



FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em
Estatística

Análise dos Factores que influenciam na Redução
da Prática do Turismo Doméstico em Moçambique.

Período de estudo: 2019 a 2020

Autor: Félix Domingos Mucavel

Maputo, Fevereiro de 2024



FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em
Estatística

Análise dos Factores que influenciam na Redução
da Prática do Turismo Doméstico em Moçambique.

Período de estudo: 2019 a 2020

Autor: Félix Domingos Mucavel

Supervisor: Bonifácio José, MSc., Kishinhev

Maputo, Fevereiro de 2024

Declaração de Honra

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

Maputo, Fevereiro de 2024

(Félix Domingos Mucavel)

AGRADECIMENTOS

Primeiro agradecer a Deus, o todo-poderoso, pela vida, saúde, por iluminar os meus caminhos neste percurso, que tanto me ajuda nos momentos de aflição e desânimo, dando-me força para nunca desistir.

Agradeço ao meu supervisor Dr. Bonifácio José, pela paciência, atenção e apoio para que eu pudesse concluir o trabalho de licenciatura e a todos docentes da seção de Estatística do DMI-UEM pela sabedoria transmitida ao longo deste percurso. Aos funcionários do DMI, pelo apoio na disponibilização do material e informação essencial durante o curso.

Aos meus pais, esposa, familiares e colegas de carteira e do curso no geral e amigos. De forma resumida, agradeço a todos aqueles que directa ou indirectamente me incentivaram, apoiaram e ajudaram a superar este desafio. Que Deus vos abençoe! Este não é o fim mas o começo de uma nova etapa.

Resumo

O turismo global sofreu uma perda significativa das suas receitas, onde vários sectores dessa indústria foram afectados devido a pandemia da Covid-19. De forma particular, O turismo moçambicano não foi apenas assolado pela pandemia, mas também pelos desastres naturais como os ciclones e cheias, os danos causados por esses eventos climáticos extremos e pela covid-19, trouxeram consequências negativas em todas as dimensões: económicas, políticas, sociais e geográficas, as quais, pela força desse impacto, sugeriram o desenvolvimento do presente estudo, cujo objectivo é analisar os factores que influenciam para a redução da prática do turismo doméstico no período de 2019 a 2020 em Moçambique. Para fazer a análise foi usada uma amostra de $n = 480$ observações, que foram filtradas da base de dados da IOF 2019/2020, tendo se usado a variável dependente “viajou” como a chave para tal filtração. Da análise constatou-se que, a maioria dos inquiridos são do sexo masculino (62.29%), onde cerca de 67,71% destes são assalariados, cerca de 33,96% têm as suas idades compreendidas dos 28 à 38 e vivem maritalmente cerca de 60,63%. Foram realizados testes de independência entre as variáveis Zona, Sexo, Civil, Nível, Emprego, Idade, Alojamento, Transporte, Cheias, Ciclones e a variável Viajou, onde constatou-se que, apenas as variáveis nível académico e o ciclone demonstraram estarem associados à prática do turismo doméstico. Para estimar o modelo que melhor explica a prática do turismo doméstico foi usada a regressão logística binária, onde constatou-se que, as variáveis independentes estatisticamente significativas para a probabilidade do inquirido praticar o turismo doméstico são a província, sexo, idade, escolaridade e o ciclone. Com os resultados alcançados espera se usar a informação para incentivar a cooperação das comunidades com as entidades empregadores da área do turismo a cultivar um ambiente que melhoraria o turismo em Moçambique, bem como procurar condições para o estabelecimento de parceiros que se identifiquem com a causa da comunidade e planificação acertiva por parte do governo em relação as vias de acesso que facilite e estimule a prática do turismo doméstico.

Palavras-chave: Turismo; Turismo doméstico; Análise de associação; Regressão logística.

Abstract

Global tourism suffered a significant loss of revenue, where several sectors of this industry were affected due to the Covid-19 pandemic. In particular, Mozambican tourism was not only devastated by the pandemic, but also by natural disasters such as cyclones and floods, the damage caused by these extreme weather events and by Covid-19, brought negative consequences in all dimensions: economic, political, social and geographical, which, due to the strength of this impact, suggested the development of the present study, whose objective is to analyze the factors that contribute to the reduction in the practice of domestic tourism in the period from 2019 to 2020 in Mozambique. To carry out the analysis, a sample of $n = 480$ observations was used, which were filtered from the IOF 2019/2020 database, using the dependent variable “travelled” as the key for such filtration. From the analysis it was found that the majority of respondents are male (62.29%), where around 67.71% of these are salaried workers, around 33.96% are aged between 28 and 38 and live together for around 60.63%. Independence tests were carried out between the variables Zone, Sex, Civil, Level, Employment, Age, Accommodation, Transport, Floods, Cyclones and the variable Traveled, where it was found that only the variables academic level and the cyclone proved to be associated with practice of domestic tourism. To estimate the model that best explains the practice of domestic tourism, binary logistic regression was used, where it was found that the statistically significant independent variables for the probability of the respondent practicing domestic tourism are the province, gender, age, education and the cyclone. With the results achieved, it is hoped to use the information to encourage cooperation between communities and employers in the tourism sector to cultivate an environment that would improve tourism in Mozambique, as well as seek conditions for the establishment of community partnerships and accurate planning on the part of the government in relation to access routes that facilitate and encourage the practice of domestic tourism.

Key-Word: Tourism; Tourism domestic; Association Analyses, Logistic regression.

Lista de Abreviaturas

- OR - Odds Ratio
- RL - Regressão Logística
- ROC - Receiver Operating Characteristic
- VIF - Variance Inflation Factor
- AIC - Critério de informação de Akaike
- BIC - Critério de Informação de Bayes
- OMT - Organização Mundial de Turismo
- UEM - Universidade Eduardo Mondlane
- INE - Instituto Nacional de Estatística

Índice

1	Introdução	1
1.1	Contextualização	1
1.2	Definição do problema	2
1.3	Objectivos	3
1.3.1	Objectivo Geral	3
1.3.2	Objectivos específicos:	3
1.4	Justificação	3
1.5	Estrutura do trabalho	4
2	Revisão da literatura	5
2.1	Turismo	5
2.1.1	Conceitos básicos sobre turismo	6
2.1.2	Classificação de turismo	7
2.1.3	Tipo de turistas	9
2.1.4	Factores que influenciam na prática do turismo	9
2.1.5	Efeitos da prática do turismo	12
2.2	Análise da associação	14
2.3	Regressão logística	15
2.3.1	Regressão Logística Múltipla	17
2.3.2	Estimação dos parâmetros do modelo	19
2.3.3	Seleção do modelo	20
2.3.4	Verificação do modelo ajustado	23
2.3.5	Significância dos coeficientes estimados	24
2.3.6	Avaliação da qualidade do modelo ajustado	25
2.3.7	Diagnóstico do modelo de regressão logística	27
2.4	Estudos relacionados	30
3	Material e Métodos	31
3.1	Material	31
3.1.1	Tipo de pesquisa	31
3.1.2	Amostragem e tamanho da amostra	32

3.1.3	Variáveis em análise	32
3.2	Métodos	33
3.2.1	Teste de independência χ^2 de Pearson	33
3.2.2	Método de estimação dos coeficientes	34
3.2.3	Significância dos coeficientes estimados	34
3.2.4	Avaliação do modelo ajustado	35
3.2.5	Diagnóstico do modelo estimado	35
4	Resultados e Discussão	36
4.1	Descrição do perfil da amostra	36
4.2	Teste de independência das variáveis	38
4.3	Regressão logística binária	42
4.3.1	Análise do modelo nulo	43
4.3.2	Análise do modelo ajustado	43
4.3.3	Avaliação de ajuste do modelo	44
4.3.4	Significância individual dos coeficientes	45
4.3.5	Interpretação e estimação do modelo ajustado	47
4.3.6	Análise residual	48
4.4	Discussão dos resultados	50
5	Conclusões e Recomendações	51
5.1	Conclusões	51
5.2	Recomendações	52
	Referências	54
	Apêndices	57

Lista de Figuras

2.1	Função da regressão logística para $\pi_j = P(Y = 1 X = x_j)$	16
4.1	Distribuição dos inquiridos segundo o sexo	36
4.2	Distribuição dos inquiridos segundo a empregabilidade	36
4.3	Distribuição dos inquiridos segundo a idade	37
4.4	Distribuição dos inquiridos segundo o estado civil	37
4.5	Distribuição dos inquiridos segundo a escolaridade	37
4.6	Representação da curva de ROC	44
4.7	Representação gráfica dos resíduos estandardizados	48
4.8	Representação gráfica dos resíduos estandardizados	49

Lista de Tabelas

2.1	Modelo generalizado das tabelas de contingências	26
2.2	Classificação da área abaixo da curva ROC	26
3.1	Descrição das variáveis em estudo	33
4.1	Cruzamento das província e zona residencial	38
4.2	Contingência entre Zona*Viajou e a significância	38
4.3	Contingência entre Emprego*Viajou e a significância	39
4.4	Contingência entre Sexo*Viajou e a significância	39
4.5	Contingência entre Civil*Viajou e a significância	39
4.6	Contingência entre Idade*Viajou e a significância	40
4.7	Contingência entre Nível*Viajou e a significância	40
4.8	Contingência entre Alojamento*Viajou e a significância	41
4.9	Contingência entre Transporte*Viajou e a significância	41
4.10	Contingência entre Cheias*Viajou e a significância	41
4.11	Contingência entre Ciclone*Viajou e a significância	42
4.12	Estatísticas de Multicolinearidade	42
4.13	Codificação de variável Viajou	42
4.14	Equação do modelo nulo	43
4.15	Tabela de classificação inicial	43
4.16	Teste de significância dos coeficientes (Omnibus)	43
4.17	Resumo do modelo ajustado	44
4.18	Teste Hosmer & Lemeshow	44
4.19	Tabela de classificação final	44
4.20	Área sob a curva	45
4.21	Significância individual dos coeficientes	45
4.22	Significância individual dos coeficientes - Continuação	46
4.23	Casos de anomalias listados	48
5.1	Tabela de contingência para teste de Hosmer e Lemeshow	57

Capítulo 1

Introdução

1.1 Contextualização

Nos países em vias de desenvolvimento, o turismo têm dado um grande contributo na luta contra a pobreza através da valorização dos recursos naturais e do património histórico e cultural que propicia a promoção de investimentos e do emprego assim como a geração de receitas em moeda externa e também nacional. Embora ainda não haja uma definição única do que seja Turismo, a Organização Mundial de Turismo/Nações Unidas (<http://www.world-tourism.org>), definem o turismo como: "as actividades que as pessoas realizam durante suas viagens e permanência em lugares distintos dos que vivem, por um período de tempo inferior a um ano consecutivo, com fins de lazer, negócios e outros".

De acordo com o Ministério da Cultura e Turismo (2015), Moçambique é dotado de um grande potencial turístico assente nos seus recursos naturais e culturais, que inclui: vida marinha, com excelentes praias, baías e lagoas ao longo de toda sua costa de 2,700 km; um mosaico cultural único que reflecte uma fusão Africana-Portuguesa Asiática-Árabe; Áreas de conservação da natureza, tais como Parques e Reservas Nacionais, e Áreas Marinhas Protegidas; proximidade do maior centro de entrada aéreo da África Austral, Joanesburgo e aos mercados regionais de turismo de relevo. O turismo em Moçambique está gradualmente a recuperar o seu lugar na economia nacional, visto que, é notório o crescimento dos investimentos ao longo dos últimos anos, que resultaram na expansão da capacidade de alojamento, dos serviços similares e no melhoramento da qualidade dos produtos, são os factores que poderão testemunhar um futuro encorajador da indústria do turismo no País.

