4%
Total		94	87	181	82,3%

Fonte: Output do SPSS. (Adaptado)

Fica evidente que pelos critérios analisados em geral e pelos critérios LR em particular os factores de risco da mortalidade causada por doenças infecciosas são: ser do estado civil separado ou divorciado (em relação aos solteiros), ser proveniente das regiões centro e norte de Moçambique (em relação aos provenientes do Grande Maputo), ser fumante e sofrer de pelo menos uma doença crónica.

Estar nos grupos etários 1 a 4 anos, 5 a 14 anos, 15 a 24 anos e 50 ou mais anos de idade (tendo como referência doentes com menos de 1 ano de idade) é significativo para correr menos risco da MCDI.

CAPÍTULO VI:
CONCLUSÕES E
CONSIDERAÇÕES FINAIS

6. Conclusões e considerações finais

6.1. Conclusões

O presente estudo procurou construir um perfil da mortalidade causada por doenças infecciosas bem como identificar factores de risco sob o mesmo fenómeno tendo como recursos principais de análise a Estatística Descritiva e a Regressão Logística.

As doenças analisadas neste estudo foram malária, tuberculose e HIV/SIDA no triénio 2020-2022 tendo o HIV/SIDA sido a doença com maior taxa de mortalidade (40%) e a malária com a menor taxa de mortalidade (4,6%). Em termos de anos houve diferenças ligeiras, no entanto teve-se a maior taxa TM em 2022 (25,7%) e a menor em 2021 (20,7%). Em geral há uma taxa de mortalidade causada por doenças infecciosas de 24%.

O género feminino (51,7%), os grupos etários 25 a 49 anos de idade (59,2%) e os com menos de 1 ano de idade (28,2%), os estados civis casado ou vivendo maritalmente (44,7%) e os solteiros (40,3%) constituem características biométricas que se destacam para os casos de óbitos. Cruzando as variáveis, destacam-se os casados ou vivendo maritalmente com 25 a 49 anos de idade e mulheres no mesmo grupo etário.

Em termos de vícios para os óbitos, considerando os maiores de 15 anos, destacam-se os consumidores de álcool (68%) e os não fumantes (70%).

Os provenientes do Grande Maputo (43,1%) e de outros pontos da província do Maputo (28,8%), os que sofriam de pelo menos uma doença crónica (69,3%) destacam-se entre os óbitos. Destacam-se ainda, os que nunca foram internados anteriores (59,4%) e os que foram internados uma vez anteriormente (28%) tendo uma média de número de internamentos anteriores de 0,56.

No entanto, o perfil da MCDI apesar de ser importante, este não é suficiente no âmbito de tomadas de decisões visto que existem categorias que se destacam no perfil da mortalidade, todavia não constituem factores de risco.

Para identificar factores de risco da MCDI realizou-se a regressão logística, usando os critérios: *Enter*, *Forward LR* e *Backward LR*. O critério *Enter* realizou-se em duas fases: uma com apenas as variáveis correlacionadas a MCDI (*Enter* reduzido) e outra com todas as variáveis em estudo

(*Enter* completo). Os critérios *Forward LP* e *Backward LP* apresentam os mesmos resultados e por isso são analisados simultaneamente.

Todos os modelos apresentam boas capacidades preditivas e de discriminação dos óbitos dos que tiveram alta, confiabilidade substancial e ajustamento adequado.

No entanto, por ser mais simples (com 5 variáveis) e apresentar razão de verosimilhança relativamente menor, elegeu-se o modelo gerado através dos critérios *Forward LR* e *Backward LR*:

$$\begin{aligned} \text{logit}(\hat{\pi}) = & 0,906 - 3,007GE(1) - 3,244GE(2) - 3,928GE(3) - 0,864GE(4) - 3,598GE(5) \\ & + 0,383EC(C) + 1,818EC(D) + 1,612EC(V) + 0,383P(M) - 0,452P(S) \\ & + 1,141P(FS) + 1,935F + 1,752DC. \end{aligned}$$

No modelo acima, feita uma validação externa, apresenta uma área significativa da curva de ROC com o valor 0,875 e em 82,3% dos acerta o desfecho (alta ou óbito), o que consolida a o desempenho do modelo.

Pelos critérios usados na presente pesquisa em geral e pelo critério *Forward LR/Backward LR* em particular tem-se os seguintes factores de risco da mortalidade causada por doenças infecciosas no Hospital Central de Maputo: ser do estado civil separado ou divorciado (em relação aos solteiros), ser proveniente das regiões centro e norte de Moçambique (em relação aos provenientes do Grande Maputo), ser fumante e sofrer de pelo menos uma doença crónica.

Os doentes nos grupos etários 1 a 4 anos, 5 a 14 anos, 15 a 24 anos e 50 ou mais anos de idade mostram-se significativamente com menor risco de MCDI em relação aos que têm menos de 1 ano de idade.

A análise bivariada mostrou previamente as variáveis correlacionadas a MCDI que posteriormente foram confirmadas pela regressão logística e por isso a análise bivariada pode ser considerada como um ponto de partida para a identificação dos factores de risco.

O critério *Enter* reduzido apresentou óptimos indicadores e na maioria dos casos próximos aos doutros critérios, para além de produzir um modelo mais simples em relação ao critério *Enter* completo. Entretanto, o modelo *Enter* reduzido analisado na secção 5.3.3.1 apresenta mais observações em relação aos outros critérios.

Quando o critério *Enter* é realizado nas mesmas observações doutros critérios e com as variáveis correlacionadas, este apresenta os mesmíssimos resultados com os dos critérios *Forward LR* e *Backward LR* em apenas uma etapa.

6.2.Considerações finais

No presente estudo verificaram-se algumas lacunas em algumas observações por falta de informação e ainda teve que se recorrer aos processos de doentes (protegida a sua identidade) para a recolha de dados o que pode fazer com que haja dados incompletos e demanda mais tempo de recolha. Por via disto, sugere-se ao HCM que crie uma base de dados com dados dos doentes para facilitar a recolha e análise dos mesmos, tanto numa perspectiva clínica como para estudos futuros.

Sugere-se que o HCM tenha em atenção os doentes que apresentam os factores considerados de risco no presente estudo e que faça um acompanhamento da possível mudança da taxa de MCDI.

Em relação a estudos futuros propõe-se pesquisas na mesma linha com mais variáveis que se mostrem pertinentes e/ou com as mesmas variáveis mas com maior discriminação das categorias. Por exemplo no consumo de álcool considerem os níveis de consumo, no ser fumante considerem níveis e/ou o tipo de substância consumida.

Verificou-se que a categoria viúvo do estado civil apresentou um nível de significância não muito acima de 0,05 (0,075 no critério de eleição e não mais que 0.08 noutros critérios) o que num nível de confiança de 90% consideraria uma categoria significativa. Por via disto, propõe-se para estudos futuros juntar-se os separados ou divorciados na mesma categoria com os viúvos e considerar a categoria como “já esteve casado” e assim fazer uma comparação com o presente estudo.

Notou-se que a análise bivariada previamente identificou as variáveis significativas relacionadas com o risco da MCDI. Por via disto, propõe-se para estudos futuros uma análise profunda de em que medida a análise bivariada pode ser usada para determinação de preditores e ainda a comparação e contribuição na regressão logística.

Os critérios *Enter* com variáveis correlacionadas (consideradas as mesmas observações dos modelos *Forward LR* e *Backward LR*), *Forward LR* e *Backward LR* neste estudo apresentam

resultados iguais, neste contexto propõe-se para estudos futuros uma comparação profunda entre estes critérios e as circunstâncias em que estes apresentam os mesmos resultados.

Referências bibliográficas

- Agência de Informação de Moçambique (2023, 28 de Novembro). *Doenças infecciosas continuam fardo pesado para África*. <https://aimnews.org/2023/11/28/doencas-infecciosas-continuam-fardo-pesado-para-africa/>
- Almeida, H. (2014). *Análise da regressão múltipla com estudo relacionado a horas de máquinas paradas na linha de produção de uma indústria de calçados*. [Monografia de Bacharelato, Universidade Estadual da Paraíba, Brasil].
- Alunos do Mestrado em Administração da UNIFACS (ADM UNIFACS 2016). *A natureza da análise de regressão*. https://madminifacs.files.wordpress.com/2026/04/metodod-quantilectures_1.pdf
- Alves, N.; Borgatto, A.; Pacheco, L. & Wangenheim, C. (2021). *Existem concordância e confiabilidade na Avaliação da Criatividade de resultados tangíveis da aprendizagem de Computação na Educação básica?*. Sociedade Brasileira de Computação. <https://sol.sbc.org.br/index.php/educomp/article/view/14467>
- Amaral, G; Silva, V. & Reis, E. (2009). *Análise de regressão linear no pacote R*. Instituto de Ciências Exactas, Universidade Federal de Minas gerais.
- Aprender Estatística Fácil (2024). *Entendendo as premissas do teste Qui-quadrado de Independência*. <https://estaticafacil.org/premissas-do-teste-qui-quadrado/>
- Assumpção, A; Fonseca, B & Sampaio, N(2018). *Estatística Descritiva*. Poisson
- Barreto, V; Batsela, C; Gaioto, M. & Simoes, D. (2016). *Regressão linear múltipla aplicada ao preço do leite*. <https://www.fc.unesp.br/Home/Departamentos/Matematica/revistacqd2228/vv07a10-regressao-linear-multipla.pdf>
- Barrios, D. (2021). *Preferência pelo tipo de parto em Portugal, uma abordagem multivariada* [Dissertação de Mestrado, Universidade Aberta de Portugal]
- Biblioteca Virtual de Saúde (2016). *HIV e Aids*. <http://bvsmms.gov.br/hiveaids>
- Biblioteca Virtual de Saúde (2019). *Malária*. <http://bvsmms.gov.br/malaria>
- Biblioteca Virtual de Saúde (2020). *Tuberculose*. <http://bvsmms.gov.br/tuberculose>
- Cabral, C. (2013). *Aplicações do modelo de regressão logística num estudo de mercado*. [Dissertação de Mestrado, Universidade de Lisboa].

- Camacho, A.; Obadias, F.; Manuele, S. & Polana, D. (2022). *Relatório de mortalidade intra-hospitalar e suas causas em Moçambique, 2018-2021*.
- Camacho, F. (2013). *Medidas de concordância para variáveis categorizadas*. [Dissertação de Mestrado, Universidade da Madeira, Portugal]
- Carvalho, J. (2022). *Identificação de alguns factores para obesidade e visualização*. [Dissertação de Mestrado, Universidade Aberta de Portugal].
- Cerqueira, R. (2016). *Factores preditores de complicações após transplantes renais: Análise estatística multivariada*. [Dissertação de Mestrado, Universidade Aberta de Portugal].
- Chein, F. (2019). *Introdução aos modelos lineares*. Escola Nacional Pública, Brasil.
- Correa, A.; Queiroz, E. & Trevisan, N. (s.d). *Teste qui-quadrado*. Universidade Federal de Paraná, Brasil.
- Correia, S. (2013). *Probabilidade e Estatística I* (2ª ed). Pontifícia Universidade Católica de Minas Gerais.
- Cunha, A. (2023). *Modelagem estatística de dados*. <https://estatjr.com.br/2023/08/11/modelagem-estatistica-de-dados-que-e-quais-sao-os-tipos-e-como-e-feita/>
- Dados Mundiais (2022). *Crescimento da população de Moçambique*. In <https://www.dadosmundiais.com/africa/mocambique/cresimento-populacao.php>
- Dutra, R. & Franco, C. (2020). *Modelos matemáticos em epidemiologia e aplicação na evolução da Covid-19 no Brasil e no Estado de Paraíba*. Educação, ciência e Saúde. <http://dx.doi.org/10.20438/ecs.v7il.269>
- Ferreira, D. (2012). *Regressão logística*. Universidade Aberta de Portugal.
- Freitas, E. C & Prodanov, C. C (2013). *Metodologia do trabalho científico: Métodos e Técnicas da Pesquisa do Trabalho Acadêmico*(2ª ed.). Novo Hamburgo: Editora Feevale..
- Gil, A. C (2008). *Métodos e técnicas de pesquisa social*, 6ª ed. São Paulo.
- Gonsalves, A. (2013). *Regressão logística aplicada a preditores de morte*. [Dissertação de Mestrado, Universidade de Coimbra, Portugal]
- Gonzalez, L. (2018). *Regressão logística e suas aplicações*. [Monografia de Graduação, Universidade Federal de Maranhão].
- Hosmer, D. & Lemshow, S. (2000). *Applied logist regression* (2ª ed). John Wiley & Sons.

Instituto nacional de Saúde Moçambique (2023, 19 de Maio). *Lançados Relatórios sobre causas de morte em Moçambique*. In <https://ins.gov.mz/lancados-relatorios-sobre-causas-de-morte-em-mocambique/>

Internacional Business Machines (IBM, s.d). IBM SPSS Statistics Base V27.

Landis, J & Koch (1977). *The Measurement of Observer Agreement for Categorical Data*. *Biometrics*. <https://doi.org/10.2307/2529310>

Lima, E. (s.d). *Modelos lineares generalizados*. Universidade Federal de Minas Gerais.

Lira, S. (2004). *Análise da correlação: Uma abordagem teórica e de construção dos coeficientes com aplicações*. [Diseertacao de Mestrado, Universidade federal do Paraná, Brasil].

Lopes, J; Ribeiro, A; Leal, G; Leitao, J; Veloso, M; Uchoa, I; Elias, J; & Pinto, E. (2023). *Factores associados mortalidade em unidades de terapia intensiva no Piauí*. <https://acervomais.com.br/index.php/medico/article/view/11858>

Luz, M. & Kufner, M. (2008). *Comparação entre a análise discriminante e regressão logística como método de melhor resultado na verificação de factores que influenciam na evasão dos alunos do Curso de Estatística da UFPR*. Universidade Federal do Paraná. http://www.leg.ufpr.br/lib/exe/fetch.php/disciplinas:ce229:tcc2008_marcelo_marcos.pdf

Macuvele, B; Simango, M & Jose, E. (2018). *Avaliação do perfil de pacientes que evoluíram para óbito no sector PNCTL do C.S Mariem Ngouaby entre Março de 2013 a Março de 2015*. Instituto Nacional de Ciências de Saúde de Moçambique.

Marôco (2011). *Análise estatística com o SPSS Statistics (5ª ed)*. Edições Silabo.

Matos, D. (2014). *Confiabilidade e concordância entre juízes: Aplicações na área educacional*. <http://www.fcc.org.br/pesquisa/publicacoes/eae/arquivos/1947/1947.pdf>

McCullagh, P. & Nelder, J. (1989). *Generalized Linear Models*. Chapman and Hall.

models. *Journal of the Royal Statistical Society*.

Moura, G. (2018). *Regressão de logística aplicada a análise de risco de crédito*. [Monografia, Universidade Federal do Rio Grande, Brasil].

Moura, G. (2018). *Regressão logística aplicada a análise de risco de crédito*. Universidade Federal do Rio Grande do Sul. https://imef.furg.br/images/stories/Monografias/Matematica_aplicada/2018/2018_Gabriela.pdf

Nelder, J., & Wedderburn, R. (1972). *Generalized Linear Models*. *Journal of the Royal Statistical Society*.

Neves, G. (2017). *Estado de Hidratação e Compressibilidade das pregas cutâneas*. [Dissertação de mestrado, Universidade Aberta de Portugal].

Nhachungue, S; Monjane, S; Mulungo, A; Kante, A; Aveika, A; Amozou, A; Macicame, I. (2022). *Perfil da mortalidade na região centro de Moçambique*. Instituto Nacional de Saúde de Moçambique.

Organização das Nações Unidas (2016). *Guia sobre desenvolvimento sustentável*.

Paes, A. & Silva (2012). *Por dentro da estatística: Teste de Concordância Kappa*. <http://apps.einstein.br/revista/arquivos/PDF/2715-165-166.pdf>

Pontificia Universidade Rio de Janeiro (PUC-Rio, s.d). *Regressão*. https://www.maxwell.vrac.puc-rio.br/9133/9133_3.PDF

Rodrigues, S. (2012). *Modelo de regressão linear e suas aplicações*.<https://ubibliorum.ubi.pt/bitstream/10400.6/1869/1/Tese%252520Sandra%252520Rodrigues.pdf>

S. a (2013). *Cálculo de Probabilidade e Estatística*.

Santiago, M. (2016). *Métodos de estimação da fiabilidade e concordância entre avaliadores*. [Dissertação de Mestrado, Universidade Aveiro, Portugal].

Santos, D. (2019). *Os determinantes da mortalidade por tuberculose e TB-HIV no Sul do Brasil: da abordagem espacial a análise de sobrevivência*. [Tese de Doutorado, Universidade de Nova Lisboa-Portugal]

Santos, V. S. *O que é infecção?* Brasil Escola. <http://brasilecola.uol.com.br/0-que-e/biologia/o-que-e-infeccao.htm>.

Silva, K. (s.d) *Cuidado no Controle de Infecção Hospitalar*. EFIVEST

Silva, M; Souza, W; Santos, M, Nunes, A & Costa, K (2023). *Perfil epidemiológico da malária no sudeste do Pará*.<https://acervomais.com.br/index.php/saude/article/download/11508/7064/>

Sondré, U. (2007). *Modelos matemáticos*. Universidade Estadual de Londrina.

Sousa, M. (2017). *Perfil da mortalidade das doenças infecciosas e parasitárias no Maranhão no período de 2003 a 2014*. [Monografia de graduação, Universidade Federal do Maranhão, Brasil]

Turkman, S. & Silva, G. (2000). *Modelos lineares generalizados: da teoria á pratica*. Universidade Técnica de Lisboa.

Uetela, D.; Cassu, A.; Monteiro, J. & Juga, A. (2018). *Determinantes da infecção por HIV em meninas adolescentes e mulheres jovens entre os 15 e 24 anos em Moçambique*. Instituto Nacional da Saúde.

Vaz, F. (2020). *Modelação linear e extensões: Aplicação da regressão logística no estudo de câncer da mama*. [Dissertação de Mestrado, Universidade Aberta de Portugal].

Vieira, L; Muando, A; Nyamula, U; Hoel, R & Sorensen, R. (2018). *Perfil clínico de pacientes HIV positivos internados nas Enfermarias do Hospital Central da Beira (HCB) com desfecho clínico de óbito de Janeiro a Junho de 2015*. Instituto Nacional de Ciências de Saúde de Moçambique.

ANEXOS

ANEXO I
DOCUMENTOS
TRATADOS PARA A
OBTENÇÃO DE DADOS

CARTA DE COBERTURA DO HOSPITAL CENTRAL DO MAPUTO



HOSPITAL CENTRAL DE MAPUTO
DIRECÇÃO CIENTÍFICA E PEDAGÓGICA

Ao
Comité Institucional de Bioética para a Saúde
Faculdade de Medicina/HCM
Maputo

Ref: nº 311.024 / DCP/HCM/24

Maputo, aos 30 de Maio de 2024

Assunto: Carta de Cobertura

O Hospital Central de Maputo autoriza a realização do trabalho de investigação intitulado: "Análise estatística: Perfil e factores de risco da mortalidade causada por doenças infecciosas em Maputo", a decorrer no Departamento Planificação e Estatística do Hospital Central de Maputo, cujo o autor é o Sr. Fernando Andre Fernando Manjate, e indica como ponto focal do HCM- Dr. Emerson Guilande.

Solicitamos a V. apreciação e aprovação Ética.

Saudações Académicas.

A Directora Científica e Pedagógica



Prof. Doutora Cesaltina Lorenzoni
(Médica Patologista MSc, MPH, PhD)

SMARL/DCP/27/05/2024

Hospital Central de Maputo. Av. Agostinho Neto 1164. Tel/fax 21320827/8

AUTORIZAÇÃO ÉTICA DA PESQUISA



Comité Institucional de Bioética em Saúde da
Faculdade de Medicina/Hospital Central de
Maputo



(CIBS FM&HCM)

Dr. Vasco António Muchanga, Presidente do Comité Institucional de Bioética em Saúde da Faculdade de
Medicina/Hospital Central de Maputo (CIBS FM&HCM)

CERTIFICA

Que este Comité avaliou a proposta do (s) Investigador (es) Principal (is):

Nome (s): **Fernando André Fernando Manjate**

Protocolo de investigação: **Sem versão, sem data**

Cosentimentos informados: **Sem versão, sem data**

Do estudo:

TÍTULO: "Análise estatística: Perfil e factores de risco da mortalidade causada por doenças infecciosas em Maputo"

Consta que:

1º Após revisão do protocolo pelos membros do comité durante a reunião do dia 05 de Outubro de 2024 e que será incluída na acta 24/2024, o CIBS FM&HCM, emite este informe notando que não há nenhuma inconveniência de ordem ética que impeça o início do estudo.

2º Que a revisão realizou-se de acordo com o Regulamento do Comité Institucional da FM&HCM emenda 2 de 28 de Julho de 2014.

3º Que o protocolo está registado com o número **CIBS FM&HCM/87/2024**.

4º Que a composição actual do CIBS FM&HCM está disponível na secretária do Comité.

5º Não foi declarado nenhum conflito de interesse pelos membros do CIBS FM&HCM.

6º O CIBS FM&HCM faz notar que a aprovação ética não substitui a aprovação científica nem a autorização administrativa.

7º A aprovação terá validade de 1 ano, até 10 de Outubro de 2025. Um mês antes dessa data, o Investigador deve enviar um pedido de renovação se necessitar.

8º Recomenda-se aos investigadores que mantenham o CIBS informado do decurso do estudo no mínimo uma vez ao ano.