A actividade turística pode contribuir para atingir alguns objectivos não somente da economia de um determinado país, impulsionado pela entrada de turistas internacionais e nacionais, mas também das diversas metas da Organização Mundial do Turismo (OMT), principalmente no que se refere à redução da pobreza, à conservação do meio ambiente, à criação de oportunidades de emprego para mulheres, comunidades e jovens, além de possuir importantes inter-relações com outros sectores produtivos, como a agricultura, o lazer e o artesanato (OMT, 2003).

Segundo Guambe *et al.* (2021), apesar da facilidade de integrar-se no processo do desenvolvimento nacional/local, é importante referir que, o turismo é uma actividade vulnerável devido a sua dependência relativamente às condições ambientais e sócio-económicas. Para tal, e em relação à Moçambique, basta referir, por exemplo, as cheias dos anos 2000, que afectaram a região sul de Moçambique, tiveram implicações directas nos empreendimentos turísticos, mas igualmente nas infraestruturas de acesso aos destinos turísticos, de abastecimento de água, energia, entre outras.

Os desastre naturais mais recentes e marcantes foram os ciclones Idai e Kenneth em 2019, as cheias na zona sul de Moçambique e o ciclone Freddy que devastaram em 2023, todos estes eventos levam a uma fraca recepção de turistas internacionais e também dificultam a prática do turismo doméstico, contribuindo assim, para a redução de oportunidades e postos de trabalho de vários moçambicanos. Nesse contexto, surge a necessidade de analisar os factores que contribuem na prática do turismo doméstico em Moçambique no período de 2019 à 2020.

1.2 Definição do problema

Moçambique é um país em via de desenvolvimento, localizado na África Austral que conta com um vasto leque de riquezas naturais, patrimoniais e culturais favoráveis ao desenvolvimento do turismo. Por essa razão, o Turismo é considerado um pilar para o desenvolvimento do país (Moçambique, 2004). Desse modo, o turismo pode ser considerado como uma actividade de grande importância na economia de um país, quando há uma crise económica a mesma sofre grande queda, reduzindo o seu nível de produção, gerando a desvalorização de activos financeiros e a falta de liquidez de diversas empresas.

Com a deflagração da pandemia de Covid-19, no ano de 2020, os ciclones e as cheias registadas em 2019 e 2023 respectivamente, o sector de turismo, de modo particular, começou a registar um decréscimo no número de chegadas de visitantes, ocasionando, como impacto, a perda de postos de emprego, redução de arrecadação de receitas e decréscimo do PIB. Ainda segundo os autores Guambe *et al.*, (2021), o sector público e privado, incluindo comunidades, foram afectados de forma negativa, gerando uma crise neste sector com respaldo em outros sectores, como o de transportes, de eventos, do comércio, entre outros.

De acordo com O País (2020), Moçambique perdeu 90% das receitas no sector do turismo nos últimos meses devido à covid-19. Actualmente, verifica-se que a crise na indústria do turismo, não deriva somente da pandemia da covid-19, mas também dos desastres naturais que vem assombrando o território nacional, esses dois fenômenos têm sido responsáveis pelo encerramento de várias estâncias turísticas e de hotelaria, como também causou perda significativa de postos de emprego dos Moçambicanos.

Diante deste cenário de retracção que se verifica no sector de turismo, com a particularidade dos moçambicanos considerados praticantes do turismo doméstico, torna-se necessário compreender vários factores que afectam a indústria do turismo, sendo que, neste estudo a questão de partida foi a seguinte: Quais são os factores que influenciam para a redução da prática do turismo doméstico pelos moçambicanos, no período de 2019 à 2020?

1.3 Objectivos

1.3.1 Objectivo Geral

Analisar os Factores que influenciam na Redução da prática do turismo doméstico no período de 2019 a 2020 em Moçambique.

1.3.2 Objectivos específicos:

- Descrever o perfil sócio-económico e demográfico dos moçambicanos que fazem parte da amostra;
- Identificar os factores que influenciam na redução da prática do turismo doméstico em Moçambique, no período em análise;
- Verificar os factores associados à prática do turismo doméstico pelos moçambicanos que fazem parte da amostra;
- Estimar o modelo logístico que explica a probabilidade de um moçambicano escolhido aleatoriamente não praticar o turismo doméstico.

1.4 Justificação

O turismo, actualmente, faz parte de uns dos principais sectores que alimenta a economia de Moçambique, assim como um dos principais geradores de renda das famílias moçambicanas. Com a actividade afectada pela pandemia, a economia do país e a das várias famílias que dependem parcial ou exclusivamente desta actividade.

No entanto, os recentes eventos relacionados com a pandemia do covid-19 e os desastres naturais deixaram a indústria moçambicana do turismo desestabilizada, sendo que, com esses eventos, advêm as dificuldades e problemas de algumas famílias que dependem exclusiva ou parcialmente dessas actividades, sendo a situação dessas famílias delicada, carecendo de uma atenção especial. Esse assunto despertou algum interesse que levou ao desenvolvimento do presente estudo, e foi sustentado pela motivação do autor, em querer introduzir as técnicas estatística multivariada na resolução de problemas de assuntos envolvendo a área do turismo.

1.5 Estrutura do trabalho

O presente trabalho é constituído por 5 capítulos, que de forma resumida passa se a apresentar os seus conteúdos:

- **Capítulo 1** (Introdução), tratou-se da contextualização, da definição do problema, questão de pesquisa, objectivos e justificação ou motivos que desencadearam a realização do presente estudo;
- No **Capítulo 2** (Revisão de Literatura), fez se uma abordagem teórica referente à problemática relacionada a redução da prática do turismo doméstico, particularmente pelos turistas moçambicanos, quais os factores por de trás da redução prática do turismo, assim como algumas informações básicas acerca das técnicas estatísticas a usar neste trabalho;
- **O Capítulo 3** (Material e Métodos), contém informações sobre a fonte de obtenção de dados, as variáveis do estudo, o processo da amostragem usado e por último foi feita uma descrição metodológica da estatística multivariada usada, assim como os passos para se alcançar os objectivos pré-definidos no trabalho;
- No **Capítulo 4** (Resultados e Discussão), neste capítulo foram apresentados os resultados obtidos através da aplicação das técnicas estatísticas de análise multivariada de dados, ilustrando-se os resultados obtidos através da aplicação da análise da associação e a regressão logística binária, também foi feita a discussão dos resultados obtidos comparando com os resultados dos outros autores;
- No **Capítulo 5** (Conclusões e Recomendações), foram apresentadas as principais conclusões obtidas nos resultados do trabalho, de seguida arroladas as recomendações como perspectivas para os futuros estudos.

Nas últimas secções são arrolados os aspectos pós-textuais, que compreendem as referências bibliográficas, apêndices e anexos.

Capítulo 2

Revisão da literatura

Neste capítulo, foram abordados aspectos relacionados com os conceitos básicos sobre o turismo, e com as técnicas da estatística multivariada usadas para realização do presente estudo, sendo que, mais detalhes passam se apresentar:

2.1 Turismo

Na segunda metade do século XVIII, passou a ser normal os jovens aristocratas ingleses fazerem uma viagem a que se chamou a “*Grand Tour*”, uma viagem de aproximadamente 3 anos pelo continente europeu com fins educativos. Desta viagem nasce o termo *touriste* designando as pessoas que faziam a “*Tour*”, introduzido em França por Stendhal nas suas “*Mémoires d’un Touriste*”.

Muitas outras línguas adoptaram posteriormente as palavras francesas *tourisme* e *touriste* com o sentido restrito de viagem feita sem fim lucrativo, por distração, repouso ou satisfação da curiosidade de conhecer outros locais e outras pessoas, embora a viagem não fosse encarada como um mero capricho, mas antes uma forma de aprendizagem ou um meio complementar de educação.

O turismo é uma actividade que envolve uma diversidade de actores, tanto nos locais de emissão como nos de recepção de turistas. Considerando o espaço de recepção, que é o foco neste estudo, evidenciam-se os seguintes actores: empreendedores do turismo (alojamento, restaurantes, agentes de viagens, transportadores, etc.), as estruturas governamentais a diferentes níveis e a população desses espaços. Os actores locais, em geral, organizam-se, de forma articulada ou não, para receberem visitantes. Isto significa que, para que haja turismo devem existir condições que assegurem, no mínimo, os movimentos de chegada e de partida de visitantes, sendo que a ausência de visitantes não cria condições para o desenvolvimento do turismo.

2.1.1 Conceitos básicos sobre turismo

Duhamel e Sacareau (1998), o **turismo** é uma actividade humana, fundada no deslocamento, isto é literalmente uma mudança de lugar e, por “extensão geográfica” uma mudança “d’habiter”. Isto significa que ser turista e fazer o turismo implica deixar temporariamente o lugar de vida habitual para ir viver num outro lugar situado fora da esfera da sua vida quotidiana.

De acordo com Mota (2001), o **turismo** é um fenómeno socioeconómico que consiste no deslocamento temporário e voluntário de um ou mais indivíduos que, por uma complexidade de factores que envolvem a motivação humana, saem do seu local de residência habitual para outro, gerando múltiplas inter-relações de importância cultural, socioeconómica e ecológica entre os núcleos emissores e receptores.

Em suma, o turismo supõe, portanto, deslocação e estadia. A estadia de um turista deve ser de no mínimo 24 horas no lugar de visita e não deve exceder um certo tempo, para além do qual o indivíduo perde a sua categoria de turista, pois, a deslocação será então assimilada a uma mudança de residência, o que pode ser considerado um outro domínio, o das migrações. Actualmente a duração dos vistos de turismo varia até um máximo de um ano.

Viajante: pessoa que visita um lugar diferente no qual têm residência fixa, com fins distintos das quais exerce em seu país;

Turista: visitante temporário que permanece no mínimo 24 horas (ou um pernoite) no lugar que visita e cujas finalidades de viagem podem ser classificadas em: férias, distração, negócios, saúde, estudo, religião, congressos etc;

Excursionista: visitante temporário que permanece menos de 24 horas (ou não realiza pernoite) no lugar que visita, e cujas finalidades são iguais às dos turistas. São comumente chamados de “visitantes de um dia” e incluem os passageiros em cruzeiros que pernoitam a bordo das embarcações.

De acordo com <http://www.esgt.ipt.pt/download/disciplina> em 1983 – Organização Mundial de Turismo (OMT) – elaborou a definição de turismo nacional.

- **Visitante nacional** – toda a pessoa, qualquer que seja a sua nacionalidade, que reside num país e que se desloca a um outro lugar situado nesse país e cujo motivo principal da visita não seja exercer uma actividade remunerada;
- **Turistas nacionais** – visitantes com uma permanência no local visitado, pelo menos de 24 horas mas não superior a um ano e cujos motivos de viagem podem ser agrupados

em prazer, férias, desportos ou negócios, visita a parentes e amigos, missão, reunião, conferência, saúde, estudos, religião;

- **Excursionistas nacionais** – visitantes que permanecem no local visitado menos de 24 horas.

2.1.2 Classificação de turismo

O turismo apresenta vários contextos da sua prática e leva as suas respectivas classificações. Neste estudo, o turismo foi classificado de acordo com as seguintes abordagens:

1. Segundo a origem dos visitantes

Geralmente essa classificação está mais relacionada com a origem dos visitantes que estão envolvidos na actividade turística.

- (a) **Turismo doméstico ou interno** – que resulta das deslocções dos residentes de um país, quer tenham a nacionalidade ou não desse país, viajando apenas dentro do próprio país;
- (b) **Turismo receptor** – abrange as visitas a um país por não residentes;
- (c) **Turismo emissor** – resulta das visitas de residentes de um país a outro ou outros países;
- (d) **Turismo interior** – abrange o turismo realizado dentro das fronteiras de um país e compreende o turismo doméstico e o receptor;
- (e) **Turismo nacional** – refere-se aos movimentos dos residentes de um dado país e compreende o turismo doméstico e emissor;
- (f) **Turismo internacional** – abrange unicamente as deslocções que obrigam atravessar uma fronteira, consiste no turismo receptor adicionado do emissor.

2. Segundo as repercussões na balança de pagamentos

As entradas de visitantes estrangeiros contribuem para o activo da balança de pagamentos de um país, na medida em que, provocam a entrada de divisas. As saídas de residentes nesse país têm um efeito passivo sobre a balança por provocarem a saída de divisas (turismo externo activo e turismo externo passivo).

- **Turismo de importação** – turismo de residentes praticado no estrangeiro (outgoing);
- **Turismo de exportação** – turismo de residentes no estrangeiro praticado no país visitado (incoming).

3. Segundo a duração da permanência

Este depende do objectivo da viagem, das condições existentes e características do local visitado (condições naturais, investimentos realizados, capacidade criativa), do país de origem, da duração das férias e das motivações.

- **Turismo de passagem** – efectuado apenas pelo período de tempo necessário para se alcançar uma outra localidade ou país, objecto da viagem;
- **Turismo de permanência** – realizado numa localidade ou num país, objectivo da viagem, por um período de tempo variável que, porém, exigirá, pelo menos uma dormida.

4. Segundo a natureza dos meios utilizados

Segundo as vias utilizadas, podemos distinguir turismo terrestre, náutico e aéreo, esta classificação é feita de acordo com os meios de transportes utilizados para se chegar ao destino turístico escolhido, desse modo, distinguem-se os seguintes grupos: o turismo por **caminho de ferro**, por **barco**, por **ar** e por **automóvel**.

5. Segundo o grau de liberdade administrativa

Os países emissores, em situações de dificuldade das respectivas balanças de pagamentos ou por razões políticas, podem limitar as saídas dos seus nacionais por vários meios: limitações na aquisição de divisas, lançamento de impostos, obrigação da constituição, prévia à saída do depósito de uma certa quantia de dinheiro, obrigação de vistos, restrições na concessão de passaportes, etc.

- **Turismo dirigido** – se existir regulamentação que limite a liberdade de deslocação dos turistas;
- **Turismo livre** – se existir inteira liberdade de movimentos.

Os países emissores também limitam, sobretudo por razões políticas, as entradas de estrangeiros ou as suas deslocações no interior do país.

6. Segundo a organização da viagem

- **Turismo individual** – quando uma pessoa ou um grupo de pessoas parte para uma viagem cujo programa é por elas próprias fixado, podendo modificá-lo livremente, com ou sem intervenção de uma agência de viagens;
- **Turismo colectivo ou de grupo (forfait ou package)** – quando um operador turístico ou uma agência de viagens oferece a qualquer pessoa, contra o pagamento de uma importância que cobre a totalidade do programa oferecido, a participação numa viagem para um determinado destino, segundo um programa previamente fixado para todo o grupo.

2.1.3 Tipo de turistas

As actividades turísticas dependem de duas partes importante que são os locais visitados e os próprios turistas, por isso, há necessidade de saber os diferentes tipos de turistas:

1. **Alocêntricos** – são os turistas exploradores, aventureiros, aqueles que adoram lugares com poucos turistas, estes turistas costumam fazer amizades com as comunidades locais. Até porque adoram explorar bastante cada local e ambiente. Mas, quando o destino começa a crescer, a ser frequentado por muitos turistas, eles começam a ir atrás de outros lugares, novos locais;
2. **Mesocêntricos** – São aqueles turistas que viajam sozinhos para lugares já conhecidos e por eles frequentados. Sua relação no local é totalmente comercial visando negócios;
3. **Psicocêntricos** – Este tipo são aqueles que viajam em grupos. Geralmente para lugares familiares, são influenciados pela sociedade e pelos modismos.

Pode-se ver que, para cada turista têm o seu jeito na escolha, vai depender de diversos factores para decidir um destino que queira. Cada momento é decisivo de cada personalidade para ter o local definido e cada turista também têm o seu período mais tranquilo ou mais agitado. Mas, aqueles que são alocêntricos, mesocêntricos ou psicocêntricos já sabem muito bem para onde ir.

2.1.4 Factores que influenciam na prática do turismo

Segundo Carvalho e Pimentel (2012), para analisar em que medida cada factor identificado têm influência exclusivamente no planeamento para a prática do turismo, ou seja, dependem apenas da elaboração e implementação de um plano turístico eficiente, associado ao empenho dos gestores e demais factores envolvidos no desenvolvimento do turismo local são:

➤ Factores físicos

Estes factores também são chamados físicos externos e estão relacionados ao ambiente natural, como as características do relevo e do clima, a presença de atractivos naturais e a distância em relação ao local de residência do turista são um dos primeiros elementos que motiva a escolha do destino de férias. Isso se deve às características buscadas pelo turista, além do tempo livre disponível que ele teria para realizar a viagem. As características geográficas de uma região conferem à oferta do destino uma forma de diferenciação, sendo um dos aspectos que estimula os deslocamentos.

Em virtude da recorrência de características consideradas distintas, como a presença de acidentes geográficos e a ocorrência de fenómenos naturais ou climáticos de uma região, se formaram centros turísticos desde a segunda metade do século XX (Acerenza, 2002).

➤ Factores Organizacionais

A existência de departamentos regionais e nacionais responsáveis por investigar sobre os fluxos turísticos de um determinado país ou local, nestes departamentos são considerados os factores que oferecem uma estrutura adequada para se efetuar estudos que auxiliem a compreender a evolução do turismo, e assim propor acções eficientes de desenvolvimento do sector.

A disponibilidade de dados estatísticos sobre a demanda (Rabahy, 2003) e os fluxos turísticos (Pearce, 2003) poderá informar de forma consistente sobre a evolução do turismo em uma região e o perfil do turista. Desde que esta ferramenta seja actualizada periodicamente, favorece a tomada de decisão baseada em dados reais, por consequência mais eficazes serão as iniciativas locais de desenvolvimento e aumento da competitividade dos destinos.

➤ Factores Socioculturais

A qualificação profissional contribui para redução do emprego informal e o subemprego (Beni, 2007). Além de estar correlacionada com a renda, a educação têm valorizado a importância do lazer e do turismo no uso do tempo livre, pelo aprendizado formal, estes factores despertam o interesse e curiosidade sobre os outros países (Rabahy, 2003).

A motivação da viagem têm origem na busca por diversão e relaxamento, se associa a religiosidade, ou à saúde. Também as viagens a negócio e estudo movimentam a indústria do turismo. Contudo, a influência dos grupos sociais é decisiva na escolha dos serviços turísticos, na seleção dos atrativos a serem visitados e no comportamento do turista no destino visitado. Cabe ressaltar que as experiências anteriores, tanto em viagens como em outras ocasiões influenciam directamente nas opções dos turistas, algumas vezes para repetir momentos agradáveis e inesquecíveis, e outras para esquecer experiências frustrantes. Um outro factor que merece ser contemplado é o ingresso no turismo das gerações nascidas pouco antes e após a 2^a Guerra Mundial. Estas pessoas estão actualmente no auge de suas actividades económicas e consequentes rendimentos. Além disso, acrescentam se certos aspectos relacionados ao lazer não priorizados anteriormente (Beni, 2007).

➤ Factores Institucionais

A obediência às leis de preservação ambiental e a recuperação das áreas degradadas deve ser um compromisso firmado por todos os actores envolvidos na indústria do turismo. No entanto, os interesses económicos são uma barreira à implementação de acções a este respeito. A existência de políticas para orientar a preservação ambiental, aliado a subsídios para garantir o cumprimento da legislação são um aspecto necessário à atração e manutenção dos fluxos turísticos.

Outro ponto a ser levantado está relacionado com a existência de uma legislação nacional e regional sobre o sector. A regulamentação e a orientação das actividades relacionadas ao turismo permite, através da Política de Turismo, um desenvolvimento com base em normas universais e esclarecimentos acerca dos direitos e das atribuições dos actores envolvidos. Os programas devem ser flexíveis e permitir a inserção de instrumentos para atingir os novos objectivos, bem como garantir a implementação dos objectivos estabelecidos (Beni, 2007).

Além desse ponto, sugere-se a criação de Unidades de Conservação que indica uma tentativa de valorização do destino. Entretanto, é preciso que sua concretização ocorra, de forma diferente do que muitas vezes se observa, quando estes espaços de preservação e conservação ambiental.

➤ **Factores aleatória**

Os factores chamados de aleatórios, ou variáveis incontroláveis contemplam o grupo de aspectos de diferentes naturezas que não podem ser previstos, controláveis ou se forem antecipados não podem ser controlados, de forma que afectam intensamente os fluxos turísticos no primeiro momento. Apenas em seguida é possível desenvolver estratégias para combatê-los (Beni, 2007). Têm-se os exemplos dos conflitos armados desde guerras civis, rebeliões e conflitos urbanos, até protestos, podem desestimular ou reduzir, temporariamente, a intensidade dos fluxos turísticos de uma região ou país.

Mas logo que passa a instabilidade, as viagens ao destino afectado, geralmente, são retomadas. O terrorismo pode ser considerado uma modalidade singular de conflito armado, pois, causa enorme insegurança aos turistas, assim, como aos moradores da região em tensão. Seu combate, ainda que seja intenso, não impede que a confiança do turista fique abalada, sendo necessário neste momento, a elaboração de planos de marketing e campanhas promocionais para estimular o retorno dos visitantes, enfatizando sempre a retomada da segurança na localidade (Palhares, 2006).

De acordo com Palhares (2006), as epidemias e pandemias podem causar mudanças nos fluxos turísticos globais. Nestes casos os países considerados mais seguros se beneficiam, e de igual forma os destinos domésticos. Mais ainda, as catástrofes climáticas, algumas vezes previsíveis, outras não previstas ou não divulgadas em tempo hábil para a evacuação das áreas de risco, arrasam localidades, regiões e países inteiros. Neste caso, os planos emergenciais de recuperação e reconstrução, aliado ao apoio de regiões não afectadas é decisivo para a retomada da normalidade.

Palhares (2006), acredita que as falhas tecnológicas também podem causar problemas irreparáveis, dentre os quais estão desastres aéreos e rodoviários, bem como acidentes nucleares. Os acidentes em grande escala, como desabamentos de construções, além das contaminações alimentares, trazem instabilidade nas regiões afectadas, e demandam acções reestruturadoras.

2.1.5 Efeitos da prática do turismo

É importante preservar o patrimônio das localidades nas quais se procura o seu desenvolvimento, é necessário que se preserve os patrimônios históricos, culturais e ambientais, não deixando que o turismo e o desenvolvimento venham a destruir desfigurar e atrapalhar o equilíbrio social existente na localidade. Deve-se buscar um desenvolvimento equilibrado, pois assim o turismo trará um crescimento na economia local, dando uma oportunidade para todos de ter uma vida melhor, ou seja, preocupando-se com o desenvolvimento sustentável do turismo. A partir da indústria do turismo pode se obter dois tipos de efeitos distintos que são:

1. Efeitos positivos gerados pelo turismo

- (a) Contribui para a valorização e conseqüentemente a preservação do patrimônio natural e cultural;
- (b) Gera empregos directos ou indirectos;
- (c) Maior arrecadação de impostos e taxas;
- (d) Possibilita a fixação do homem à terra, evitando a migração em regiões em opção de trabalho;
- (e) Viabiliza intercâmbio social e cultural, divulgando assim, a promoção de ligações entre povos, línguas, hábitos e gostos diferentes;
- (f) O turismo internacional é de particular importância como promotor de uma compreensão internacional, e como redutor das tensões políticas, pois como consequência do encontro e observação do modo de vida de pessoas de nacionalidades diferentes, é bem mais provável que elas se entendam melhor;
- (g) Contribui para melhorar a imagem dos lugares visitados;
- (h) O turismo também desempenha um papel relacionado com a saúde;
- (i) O abandono da rotina, a mudança do local, do clima regeneram a resistência física e reduzem significativamente na pressão nervosa;
- (j) É um importante factor de desenvolvimento econômico na medida em que serve como motivador do desenvolvimento de vários sectores da economia, a saber:
 - Aumento da urbanização através do contínuo crescimento da construção e renovação das instalações turísticas e modificações urbanas em benefício da população como um todo;

- Incrementa as indústrias associadas aos serviços de turismo, por exemplo transportes e alojamento;
- Aumenta a demanda de produtos do primário através do aumento do consumo;
- Expande o mercado para produtos locais;
- Leva a um aumento de divisas, necessários à redução do déficit na balança de pagamentos, fortalecendo desta maneira a economia local.
- Têm impacto favorável sobre o número de empregos, pois aumenta as oportunidades disponíveis (artesanato, transportes, comércio, guias, agências, meios de hospedagens, etc.);
- É um dos factores redistributivos mais efetivos nas relações econômicas. Os viajantes normalmente partem de regiões mais ricas para regiões menos privilegiadas, portanto, o turismo redistribui renda entre regiões;
- Ajuda o desenvolvimento de regiões remotas do país quando apresentam algum interesse turístico.