9º Solicitamos aos investigadores que enviem no final de estudo um relatório dos resultados obtidos

E emite

RESULTADO: **APROVADO**

Vasco António Muchanga
Assinado em Maputo aos 07 de Outubro de 2024



ANEXO II:
DADOS DE ALTAS OBITOS
POR ANO E DIAGNOSTICO

Tabela 2.1: Altas por ano

HOSPITAL CENTRAL DE MAPUTO			
DEPARTAMENTO DE PLANIFICAÇÃO E ESTATÍSTICA			
DADOS ANUAIS			
TOTAL DE ALTAS POR DIAGNÓSTICO E POR ANO			
Diagnóstico	2020	2021	2022
Malária	401	387	463
HIV	724	460	301
Tuberculose	251	277	254

Fonte: HCM

Tabela 2.2: Óbitos por ano

HIV 2020						HIV 2021						HIV 2022					
Código	Doença	Masculino	Feminino	Indetermi	total	Código	Doença	Masculino	Feminino	total	Código	Doença	Masculino	Feminino	total		
B20	Doença pé	44	40	0	84	B20	Doença pé	27	23	50	B20	Doença pé	78	83	161		
B22	Doença pé	10	6	0	16	B22	Doença pé	4	3	7	B22	Doença pé	4	3	7		
B23	Síndrome	3	3	0	6	B23	Doença pé	4	2	6	B23	Doença pé	4	3	7		
B24	Doença pé	169	150	0	319	B24	Doença pé	115	79	194	B24	Doença pé	66	66	132		
TUBERCULOSE 2020						TUBERCULOSE 2021						TUBERCULOSE 2022					
A15	Tuberculo	8	4	0	12	A15	Tuberculo	3	0	3	A15	Tuberculo	4	1	5		
A16	Tuberculo	1	1	0	2	A16	Tuberculo	5	2	7	A16	Tuberculo	2	4	6		
A17	Meningite	0	1	0	1	A18	Tuberculo	4	0	4	A17	Tuberculo	1	0	1		
A18	Tuberculo	2	1	0	3	A19	Tuberculo	0	3	3	A18	Tuberculo	2	1	3		
A19	Tuberculo	3	1	0	4						A19	Tuberculo	4	3	7		
MALÁRIA 2020						MALÁRIA 2021						MALÁRIA 2022					
B50	Malária p	6	3	0	9	B50	Malária p	4	4	8	B50	Outras for	7	9	16		
B52	Malária p	0	1	0	1	B54	Malária n	8	4	12	B52	Malária p	1	0	1		
B54	Malária n	1	6	0	7						B54	Malária n	2	3	5		
											B58	Toxoplasn	0	1	1		

Fonte: HCM

ANEXO III: ANÁLISE UNIVARIADA

Tabela 3.1: Mortalidade ou sobrevivência

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Alta	364	53.0	53.0	53.0
	Obito	323	47.0	47.0	100.0
	Total	687	100.0	100.0	

Fonte: Output do SPSS

Tabela 3.2: Género

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Feminino	355	51.7	51.7	51.7
	Masculino	332	48.3	48.3	100.0
	Total	687	100.0	100.0	

Fonte: Output do SPSS

Tabela 3.3: Grupo etário

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Menos de 1 ano	118	17.2	17.2	17.2
	1 a 4 anos	78	11.4	11.4	28.5
	5 a 14 anos	85	12.4	12.4	40.9
	15 a 24 anos	144	21.0	21.0	61.9
	25 a 49 anos	215	31.3	31.3	93.2
	50 ou mais anos	47	6.8	6.8	100.0
	Total	687	100.0	100.0	

Fonte: Output do SPSS

Tabela 3.4: Estado civil

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Solteiro	376	54.7	56.2	56.2
	Casado ou vivendo maritalmente	235	34.2	35.1	91.3

	Divorciado ou separado	37	5.4	5.5	96.9
	Viuvo	21	3.1	3.1	100.0
	Total	669	97.4	100.0	
Missing	System	18	2.6		
	Total	687	100.0		

Tabela Proveniência 3.5 : proveniência do internado

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Grande Maputo	286	41.6	42.9	42.9
	Provincia de Maputo	190	27.7	28.5	71.4
	Na regioao sul do pais	127	18.5	19.0	90.4
	Fora da regioao sul	64	9.3	9.6	100.0
	Total	667	97.1	100.0	
Missing	System	20	2.9		
	Total	687	100.0		

Fonte: Output do SPSS

Tabela 3.6: Consumo de álcool

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Nao	393	57.2	58.8	58.8
	Sim	275	40.0	41.2	100.0
	Total	668	97.2	100.0	
Missing	System	19	2.8		
	Total	687	100.0		

Fonte: Output do SPSS

Tabela 3.7: Fumante ou nao fumante

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Nao	587	85.4	89.1	89.1
	Sim	72	10.5	10.9	100.0
	Total	659	95.9	100.0	

Missing	System	28	4.1		
Total		687	100.0		

Fonte: Output do SPSS

Tabela 3.8: Doença crónica associada

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Nao	548	79.8	81.2	81.2
	Sim	127	18.5	18.8	100.0
	Total	675	98.3	100.0	
Missing	System	12	1.7		
Total		687	100.0		

Fonte: Output do SPSS

Tabela 3.9: Número de internamentos anteriores

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	423	61.6	62.6	62.6
	1	167	24.3	24.7	87.3
	2	72	10.5	10.7	97.9
	3	11	1.6	1.6	99.6
	4	3	.4	.4	100.0
	Total	676	98.4	100.0	
Missing	System	11	1.6		
Total		687	100.0		

Fonte: Output do SPSS

Tabela 3.10: Outras medidas descritivas para número de internamentos anteriores

	N	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Numero de internamentos anteriores	676	0	4	.53	.784	.614	1.477	.094
Valid N (listwise)	676							

Fonte: Output do SPSS

ANEXO IV: ANÁLISE BIVARIADA

Anexos 4.1. Mortalidade ou sobrevivência e gênero

Tabela 4.1.1: Valores observados

		Count		
		Genero do internado		Total
		Feminino	Masculino	
Mortalidade ou sobrevivencia	Alta	188	176	364
	Obito	167	156	323
Total		355	332	687

Fonte: Output do SPSS

Tabela 4.1.2: Teste Qui-quadrado de independência

	Value	df	Asymptotic Significance (2- sided)	Exact Sig. (2- sided)	Exact Sig. (1- sided)
Pearson Chi-Square	.000 ^a	1	.989		
Continuity Correction ^b	.000	1	1.000		
Likelihood Ratio	.000	1	.989		
Fisher's Exact Test				1.000	.525
Linear-by-Linear Association	.000	1	.989		
N of Valid Cases	687				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 156.09.

b. Computed only for a 2x2 table

Fonte: Output do SPSS

Tabela 4.1.3: Correlação Phi e Continência

		Value	Approximate Significance
Nominal by Nominal	Phi	-.001	.989
	Cramer's V	.001	.989
	Contingency Coefficient	.001	.989
N of Valid Cases		687	

Fonte: Output do SPSS

Anexos 4.2. Mortalidade ou sobrevivência e grupo etário

Tabela 4.2.1: Valores observados

		Count						Total
		Grupo etario do internado						
		Menos de 1 ano	1 a 4 anos	5 a 14 anos	15 a 24 anos	25 a 49 anos	50 ou mais anos	
Mortalidade ou sobrevivencia	Alta	27	63	73	124	44	33	364
	Obito	91	15	12	20	171	14	323
Total		118	78	85	144	215	47	687

Fonte: Output do SPSS

Tabela 4.2.2: Teste Qui-quadrado de independência

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	264.332 ^a	5	.000
Likelihood Ratio	286.215	5	.000
Linear-by-Linear Association	.784	1	.376
N of Valid Cases	687		

- a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 22.10.

Fonte: Output do SPSS

Tabela 4.2.3: Correlação V de Cramer

		Value	Approximate Significance
Nominal by Nominal	Phi	.620	.000
	Cramer's V	.620	.000
N of Valid Cases		687	

Fonte: Output do SPSS

Anexos 4.3: Mortalidade ou sobrevivência e estado civil

Tabela 4.3.1: Valores observados

		Count	Total
		Estado Civil do internado	

		Solteiro	Casado ou vivendo maritalmente	Divorciado ou separado	Viuvo	
Mortalidade ou sobrevivencia	Alta	247	92	6	4	349
	Obito	129	143	31	17	320
Total		376	235	37	21	669

Fonte: Output do SPSS

Tabela 4.3.2: Teste Qui-quadrado de Independência

	Value	df	Asymptotic Significance (2- sided)
Pearson Chi-Square	71.918 ^a	3	.000
Likelihood Ratio	74.719	3	.000
Linear-by-Linear Association	66.780	1	.000
N of Valid Cases	669		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 10.04.

Fonte: Output do SPSS

Tabela 4.3.3: Correlação V de Cramer

		Value	Approximate Significance
Nominal by Nominal	Phi	.328	.000
	Cramer's V	.328	.000
N of Valid Cases		669	

Fonte: Output do SPSS

Anexos 4.4: Mortalidade ou sobrevivência e proveniência

Tabela 4.4.1: Valores observados

Count				Total
Provinciãcia do internado				
	Provincia de Maputo	Na regioã sul do pais	Fora da regioã sul	
Grande Maputo				

Mortalidade ou sobrevivencia	Alta	151	100	78	25	354
	Obito	135	90	49	39	313
Total		286	190	127	64	667

Fonte: Output do SPSS

Tabela 4.4.2: Teste Qui-quadrado de independência

	Value	df	Asymptotic Significance (2- sided)
Pearson Chi-Square	8.618 ^a	3	.035
Likelihood Ratio	8.669	3	.034
Linear-by-Linear Association	.257	1	.612
N of Valid Cases	667		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 30.03.

Fonte: Output do SPSS

Anexos 4.5. Mortalidade ou sobrevivência e consumo de álcool

Tabela 4.5.1: Valores observados

		Count		Total
		Consumo de alcool no internado		
		Nao	Sim	
Mortalidade ou sobrevivencia	Alta	214	140	354
	Obito	179	135	314
Total		393	275	668

Fonte: Output do SPSS

Tabela 4.5.2: Teste qui-quadrado de Independência

	Value	df	Asymptotic Significance (2- sided)	Exact Sig. (2- sided)	Exact Sig. (1- sided)
Pearson Chi-Square	.816 ^a	1	.366		
Continuity Correction ^b	.680	1	.410		
Likelihood Ratio	.815	1	.367		
Fisher's Exact Test				.387	.205

Linear-by-Linear Association	.814	1	.367		
N of Valid Cases	668				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 129.27.

b. Computed only for a 2x2 table

Fonte: Output do SPSS

Tabela 4.5.3: Correlação Phi e contingência

		Value	Approximate Significance
Nominal by Nominal	Phi	.035	.366
	Cramer's V	.035	.366
	Contingency Coefficient	.035	.366
N of Valid Cases		668	

Fonte: Output do SPSS

Anexos 4.6. Mortalidade ou sobrevivência e ser ou não fumante

Tabela 4.6.1: Valores observados

		Count		Total
		Fumante ou nao fumante		
		Nao	Sim	
Mortalidade ou sobrevivencia	Alta	342	17	359
	Obito	245	55	300
Total		587	72	659

Fonte: Output do SPSS

Tabela 4.6.2: Teste qui-quadrado de independência

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	31.051 ^a	1	.000		
Continuity Correction ^b	29.670	1	.000		
Likelihood Ratio	31.923	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	31.004	1	.000		

N of Valid Cases	659			
------------------	-----	--	--	--

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 32.78.

b. Computed only for a 2x2 table

Fonte: Output do SPSS

Tabela 4.6.3: Correlação Phi e de Contingência

		Value	Approximate Significance
Nominal by Nominal	Phi	.217	.000
	Cramer's V	.217	.000
	Contingency Coefficient	.212	.000
N of Valid Cases		659	

Fonte: Output do SPSS

Anexos 4.7: Mortalidade ou sobrevivência e Doença crônica associada

Tabela 4.7.1: Valores observados

		Doença crônica associada		Total
		Nao	Sim	
Mortalidade ou sobrevivencia	Alta	329	30	359
	Obito	219	97	316
Total		548	127	675

Fonte: Output do SPSS

Tabela 4.7.2: Teste qui-quadrado de Independência

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	54.910 ^a	1	.000		
Continuity Correction ^b	53.458	1	.000		
Likelihood Ratio	56.692	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	54.829	1	.000		
N of Valid Cases	675				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 59.45.

b. Computed only for a 2x2 table

Fonte: Output do SPSS

Tabela 4.7.3: Correlação Phi e de Contingência

		Value	Approximate Significance
Nominal by Nominal	Phi	.285	.000
	Cramer's V	.285	.000
	Contingency Coefficient	.274	.000
N of Valid Cases		675	

Fonte: Output do SPSS

Anexos 4.8: Mortalidade ou sobrevivência e número de internamentos anteriores

Tabela 4.8.1: Coeficiente Ponto bisserial

		Mortalidade ou sobrevivencia	Numero de internamentos anteriores
Mortalidade ou sobrevivencia	Pearson Correlation	1	.036
	Sig. (2-tailed)		.349
	N	687	676
Numero de internamentos anteriores	Pearson Correlation	.036	1
	Sig. (2-tailed)	.349	
	N	676	676

Fonte: Output do SPSS

Tabela 4.8.2: Medida Eta

			Value
Nominal by Interval	Eta	Mortalidade ou sobrevivencia Dependent	.082
		Numero de internamentos anteriores Dependent	.036

Fonte: Output do SPSS

**ANEXO V:
PERFIL DA
MORTALIDADE
CAUSADA POR
DOENÇAS
INFECCIOSAS**

Tabela 5.1: Género dos óbitos

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Feminino	167	51.7	51.7	51.7
	Masculino	156	48.3	48.3	100.0
	Total	323	100.0	100.0	

Fonte: Output do SPSS

Tabela 5.2: Grupo etário dos óbitos

	N	%
Menos de 1 ano	91	28.2%
1 a 4 anos	15	4.6%
5 a 14 anos	12	3.7%
15 a 24 anos	20	6.2%
25 a 49 anos	171	52.9%
50 ou mais anos	14	4.3%

Fonte: Output do SPSS

Tabela 5.3: estado civil dos óbitos

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Solteiro	129	39.9	40.3	40.3
	Casado ou vivendo maritalmente	143	44.3	44.7	85.0
	Divorciado ou separado	31	9.6	9.7	94.7
	Viuvo	17	5.3	5.3	100.0
	Total	320	99.1	100.0	
Missing	System	3	.9		
	Total	323	100.0		

Fonte: Output do SPSS

Tabela 5.4: Consumo de álcool nos óbitos

	Frequency	Percent	Valid Percent	Cumulative Percent
--	-----------	---------	---------------	--------------------

Valid	Nao	179	55.4	57.0	57.0
	Sim	135	41.8	43.0	100.0
	Total	314	97.2	100.0	
Missing	System	9	2.8		
	Total	323	100.0		

Fonte: Output do SPSS

Tabela 5.5: Cruzamento grupo etário e consumo de álcool

		Count		
		Consumo de alcool no internado		Total
		Não	Sim	
Grupo etario do internado	Menos de 1 ano	91	0	91
	1 a 4 anos	14	0	14
	5 a 14 anos	11	1	12
	15 a 24 anos	9	11	20
	25 a 49 anos	50	115	165
	50 ou mais anos	4	8	12
Total		179	135	314

Fonte: Output do SPSS

Tabela 5.6: Ser ou não fumante em óbitos

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Nao	245	75.9	81.7	81.7
	Sim	55	17.0	18.3	100.0
	Total	300	92.9	100.0	
Missing	System	23	7.1		
	Total	323	100.0		

Fonte: Output do SPSS

Tabela 5.7: Cruzamento grupo etário e ser ou não fumante em óbitos

		Count		
		Fumante ou nao fumante		Total
		Nao	Sim	
Grupo etario do internado	Menos de 1 ano	91	0	91

	1 a 4 anos	13	0	13
	5 a 14 anos	12	0	12
	15 a 24 anos	13	6	19
	25 a 49 anos	110	41	151
	50 ou mais anos	6	8	14
	Total	245	55	300

Fonte: Output do SPSS

Tabela 5.8 Proveniência do óbito

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Grande Maputo	135	41.8	43.1	43.1
	Provincia de Maputo	90	27.9	28.8	71.9
	Na regioao sul do pais	49	15.2	15.7	87.5
	Fora da regioao sul	39	12.1	12.5	100.0
	Total	313	96.9	100.0	
Missing	System	10	3.1		
	Total	323	100.0		

Fonte: Output do SPSS

Tabela 5.9: Doença associada no óbito

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Nao	219	67.8	69.3	69.3
	Sim	97	30.0	30.7	100.0
	Total	316	97.8	100.0	
Missing	System	7	2.2		
	Total	323	100.0		

Fonte: Output do SPSS

Tabela 5.10: Medidas de tendência central e separatrizes do número internamentos anteriores dos óbitos

Numero de internamentos anteriores

N	Valid	318
	Missing	5
Mean		.56
Median		.00
Mode		0
Minimum		0
Maximum		4
Percentiles	25	.00
	50	.00
	75	1.00

Fonte: Output do SPSS

Tabela 5.11: Frequências de número internamentos anteriores dos óbitos

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	189	58.5	59.4	59.4
	1	89	27.6	28.0	87.4
	2	34	10.5	10.7	98.1
	3	4	1.2	1.3	99.4
	4	2	.6	.6	100.0
	Total	318	98.5	100.0	
Missing	System	5	1.5		
Total		323	100.0		

Fonte: Output do SPSS

**ANEXO VI:
MULTICOLINEARI
DADE E MODELO
SEM VARIÁVEIS
INDEPENDENTES**

Tabela 6.1: Tolerância e VIF

Model		Unstandardized Coefficients		Standardized	t	Sig.	Collinearity Statistics	
		B	Std. Error	Coefficients Beta			Tolerance	VIF
1	(Constant)	.263	.064		4.120	.000		
	Genero do internado	.006	.036	.006	.168	.867	.982	1.018
	Grupo etario do internado	-.132	.018	-.423	-7.432	.000	.387	2.586
	Estado Civil do internado	.318	.033	.476	9.596	.000	.510	1.962
	Provincia do internado	.032	.017	.066	1.855	.064	.991	1.009
	Consumo de alcool no internado	.020	.046	.019	.426	.670	.610	1.640
	Fumante ou nao fumante	.312	.063	.187	4.983	.000	.894	1.119
	Doenca cronica associada	.287	.046	.228	6.229	.000	.940	1.064
	Numero de internamentos anteriores	.010	.024	.015	.417	.677	.920	1.087

a. Dependent Variable: Mortalidade ou sobrevivencia

Tabela 6.2: Correlações bivariadas

	Constant	Genero do internado(1)	Grupo etario do internado (1)	Grupo etario do internado (2)	Grupo etario do internado (3)	Grupo etario do internado (4)	Grupo etario do internado (5)	Estado Civil do internado (1)	Estado Civil do internado (2)	Estado Civil do internado (3)	Pr do
1 Constant	1.000	-.460	-.409	-.419	-.287	-.207	-.137	-.036	-.046	-.044	
Genero do internado(1)	-.460	1.000	.033	.050	.023	.033	.014	.005	-.012	.004	
Grupo etario do internado (1)	-.409	.033	1.000	.339	.290	.194	.176	-.012	-.022	-.018	
Grupo etario do internado (2)	-.419	.050	.339	1.000	.305	.216	.199	-.028	-.033	-.024	
Grupo etario do internado (3)	-.287	.023	.290	.305	1.000	.666	.611	-.419	-.356	-.252	
Grupo etario do internado (4)	-.207	.033	.194	.216	.666	1.000	.796	-.833	-.595	-.540	
Grupo etario do internado (5)	-.137	.014	.176	.199	.611	.796	1.000	-.701	-.542	-.527	
Estado Civil do internado (1)	-.036	.005	-.012	-.028	-.419	-.833	-.701	1.000	.634	.624	
Estado Civil do internado (2)	-.046	-.012	-.022	-.033	-.356	-.595	-.542	.634	1.000	.445	
Estado Civil do internado (3)	-.044	.004	-.018	-.024	-.252	-.540	-.527	.624	.445	1.000	
Provincia do internado (1)	-.334	-.014	-.098	-.077	-.121	-.093	-.130	.077	.096	.109	
Provincia do internado (2)	-.307	.039	.017	.018	.029	.010	.031	-.047	-.015	.011	
Provincia do internado (3)	-.266	.047	-.091	-.071	-.154	-.104	-.081	.121	.127	.124	
Consumo de alcool no internado(1)	.003	-.004	-.003	-.014	-.425	-.284	-.233	-.036	.013	-.074	
Fumante ou nao fumante (1)	-.012	-.008	-.008	-.009	-.233	-.053	-.128	-.009	-.012	-.073	
Doenca cronica associada(1)	-.093	.075	-.131	-.138	-.174	-.028	-.117	-.030	.046	-.004	
Numero de	-.169	.017	.023	-.017	-.143	-.152	-.262	.082	.100	.007	

Grupo etario do internado (4)	Grupo etario do internado (5)	Estado Civil do internado (1)	Estado Civil do internado (2)	Estado Civil do internado (3)	Provincia do internado (1)	Provincia do internado (2)	Provincia do internado (3)	Consumo de alcool no internado(1)	Fumante ou nao fumante (1)	Doenca cronica associada(1)	Numero de internamentos anteriores
-207	-137	-036	-046	-044	-334	-307	-266	.003	-012	-093	-169
.033	.014	.005	-.012	.004	-.014	.039	.047	-.004	-.008	.075	.017
.194	.176	-.012	-.022	-.018	-.098	.017	-.091	-.003	-.008	-.131	.023
.216	.199	-.028	-.033	-.024	-.077	.018	-.071	-.014	-.009	-.138	-.017
.666	.611	-.419	-.356	-.252	-.121	.029	-.154	-.425	-.233	-.174	-.143
1.000	.796	-.833	-.595	-.540	-.093	.010	-.104	-.284	-.053	-.028	-.152
.796	1.000	-.701	-.542	-.527	-.130	.031	-.081	-.233	-.128	-.117	-.262
-.833	-.701	1.000	.634	.624	.077	-.047	.121	-.036	-.009	-.030	.082
-.595	-.542	.634	1.000	.445	.096	-.015	.127	.013	-.012	.046	.100
-.540	-.527	.624	.445	1.000	.109	.011	.124	-.074	-.073	-.004	.007
-.093	-.130	.077	.096	.109	1.000	.335	.307	-.007	.057	.013	.085
.010	.031	-.047	-.015	.011	.335	1.000	.228	.008	-.017	-.037	-.073
-.104	-.081	.121	.127	.124	.307	.228	1.000	-.010	-.014	.081	.025
-.284	-.233	-.036	.013	-.074	-.007	.008	-.010	1.000	.036	.032	-.015
-.053	-.128	-.009	-.012	-.073	.057	-.017	-.014	.036	1.000	.028	.042
-.028	-.117	-.030	.046	-.004	.013	-.037	.081	.032	.028	1.000	.011
-.152	-.262	.082	.100	.007	.085	-.073	.025	-.015	.042	.011	1.000

Tabela 6.3: Variaveis que nao constam no modelo nulo

Step 0	Variables	Score	Df	Sig.
	Genero do internado(1)	.105	1	.745
	Grupo etario do internado	241.803	5	.000
	Grupo etario do internado(1)	54.074	1	.000
	Grupo etario do internado(2)	26.324	1	.000
	Grupo etario do internado(3)	39.283	1	.000
	Grupo etario do internado(4)	67.250	1	.000
	Grupo etario do internado(5)	115.152	1	.000
	Estado Civil do internado	57.469	3	.000
	Estado Civil do internado(1)	47.950	1	.000
	Estado Civil do internado(2)	16.217	1	.000
	Estado Civil do internado(3)	18.141	1	.000
	Provincia do internado	9.043	3	.029
	Provincia do internado(1)	.006	1	.938
	Provincia do internado(2)	.034	1	.855
	Provincia do internado(3)	3.117	1	.077
	Consumo de alcool no internado(1)	.448	1	.503
	Fumante ou nao fumante(1)	34.318	1	.000
	Doenca cronica associada(1)	48.529	1	.000

Numero de internamentos anteriores	.102	1	.749
Overall Statistics	291.191	16	.000

**ANEXO VII:
CRITÉRIO ENTER
REDUZIDO**

Tabela 7.1: Resumo dos casos processados

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	624	90.8
	Missing Cases	63	9.2
	Total	687	100.0
Unselected Cases		0	.0
Total		687	100.0

a. If weight is in effect, see classification table for the total number of cases.