2. Efeitos negativos gerados pelo turismo:

- (a) A chegada de turistas de locais mais desenvolvidos, violando hábitos e tradições antigas são uma constante, padronizando os costumes e os lugares, tirando-lhes a originalidade e autenticidade, principais motivos de atração das correntes turísticas, ou seja, perda da identidade cultural;
- (b) Risco da introdução do consumo de drogas;
- (c) Possibilita o incremento da prostituição;
- (d) Gera Inflação e aumento temporário dos preços no núcleo receptor;
- (e) Em muitos lugares, o turismo pode trazer a degradação ambiental;
- (f) Estimula o processo de especulação imobiliária que, ao valorizar novas áreas, faz com que as características ambientais se tornem totalmente secundárias.

2.2 Análise da associação

Segundo Agresti e Franklin (2013), diz se existir uma associação entre duas variáveis se os valores de uma variável influenciam na ocorrência dos valores da outra variável em estudo. O teste não paramétrico usado como ferramenta nas condições expostas aqui deve-se ao matemático e estatístico inglês Karl Pearson (1857-1936), e é conhecido por Teste de qui-quadrado (χ^2). Geralmente têm sido usados 3 tipos de testes Qui-quadrado, que são: o teste de aderência, de homogeneidade e de independência, neste estudo apenas vai se abordar sobre o teste de independência Qui-quadrado, pois, têm-se o interesse em avaliar a relação de dependência entre duas variáveis categóricas coletadas a partir de uma amostra aleatória.

Teste de independência χ^2

Segundo Assis (2020), este teste é usado quando duas variáveis categóricas estão classificadas segundo atributos, categorias, eventos ou variáveis qualitativas que necessariamente não identifiquem distintas populações. Nesse caso, o pesquisador está preocupado em medir o grau de associação entre as variáveis e colocará à prova ou teste as seguintes hipóteses: $H_0 =$ As variáveis são independentes; vs. $H_1 =$ As variáveis não são independentes.

Se X e Y são duas variáveis aleatórias e independentes, em que cada linha representa uma categoria da variável X e cada coluna a categoria da variável Y , assumindo valores nos conjuntos enumeráveis $X = \{x_1, x_2, \dots\}$ e $Y = \{y_1, y_2, \dots\}$ respectivamente, então:

$$P(X = x, Y = y) = P(X = x) \times P(Y = y) \quad (2.1)$$

Para todo $x \in X$ e todo $y \in Y$. Ou seja, a probabilidade conjunta é dada pelo produto das marginais, que também são usados para calcular os valores esperados, pois, pela definição o valor esperado (E_{ij}) é dado pelo produto dos dois respectivos totais marginais $L_i = \sum_{i=1}^L O_{ij}$ e $C_j = \sum_{j=1}^C O_{ij}$, dividido pelo total das observações N , como descrito na equação (2.2).

$$E_{ij} = \frac{L_i \times C_j}{N} \quad (2.2)$$

No entanto, depois de obter os valores esperados de cada célula da tabela de contingência, segue-se com o cálculo da estatística de teste de independência Qui-quadrado, que segundo Maroco (2007), é a medida das diferenças entre as frequências observadas e esperadas, a sua estatística é definida por:

$$\chi^2 = \sum_{j=1}^C \sum_{i=1}^L \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (2.3)$$

Cuja distribuição é Qui-quadrado com $(C-1)(L-1)$ graus de liberdade, em que C representa o número de colunas e L representa o número de linhas na tabela de contingência, o E_{ij} é a frequência esperada da categoria i , na população j .

Para validação do teste de independência Qui-quadrado (χ^2), se faz necessário que sejam respeitados alguns critérios que são:

- Os dados serem selecionados aleatoriamente;
- Todas as frequências esperadas sejam maiores ou igual a 1;
- Não mais de 20% das frequências esperadas sejam inferiores a 5.

Observações: O teste está baseado na comparação entre duas hipóteses, denominadas, respectivamente de, hipótese nula e hipótese alternativa. A hipótese nula é de que as variáveis não estão associadas, em outras palavras, eles são independentes.

2.3 Regressão logística

A regressão logística envolve uma equação de predição na qual uma ou mais variáveis explicativas (preditoras) são usadas para fornecer informações sobre os valores esperados de uma variável de resposta binária (dependente). A técnica de regressão logística, desenvolvida no século XIX, obteve maior visibilidade após 1950 ficando então, mais conhecida. Autores como: Cox e Snell (1989) e Hosmer e Lemeshow (1989), concordam que o modelo de regressão logística ganhou reconhecimento após o trabalho de Truett, Cornfield e Kennel (1967) que analisava o risco de doença coronária em um grande projecto conhecido por “*Framingham heart study*”. Esse trabalho ganhou fama e até hoje é considerado um marco inicial dos estudos envolvendo regressão logística nas áreas da saúde.

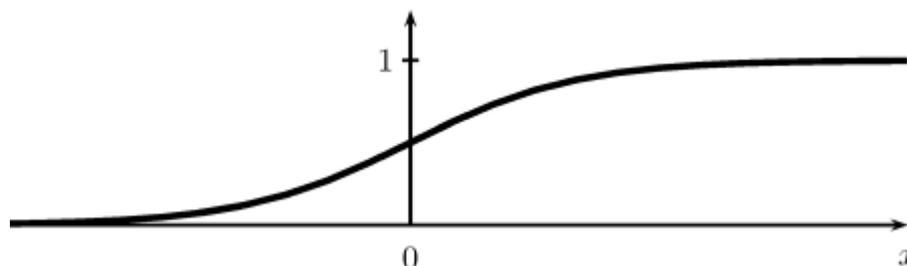
Em geral, o objetivo ao ajustar um modelo de regressão logística é o de descrever a relação entre uma variável resposta e um conjunto de variáveis explicativas (preditoras ou independentes). Para isso, a regressão logística modela o valor esperado da variável resposta condicionado aos valores de p variáveis explicativas $X = (x_1, \dots, x_p)$, isto é, $E(Y|X) = \pi(x)$, em que $E(Y|X)$ pertence ao intervalo $[0,1]$ (Hosmer e Lemeshow, 2000).

Segundo Bittencourt (2003), o modelo de regressão logística, originalmente desenvolvido para variáveis resposta binárias, é extensível para variáveis de resposta politômicas (três ou mais categorias), que por sua vez podem ser distintas por regressão logística multinomial ou Ordinal segundo o tipo de variável dependente ou seja os valores assumidos pela variável dependente.

A regressão logística binária resulta de uma função não linear, devido a natureza da variável resposta, sua média condicional deve ser maior ou igual a zero e menor ou igual a um, isto é, $0 \leq E(Y|X = x_i) \leq 1$.

A representação gráfica da regressão logística assume a forma parecida com um “S”, havendo áreas onde a mudança é mais acentuada e outras onde ela nem ocorre. As variações nos valores de x causam grandes mudanças nos valores y e de representam áreas de maior probabilidade de mudança de estado da variável y em função de x .

Figura 2.1: Função da regressão logística para $\pi_i = P(Y = 1|X = x_j)$



Fonte: Wasserman (2004)

De maneira mais ampla, a regressão logística atribui à variável resposta o intervalo compreendido entre zero e um, que pode ser interpretado como a probabilidade de ocorrência de certo evento. Conforme Corrar *et al.* (2007), baseado na regra de decisão, resultados superiores a 0.5 são assumidos como possível sucesso, e os inferiores como possível fracasso.

Suposições do modelo de regressão logística

Segundo Hair *et al.* (2009), vários investigadores usam o modelo de regressão logística para as suas análises, pois, ela não têm pressupostos rígidos e quanto mais violam-se os pressupostos dos modelos clássicos como da regressão Múltipla, análise discriminante e outros que poderiam ser usados para resolver o problema em estudo, isso se os seus pressupostos fossem validados para a análise, mas, com os pressupostos violados a análise de regressão logística tende a ser mais robusta e eficaz por usar.

Embora a técnica de regressão logística não seja rigorosa em relação aos pressupostos clássicos ou aplicados em modelos de regressão linear e discriminante, mas sua solução pode ser mais estável se seus preditores tiverem uma distribuição normal multivariada. Existem algumas condições indispensáveis para a execução dessa técnica, que são:

- **A variável dependente:** Deve ser dicotômica, categórica (para regressão logística multinomial) ou ordinal (para regressão logística ordinal);
- **As variáveis independentes:** Podem ser qualitativas ou quantitativas;
- **Linearidade e aditividade** – A escala que resulta da transformação $Logit(\pi)$ é aditiva e linear (mas a de π não);

- **Multicolinearidade:** Se houver a multicolinearidade entre os preditores podem conduzir as estimativas parciais de forma errada.

Os modelos de regressão logística com variável resposta binária podem se dividir em simples quando o modelo apresenta apenas uma variável independente e a mais generalizada é a regressão logística múltipla, a que contém mais de uma variável independente. Neste estudo vais se abordar sobre o modelo de regressão logística múltipla, pois, foram analisadas mais de uma variável independente para descrever a probabilidade de um indivíduo seleccionado praticar turismo doméstico.

2.3.1 Regressão Logística Múltipla

No caso da variável dependente Y assumir apenas dois possíveis estados (1 ou 0) e haver um conjunto de p variáveis independentes $\mathbf{X} = \{x_1, x_2, \dots, x_p\}$, então esse modelo denomina-se regressão logística múltipla, como não se têm acesso ao universo do Y , usam-se valores estimados a partir de dados amostrais, assim sendo, a probabilidade da ocorrência do sucesso ($P(Y = 1|\mathbf{X})$), pode ser escrito da seguinte forma:

$$\hat{\pi}(x) = P(Y = 1|\mathbf{X}) = \frac{e^{(\hat{\beta}_0 + \sum_{i=1}^p \hat{\beta}_i x_i)}}{1 + e^{(\hat{\beta}_0 + \sum_{i=1}^p \hat{\beta}_i x_i)}} \quad (2.4)$$

Para que os parâmetros do modelo estimado estejam linearmente distribuídos é necessária a execução de uma transformação logarítmica da função logística, e esse processo é designado por **transformação Logit** e pode ser expressa da seguinte maneira.

$$\text{Logit}(\hat{\pi}(x)) = \ln \left(\frac{e^{(\hat{\beta}_0 + \sum_{i=1}^p \hat{\beta}_i x_i)}}{1 + e^{(\hat{\beta}_0 + \sum_{i=1}^p \hat{\beta}_i x_i)}} \right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_p x_p \quad (2.5)$$

Onde:

$\text{Logit}(\hat{\pi}(x))$ – Representa a variável resposta ou dependente;

x_j – Representam as variáveis independentes;

$\hat{\beta}_i$ - Representam a variação media do Logit ($\hat{\pi}(x)$) quando x_j variável em uma unidade e as

outras mante-se fixas.