Tabela 7.2 Resumo do modelo

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	508.296 ^a	.432	.577

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Fonte: Output do SPSS

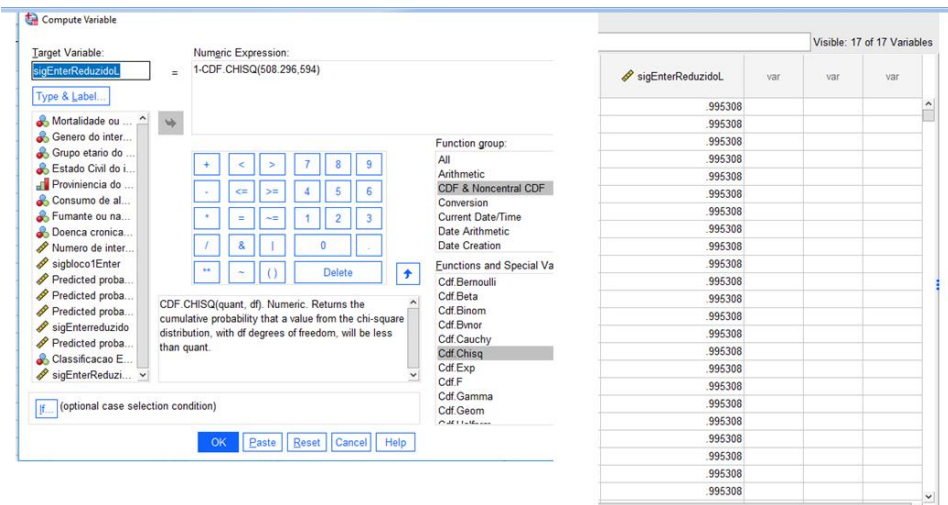


Imagem A: cálculo da significância do -2LL do Enter reduzido

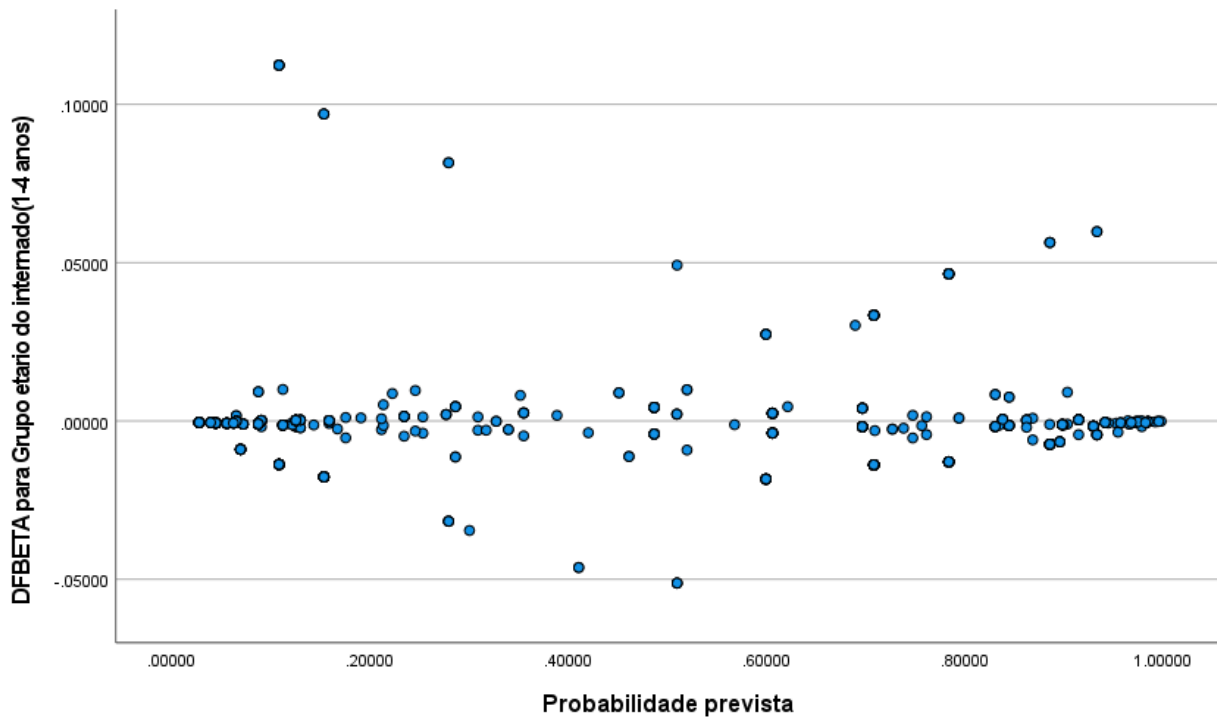
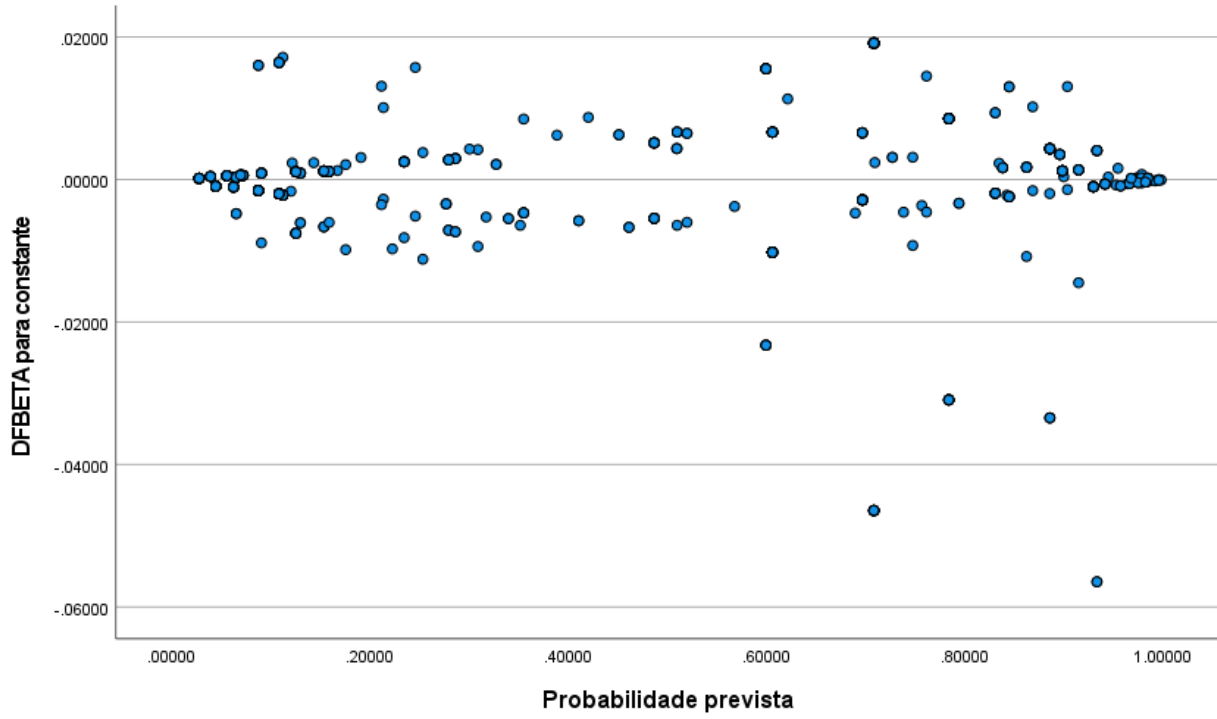
Tabela 7.3: Observacoes candidatas a outliers^b

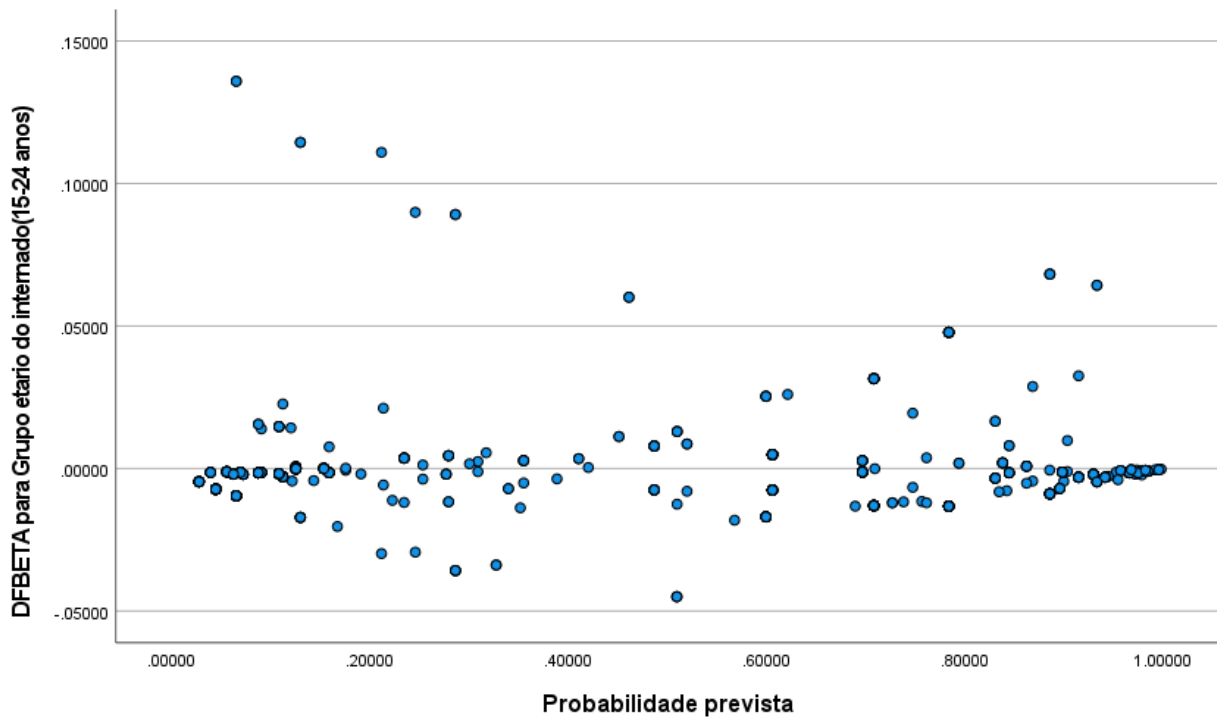
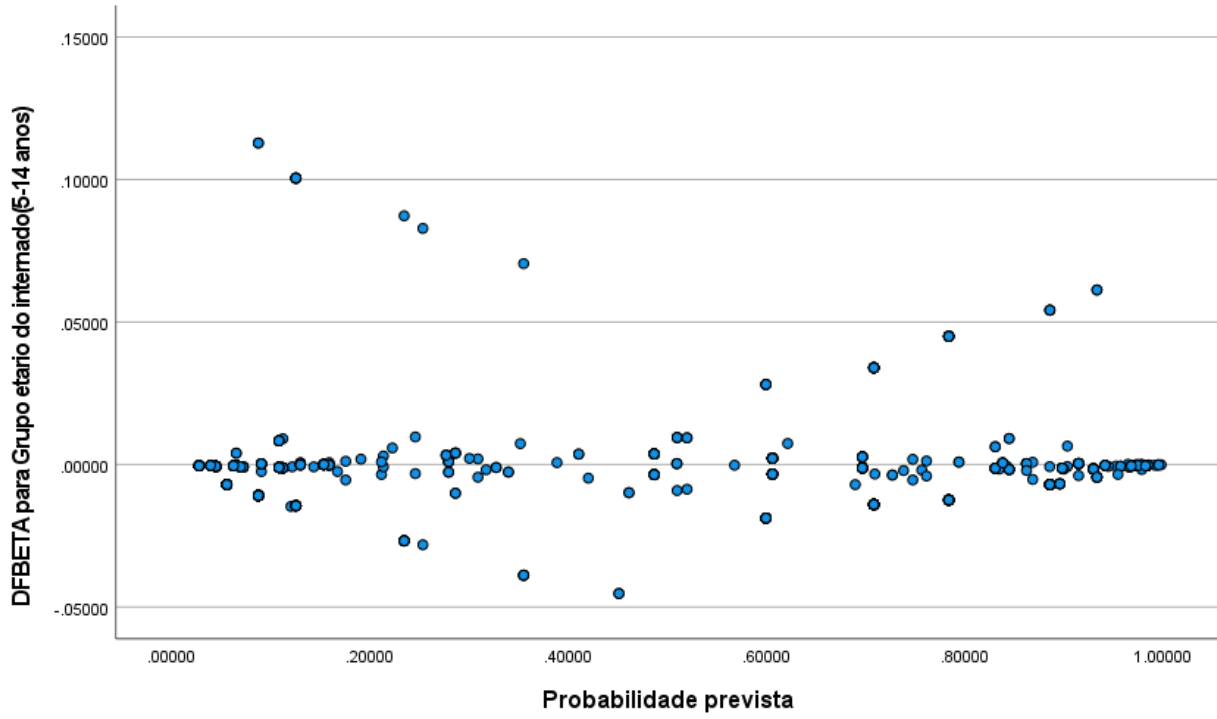
Case	Selected Status ^a	Observed	Predicted	Predicted Group	Temporary Variable		
		Mortalidade ou sobrevivencia			Resid	ZResid	SResid

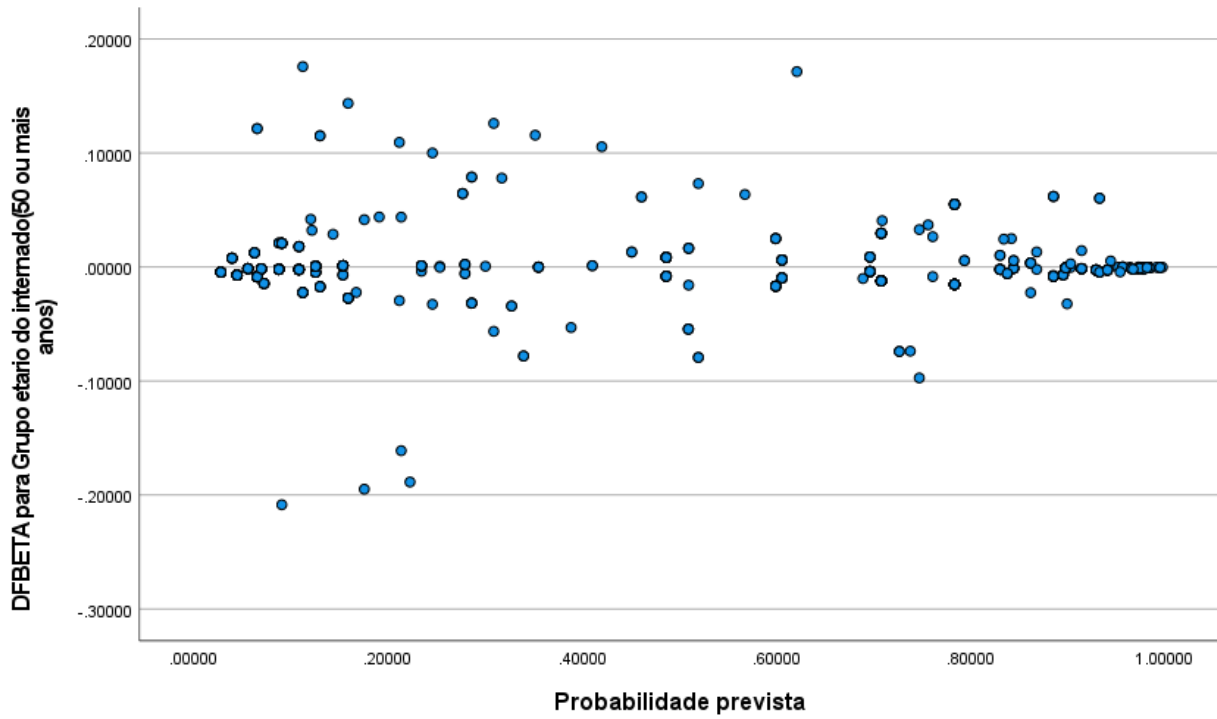
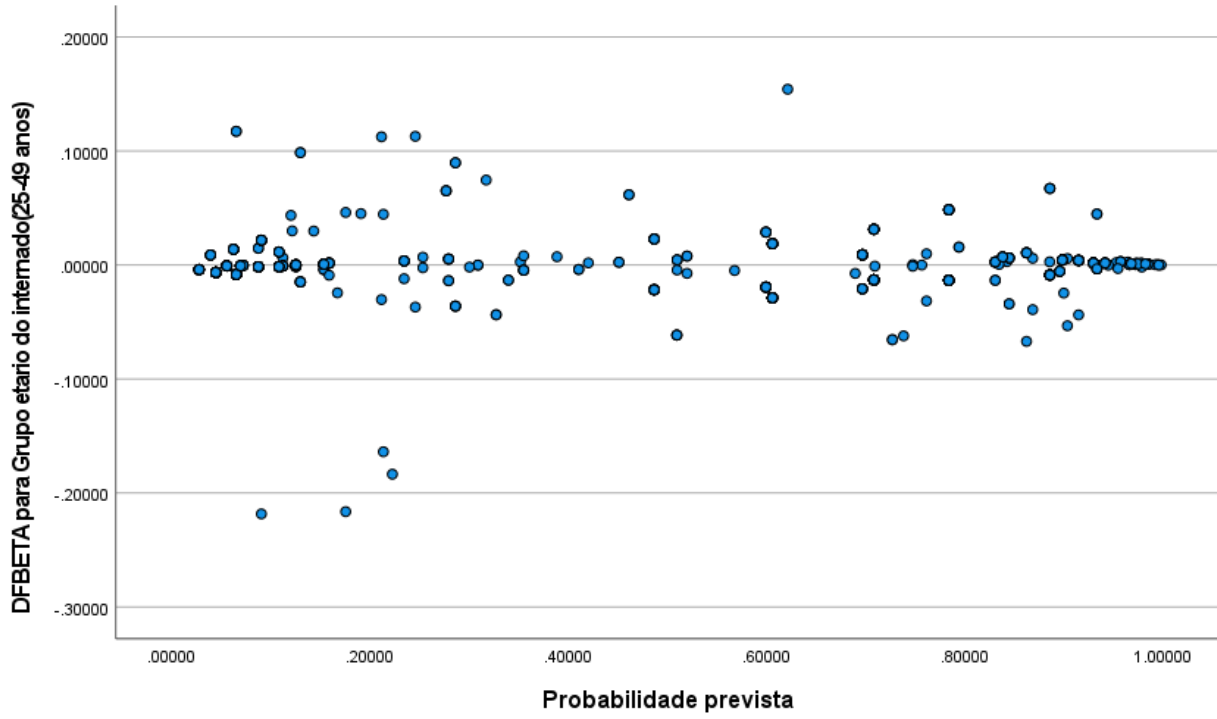
2	S	O**	.109	A	.891	2.864	2.122
22	S	O**	.126	A	.874	2.637	2.053
31	S	O**	.066	A	.934	3.774	2.347
32	S	O**	.113	A	.887	2.807	2.114
37	S	O**	.130	A	.870	2.584	2.045
49	S	O**	.109	A	.891	2.864	2.122
53	S	O**	.066	A	.934	3.774	2.347
67	S	O**	.126	A	.874	2.637	2.053
73	S	O**	.088	A	.912	3.221	2.218
85	S	O**	.126	A	.874	2.637	2.053
87	S	O**	.091	A	.909	3.161	2.217
99	S	O**	.126	A	.874	2.637	2.053
137	S	O**	.109	A	.891	2.864	2.122
159	S	O**	.126	A	.874	2.637	2.053
196	S	O**	.109	A	.891	2.864	2.122
284	S	O**	.109	A	.891	2.864	2.122
285	S	O**	.126	A	.874	2.637	2.053
298	S	O**	.130	A	.870	2.584	2.045
319	S	O**	.088	A	.912	3.221	2.218
337	S	A**	.885	O	-.885	-2.777	-2.099
382	S	A**	.862	O	-.862	-2.497	-2.037
392	S	A**	.933	O	-.933	-3.722	-2.334
399	S	A**	.868	O	-.868	-2.564	-2.044
401	S	A**	.885	O	-.885	-2.777	-2.099
415	S	A**	.933	O	-.933	-3.722	-2.334
425	S	A**	.903	O	-.903	-3.050	-2.201
465	S	A**	.885	O	-.885	-2.777	-2.099
514	S	A**	.914	O	-.914	-3.264	-2.237

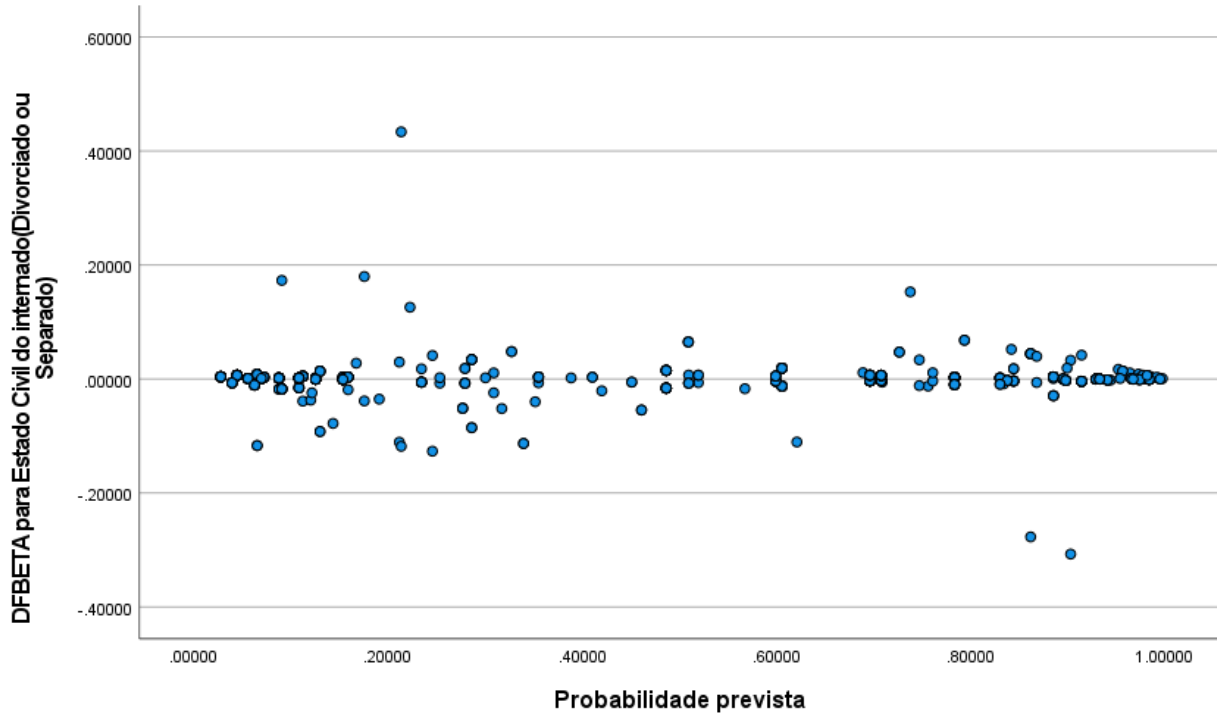
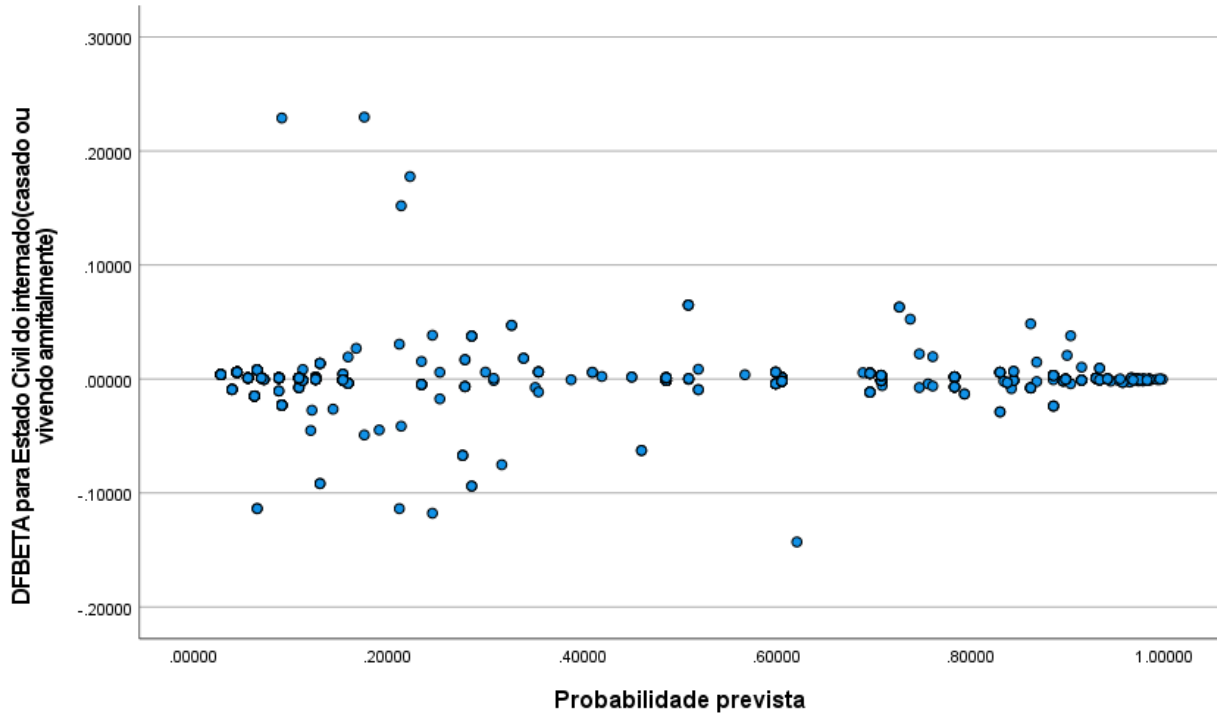
a. S = Selected, U = Unselected cases, and ** = Misclassified cases.