De forma semelhante, pode se obter a probabilidade da ocorrência do fracasso ($P(Y = 0|\mathbf{X})$) é dada pela seguinte equação:

$$1 - \hat{\pi}(x) = P(Y = 0|\mathbf{X}) = \frac{1}{1 + e^{(\hat{\beta}_0 + \sum_{i=1}^p \hat{\beta}_i x_i)}} \quad (2.6)$$

Interpretação do modelo logístico

A interpretação dos parâmetros estimados de um modelo de regressão logística é complexa, por isso, a forma mais eficaz obtida para fazer a interpretação desses modelos foi comparando as probabilidades de sucesso e do fracasso, através da razão de chances (RC). Antes de chegar a essa medida há necessidade de ilustrar a forma de chance:

$$\text{Chance} = \frac{\hat{\pi}(x)}{1 - \hat{\pi}(x)} = \frac{e^{(\hat{\beta}_0 + \sum_{i=1}^p \hat{\beta}_i x_i)}}{1 + e^{(\hat{\beta}_0 + \sum_{i=1}^p \hat{\beta}_i x_i)}} = e^{(\hat{\beta}_0 + \sum_{i=1}^p \hat{\beta}_i x_i)} \quad (2.7)$$

Substituído os valores de sucesso ($Y=1$) e fracasso ($Y=0$) nas equações (2.4) e (2.6), levando em consideração que, ao analisar cada variável independente de forma singular as outras variáveis são consideradas constantes (fixas), então, no cálculo da razão de chances teria-se:

$$\text{RC} = \frac{\frac{\hat{\pi}(1)}{1 - \hat{\pi}(1)}}{\frac{\hat{\pi}(0)}{1 - \hat{\pi}(0)}} = \frac{\frac{e^{(\hat{\beta}_0 + \hat{\beta}_i)}}{1 + e^{(\hat{\beta}_0 + \hat{\beta}_i)}}}{\frac{1}{1 + e^{\hat{\beta}_0}}} \div \frac{1}{e^{\hat{\beta}_0}} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_i}}{e^{\hat{\beta}_0}} = \frac{e^{\hat{\beta}_0} \times e^{\hat{\beta}_i}}{e^{\hat{\beta}_0}} = e^{\hat{\beta}_i} \quad (2.8)$$

Fazendo o logaritmo da RC, poderia se obter o seguinte:

$$\ln(RC) = \ln e^{\hat{\beta}_i} = \hat{\beta}_i \ln e = \hat{\beta}_i$$

O $\hat{\beta}_i$ pode ser interpretado como sendo o logaritmo da razão de chances e define o incremento no *logit* que é devido ao aumento em uma unidade na variável explicativa x . Mas devido a complexidade dessa interpretação, usa-se a RC para verificar o quão provável o resultado ocorrerá entre as categorias em relação a categoria de referência da mesma variável, fazendo, portanto, as comparações podem ser descritas da seguinte maneira:

- $\beta_i > 0 \iff RC > 1$: O que significa que as chances da categoria em análise pertencer ao evento de interesse, são RC vezes superiores as chances da categoria de referência dessa variável;
- $\beta_i < 0 \iff RC < 1$: O que significa que as chances da categoria em análise pertencer ao evento de interesse, são RC vezes inferiores as chances da categoria de referência dessa variável;
- O caso menos provável é aquele em que, o $\beta_i = 0 \iff RC = 1$: O que levaria a considerar que, as chances do sucesso e fracasso na ocorrência de um determinado é a mesma.

2.3.2 Estimação dos parâmetros do modelo

Regressão Logística Binária, os valores atribuídos a variável dependente y podem assumir apenas dois valores que são $P(Y_i = 1|\mathbf{X}) = \hat{\pi}$ ou $P(Y_i = 0|\mathbf{X}) = 1 - \hat{\pi}$. Assim sendo, a distribuição de y pode ser descrita como uma distribuição de Bernoulli, como ilustra a equação seguinte:

$$P(Y_i|\mathbf{X} = x_i) \sim Ber(\hat{\pi}(x)) \quad (2.9)$$

Segundo Hosmer e Lemeshow (2000), a forma mais conveniente de expressar a distribuição de cada par da variável dependente e das independentes da regressão logística é descrita de seguida:

$$f(y_i|x_i) = \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad y_i = \{0, 1\} \quad (2.10)$$

Considerando que as realizações y são independentes, a função de verossimilhança é obtida pelo produto de todos os n termos ou realizações, como descrito na equação (2.11) que segue:

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (2.11)$$

O método de máxima verossimilhança consiste em encontrar o valor de $\hat{\beta}_i$ que maximizam a função de verossimilhança para uma determinada amostra. Para tal, é necessário que se faça o logaritmo da função de verossimilhança descrita em (2.11). Deste modo, a equação (2.12) descreve o logaritmo da função de verossimilhança:

$$L(\beta) = \ln [l(\beta)] = \sum_{i=1}^n \{y_i \ln [\pi(x_i)] + (1 - y_i) \ln [1 - \pi(x_i)]\} \quad (2.12)$$

No entanto, para encontrar os valores de β que maximizam o logaritmo da função de verossimilhança ($L(\beta)$), faz a primeira diferença para equação (2.12) em função dos coeficientes β_0 e β_i , como descrevem as equações (2.13) e (2.14):

Derivada em relação ao β_0

$$\frac{\partial[L(\beta_i)]}{\partial\beta_0} = 0 \leftrightarrow \sum_{i=1}^n y_i - \pi(x_i) = 0 \quad (2.13)$$

Derivada em relação ao β_i

$$\frac{\partial[L(\beta_i)]}{\partial\beta_i} = 0 \leftrightarrow \sum_{i=1}^n x_i [y_i - \pi(x_i)] = 0 \quad (2.14)$$

Já que os coeficientes β_i não são lineares, por isso, esse processo deve ser iterativo para encontrar a discrepância mínima entre a resposta observada Y , e a resposta prevista \hat{Y} , precisando de auxílio computacional.

2.3.3 Seleção do modelo

A seleção de modelos é uma parte importante da modelação estatística que têm como objetivo encontrar o melhor modelo, ou melhores modelos, que descreva bem os dados observados para explicar a variabilidade da variável resposta.

Método de seleção do modelo

De acordo com Maroco (2007), no processo de seleção, é frequentemente utilizado dois modelos de referência:

- **Modelo completo ou saturado** – este modelo contém parâmetros linearmente independentes e atribui toda a variação dos dados à componente sistemática. O modelo saturado serve de referência para medir a discrepância de um modelo intermédio com $p+1$ parâmetros;
- **Modelo Nulo** – este modelo é o mais simples por ter apenas um único parâmetro. Neste modelo assume-se que todas as variáveis têm o mesmo valor médio, neste caso, atribui toda a variação dos dados à componente aleatória.

Existem duas questões relacionadas com a seleção do modelo, no entanto, por conseguinte apresentam-se os métodos *stepwise*, e de seguida, os critérios e procedimentos comumente usados na seleção do modelo.

O **método stepwise** é um dos métodos mais aplicado em regressão logística. O método baseia-se na selecção automática das variáveis importantes para o modelo. Este método é usado para seleccionar as variáveis que mais influenciam no conjunto de saída podendo, assim, diminuir o número de variáveis a compor a equação de regressão. Existem três formas de realizar uma regressão através do método stepwise:

1. **Forward** - quando a equação começa vazia e cada preditor entra, um por um, na equação;
2. **Backward** - quando todos os preditores são incluídos de uma só vez na equação, e depois são retirados, um a um, até que identifiquem se apenas os preditores significativos para a ocorrência do evento de interesse;
3. **Blockwise ou Setwise** – assemelha-se à regressão stepwise forward, mas, ao invés dos preditores serem incluídos individualmente.

Critérios para a seleção do modelo

Escolher o melhor modelo é controverso, mas um bom modelo deve conseguir equilibrar a qualidade do ajuste e a complexidade, sendo esta, em geral, medida pelo número de parâmetros

presentes no modelo; quanto mais parâmetros, mais complexo o modelo, sendo pois mais difícil interpretar o modelo. A seleção do “melhor” modelo torna-se então necessário.

Segundo Corrar *et al.* (2007), o teste Nagelkerke R^2 situa-se numa escala que vai de 0 a 1 e têm a mesma finalidade do indicador Cox-Snell R^2 , isso torna o Pseudo R^2 de Nagelkerke mais aconselhado a usar no momento da escolha do melhor modelo e é dado por:

$$R^2_{Nagelkerke} = R^2_{C\&U} = \frac{1 - \left[\frac{L(\beta)_0}{L(\beta)_M} \right]^{\frac{2}{n}}}{1 - L(\beta)_0^{\frac{2}{n}}}, 0 \leq R^2_{C\&U} \leq 1 \quad (2.15)$$

Embora existam outros valores do **pseudo- R^2** , apenas descreve-se este pelo facto deste ser o que será interpretado diante dos de mais pseudo- R^2 .

Segundo Agresti (2007), o melhor critério conhecido é o de informação de Akaike (AIC), pois, ele julga um modelo pela distância, onde os seus valores ajustados tendem a ser os verdadeiros valores esperados, e o modelo ideal é aquele que tende a ter seus valores ajustados mais próximos das verdadeiras probabilidades de resultado, e este critério é descrita pela equação (2.16).

$$CIA = e^{\frac{2k}{n}} \frac{\sum \hat{u}_i^2}{n} = e^{\frac{2k}{n}} \frac{SQR}{n} \quad (2.16)$$

Onde: O k é o número de repressores (incluindo o intercepto) e n é o número de observações. Por conveniência matemática, a equação (2.16) pode ser logaritimizada, pois, Alguns textos e programas definem CIA apenas em termos de sua transformação logarítmica, ao fazer essa transformação teria-se a equação (2.17)

$$\ln CIA = \left(\frac{2k}{n} \right) + \ln \frac{\sum \hat{u}_i^2}{n} = \left(\frac{2k}{n} \right) + \ln \frac{SQR}{n} \quad (2.17)$$

Onde: $\ln CIA$ = logaritmo natural de CIA e $2k/n$ D factor de correção.

Critério de informação de Schwarz (CIS)

Semelhante ao CIA, mas O CIS impõe medidas correctivas mais duras que o CIA, o critério CIS é definido como:

$$CIS = n^{k/n} \frac{\sum \hat{u}_i^2}{n} = n^{2k/n} \frac{SQR}{n} \quad (2.18)$$

ou na forma logarítmica:

$$\ln CIS = \left(\frac{k}{n} \right) \ln n + \ln \frac{\sum \hat{u}_i^2}{n} = \left(\frac{k}{n} \right) \ln n + \ln \frac{SQR}{n} \quad (2.19)$$

Segundo Agresti (2007), o melhor critério conhecido é o de informação de Akaike (AIC), pois, ele julga um modelo pela distancia, onde os seus valores ajustados tendem a ser os verdadeiros valores esperados, e o modelo ideal é aquele que tende a ter seus valores ajustados mais próximos das verdadeiras probabilidades de resultado.