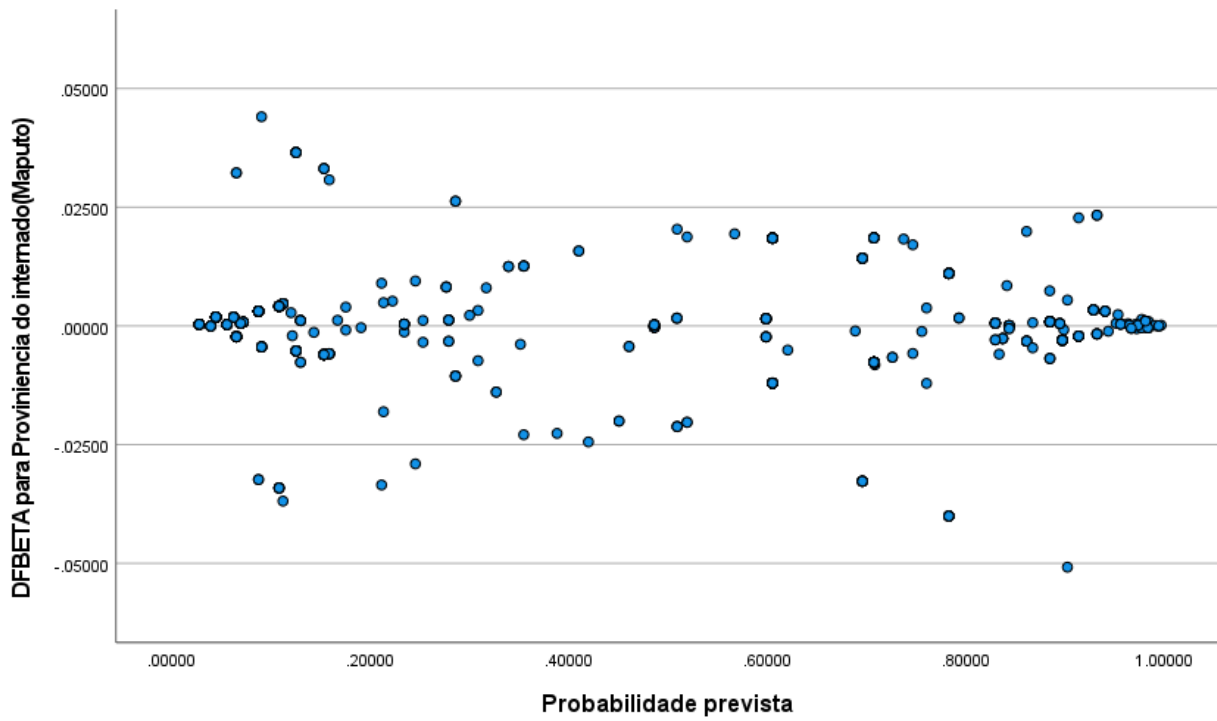
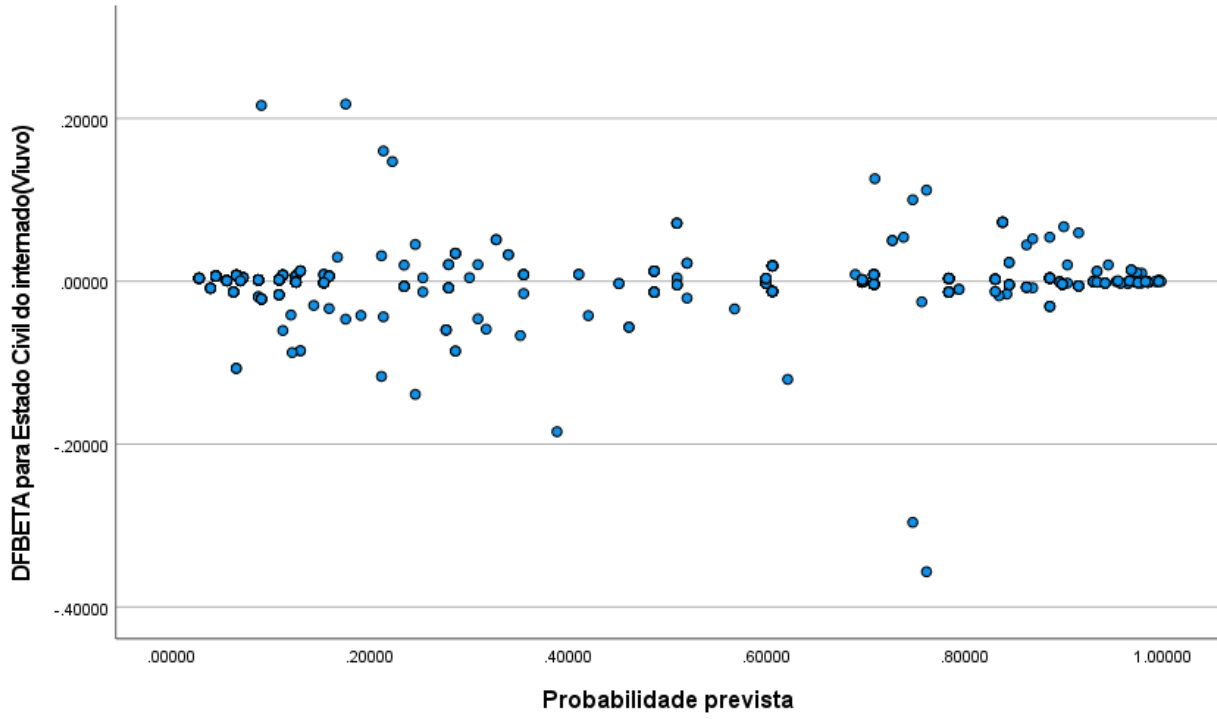
b. Cases with studentized residuals greater than 2.000 are listed.

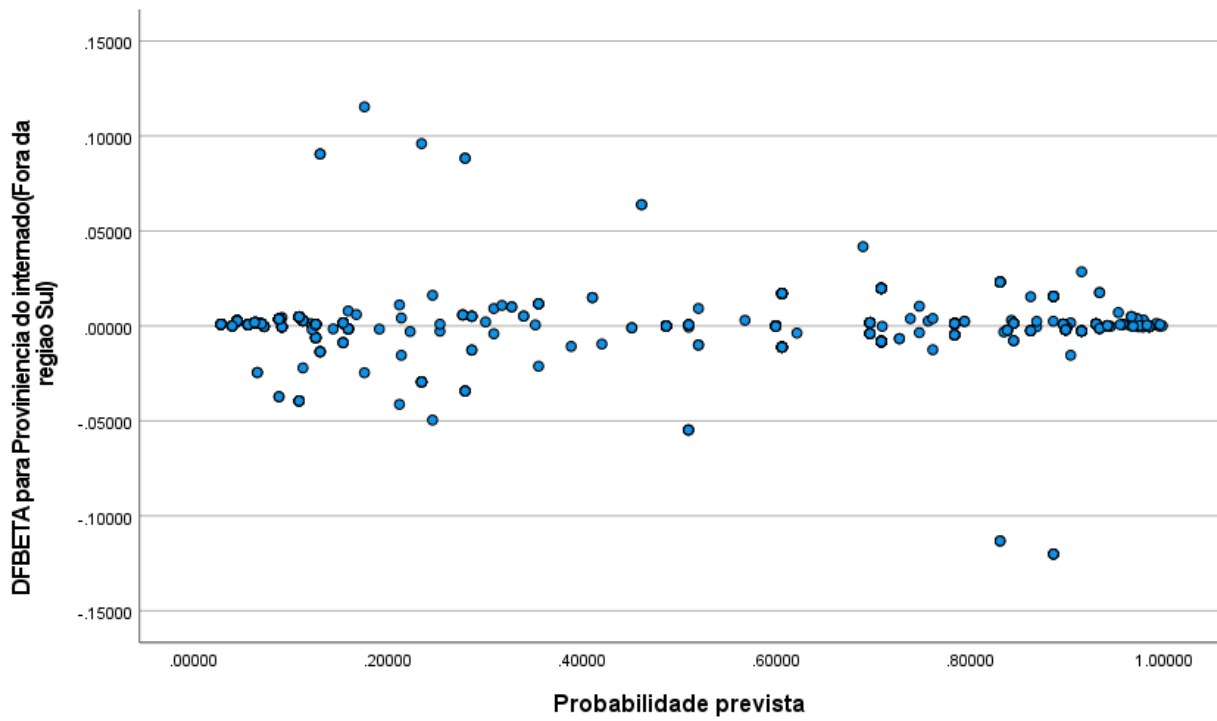
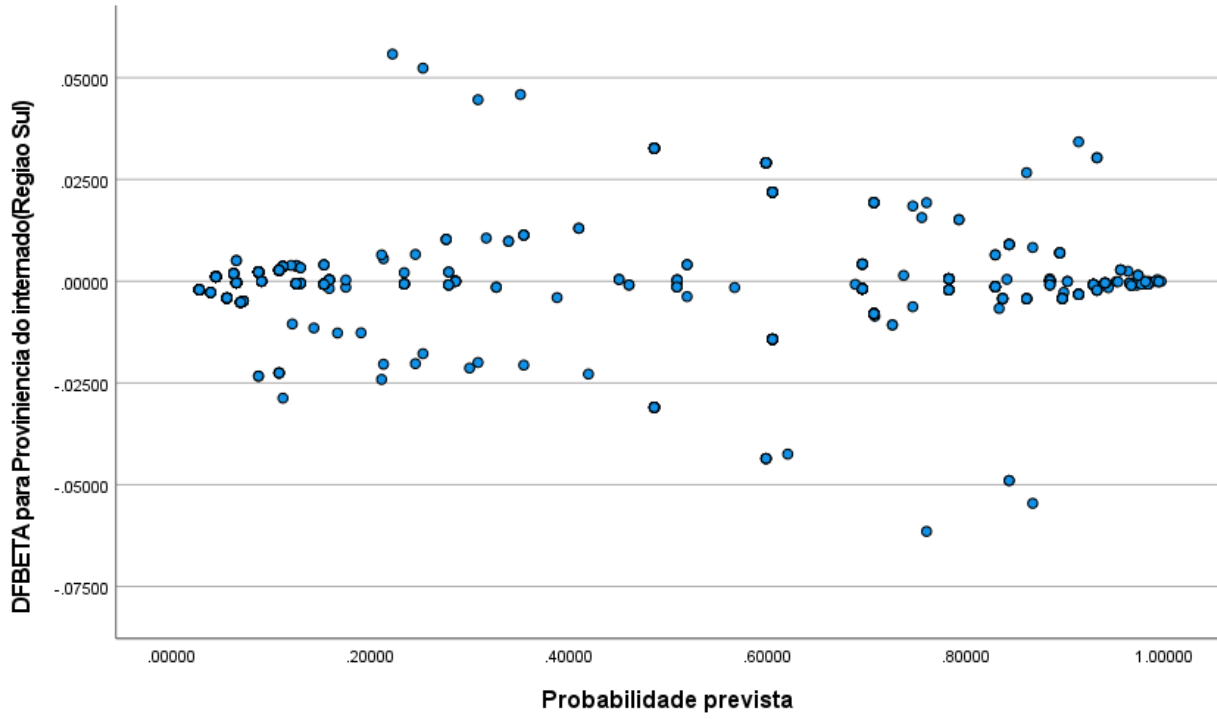


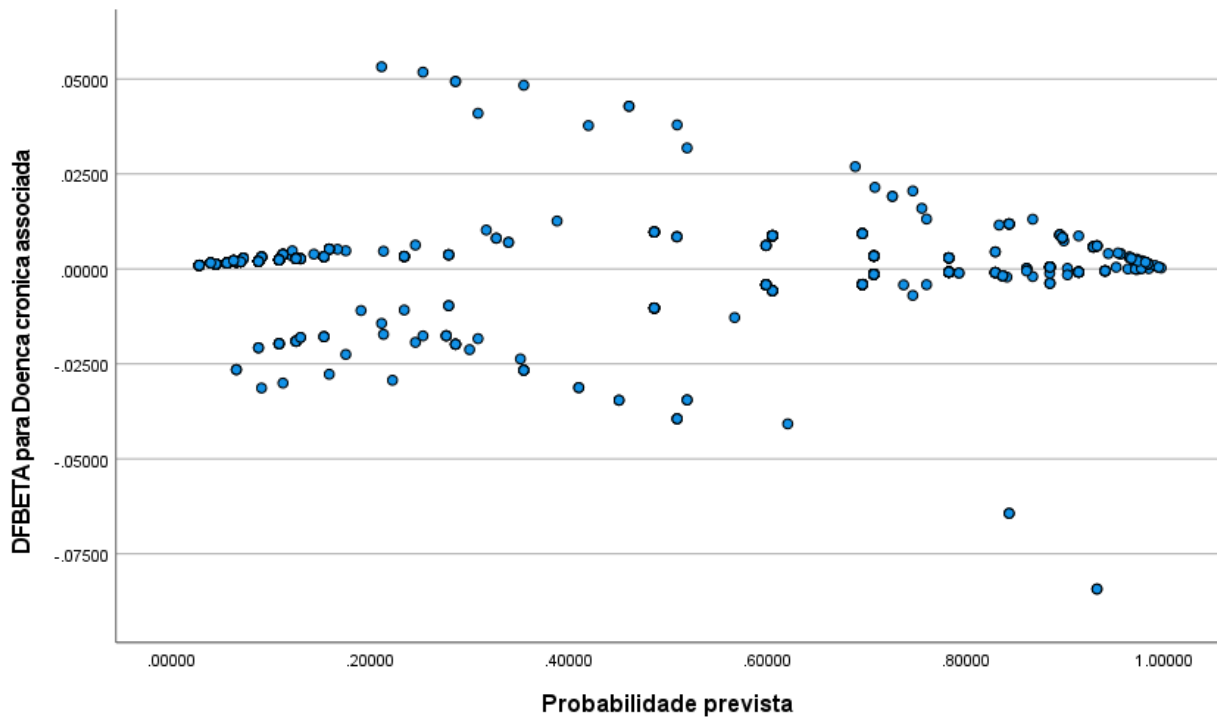
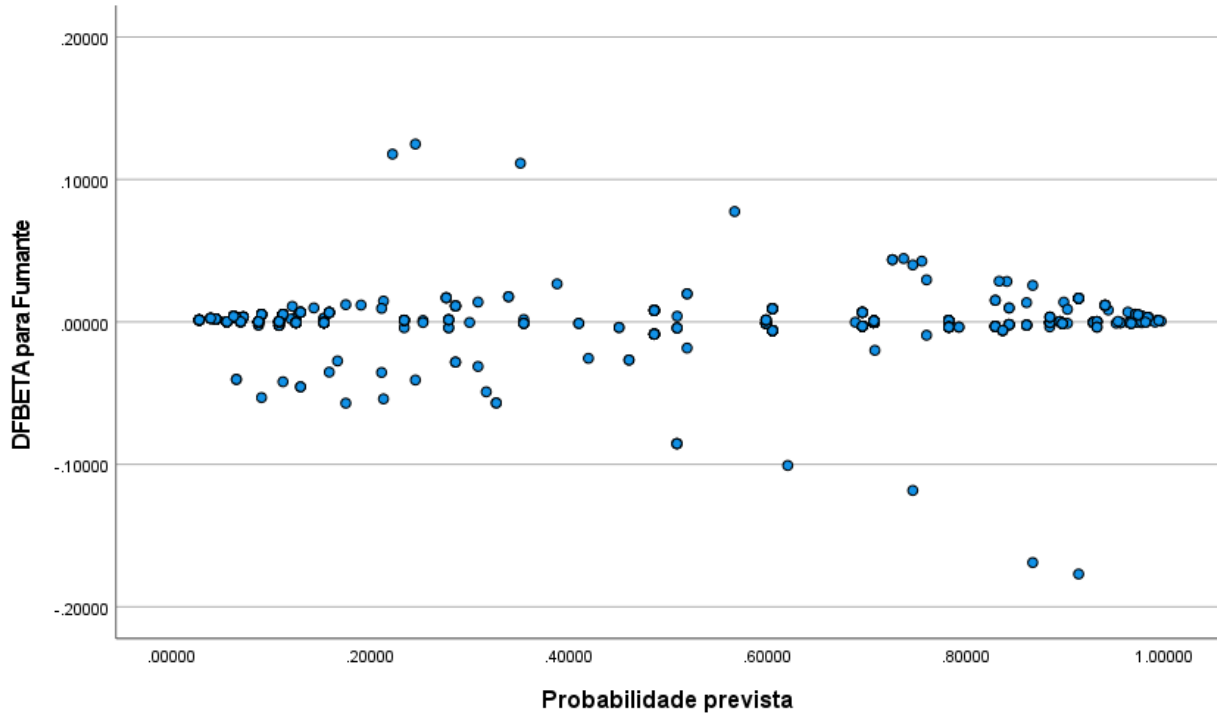












**ANEXO VIII:
CRITÉRIO ENTER
COMPLETO**

Tabela 8.1: Resumo do modelo do Critério Enter Completo

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	492.233 ^a	.434	.581

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Fonte: Output do SPSS

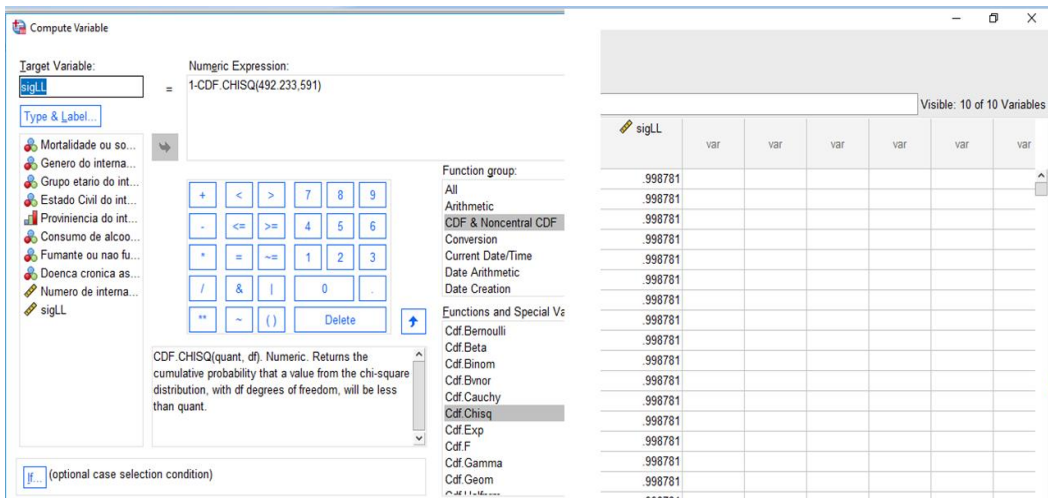


Imagem B: calculo da significancia de -2LL do Critério Enter Completo

Tabela 8.2: Candidatas a outliers do Critério Enter Completo

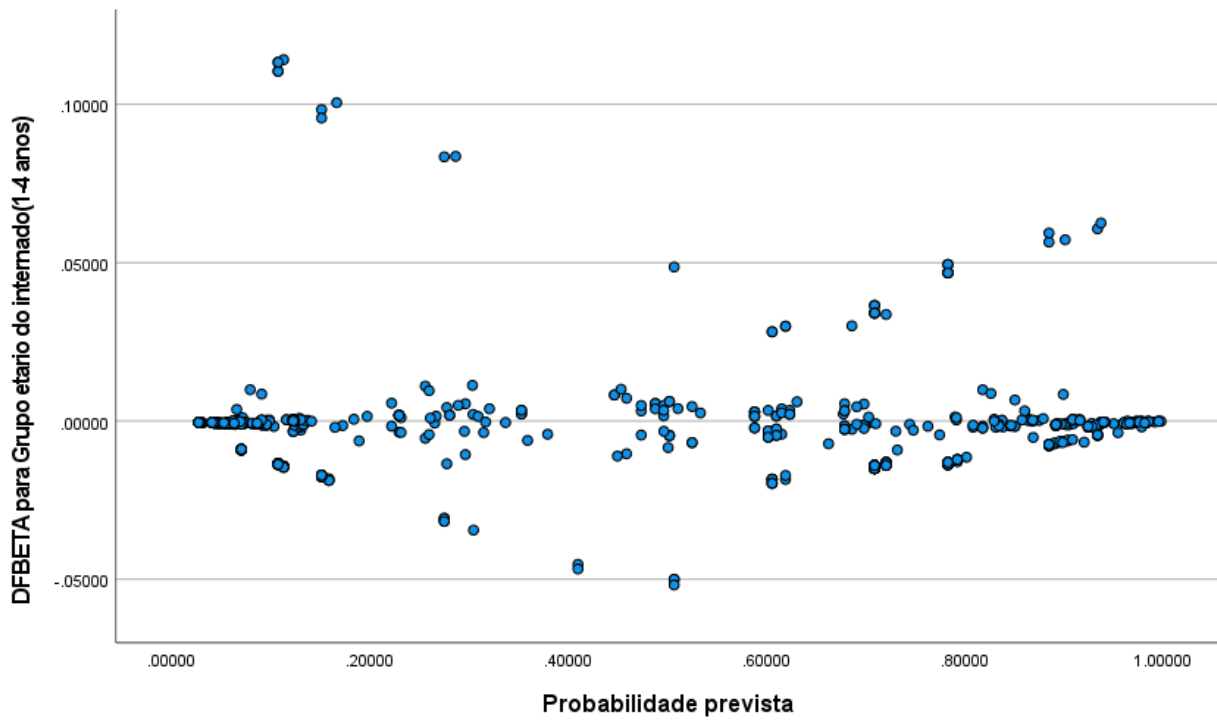
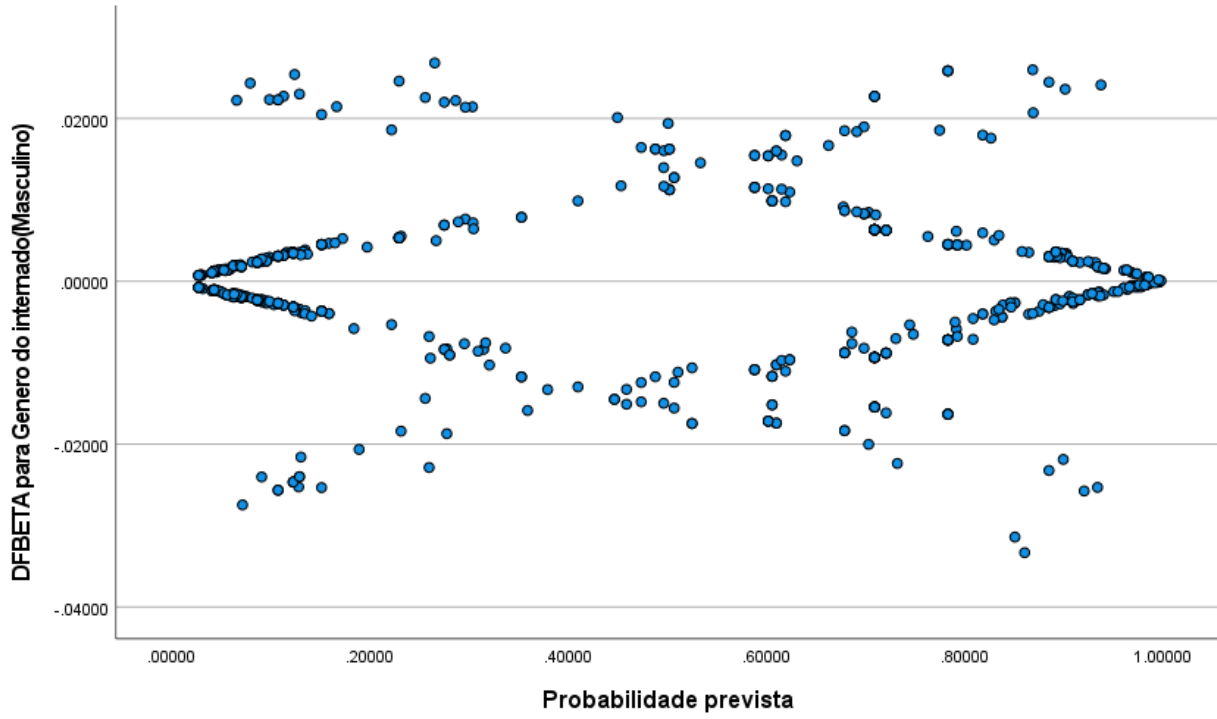
Case	Selected Status ^a	Observed	Predicted	Predicted Group	Temporary Variable		
		Mortalidade ou sobrevivencia			Resid	ZResid	SResid
2	S	O**	.108	A	.892	2.877	2.127
22	S	O**	.123	A	.877	2.669	2.065
31	S	O**	.066	A	.934	3.758	2.345
37	S	O**	.131	A	.869	2.578	2.046
49	S	O**	.113	A	.887	2.797	2.105
53	S	O**	.072	A	.928	3.590	2.313
67	S	O**	.123	A	.877	2.669	2.065
73	S	O**	.091	A	.909	3.156	2.203
85	S	O**	.129	A	.871	2.594	2.043
87	S	O**	.099	A	.901	3.015	2.187
99	S	O**	.129	A	.871	2.594	2.043