2.3.4 Verificação do modelo ajustado

Depois de estimados os coeficientes do modelo ajustado, segue-se a verificação da qualidade do próprio modelo de regressão logístico ajustado em que além do valor de Pseudo R^2 , deve se analisar o próprio teste de ajuste do modelo. Assim sendo passa-se a apresentar as medidas da qualidade de ajuste:

➤ Razão de verossimilhança

Para realizar o teste da razão da verossimilhança, compara-se os valores observados da variável resposta com os valores preditos obtidos dos modelos com e sem as variáveis em questão. Segundo Hosmer e Lemeshow (2000), a comparação dos valores observados e previstos usando a função de verossimilhança é baseada na seguinte expressão:

$$D = -2 \ln \left[\frac{\text{Verossimilhança do modelo ajustado}}{\text{Verossimilhança do modelo saturado}} \right] \quad (2.20)$$

Para obter um ajuste adequado é necessário verificar o valor de $-2 \log$ da função de verossimilhança, que é descrita como o teste de razão de verossimilhança, esse teste é obtido usando a equação (2.12) e (2.20), obtêm se a seguinte equação:

$$D = -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{\pi}(x_i)}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}(x_i)}{1 - y_i} \right) \right] \quad (2.21)$$

A estatística D, também é denominada **Deviance** por alguns autores, sendo que, essa estatística é semelhante ao termo da soma dos quadrados dos resíduos que são usadas para avaliar a qualidade do modelo ajustado. A alteração no valor de β_i esperada pela inclusão da variável independente no modelo é obtida através de $G = D(\text{para o modelo sem a variável}) - D(\text{para o modelo com a variável})$. Essa estatística faz análise da significância individual das variáveis e pode ser descrita pela equação:

$$G = -2 \ln \left[\frac{\text{Verossimilhança sem a variável}}{\text{Verossimilhança com a variável}} \right] \quad (2.22)$$

Geralmente para um caso específico de uma variável independente, é fácil demonstrar quando a variável não faz parte do modelo estimado, a estimativa $\hat{\beta}_0$ é $\ln(n_1/n_0)$ onde: $n_1 = \sum y_i$, $n_0 = \sum(1 - y_i)$, desse modo, o valor da estatística G é dada por:

$$G = -2 \ln \left[\frac{\left(\frac{n_1}{n} \right)^{n_1} \left(\frac{n_0}{n} \right)^{n_0}}{\prod_{i=1}^n \hat{\pi}(x_i)^{y_i} [1 - \hat{\pi}(x_i)]^{(1-y_i)}} \right] \quad (2.23)$$

A estatística G segue uma distribuição Qui-quadrado com 1 grau de liberdade, havendo necessidade de usar uma amostra suficientemente grande, com as respostas de $y=0$ e $y=1$.

➤ Teste de Hosmer e Lemeshow

Segundo Maroco (2007), os autores Hosmer e Lemeshow propuseram uma outra estatística para testar a hipótese de ajustamento do modelo aos dados, baseada na estimação das probabilidades de sucesso de cada uma das n observações. A estatística de teste é obtida com um teste de Qui-quadrado a uma tabela de contingência $2 \times g$.

Para avaliar o ajuste geral do modelo usada é a estatística de Hosmer e Lemeshow, as probabilidades estimadas foram ordenadas de mais baixa a mais alta e depois faz-se teste qui-quadrado para determinar se as frequências previstas estão próximas das frequências observadas, e a estatística do teste é dada pela equação:

$$X_{HL}^2 = \sum_{i=1}^g \frac{(O_i - E_i)^2}{E_i} \sim X_{(g-2)}^2 \quad (2.24)$$

Onde:

g - é o número de grupos criados; O_{i-} é o número de “sucessos” observados da variável dependente em cada classe; E_i - é o número de “sucessos” esperados em cada classe.

2.3.5 Significância dos coeficientes estimados

Após a estimação dos coeficientes do modelo, avalia-se a qualidade da estimação, ou seja, testa-se se as variáveis explicativas pertencentes ao modelo são significativas para explicar o comportamento da variável resposta. De acordo com Paula (2013), os testes de hipóteses para os modelos lineares generalizados baseiam-se em três estatísticas: teste de razão de verossimilhança, teste de Wald e o teste de Escore:

1. Teste de Razão de Verossimilhanças

Pela teoria de estimação de máxima verossimilhança, sabe-se que os estimadores de máxima verossimilhança maximizam a função log-verossimilhança. No entanto, o teste de razão de verossimilhança avalia se o valor de log-verossimilhança é suficientemente grande para concluir que as variáveis retiradas são importantes para o modelo.

A estatística de razão de verossimilhança é o critério mais poderoso, que testa simultaneamente, se os coeficientes estimados (β_i), associados as variáveis são todos nulos com exceção do $\hat{\beta}_0$. Esse teste realiza comparações entre os valores observados e esperados usando a função de verossimilhança, esses testes são descritos pelas equações (2.20) e (2.23) que também pode ser descrita pela equação seguinte:

$$\mathbf{G} = -2 \left\{ \sum_{i=1}^n [y_i \ln(\hat{\pi}_i) + (1 - y_i) \ln(1 - \hat{\pi}_i)] - [n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n)] \right\} \quad (2.25)$$

2. Teste de Wald

O teste de Wald avalia se cada coeficiente (β_i) do modelo é significativamente diferente de zero. Neste sentido, pode afirmar que nesse teste avalia se a relação uma determinada variável independente com a variável dependente é estatisticamente significativa. A estatística de Wald apresenta uma distribuição qui-quadrado com número de graus de liberdade igual ao número de restrições, sendo que essa estatística é obtida comparando a estimativa do coeficiente estimado pelo método de máxima verossimilhança e a estimativa do seu erro padrão como descrito na equação (2.26):

$$\mathbf{Wald} = \frac{\hat{\beta}_i}{\text{EP}(\hat{\beta}_i)} \quad (2.26)$$

3. Teste de Escore

O teste de significância das variáveis que não interações computacionais, sendo esta uma das vantagens do teste de escores. O uso desse teste têm a limitação pelo facto de alguns pacotes ou software estatísticos não disporem da execução desse teste. Deste modo, a estatística de teste para a execução do teste score é:

$$ST = \frac{\sum_{i=1}^n x_i(y_i - \bar{y}_i)}{\sqrt{\bar{y}(1 - \bar{y}) \sum_{i=1}^2 (x_i - \bar{x})^2}} \quad (2.27)$$

Em que $\hat{y} = \hat{\pi}_i$ é a proporção de sucessos na amostra, nesse teste também têm-se o interesse em testar as hipóteses $H_0: \hat{\beta}_i = 0$ vs. $H_1: \hat{\beta}_i \neq 0$

2.3.6 Avaliação da qualidade do modelo ajustado

Depois de ajustar o modelo torna-se necessário avaliar a qualidade do modelo ajustado, para tal, são analisadas a tabela de classificação do modelo e a área abaixo da curva do ROC¹.

1. Tabela de classificação

Segundo Maroco (2007), para avaliar a qualidade da classificação feita pelo modelo é necessário comparar a percentagem global de classificações correctas obtidas com o modelo, sendo que, essa percentagem é calculada a partir do número de sujeitos observados em cada uma das k classes da variável dependente (C_i) pela expressão:

$$\mathbf{Classificação\ correcta\ proporcional\ por\ acaso\ (\%)} = \sum_{i=1}^k \left(\frac{C_i}{N} \right)^2 \times 100\% \quad (2.28)$$

Se a percentagem de casos classificados correctamente pelo modelo for superior em pelo 25%, este modelo pode ser considerado ter o mínimo de boas propriedades classificativas.

¹Receiver Operating Characteristic

A eficiência classificativa do modelo pode também ser avaliada pela **sensibilidade** - que percentagem de classificações correctas na classe de referência da variável dependente "1 - Sucesso" e pela **especificidade** - que é percentagem das classificações correctas da classe do insucesso do respondentes "0 - Insucesso". A tabela de classificação que vem como a tabela de contingência 2×2 dos eventos observados e previstos, como ilustrado na tabela abaixo:

Tabela 2.1: Modelo generalizado das tabelas de contingências

		Eventos previstos		Percentagem correcta
		Positivo	Negativo	
Eventos observados	Verdadeiros	VP	VN	Sensibilidade
	Falsos	FP	FN	Especificidade
Totais		P	N	Percentagem global

Deste modo, com base na tabela (2.1) as medidas de sensibilidade e especificidade podem ser calculadas pelas equações (2.29) e (2.30) respectivamente:

$$\text{Sensibilidade} = P(\hat{Y} = 1|y = 1) = \frac{VP}{VP + VN} \times 100\% \quad (2.29)$$

Em que:

VP - Significa verdadeiros positivos; e

VN - Significa verdadeiros negativos.

$$\text{Especificidade} = P(\hat{Y} = 0|y = 0) = \frac{FN}{FP + FN} \times 100\% \quad (2.30)$$

Onde:

FP - Significa falsos positivos; e

FN - Significa falsos negativos.

2. Área abaixo da curva do ROC

Uma das melhores formas de verificar a capacidade do modelo de classificar é a área abaixo da curva do ROC. Segundo Hosmer e Lemeshow (2000), a regra para avaliação da capacidade do modelo logístico, através da área abaixo da curva do ROC pode ser classificada de acordo com a seguinte tabela:

Tabela 2.2: Classificação da área abaixo da curva ROC

Área ROC	Classificação do modelo logístico
Área < 0.7	Discriminação fraca
[0.7 à 0.8 [Discriminação aceitável
[0.8 à 0.9 [Discriminação boa
≥ 0.9	Discriminação muito boa

Fonte: Hosmer e Lemeshow (2000)

2.3.7 Diagnóstico do modelo de regressão logística

O diagnóstico dos modelos de regressão, geralmente estão associados a verificação das suposições dos mesmo, sendo que, para a regressão logística será verificado o seguinte:

Multicolinearidade

Segundo Maroco (2007), O termo colinearidade é utilizado para expressar a existência de correlação elevada entre duas variáveis independentes, enquanto o termo multicolinearidade é utilizado quando se trata de mais do que duas variáveis independentes fortemente correlacionadas, o uso de variáveis com alguma colinearidade podem levar a má especificação do modelo logístico ajustado.

A multicolinearidade pode ser diagnosticada de várias formas. Segundo Salvian (2016), a multicolinearidade pode ser diagnosticada através da Matriz de correlações, Factor de Inflação da Variância (VIF), Tolerância (T) e Número de condição, mas nesse estudo serão abordadas apenas duas maneiras de diagnosticar a multicolinearidade que são:

1. Factor de Inflação da Variância (VIF)

Para Maroco e Bispo (2003), Este é um dos indicadores usado com frequência para detectar a multicolinearidade, onde a variância de cada um dos coeficientes de regressão associados às variáveis independentes é dada por:

$$var(\beta_i) = \sigma^2 \left(\frac{1}{1 - R_i^2} \right) \times \frac{1}{\sum_{j=1}^n (x_{ij} - \bar{x}_i)} \quad (2.31)$$

Esta variância é tanto maior quanto maior for a correlação múltipla entre e as variáveis independentes. O termo $\frac{1}{1 - R_i^2}$ designa-se, em concreto, por VIF para o coeficiente de regressão associado à variável. Segundo Maroco e Bispo (2003), caso se obtenham valores de VIF > 5 conclui-se que, está se perante problemas da estimação devido à presença de multicolinearidade nas variáveis independentes.

2. Tolerância

Outra medida utilizada pelo SPSS para diagnosticar a presença de multicolinearidade é a “**Tolerância (T)**” da variável X_i . Quando T é próximo de 0 a variável X_i pode escrever-se como combinação quase-linear das outras variáveis independentes e, como consequência o respectivo coeficiente de regressão é instável quer na magnitude quer no sinal. Portanto, $VIF = 1/T$, variáveis com baixa tolerância têm valores elevados de VIF e vice-versa. No entanto, valores de T menores ou iguais a 0,1 indicam indícios de presença de multicolinearidade nas variáveis.

Análise residual

Em qualquer análise de regressão é importante verificar o comportamento do modelo ajustado através dos resíduos. Um resíduo pode ser definido como a distância entre o valor estimado e o valor observado correspondente a variável dependente Y (Agranonik, 2005). De seguida são apresentados os Resíduos estandardizados de Pearson, Deviance e estandardizados, que são úteis para identificar observações que não estão a ser bem explicadas pelo modelo:

➤ Resíduos estandardizados de Pearson

De acordo com Maroco (2007), para obter estes resíduos é necessário dividir os resíduos não estandardizados ($e_j = y_i - \hat{y}_i$), pela estimativa do desvio padrão dos valores estimados, os resíduos estandardizados de Pearson é dado por:

$$e'_i = \frac{e_i}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}} \quad (2.32)$$

Estes resíduos apresentam a média 0 e desvio padrão aproximadamente a 1, onde para amostras grandes ela segue uma distribuição normal $e'_i \sim N(0, 1)$.

➤ Resíduos Deviance

Esses resíduos foram descritos pela equação (2.21), elas medem o grau de discordância o máximo da função de verossimilhança observado e o estimado, como na regressão logística usa se o método da máxima verossimilhança, que têm como objectivo maximizar o poder preditivo dos coeficientes estimados (Agranonik, 2005).

➤ Resíduos Estandarizados

Estes resíduos são dados pelo quociente entre o resíduo e a a estimativa do seu desvio padrão, mas como esses valores do desvio padrão não é constante. Nesse contexto, para obter os resíduos estandardizados é necessário garantir a independência entre o numerador e denominador na padronização dos resíduos. Deste modo, define-se o resíduo Estandarizado por:

$$r_i = \frac{e'_i}{\sqrt{1 - h_i}} = \frac{y_i - \hat{y}_i}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i) \sqrt{1 - h_i}}} \quad (2.33)$$

Onde a medida h_j podem ser estimadas pelo *Leverage* que são importantes para identificar valores extremos.

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

mas infelizmente eles não mostram qual é o impacto que esses valores têm nos vários aspectos do ajuste do modelo.

Medidas de influência

As medidas de influência são responsáveis por verificar as observações que contribuem para a modificação das estimativas dos parâmetros do modelo de regressão. Os resíduos estandarizados e a *Leverage* podem também ser utilizadas para a influência de uma observação no ajuste do modelo (Maroco, 2007).

Hosmer e Lemeshow (2000), demonstram a variação da estatística χ^2 de Pearson (Usado para testar o ajuste do modelo) por eliminação da observação i é denominado distância de Cook e é simplesmente dada por:

$$\Delta\chi_i^2 = \frac{(e'_i)^2}{1 - h_i} \quad (2.34)$$

Estas medidas descrevem as influências de Cook, também podem ser representadas graficamente, através dos valores previstos e as respectivas observações. A outra medida de influência usada é a Deviance (d_i) da observação em análise, esta estatística é dada por:

$$d_i = \text{Sign}(e_i) \sqrt{2 \left[y_i \ln \left(\frac{y_i}{n_i \hat{\pi}_i} \right) + (n_i - y_i) \ln \left(\frac{n_i - y_i}{n_i (1 - \hat{\pi}_i)} \right) \right]} \quad (2.35)$$

Onde a função $\text{Sign}(e_i)$ devolve o sinal (+ ou -) de e_i , a estatística D descrita em (2.21) avalia a qualidade do ajuste do modelo, ela é o somatório dos quadrados de d_i .

Segundo Agranonik (2005), além das duas medidas citadas acima (distância de Cook e Deviance), também têm o **DFFFITS**, têm como objectivo de medir a influência da i -ésima observação nos parâmetros de localização e a de modelo. O nome DFFITS vem do inglês *Difference in Fit* é calculada em função do resíduo studentizado r_i e a média alavancagem h_i , e é dada por:

$$\mathbf{DFFFITS} = r_i \left\{ \frac{h_i}{(1 - h_i)} \right\}^{\frac{1}{2}} \quad (2.36)$$

A próxima medida a saber é **DFBETA** - que mede para cada coeficiente estimado $\hat{\beta}$ da regressão, relacionada com a variável x_i , verificando o quanto se modifica a regressão quando a observação i é excluída da análise. A estatística DFBETA é definida por:

$$\hat{\beta} - \hat{\beta}_i = (X'X)^{-1} x'_i (1 - h_i)^{-1} \hat{\varepsilon}_i \quad (2.37)$$

O sinal do DFBETA indica se a inclusão de uma observação leva a um acréscimo ou decréscimo dos coeficientes estimados para a regressão e seu valor absoluto mostra o futuro dessa diferença em relação ao seu desvio padrão estimado.

2.4 Estudos relacionados

Segundo Gêmo (2017), da sua investigação com o identificar os factores críticos de sucesso do turismo em Moçambique. Uma vez que o país permanece muito endividado e por isso o turismo aparece como uma actividade de grande importância. Realizou-se um inquérito online, com 8 perguntas, recorrendo-se ao método quantitativo, que contou com a colaboração de 155 inquiridos moçambicanos, que deram o seu ponto de vista de forma eficaz e constatou se que, muitas pessoas especialmente as de posse significativas, preferem na maioria das vezes viajar ao exterior do que no interior de Moçambique, por vários motivos tais como; a falta de infra-estruturas no país, falta de um intenso trabalho de estratégia de marketing sobre as belezas naturais da terra, bem como, o facto de alguns pacotes turísticos para as viagens nacionais serem mais caros do que os pacotes internacionais e principalmente por falta de recurso financeiro para se aplicar na área de lazer.

Wache (2021) realizou um estudo, cujo objectivo era avaliar os factores determinantes dos turistas na escolha do Município de Inhambane como destino turístico, para melhorar a performance da oferta turística local. Para a concretização da sua investigação, recorreu-se a pesquisa uma bibliográfica e documental através do inquérito por questionário. Portanto, conclui-se que os factores identificados na escolha de Município de Inhambane como destino turístico são: preços de produto e serviços, alojamento turístico, estabilidade política e segurança, infra-estruturas e acessibilidade, património natural, património histórico e cultural, hospitalidade para com os turistas, turismo de sol e mar. O factor “sol e mar” constitui o factor predominante na escolha do Município de Inhambane como destino turístico pelos turistas inqueridos.

Segundo Silva (2019), Sob o título Turismo em Moçambique: oportunidades desafios e riscos, desenvolve-se um estudo com o propósito de demonstrar evidências das oportunidades, desafios e riscos associados ao desenvolvimento do turismo. Para o efeito recorreu-se a uma revisão bibliográfica e documental e procedeu-se à interpretação dos mesmos. Os resultados do estudo revelam que, o potencial turístico de Moçambique constituído por elementos naturais associado à sua localização geográfica. As oportunidades para o desenvolvimento do turismo são imensas e foram percebidas pelos turistas, agentes económicos e pelo estado. No entanto, a mesma localização geográfica que constitui uma das principais forças de Moçambique associada à disposição do relevo, constitui igualmente um desafio para o turismo, uma vez que, pode ser visto como um importante factor de vulnerabilidade, pois, o país está sujeito à influência do canal de Moçambique, responsável pela ocorrência de situações adversas como ciclones, cheias e inundações, o que têm posto em causa os investimentos feitos em infraestruturas em geral e nas facilidades turísticas em particular, colocando em risco os investimentos e o desenvolvimento da actividade.

Capítulo 3

Material e Métodos

Neste capítulo, foram descritos o material e métodos estatísticos usados para se alcançar os objectivos deste trabalho.

3.1 Material

Neste estudo foram usados dados secundários, pois, foram dados fornecidos pelo Instituto Nacional de Estatística (INE), que resultaram do Inquérito sobre Orçamento Familiar realizado entre 2019 e 2020 (IOF 2019/2020). Este inquérito, teve como objectivo obter informações sobre a natureza e o destino das despesas de consumo, bem como informações sobre diversos recursos relacionados com as condições de vida dos agregados familiares.

Das diversas informações disponibilizadas nas bases de dados resultantes do Inquérito do Orçamento Familiar (IOF 2019/2020), neste estudo, apenas vai se levar em consideração variáveis relacionadas com o perfil sócio-económico e demográfica, o quintil de riqueza, variáveis que descrevem as calamidades e a informação sobre o próprio turismo doméstico em Moçambique. Para juntar, filtrar e processar as bases de dados de interesse, será usado o pacote estatístico SPSS¹ na sua versão 20.0, e para redigir tanto o relatório do trabalho, como os slides de apresentação será utilizado o software de digitação denominado L^AT_EX.

3.1.1 Tipo de pesquisa

Existem várias formas de classificar as pesquisas, no entanto, esse estudo pode ser classificado da seguinte forma:

- Sub ponto de vista da sua **natureza**: o estudo enquadra-se nas Pesquisas Aplicadas: Pelo facto deste ter como objetivo gerar conhecimentos para aplicação prática dirigidos à solução de problemas específicos, ela envolve verdades e interesses locais;

¹Statistical Package for Social Science

- ponto de vista da forma de **abordagem do problema**: Esse estudo pode ser classificado como quantitativa, pois, considera que tudo pode ser quantificável, o que significa traduzir em números opiniões e informações para classificá-las e analisá-las;
- Do ponto de vista **dos seus objetivos**: O presente estudo, pode ser visto como uma Pesquisa Explicativa: visa identificar os factores que determinam ou influenciam para a ocorrência dos fenômenos. Aprofunda o conhecimento da realidade porque explica a razão, o “porquê” das coisas;
- Sub ponto de vista dos **procedimentos técnicos**: Essa é uma Pesquisa Experimental, pois, ela determina um objecto de estudo, onde seleccionam-se as variáveis que seriam capazes de influenciá-lo, definem-se as formas de controle e de observação dos efeitos que a variável produz no objecto do estudo.

3.1.2 Amostragem e tamanho da amostra

O universo do IOF 2019/20 corresponde à população residente em território nacional. No entanto, a amostra do IOF 2019/20 foi seleccionada a partir de uma base de amostragem, conhecida por “Amostra Mãe 2017-2026”, que é construída a partir dos dados do Recenseamento Geral da População e Habitação de 2017, em que nesta amostra mãe, foi feita uma amostragem probabilística, estratificada e multi-etápica, em que a partir destas amostragens, foi extraída a amostra constituída pelos agregados familiares inquiridos pelo INE.

O tamanho da amostra foi obtido a partir de vários filtros feitos ao universo de 64.519 inquirido, onde foi usado o SPSS para remover as observações, cuja variável dependente “**Viajou**” não apresentava alguma resposta por parte dos inquiridos, desta feita, o tamanho da amostra usado para o presente estudo foi de $n = 480$.

3.1.3 Variáveis em análise

Para que fosse possível a realização do presente estudo, foram feitas junções de 4 bases de dados do IOF 2019/2023 que são: Membros, Calamidades, Quintil de riqueza e turismo doméstico, onde usou-se a variável `interview_key` como a chave, depois de ordená-la em ordem crescente em todas as bases de INE.

Com essas informações, foram extraídas as variáveis de perfil da amostra (Sexo, Idade, Nível académico, Estado civil, situação de empregabilidade, província e zona residencial), sobre as calamidades (Cheias e ciclones), quintil de riqueza, e sobre o turismo doméstico (Viajou, Transporte, Alojamento).

Deste modo, passam-se a apresentar na tabela (3.1), as variáveis em análise e as suas respectivas classificações.