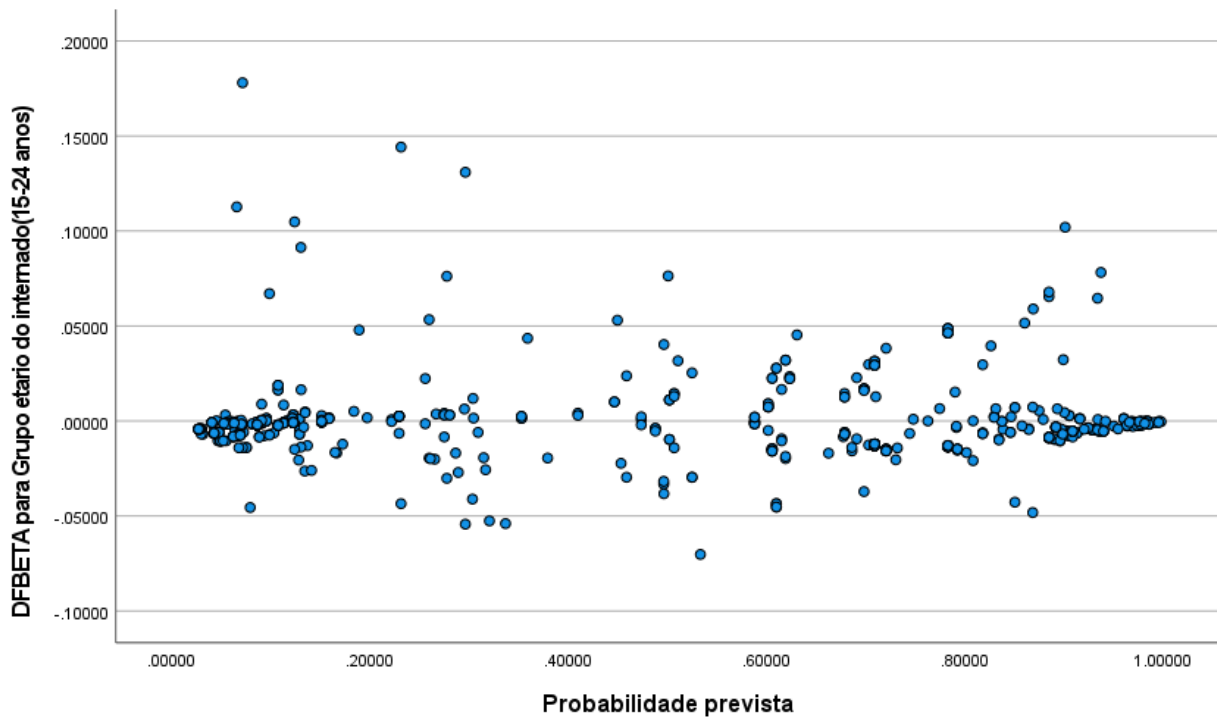
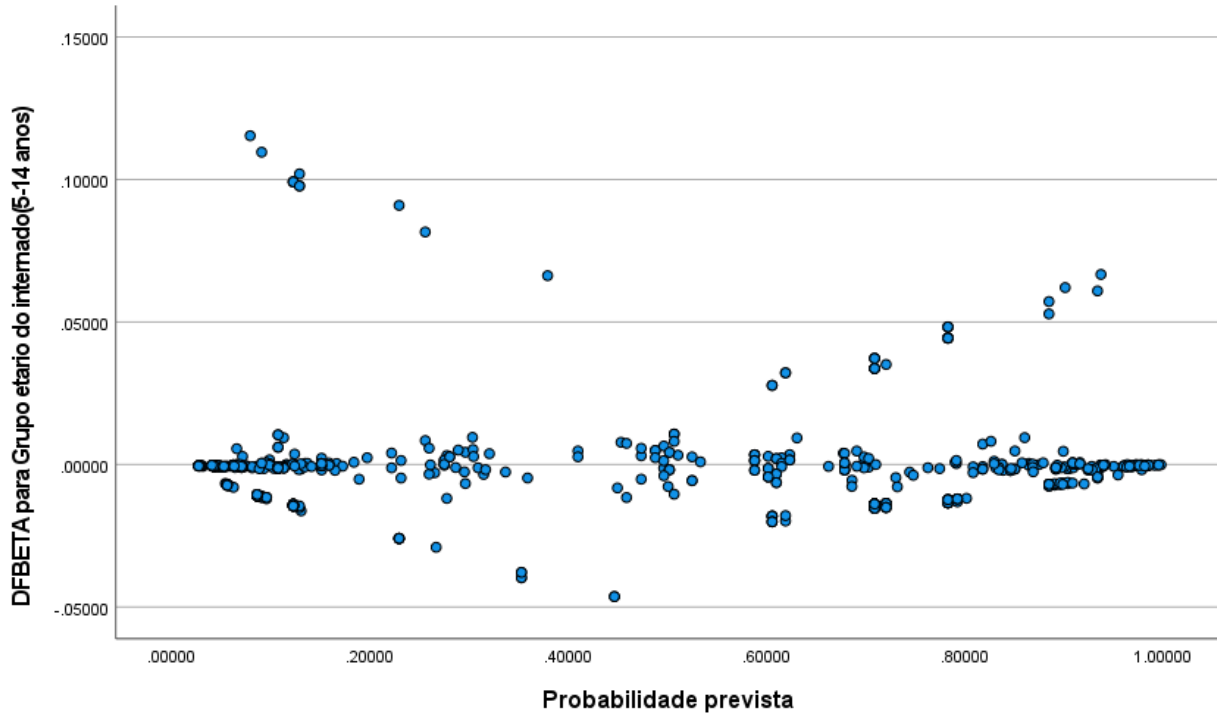
114	S	O**	.129	A	.871	2.604	2.060
137	S	O**	.108	A	.892	2.877	2.127
159	S	O**	.123	A	.877	2.669	2.065
196	S	O**	.108	A	.892	2.877	2.127
284	S	O**	.108	A	.892	2.877	2.127
285	S	O**	.129	A	.871	2.594	2.043
298	S	O**	.124	A	.876	2.653	2.072
319	S	O**	.080	A	.920	3.398	2.271
337	S	A**	.884	O	-.884	-2.765	-2.097
382	S	A**	.868	O	-.868	-2.564	-2.067
392	S	A**	.933	O	-.933	-3.740	-2.339
399	S	A**	.868	O	-.868	-2.569	-2.050
401	S	A**	.884	O	-.884	-2.765	-2.097
415	S	A**	.937	O	-.937	-3.847	-2.362
425	S	A**	.899	O	-.899	-2.979	-2.187
465	S	A**	.901	O	-.901	-3.011	-2.185
514	S	A**	.920	O	-.920	-3.386	-2.273
582	S	A**	.860	O	-.860	-2.477	-2.020

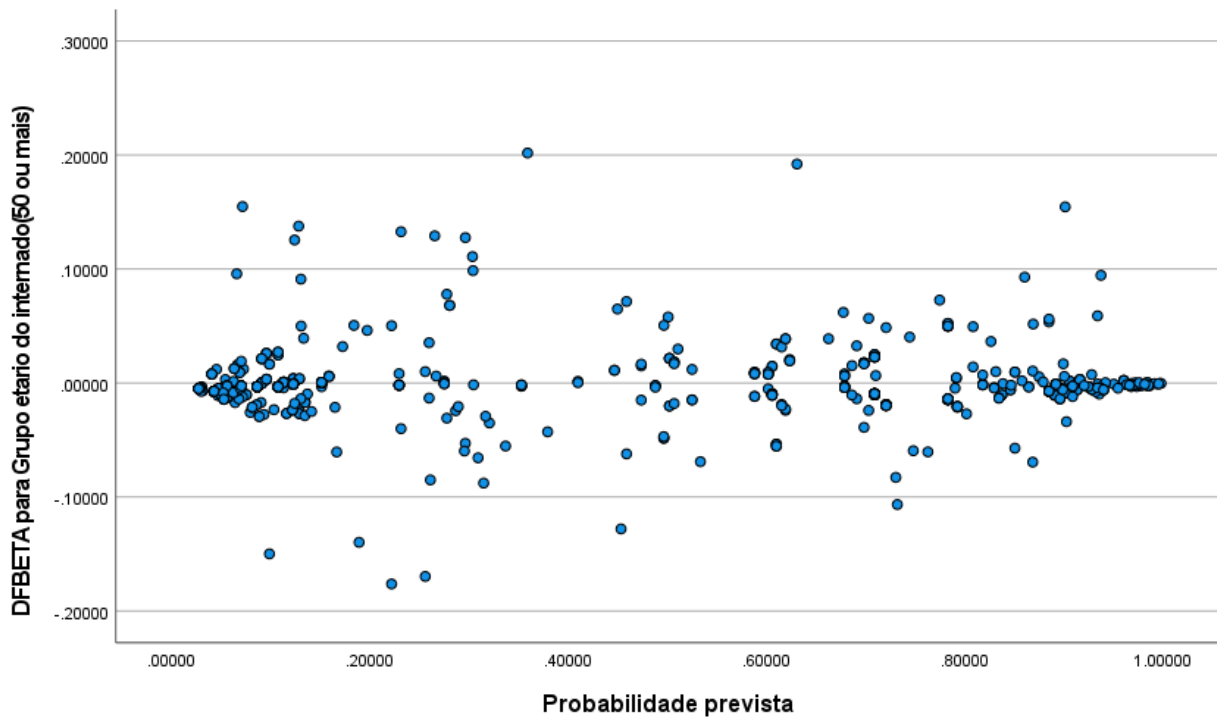
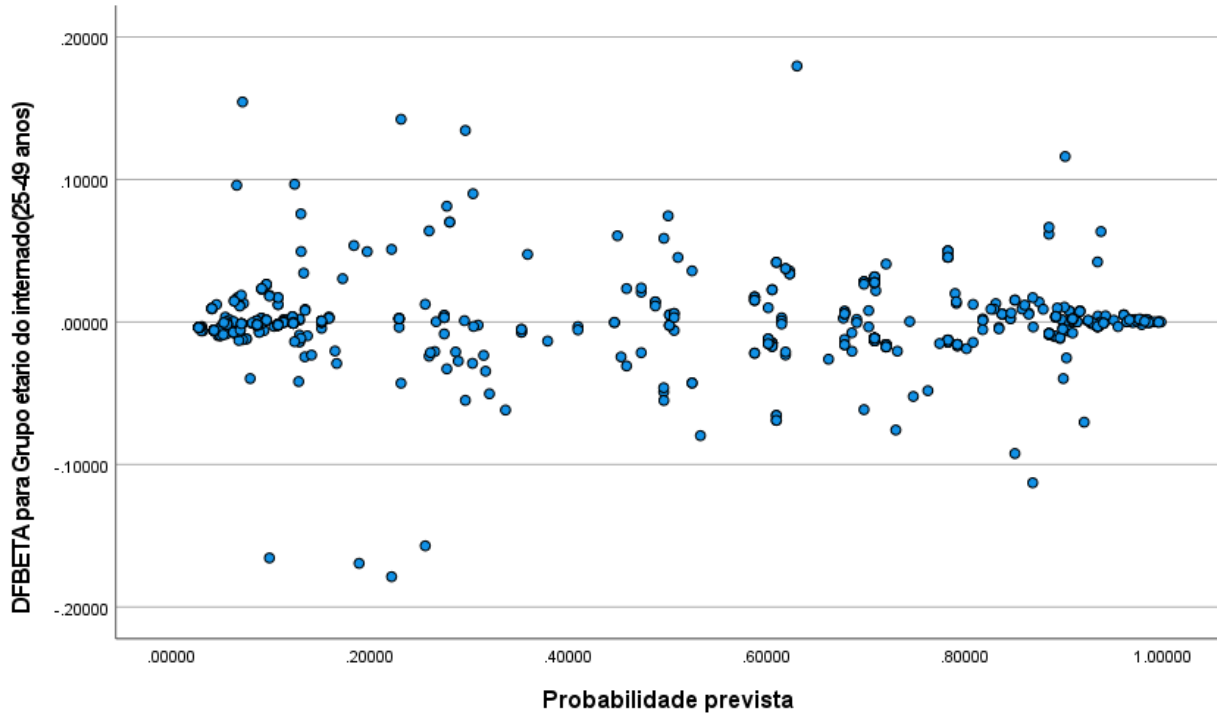
a. S = Selected, U = Unselected cases, and ** = Misclassified cases.

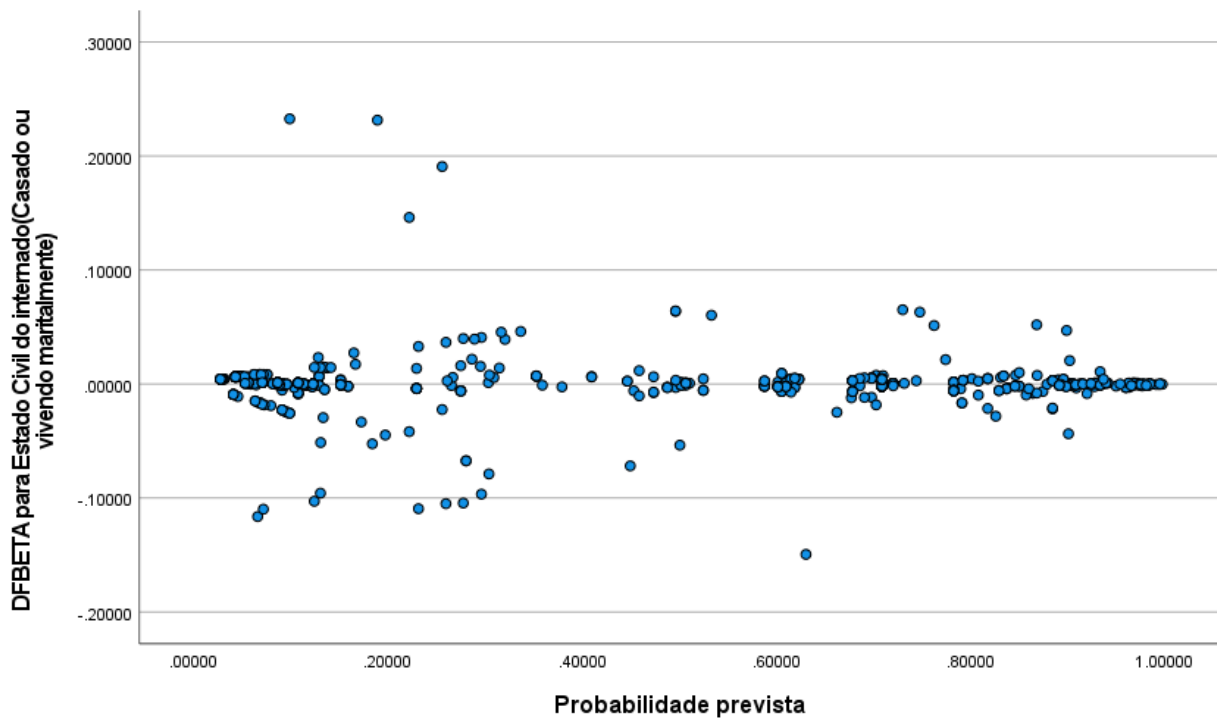
b. Cases with studentized residuals greater than 2.000 are listed.

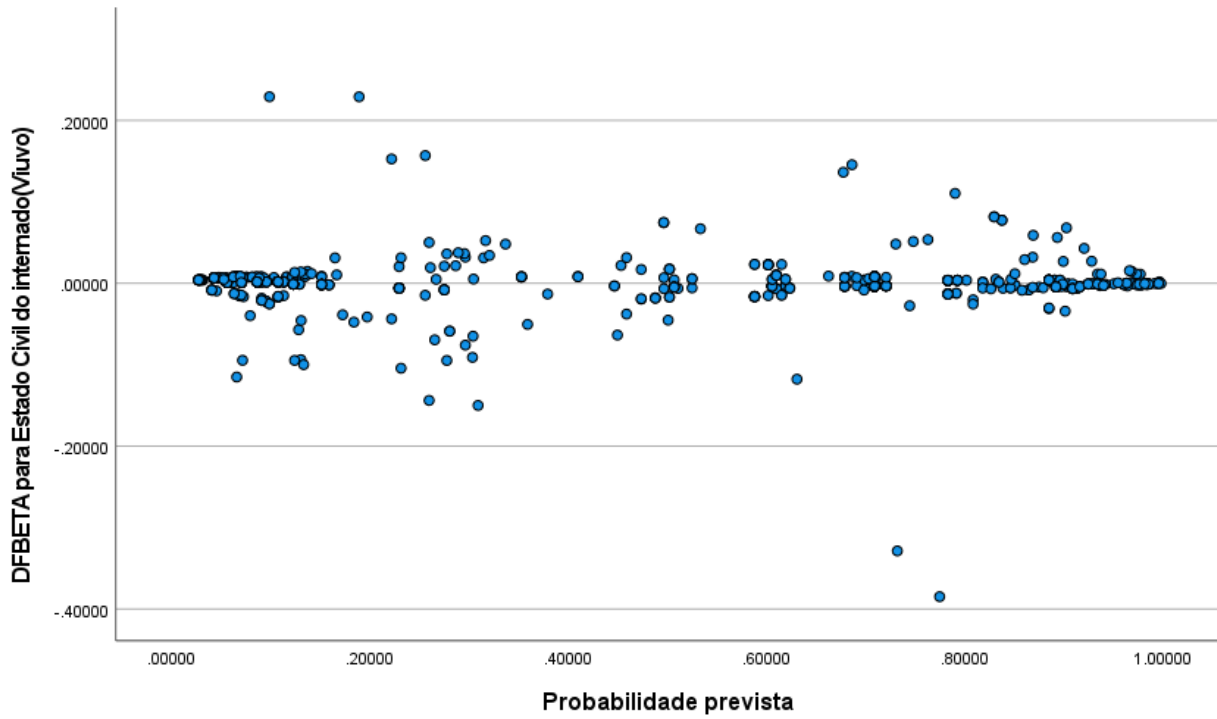
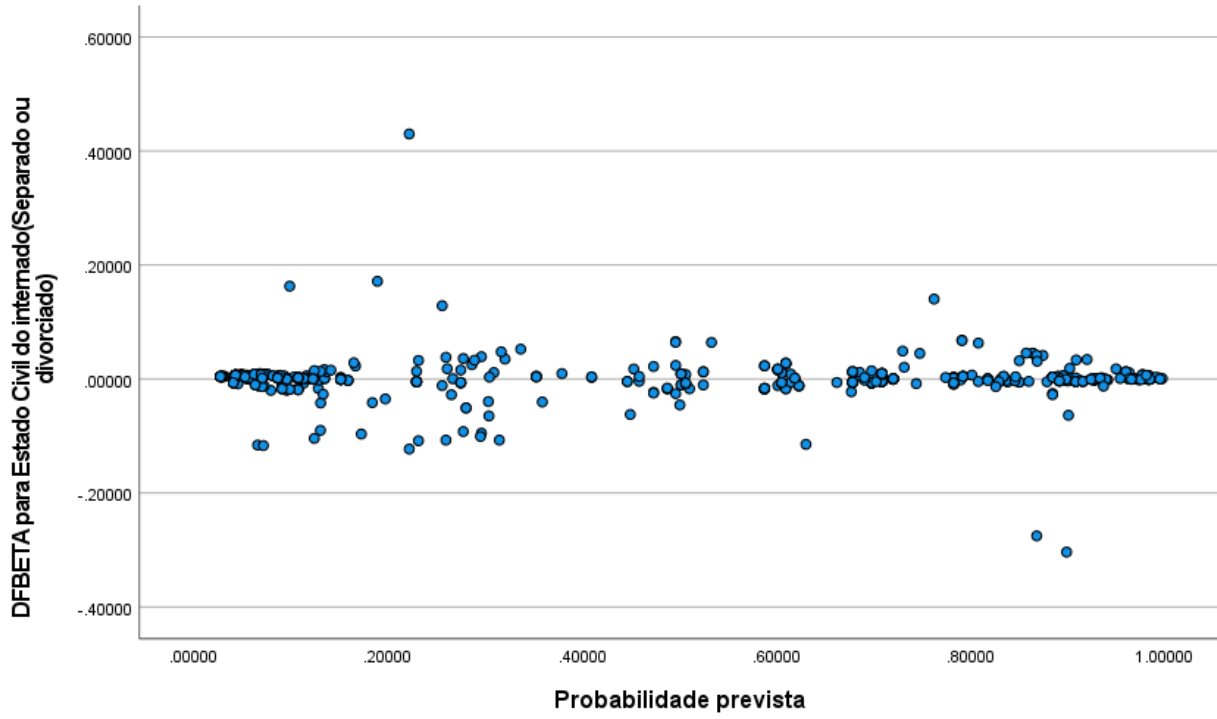
A seguir tem-se os graficos de Dfbetas referentes a ao criterio Enter Completo.

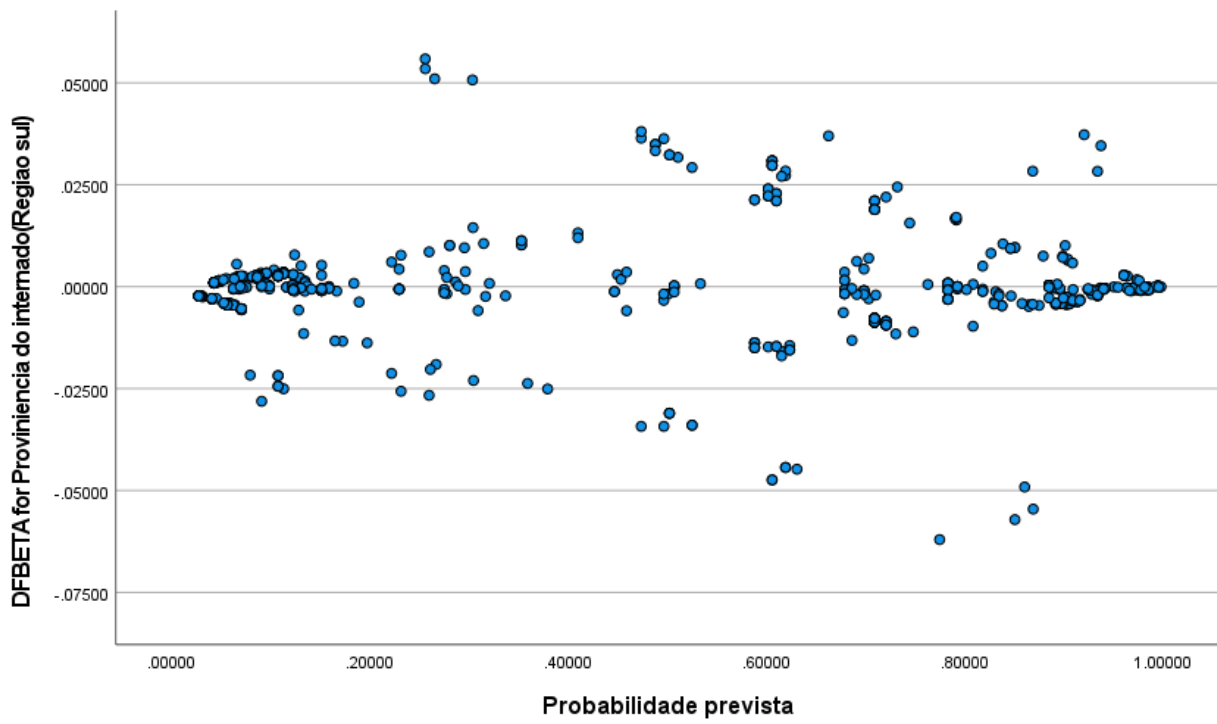
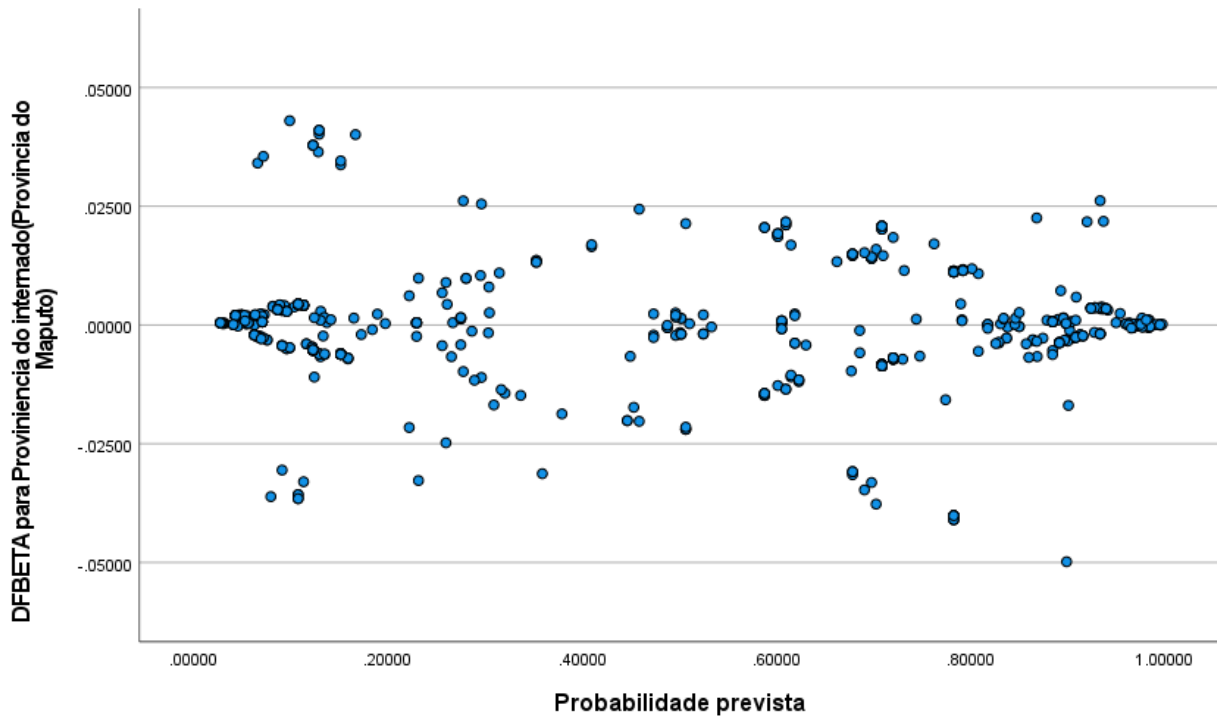


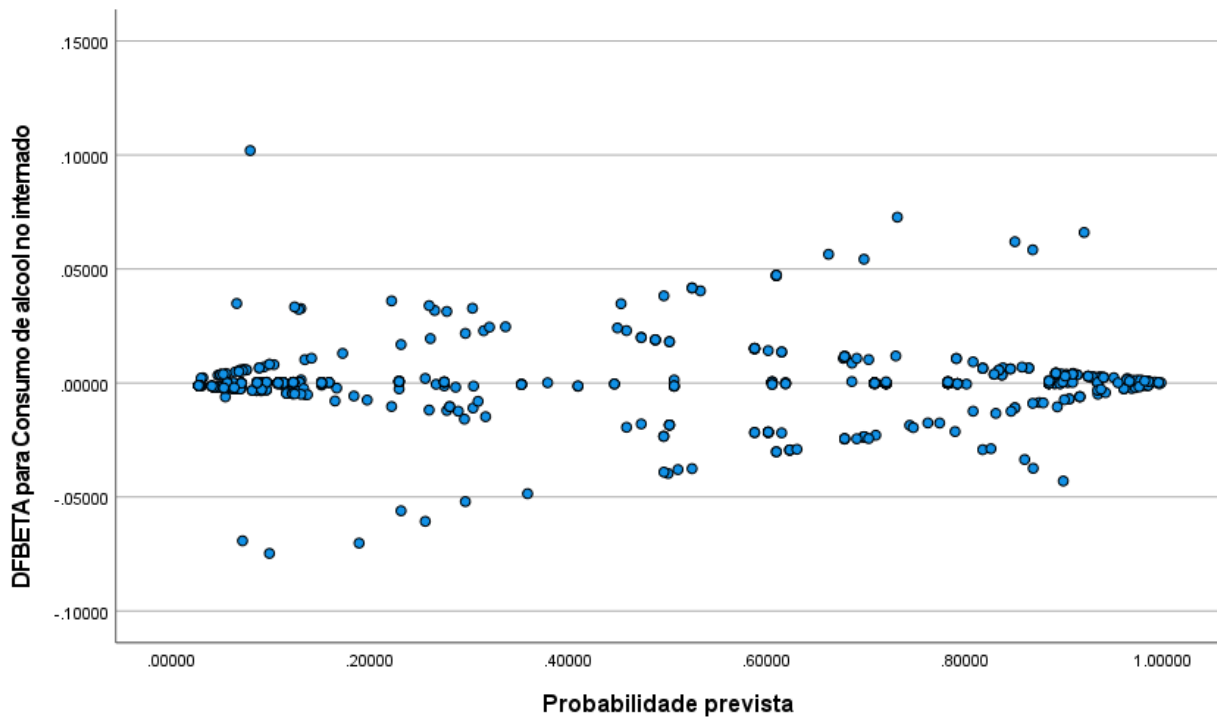
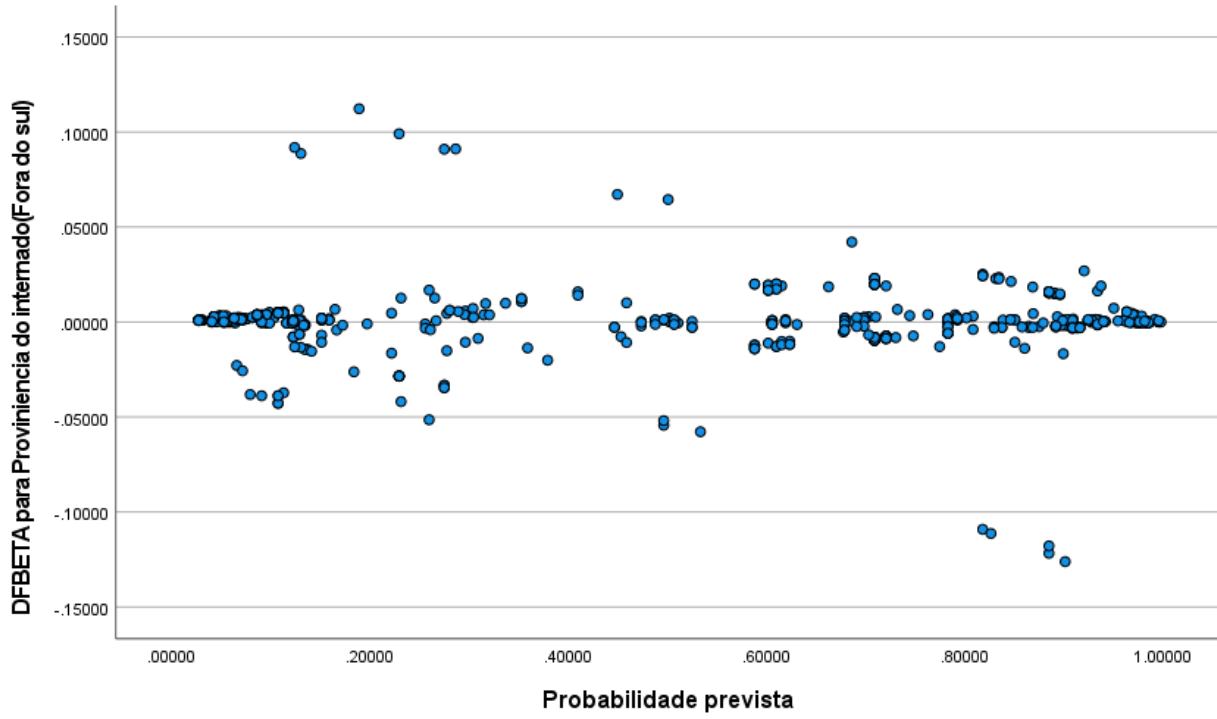


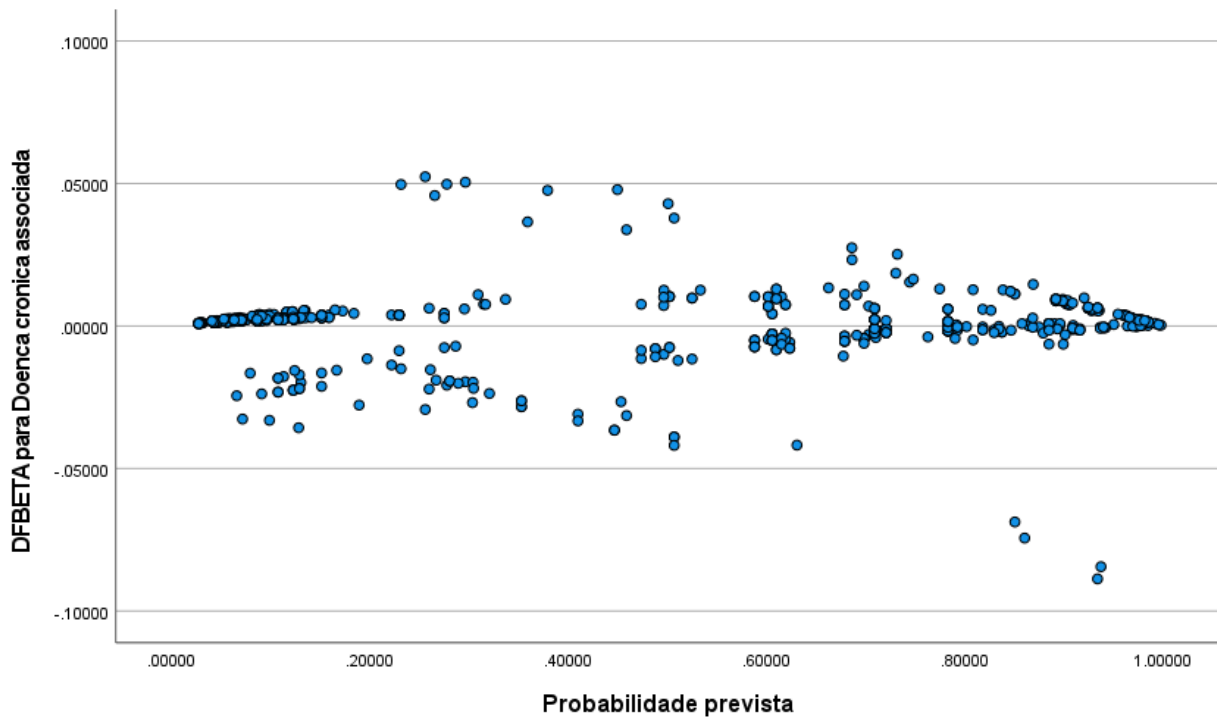
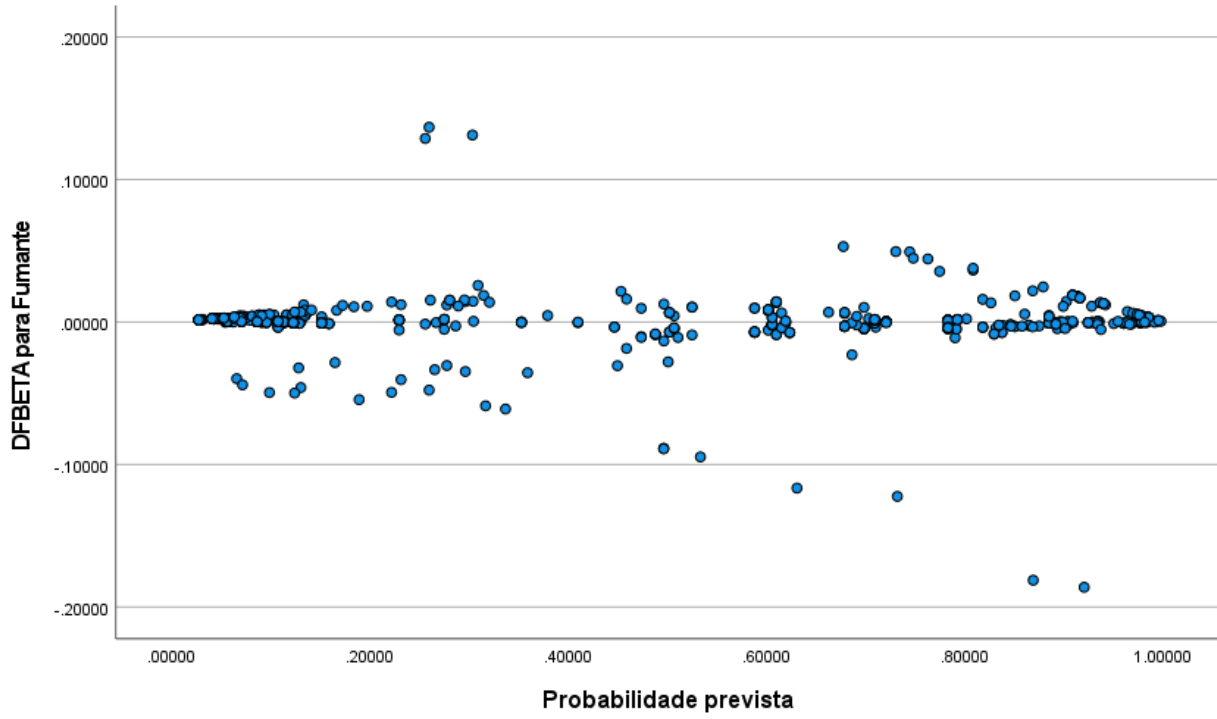


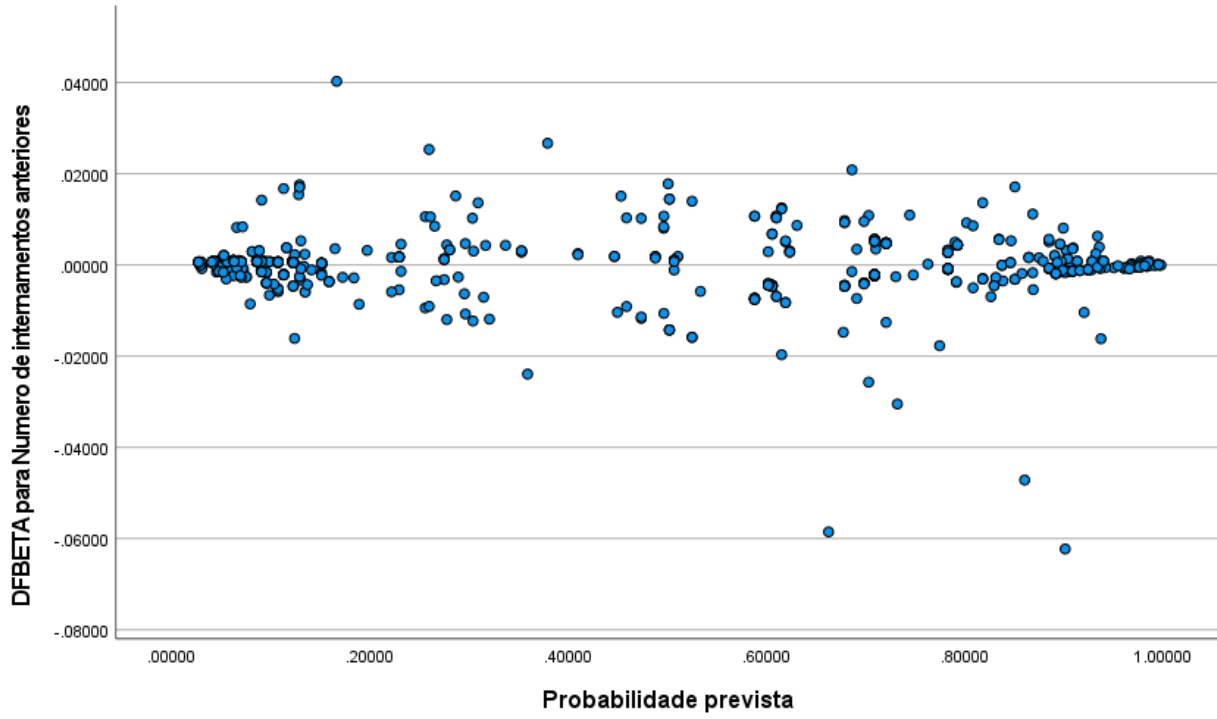












**ANEXO IX:
CRITERIO FORWARD LR**

Tabela 9.1: Resumo do modelo pelo Critério Forward LR

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	576.855 ^a	.350	.468
2	540.214 ^b	.388	.519
3	515.490 ^b	.412	.551
4	501.924 ^b	.425	.568
5	492.439 ^b	.434	.580

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

b. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

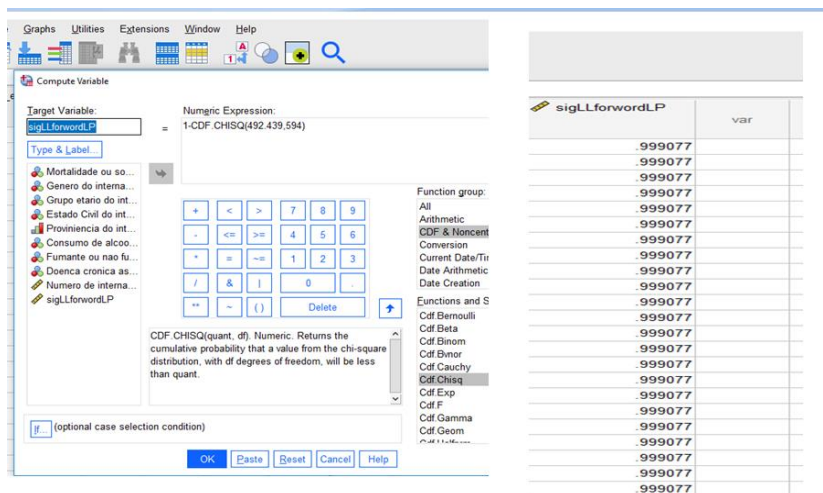


Imagem C: calculo da significância de -2LL do Critério Forward LR

Tabela 9.2: Significancia em etapas do Criterio Forward LR

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	Grupo etario do internado			198.582	5	.000			
	Grupo etario do internado(1)	-2.783	.388	51.465	1	.000	.062	.029	.132
	Grupo etario do internado(2)	-3.058	.394	60.289	1	.000	.047	.022	.102
	Grupo etario do internado(3)	-3.027	.338	80.008	1	.000	.048	.025	.094
	Grupo etario do internado(4)	.178	.293	.372	1	.542	1.195	.674	2.122
	Grupo etario do internado(5)	-2.177	.419	27.033	1	.000	.113	.050	.258
	Constant	1.208	.224	29.202	1	.000	3.346		
Step 2 ^b	Grupo etario do internado			184.063	5	.000			
	Grupo etario do internado(1)	-2.940	.403	53.126	1	.000	.053	.024	.117
	Grupo etario do internado(2)	-3.228	.409	62.157	1	.000	.040	.018	.088
	Grupo etario do internado(3)	-3.267	.356	84.070	1	.000	.038	.019	.077
	Grupo etario do internado(4)	-.079	.300	.070	1	.792	.924	.513	1.663
	Grupo etario do internado(5)	-2.572	.451	32.505	1	.000	.076	.032	.185
	Doenca cronica associada(1)	1.703	.295	33.397	1	.000	5.488	3.081	9.778
	Constant	1.076	.226	22.552	1	.000	2.932		
Step 3 ^c	Grupo etario do internado			176.304	5	.000			
	Grupo etario do internado(1)	-2.934	.403	53.027	1	.000	.053	.024	.117

	Grupo etario do internado(2)	-3.221	.409	62.024	1	.000	.040	.018	.089
	Grupo etario do internado(3)	-3.655	.387	89.320	1	.000	.026	.012	.055
	Grupo etario do internado(4)	-.351	.305	1.322	1	.250	.704	.387	1.280
	Grupo etario do internado(5)	-3.021	.491	37.899	1	.000	.049	.019	.128
	Fumante ou nao fumante(1)	2.086	.458	20.717	1	.000	8.054	3.280	19.777
	Doenca cronica associada(1)	1.669	.303	30.391	1	.000	5.309	2.933	9.612
	Constant	1.077	.226	22.612	1	.000	2.935		
Step 4 ^d	Grupo etario do internado			171.956	5	.000			
	Grupo etario do internado(1)	-2.989	.413	52.460	1	.000	.050	.022	.113
	Grupo etario do internado(2)	-3.226	.412	61.197	1	.000	.040	.018	.089
	Grupo etario do internado(3)	-3.737	.398	88.060	1	.000	.024	.011	.052
	Grupo etario do internado(4)	-.265	.310	.730	1	.393	.767	.417	1.409
	Grupo etario do internado(5)	-2.904	.495	34.429	1	.000	.055	.021	.145
	Provincia do internado			12.896	3	.005			
	Provincia do internado(1)	.308	.275	1.251	1	.263	1.361	.793	2.334
	Provincia do internado(2)	-.502	.314	2.561	1	.110	.605	.327	1.119
	Provincia do internado(3)	1.019	.396	6.623	1	.010	2.771	1.275	6.021
	Fumante ou nao fumante(1)	2.099	.461	20.683	1	.000	8.156	3.301	20.151
	Doenca cronica associada(1)	1.729	.305	32.200	1	.000	5.636	3.102	10.241

	Constant	.946	.259	13.326	1	.000	2.576		
Step 5 ^e	Grupo etario do internado			157.116	5	.000			
	Grupo etario do internado(1)	-3.007	.415	52.511	1	.000	.049	.022	.112
	Grupo etario do internado(2)	-3.244	.414	61.495	1	.000	.039	.017	.088
	Grupo etario do internado(3)	-3.928	.456	74.357	1	.000	.020	.008	.048
	Grupo etario do internado(4)	-.864	.677	1.631	1	.202	.421	.112	1.587
	Grupo etario do internado(5)	-3.598	.790	20.746	1	.000	.027	.006	.129
	Estado Civil do internado			8.594	3	.035			
	Estado Civil do internado(1)	.383	.609	.394	1	.530	1.466	.444	4.841
	Estado Civil do internado(2)	1.818	.806	5.081	1	.024	6.158	1.268	29.911
	Estado Civil do internado(3)	1.612	.906	3.167	1	.075	5.012	.849	29.582
	Provincia do internado			14.037	3	.003			
	Provincia do internado(1)	.383	.281	1.858	1	.173	1.466	.846	2.541
	Provincia do internado(2)	-.452	.318	2.016	1	.156	.637	.341	1.187
	Provincia do internado(3)	1.141	.402	8.050	1	.005	3.131	1.423	6.888
	Fumante ou nao fumante(1)	1.935	.466	17.276	1	.000	6.923	2.780	17.239
	Doenca cronica associada(1)	1.752	.311	31.753	1	.000	5.763	3.134	10.598
	Constant	.906	.260	12.187	1	.000	2.476		

a. Variable(s) entered on step 1: Grupo etario do internado.

b. Variable(s) entered on step 2: Doenca cronica associada.

c. Variable(s) entered on step 3: Fumante ou nao fumante.

d. Variable(s) entered on step 4: Provincia do internado.

e. Variable(s) entered on step 5: Estado Civil do internado.

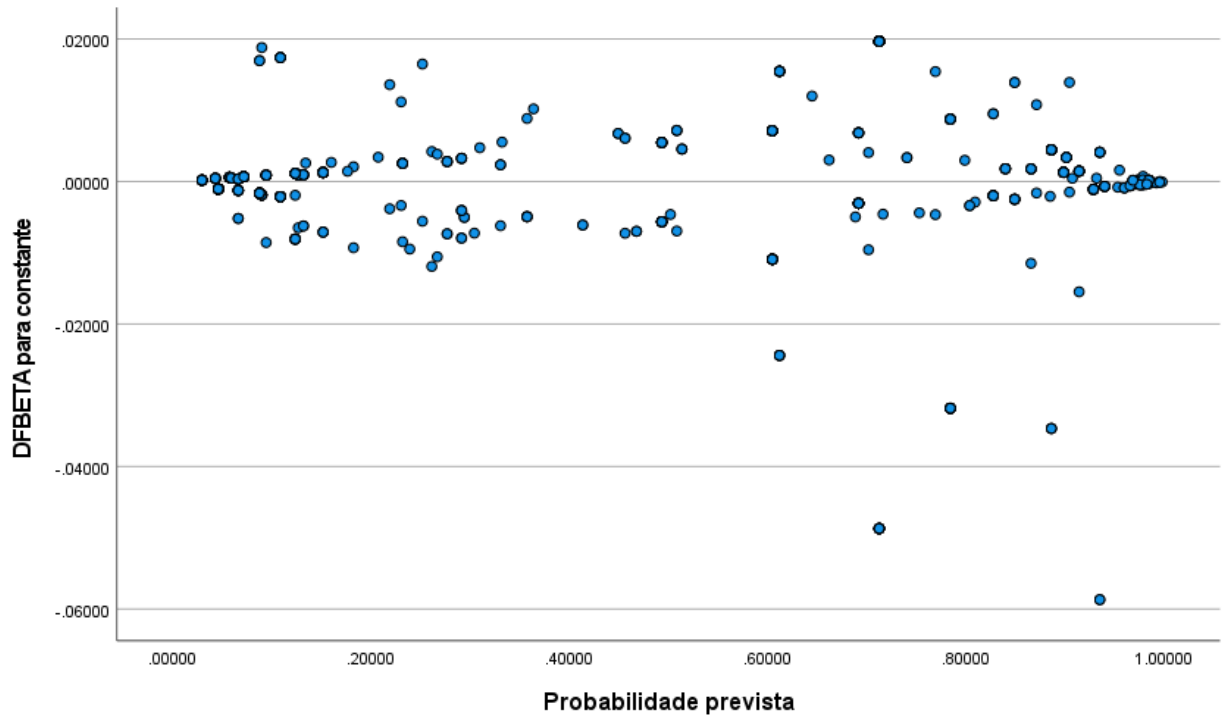
Tabela 9.3: Candidatas a outlier no Critério Forward LP

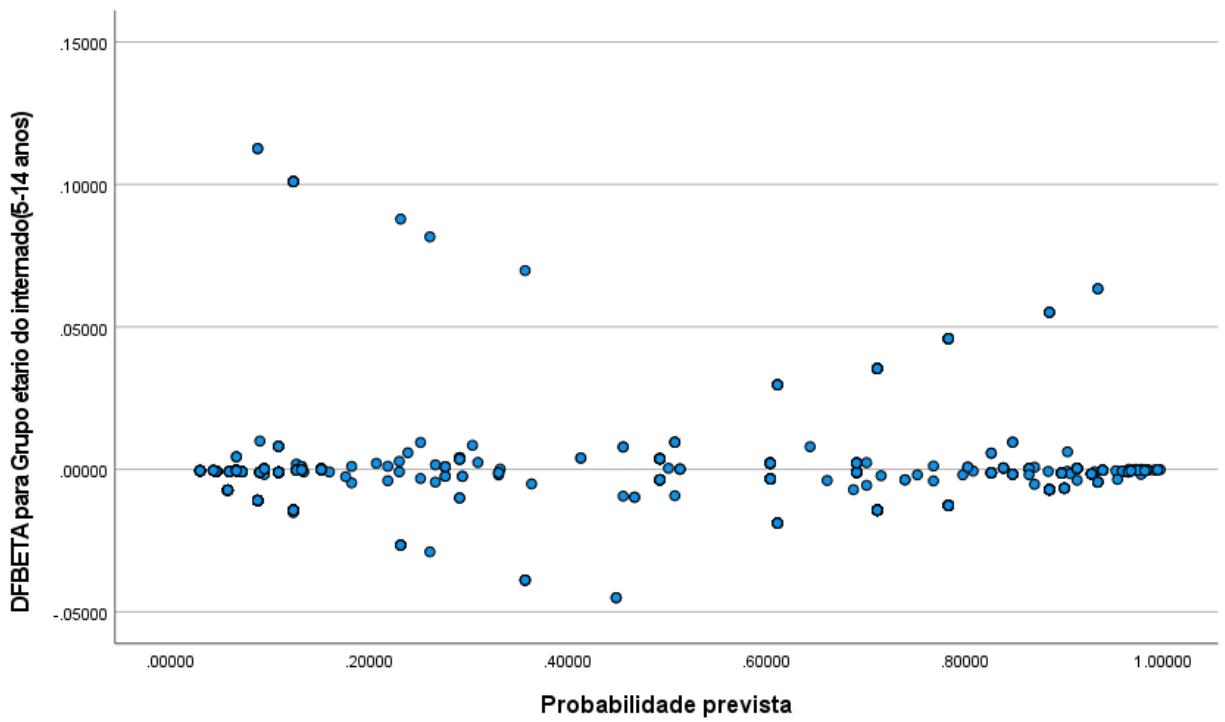
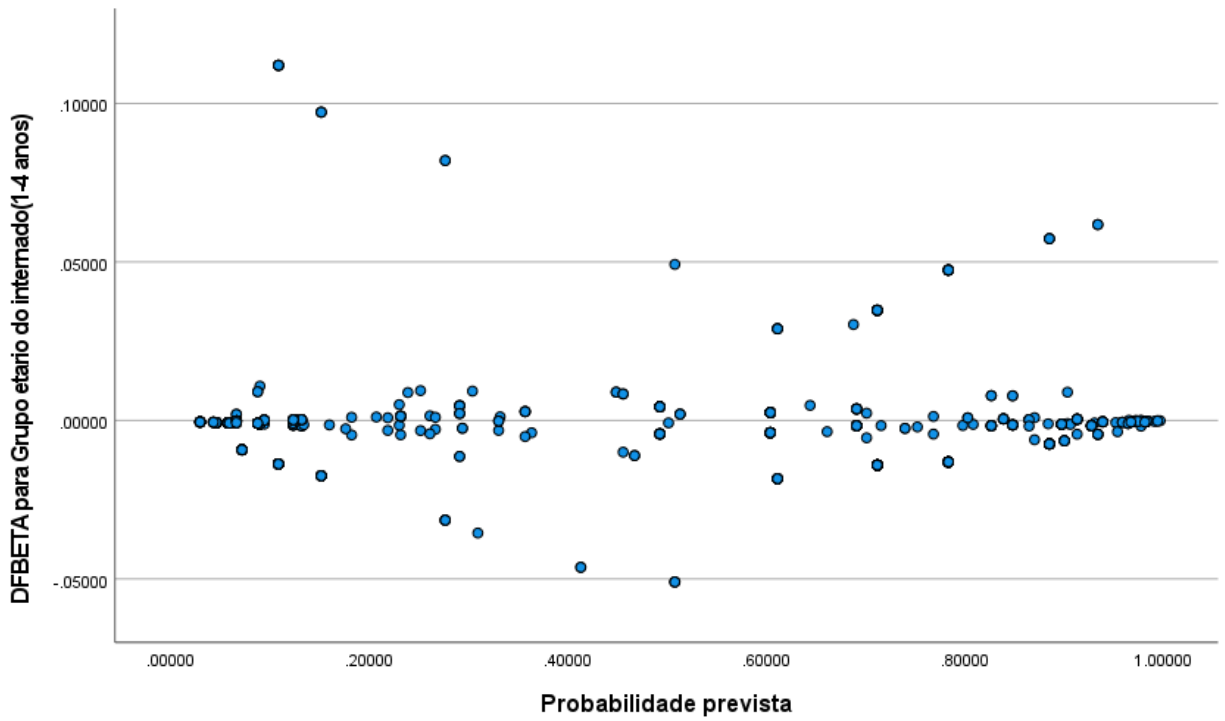
Case	Selected Status ^a	Observed	Predicted	Predicted Group	Temporary Variable		
		Mortalidade ou sobrevivencia			Resid	ZResid	SResid
2	S	O**	.109	A	.891	2.858	2.120
22	S	O**	.124	A	.876	2.658	2.060
31	S	O**	.067	A	.933	3.742	2.340
32	S	O**	.090	A	.910	3.173	2.216
37	S	O**	.132	A	.868	2.560	2.038
49	S	O**	.109	A	.891	2.858	2.120
53	S	O**	.067	A	.933	3.742	2.340
67	S	O**	.124	A	.876	2.658	2.060
73	S	O**	.088	A	.912	3.218	2.217
85	S	O**	.124	A	.876	2.658	2.060
87	S	O**	.095	A	.905	3.090	2.199
99	S	O**	.124	A	.876	2.658	2.060
114	S	O**	.127	A	.873	2.620	2.060
137	S	O**	.109	A	.891	2.858	2.120
159	S	O**	.124	A	.876	2.658	2.060
196	S	O**	.109	A	.891	2.858	2.120
284	S	O**	.109	A	.891	2.858	2.120
285	S	O**	.124	A	.876	2.658	2.060
298	S	O**	.132	A	.868	2.560	2.038
319	S	O**	.088	A	.912	3.218	2.217
337	S	A**	.886	O	-.886	-2.784	-2.101
382	S	A**	.865	O	-.865	-2.534	-2.051
392	S	A**	.934	O	-.934	-3.777	-2.346
399	S	A**	.871	O	-.871	-2.596	-2.055
401	S	A**	.886	O	-.886	-2.784	-2.101
415	S	A**	.934	O	-.934	-3.777	-2.346
425	S	A**	.904	O	-.904	-3.068	-2.207
465	S	A**	.886	O	-.886	-2.784	-2.101
514	S	A**	.914	O	-.914	-3.253	-2.236

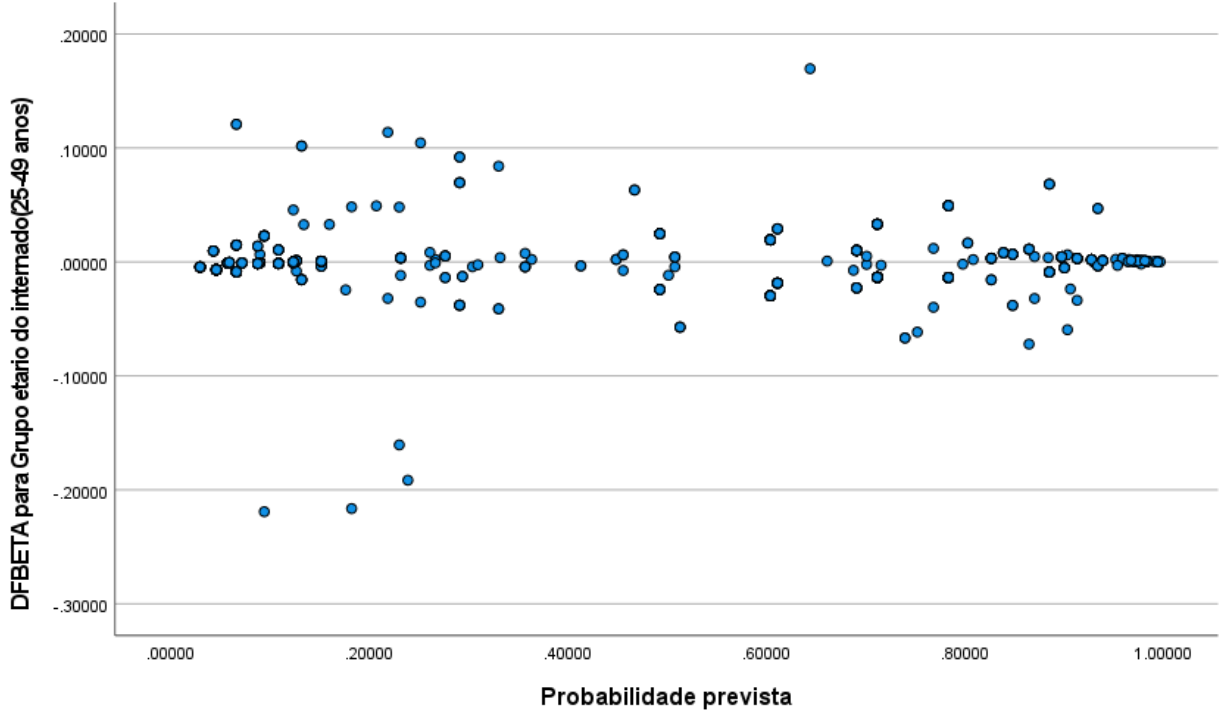
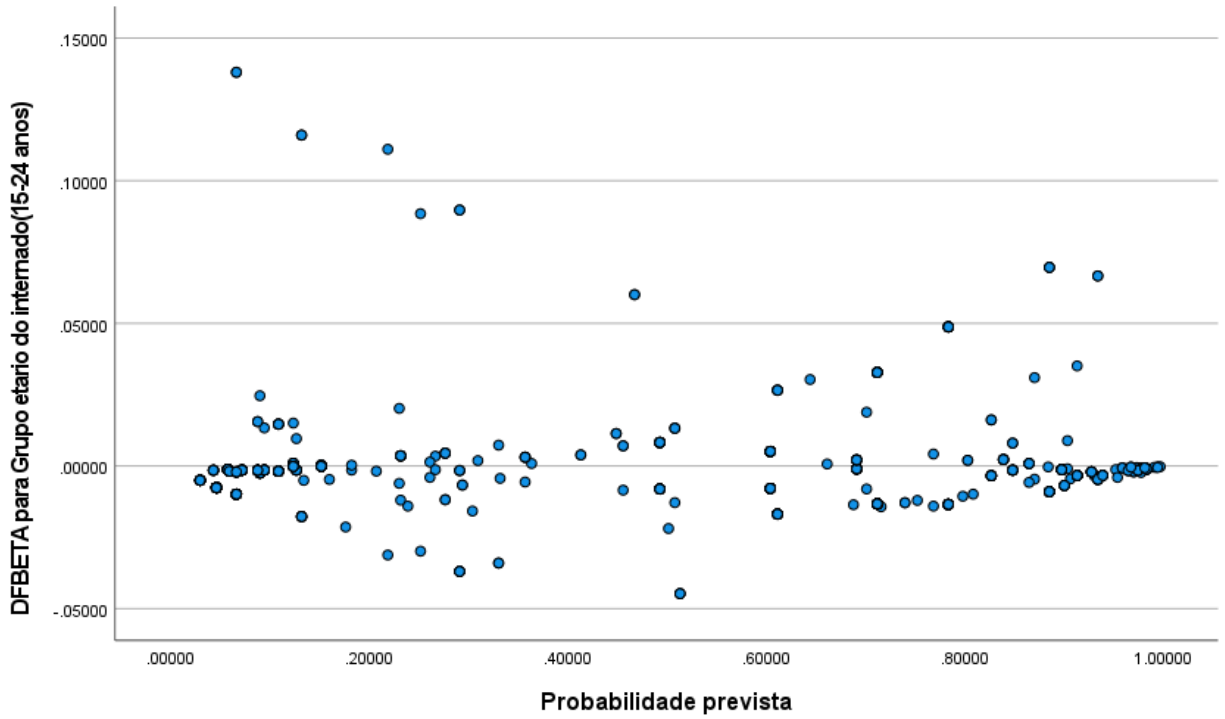
a. S = Selected, U = Unselected cases, and ** = Misclassified cases.

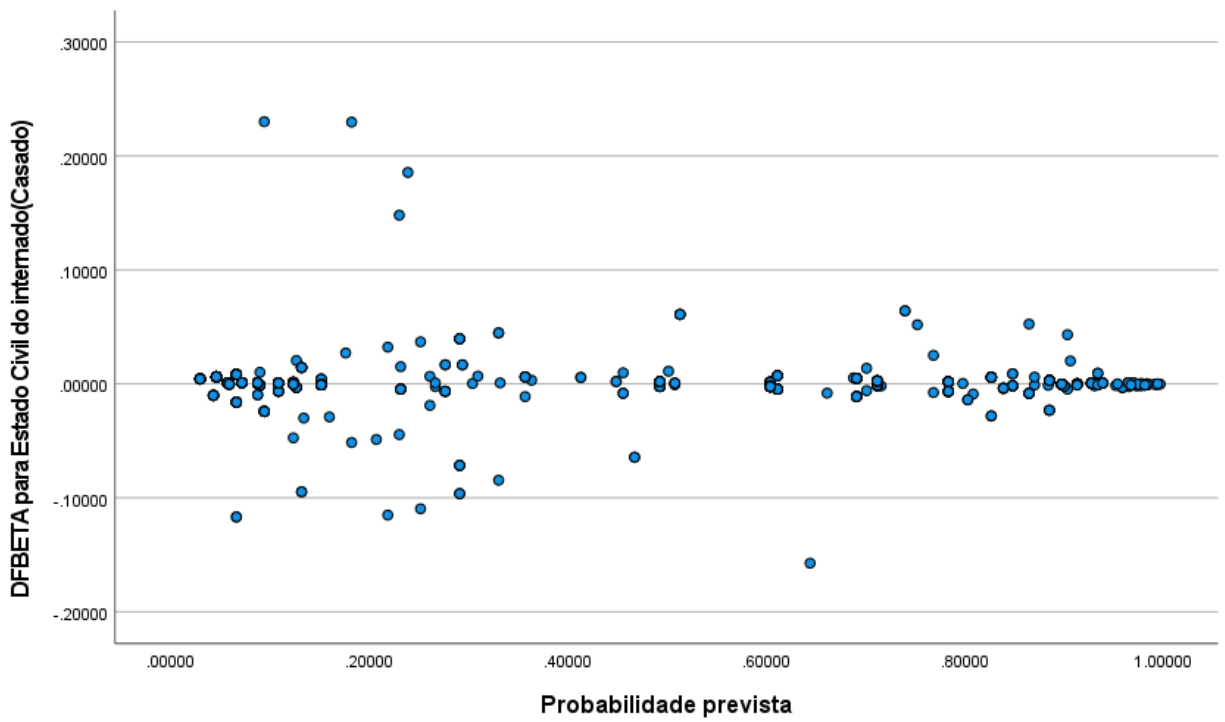
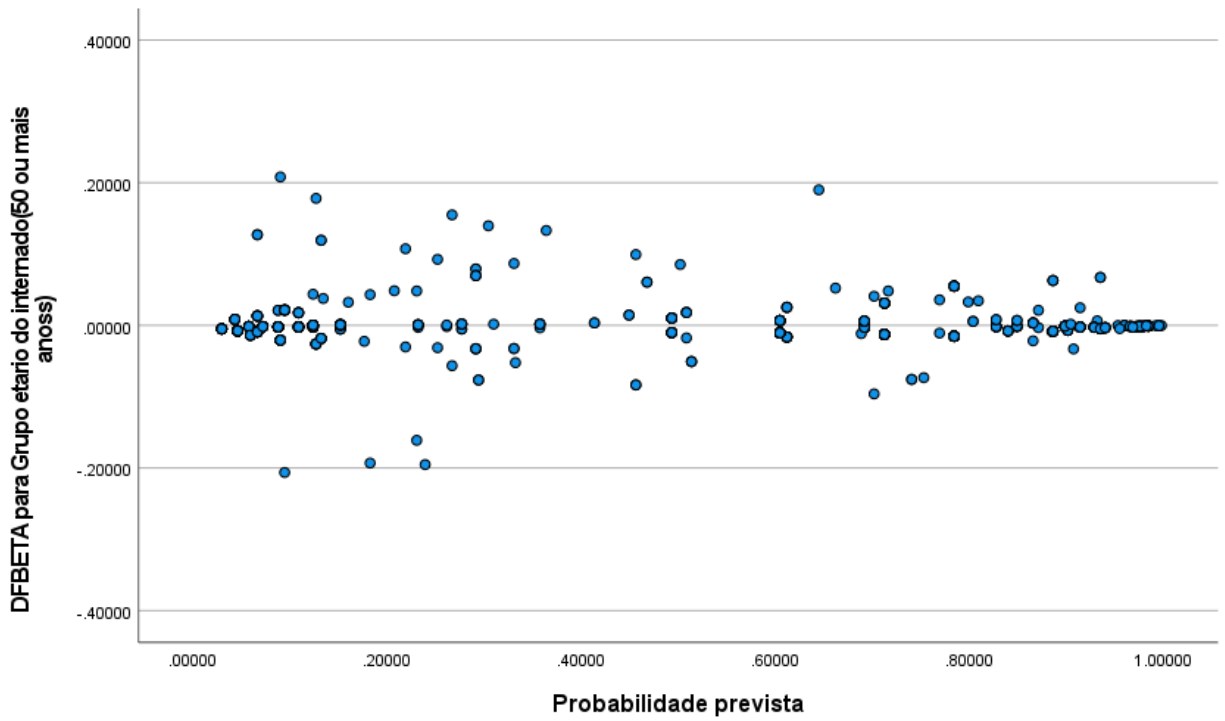
b. Cases with studentized residuals greater than 2.000 are listed.

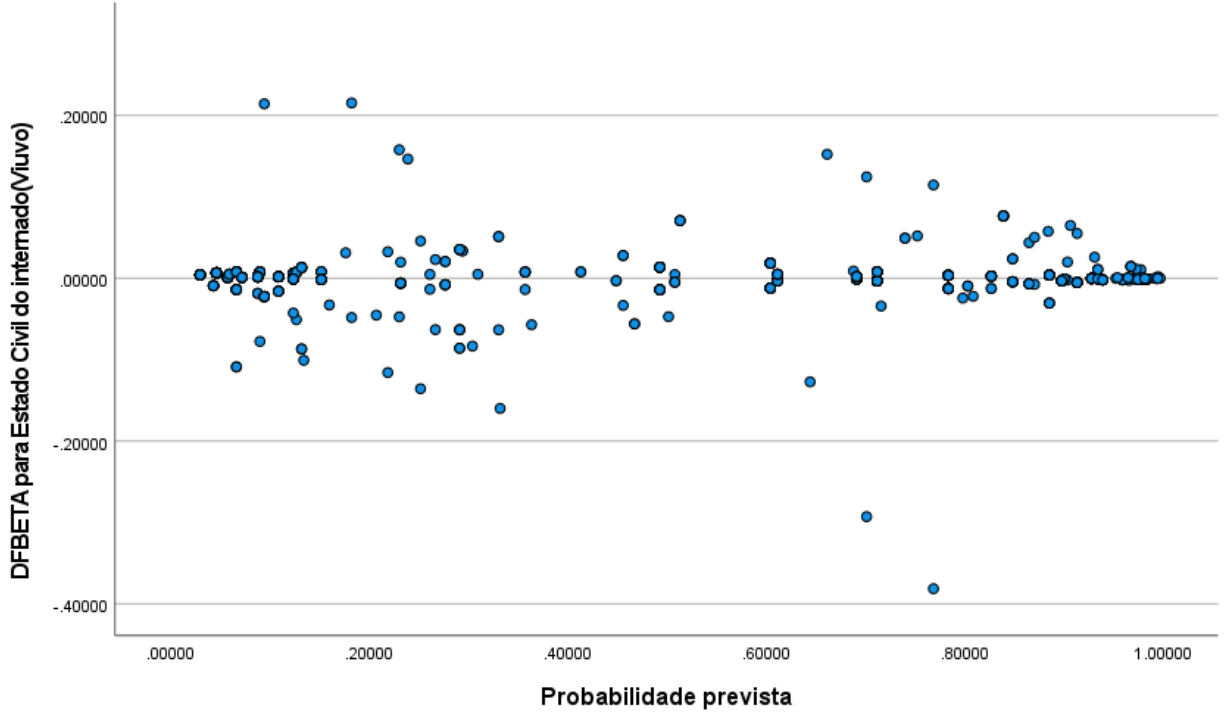
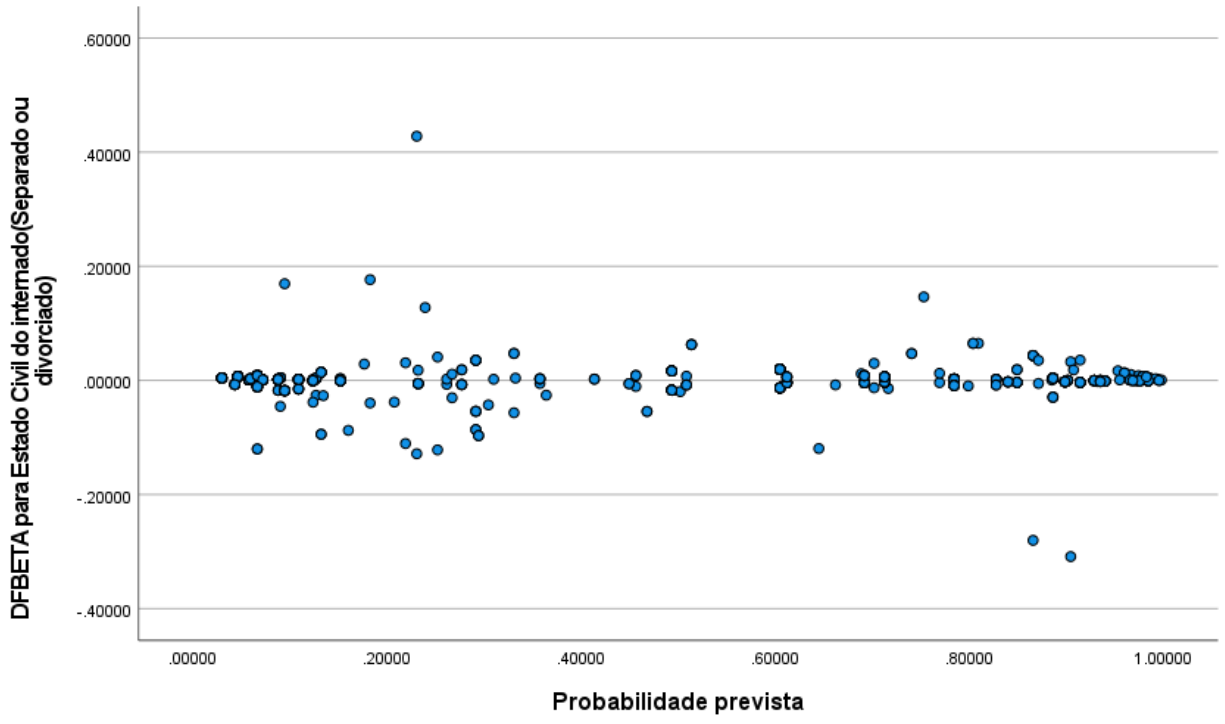
A seguir tem-se os gráficos de DfBetas do Critério Forward LR

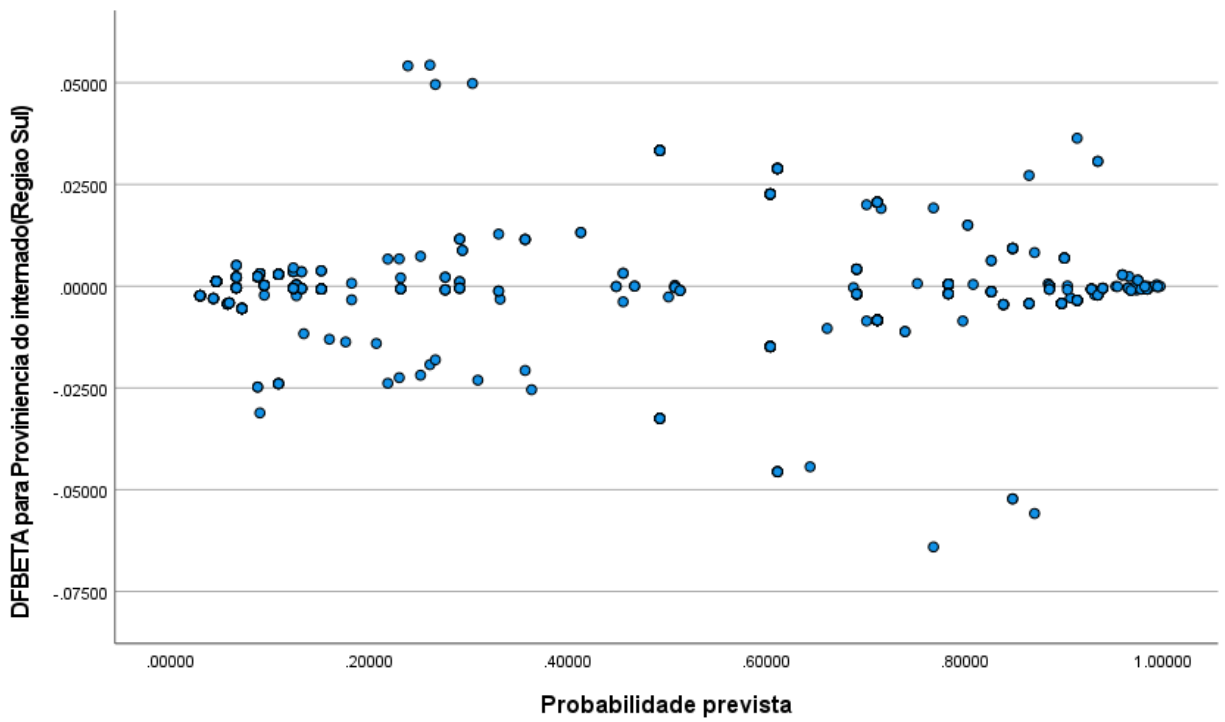
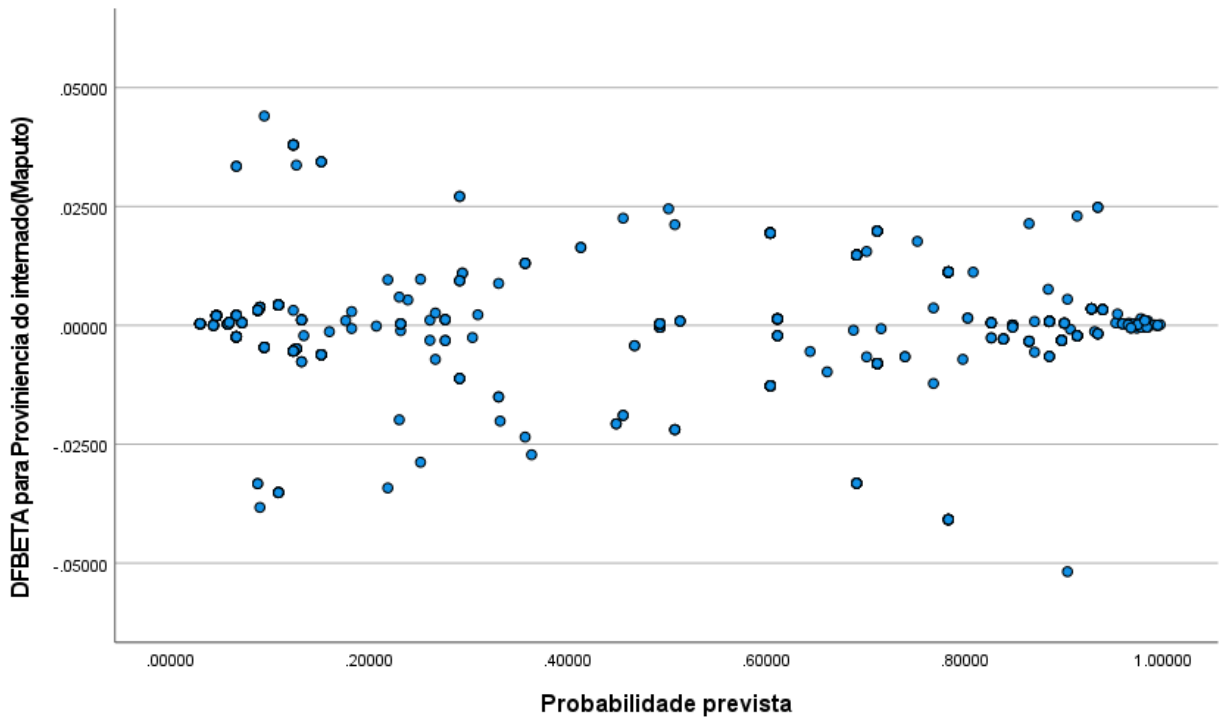


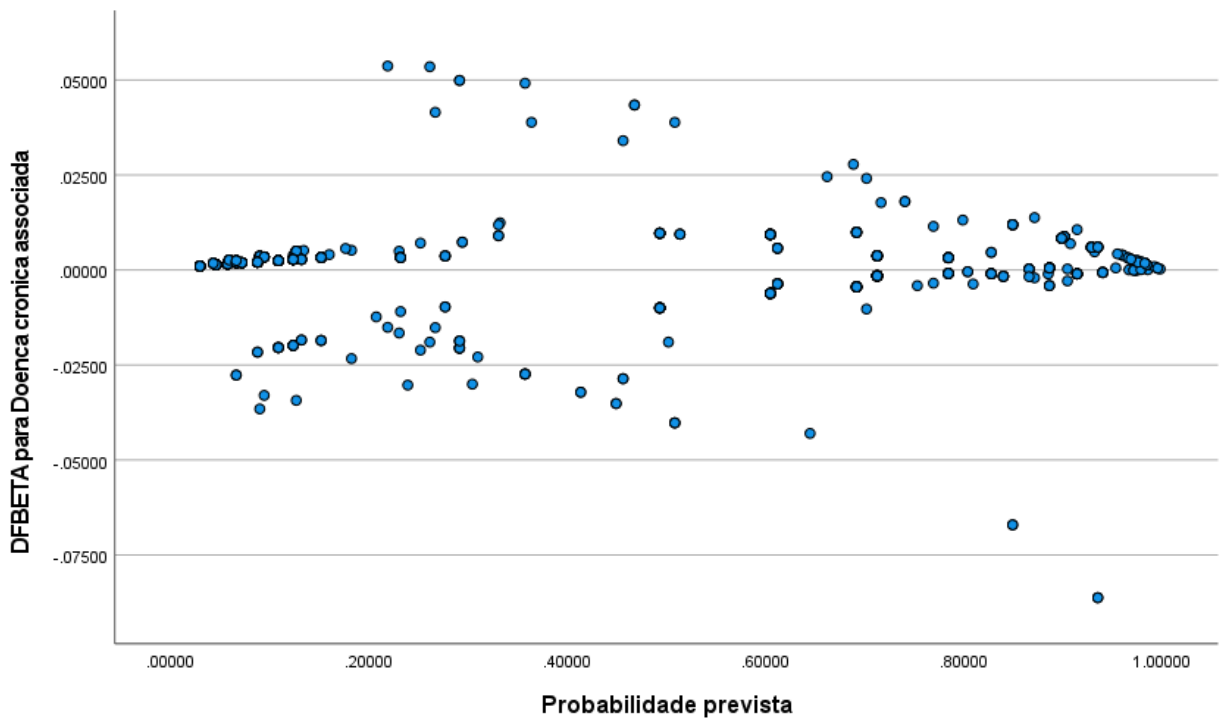
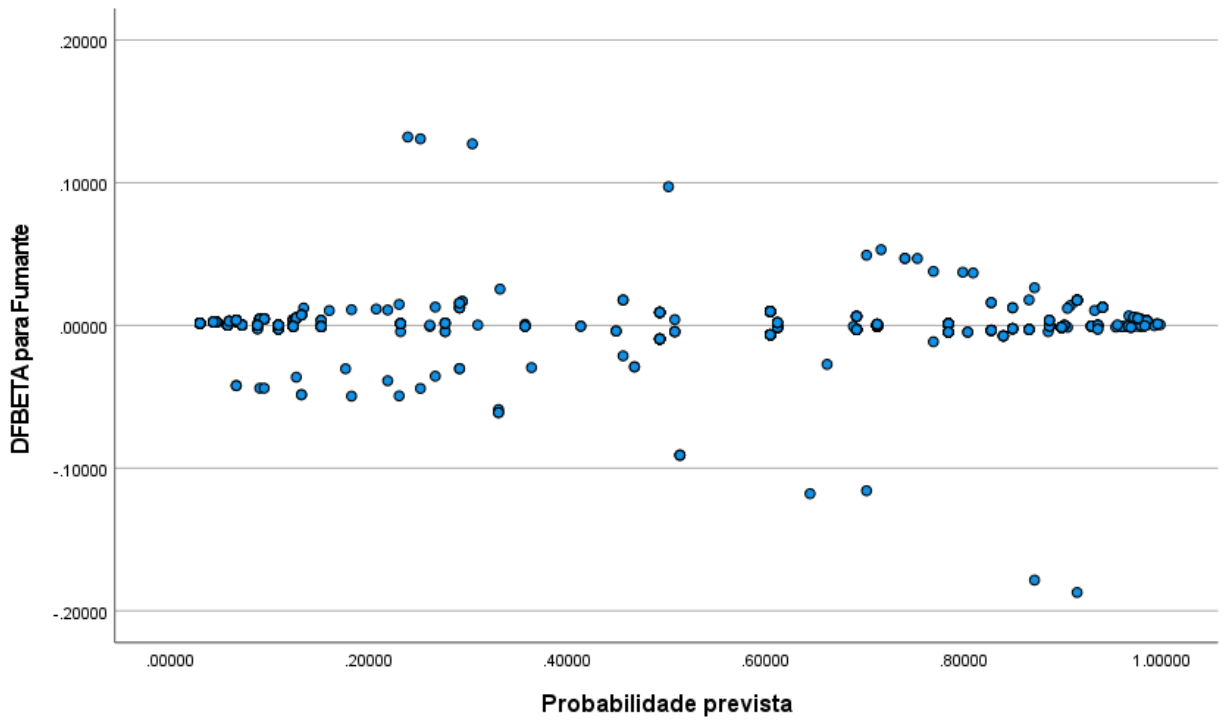












**ANEXO X:
CRITÉRIO BACKWARD
LR**

**Tabela do Teste de rácio de verossimilhança
(Omnibus) do Criterio Backward LR**

		Chi-square	df	Sig.
Step 1	Step	346.518	16	.000
	Block	346.518	16	.000
	Model	346.518	16	.000
Step 2 ^a	Step	.000	1	1.000
	Block	346.518	15	.000
	Model	346.518	15	.000
Step 3 ^a	Step	-.075	1	.784
	Block	346.442	14	.000
	Model	346.442	14	.000
Step 4 ^a	Step	-.130	1	.718
	Block	346.312	13	.000
	Model	346.312	13	.000

a. A negative Chi-squares value indicates that the Chi-squares value has decreased from the previous step.

**Tabela do Teste de Hosmer e
Lemshow do Criterio Backward LR**

Step	Chi-square	df	Sig.
1	8.791	8	.360
2	5.910	8	.657
3	12.400	8	.134
4	12.044	8	.149

Tabela da Classificacao do Criterio Backward LR

Observed		Predicted		Percentage Correct
		Mortalidade ou sobrevivencia Alta	Obito	
Mortalidade ou sobrevivencia	Alta	264	65	80.2
	Obito	45	234	83.9
Overall Percentage				81.9
Mortalidade ou sobrevivencia	Alta	264	65	80.2
	Obito	45	234	83.9

Overall Percentage				81.9
Mortalidade ou sobrevivencia	Alta	262	67	79.6
	Obito	47	232	83.2
Overall Percentage				81.3
Mortalidade ou sobrevivencia	Alta	267	62	81.2
	Obito	50	229	82.1
Overall Percentage				81.6

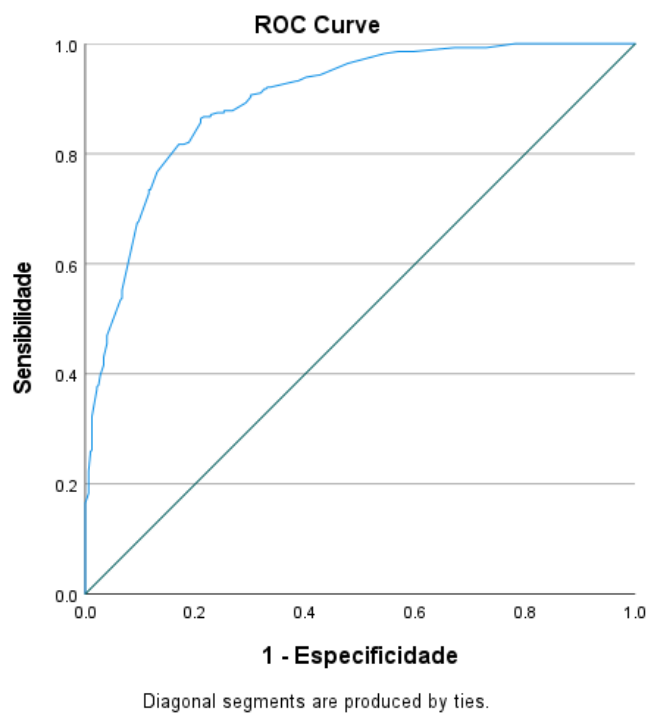


Grafico da curava ROC do Criterio Backward LR

Tabela da Area da Curva ROC do Criterio Backward LR

Test Result Variable(s): Probabilidade prevista

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.898	.012	.000	.874	.922

The test result variable(s): Probabilidade prevista has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Tabela da Confiabilidade do Critério Backward LR

		Value	Asymptotic Standard Error ^a	Approximate T ^b	Approximate Significance
Measure of Agreement	Kappa	.630	.032	15.553	.000
N of Valid Cases		608			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Tabela de Significancia das variaveis pelo Critério Backward LR

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	Genero do internado(1)	.000	.230	.000	1	1.000	1.000	.637	1.570
	Grupo etario do internado			149.275	5	.000			
	Grupo etario do internado(1)	-3.003	.415	52.275	1	.000	.050	.022	.112
	Grupo etario do internado(2)	-3.244	.414	61.300	1	.000	.039	.017	.088
	Grupo etario do internado(3)	-3.894	.511	58.055	1	.000	.020	.007	.055
	Grupo etario do internado(4)	-.849	.715	1.408	1	.235	.428	.105	1.739
	Grupo etario do internado(5)	-3.623	.846	18.356	1	.000	.027	.005	.140
	Estado Civil do internado			8.660	3	.034			
	Estado Civil do internado(1)	.406	.611	.442	1	.506	1.501	.453	4.969
	Estado Civil do internado(2)	1.843	.810	5.178	1	.023	6.315	1.291	30.883
	Estado Civil do internado(3)	1.630	.909	3.217	1	.073	5.104	.860	30.309
	Provincia do internado			14.123	3	.003			
	Provincia do internado(1)	.392	.282	1.933	1	.164	1.479	.852	2.569

	Provinciãncia do internado(2)	-.461	.320	2.076	1	.150	.631	.337	1.181
	Provinciãncia do internado(3)	1.145	.403	8.067	1	.005	3.143	1.426	6.926
	Consumo de alcool no internado(1)	-.091	.332	.075	1	.784	.913	.476	1.751
	Fumante ou nao fumante(1)	1.936	.466	17.267	1	.000	6.932	2.781	17.278
	Doença cronica associada(1)	1.749	.312	31.462	1	.000	5.749	3.120	10.593
	Numero de internamentos anteriores	.057	.156	.133	1	.716	1.058	.779	1.437
	Constant	.889	.297	8.935	1	.003	2.432		
Step 2 ^a	Grupo etario do internado			149.364	5	.000			
	Grupo etario do internado(1)	-3.003	.415	52.333	1	.000	.050	.022	.112
	Grupo etario do internado(2)	-3.244	.414	61.456	1	.000	.039	.017	.088
	Grupo etario do internado(3)	-3.894	.511	58.085	1	.000	.020	.007	.055
	Grupo etario do internado(4)	-.849	.715	1.410	1	.235	.428	.105	1.737
	Grupo etario do internado(5)	-3.623	.846	18.359	1	.000	.027	.005	.140
	Estado Civil do internado			8.662	3	.034			
	Estado Civil do internado(1)	.406	.611	.442	1	.506	1.501	.453	4.969
	Estado Civil do internado(2)	1.843	.810	5.179	1	.023	6.315	1.291	30.879
	Estado Civil do internado(3)	1.630	.909	3.217	1	.073	5.104	.860	30.308
	Provinciãncia do internado			14.124	3	.003			
	Provinciãncia do internado(1)	.392	.282	1.934	1	.164	1.479	.852	2.569
	Provinciãncia do internado(2)	-.461	.319	2.079	1	.149	.631	.337	1.180

	Proviencia do internado(3)	1.145	.403	8.084	1	.004	3.143	1.427	6.920
	Consumo de alcool no internado(1)	-.091	.332	.075	1	.784	.913	.476	1.751
	Fumante ou nao fumante(1)	1.936	.466	17.268	1	.000	6.932	2.781	17.277
	Doenca cronica associada(1)	1.749	.311	31.641	1	.000	5.749	3.126	10.575
	Numero de internamentos anteriores	.057	.156	.133	1	.716	1.058	.780	1.437
	Constant	.889	.264	11.336	1	.001	2.432		
Step 3 ^a	Grupo etario do internado			155.748	5	.000			
	Grupo etario do internado(1)	-3.003	.415	52.335	1	.000	.050	.022	.112
	Grupo etario do internado(2)	-3.246	.414	61.536	1	.000	.039	.017	.088
	Grupo etario do internado(3)	-3.954	.462	73.282	1	.000	.019	.008	.047
	Grupo etario do internado(4)	-.904	.685	1.743	1	.187	.405	.106	1.550
	Grupo etario do internado(5)	-3.678	.822	20.041	1	.000	.025	.005	.126
	Estado Civil do internado			8.664	3	.034			
	Estado Civil do internado(1)	.400	.610	.430	1	.512	1.492	.451	4.934
	Estado Civil do internado(2)	1.846	.809	5.207	1	.022	6.336	1.297	30.937
	Estado Civil do internado(3)	1.612	.906	3.169	1	.075	5.014	.850	29.584
	Proviencia do internado			14.101	3	.003			
	Proviencia do internado(1)	.391	.282	1.929	1	.165	1.479	.851	2.568
	Proviencia do internado(2)	-.460	.319	2.076	1	.150	.631	.338	1.180
	Proviencia do internado(3)	1.144	.403	8.065	1	.005	3.140	1.425	6.916

	Fumante ou nao fumante(1)	1.941	.465	17.398	1	.000	6.968	2.799	17.349
	Doenca cronica associada(1)	1.752	.311	31.736	1	.000	5.767	3.135	10.608
	Numero de internamentos anteriores	.056	.156	.130	1	.719	1.058	.779	1.436
	Constant	.889	.264	11.338	1	.001	2.433		
Step 4 ^a	Grupo etario do internado			157.116	5	.000			
	Grupo etario do internado(1)	-3.007	.415	52.511	1	.000	.049	.022	.112
	Grupo etario do internado(2)	-3.244	.414	61.495	1	.000	.039	.017	.088
	Grupo etario do internado(3)	-3.928	.456	74.357	1	.000	.020	.008	.048
	Grupo etario do internado(4)	-.864	.677	1.631	1	.202	.421	.112	1.587
	Grupo etario do internado(5)	-3.598	.790	20.746	1	.000	.027	.006	.129
	Estado Civil do internado			8.594	3	.035			
	Estado Civil do internado(1)	.383	.609	.394	1	.530	1.466	.444	4.841
	Estado Civil do internado(2)	1.818	.806	5.081	1	.024	6.158	1.268	29.911
	Estado Civil do internado(3)	1.612	.906	3.167	1	.075	5.012	.849	29.582
	Provincia do internado			14.037	3	.003			
	Provincia do internado(1)	.383	.281	1.858	1	.173	1.466	.846	2.541
	Provincia do internado(2)	-.452	.318	2.016	1	.156	.637	.341	1.187
	Provincia do internado(3)	1.141	.402	8.050	1	.005	3.131	1.423	6.888
	Fumante ou nao fumante(1)	1.935	.466	17.276	1	.000	6.923	2.780	17.239
	Doenca cronica associada(1)	1.752	.311	31.753	1	.000	5.763	3.134	10.598
	Constant	.906	.260	12.187	1	.000	2.476		

a. Variable(s) entered on step 1: Genero do internado, Grupo etario do internado, Estado Civil do internado, Proviencia do internado, Consumo de alcool no internado, Fumante ou nao fumante, Doenca cronica associada, Numero de internamentos anteriores.

Observacoes candidatas a outlier do Criterio Backward LR

Case	Selected Status ^a	Observed	Predicted	Predicted Group	Temporary Variable		
		Mortalidade ou sobrevivencia			Resid	ZResid	SResid
2	S	O**	.109	A	.891	2.858	2.120
22	S	O**	.124	A	.876	2.658	2.060
31	S	O**	.067	A	.933	3.742	2.340
32	S	O**	.090	A	.910	3.173	2.216
37	S	O**	.132	A	.868	2.560	2.038
49	S	O**	.109	A	.891	2.858	2.120
53	S	O**	.067	A	.933	3.742	2.340
67	S	O**	.124	A	.876	2.658	2.060
73	S	O**	.088	A	.912	3.218	2.217
85	S	O**	.124	A	.876	2.658	2.060
87	S	O**	.095	A	.905	3.090	2.199
99	S	O**	.124	A	.876	2.658	2.060
114	S	O**	.127	A	.873	2.620	2.060
137	S	O**	.109	A	.891	2.858	2.120
159	S	O**	.124	A	.876	2.658	2.060
196	S	O**	.109	A	.891	2.858	2.120
284	S	O**	.109	A	.891	2.858	2.120
285	S	O**	.124	A	.876	2.658	2.060
298	S	O**	.132	A	.868	2.560	2.038
319	S	O**	.088	A	.912	3.218	2.217
337	S	A**	.886	O	-.886	-2.784	-2.101
382	S	A**	.865	O	-.865	-2.534	-2.051
392	S	A**	.934	O	-.934	-3.777	-2.346
399	S	A**	.871	O	-.871	-2.596	-2.055
401	S	A**	.886	O	-.886	-2.784	-2.101
415	S	A**	.934	O	-.934	-3.777	-2.346
425	S	A**	.904	O	-.904	-3.068	-2.207
465	S	A**	.886	O	-.886	-2.784	-2.101
514	S	A**	.914	O	-.914	-3.253	-2.236

- a. S = Selected, U = Unselected cases, and ** = Misclassified cases.
- b. Cases with studentized residuals greater than 2.000 are listed.

**ANEXO XI:
CRITÉRIO ENTER
(OBSERVAÇÕES E
VARIÁVEIS IGUAIS AOS
CRITÉRIOS LR)**

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	346.312	13	.000
	Block	346.312	13	.000
	Model	346.312	13	.000

Fonte: Output do SPSS

Model Summary

Step	-2 Log likelihood	Cox & Snell R	Nagelkerke R
		Square	Square
1	492.439 ^a	.434	.580

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Fonte: Output do SPSS

Observed		Predicted		Percentage Correct
		Mortalidade ou sobrevivencia Alta	Obito	
Mortalidade ou sobrevivencia	Alta	267	62	81.2
	Obito	50	229	82.1
Overall Percentage				81.6

Fonte: Output do SPSS

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Grupo etario do internado			157.116	5	.000	
	Grupo etario do internado(1)	-3.007	.415	52.511	1	.000	.049
	Grupo etario do internado(2)	-3.244	.414	61.495	1	.000	.039
	Grupo etario do internado(3)	-3.928	.456	74.357	1	.000	.020
	Grupo etario do internado(4)	-.864	.677	1.631	1	.202	.421
	Grupo etario do internado(5)	-3.598	.790	20.746	1	.000	.027
	Estado Civil do internado			8.594	3	.035	
	Estado Civil do internado(1)	.383	.609	.394	1	.530	1.466

Estado Civil do internado(2)	1.818	.806	5.081	1	.024	6.158
Estado Civil do internado(3)	1.612	.906	3.167	1	.075	5.012
Provincia do internado			14.037	3	.003	
Provincia do internado(1)	.383	.281	1.858	1	.173	1.466
Provincia do internado(2)	-.452	.318	2.016	1	.156	.637
Provincia do internado(3)	1.141	.402	8.050	1	.005	3.131
Fumante ou nao fumante(1)	1.935	.466	17.276	1	.000	6.923
Doenca cronica associada(1)	1.752	.311	31.753	1	.000	5.763
Constant	.906	.260	12.187	1	.000	2.476

a. Variable(s) entered on step 1: Grupo etario do internado, Estado Civil do internado, Provincia do internado, Fumante ou nao fumante, Doenca cronica associada.

Fonte: Output do SPSS