Tabela 3.1: Descrição das variáveis em estudo

Variáveis	Descrição da variável	Classificação
Sexo	Sexo do indivíduo	Qualitativa nominal
Idade	Idade indivíduo	Quantitativa discreta
Civil	Estado civil do indivíduo	Qualitativa categórica
Nível	Nível académico do indivíduo	Qualitativa ordinal
Emprego	Se é trabalhador assalariado	Qualitativa categórica
Província	Província em que vive	Qualitativa categórica
Zona	Zona residencial do indivíduo	Quantitativa contínua
Riqueza	Em que quintil de riqueza pertence	Qualitativa Nominal
Transporte	Meio de transporte usado	Quantitativa contínua
Alojamento	Alojamento ocupado durante a viagem	Qualitativa categórica
Cheias	Foi afectado pelas cheias recentemente	Qualitativa categórica
Ciclone	Foi afectado pelo ciclone recentemente	Qualitativa categórica
Viajou	Nos últimos 3 meses viajou ou dormiu distante de casa	Qualitativa categórica

3.2 Métodos

Nesta secção do trabalho, são arrolados todas as técnicas estatísticas usadas para o alcance dos objectivos previamente definidos, deste modo, para fazer a descrição do perfil da amostra foram usados gráficos circulares que expressam de forma percentual cada categoria da variável e a tabela cruzada das variáveis. Por conseguinte foram realizados os seguintes testes da estatística multivariada:

3.2.1 Teste de independência χ^2 de Pearson

De acordo com Triola (2005), às vezes interessa ao investigador saber se duas variáveis cujas medidas foram obtidas nos mesmos indivíduos estão ou não associadas. Deste modo, neste estudo será usado o teste de independência qui-quadrado (Através da estatística χ^2 de Pearson) que nos permite determinar se duas variáveis são independentes ou não. Este teste é realizado sub as seguintes hipóteses:

H_0 : A variável em análise e a prática do turismo doméstico são independentes;

H_1 : A variável em análise e a prática do turismo doméstico não são independentes;

Regra de decisão:

Se o valor de χ^2 calculado for maior que o valor crítico ou o p-valor associado ao teste for inferior ao nível de significância ($\alpha = 5\% = 0.05$), então rejeita-se a hipótese nula e conclui-se que as duas variáveis não são independentes, ou seja, as variáveis estão associadas.

3.2.2 Método de estimação dos coeficientes

Como os parâmetros são desconhecidos, é necessário a estimação dos coeficientes da regressão logística, neste contexto, foi usado o método de máxima verossimilhança, pois, este método maximiza o poder preditivo dos coeficientes estimados, o que ajuda na determinação das probabilidades de ocorrência de um determinado evento (para esse estudo, seria a probabilidade do indivíduo praticar o turismo doméstico).

Após a estimação do modelo inicial, que é constituído pelo modelo nulo (aquele que contém apenas a constante) passa-se então, a considerar todas variáveis selecionadas, com objectivo de identificar quais são as variáveis que influenciam na probabilidade do indivíduo praticar o turismo doméstico.

3.2.3 Significância dos coeficientes estimados

Neste estudo, a verificação da significância dos parâmetros estimados foi feita pelos testes de razão de verossimilhança e Wald.

1. Teste de razão de verossimilhança

Para a realização deste teste, foi usada a função de verossimilhança para fazer as comparações entre os valores observados e esperados, em que através das equações (2.20) e (2.21), foi feita a análise conjunta da significância dos coeficientes sob as seguintes hipóteses:

$$H_0 : \hat{\beta}_{1j} = \hat{\beta}_{2j} = \dots = \hat{\beta}_{ij} = 0$$

$$H_1 : \exists \hat{\beta}_{ij} \neq 0$$

2. Teste de Wald

A estatística utilizada para a inclusão definitiva das variáveis no modelo final é a estatística de Wald, que faz a análise de significância individual de cada variável, essa estatística é dada pela equação (2.26) e realiza-se sob as seguintes hipóteses:

$$H_0 : \hat{\beta}_{1j} = 0, \hat{\beta}_{2j} = 0 \dots \hat{\beta}_{ij} = 0;$$

$$H_1 : \exists \hat{\beta}_{ij} \neq 0$$

Para a tomada de decisão, foram usados os p-valores associados aos testes de razão de verossimilhança e de Wald, em que, a um nível de significância de 0.05, rejeita-se a hipótese nula quando o p-valor é superior a $\alpha = 0.05$.

3.2.4 Avaliação do modelo ajustado

Para a avaliação do modelo ajustado primeiro recorre-se a -2LL (Log da função de verossimilhança). O procedimento consiste em comparar o valor final desta medida com o valor do modelo nulo (modelo com apenas a constante). Uma maior redução deste valor indica um bom ajuste do modelo. O poder explicativo do modelo é dado pelo valor de R^2 de Nagelkerke e quanto maior este valor, melhor será o ajuste (Hair *et al.*, 2005).

A outra medida usada para avaliar modelo ajustado é a estatística de Hosmer e Lemeshow. Onde as probabilidades estimadas foram ordenadas de mais baixa a mais alta e depois, fez-se o teste qui-quadrado para determinar se as frequências previstas estão próximas das frequências observadas calculado pela equação (2.24). Esse teste é realizado sob as seguintes hipóteses:

H_0 : O modelo estimado se ajusta aos dados;

H_1 : O modelo estimado não se ajusta aos dados.

Onde rejeita-se a hipótese nula, em que, no caso do p-valor associado ao teste ser inferior ao nível de significância de $\alpha = 0.05$. De seguida, foi usada a tabela de classificação para quantificar percentualmente as classificações correctas feitas pelo modelo estimado. Além disso, foi verificado o poder discriminativo do modelo estimado através da área abaixo da curva de ROC, conforme a tabela (2.2) que classifica o poder discriminativo do modelo ajustado.

3.2.5 Diagnóstico do modelo estimado

Geralmente diagnóstico dos modelos de regressão consiste em verificar os pressupostos destes, sendo que, nesse estudo o diagnóstico será feito certificando-se dos seguintes aspectos:

➤ **Multicolinearidade**

Segundo Maroco (2007), de forma geral os valores de VIF superiores a 10 e os da Tolerância menores ou iguais a 0,1 demonstram a presença de multicolinearidade, deste modo, essas variáveis têm problemas com a estimação dos $\hat{\beta}_i$ das suas variáveis independentes.

➤ **Análise residual**

Esta análise é feita levando em conta que, cerca de 95% dos valores $|e'_i|$ devem ser inferiores a 1.96, e qualquer observação com $|e'_i|$ superior a 1.96 $\cong 2$ pode ser classificadas como um "outlier", a um nível de significância de 5% ($\alpha = 0.05$), pois, com esta confiança o ($z_{0.975} = 1.96$). Ainda na análise residual, verifica-se a influência das observações no modelo estimado, onde são considerados observações influentes no modelo ajustado, aqueles cujos seus valores de $\Delta\chi_i^2$ são superiores a 3.84, pois, para $\alpha = 0.05$, uma vez que $\Delta\chi_i^2 \sim \chi_{(1)}^2$ e que o $\chi_{0.95;(1)}^2 = 3.84$.

Capítulo 4

Resultados e Discussão

Nesta secção, foram apresentados os resultados obtidos com o processamento das 480 observações filtradas da base de dados do IOF 2019/20, elaborado pela INE. Com esses resultados, espera-se responder os objectivos previamente definidos na introdução do presente trabalho académico.

4.1 Descrição do perfil da amostra

De acordo com a figura (4.1), pode-se constatar que, cerca de (62,29%) dos inquiridos que fizeram parte da amostra são do sexo masculino e a percentagem complementar ao sexo oposto. Destes inquiridos, foi verificada a situação de empregabilidade, onde a maior percentagem é dos que têm alguma ocupação remunerada (emprego), com um percentual de 67.71%.

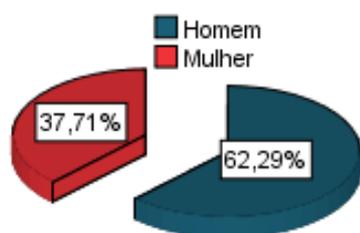


Figura 4.1: Distribuição dos inquiridos segundo o sexo

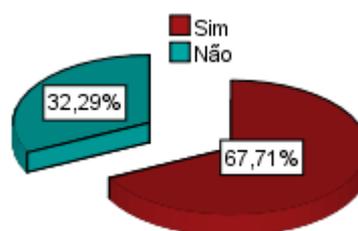


Figura 4.2: Distribuição dos inquiridos segundo a empregabilidade

Em relação as suas idades, os destaques foram para os inquiridos com as idades compreendidas dos 28 à 38 anos de idade, que detêm um percentual de 33,96%, seguido dos que têm as suas idades nos intervalos dos 18 à 28 anos (Com cerca de 24,58%), o terceiro colocado têm cerca de 24,38% dos inquiridos com idades compreendidas dos 38 à 48 anos de idade, conforme descrito na figura (4.1).

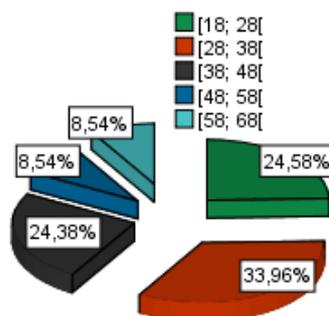


Figura 4.3: Distribuição dos inquiridos segundo a idade

De acordo com a distribuição do estado marital, verificou-se que, cerca de 60,63% dos inquiridos vive maritalmente com os seus parceiros, e os viúvos(as) representam a menor parte com cerca de 4,79% dos inquiridos.

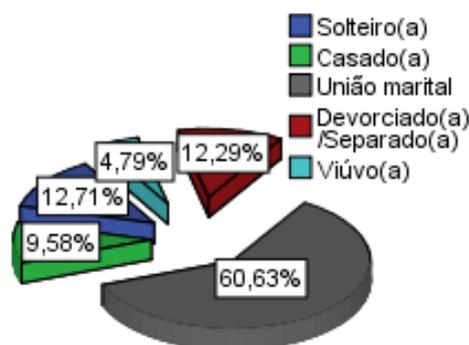


Figura 4.4: Distribuição dos inquiridos segundo o estado civil

Por conseguinte, na figura (4.5) foi apresentada a distribuição dos inquiridos segundo o seu nível acadêmico, onde verificou-se que, cerca de 43,96% dos inquiridos tinham o seu ensino primário completo, 30% destes têm com o nível médio, cerca de 16,46% dos inquiridos tinham o nível básico, sendo que, o nível superior apresenta 8,96% dos inquiridos e a menor proporção é dos pós-graduados com um percentual de 0,63%.

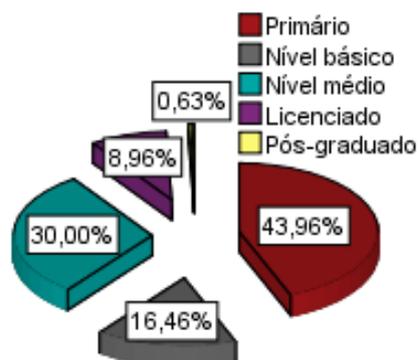


Figura 4.5: Distribuição dos inquiridos segundo a escolaridade

Em forma conjunta, foram analisada distribuição dos inquiridos pelos seus locais de residência (Zona urbana ou rural) de acordo com as suas províncias, onde verificou-se que, a zona urbana de destaque foi para a cidade de Maputo com 100%, seguido da província de Manica com 78,4% dos inquiridos, por sua vez a zona rural destacou-se na província de Zambézia, seguidos da província de Tete.

Tabela 4.1: Cruzamento das província e zona residencial

Variáveis	Zona residencial	
	Urbano	Rural
Nomes das províncias		
Cabo Delegado	62,5%	37,5%
Gaza	75,0%	25,0%
Inhambane	62,9%	37,1%
Manica	78,4%	21,6%
Maputo Cidade	100,0%	
Maputo Província	71,7%	28,3%
Nampula	67,5%	32,5%
Niassa	74,1%	25,9%
Sofala	63,3%	36,7%
Tete	55,6%	44,4%
Zambézia	42,1%	57,9%
Totais	70,6%	29,4%

Desta forma, prossegue-se com a verificação da associação entre as variáveis em estudo e a variável dependente **Viajou**, que faz menção destes terem realizado alguma actividade turística no período em análise.

4.2 Teste de independência das variáveis

Foi verificada a existência de associação entre as variáveis Zona residencial e a variável viajou, onde criou-se uma tabela de contingência (4.2), onde verificou-se que cerca de 286 inquiridos que residem na zona urbana não viajou, e na zona rural cerca de 119 deles não viajou.

Tabela 4.2: Contingência entre Zona*Viajou e a significância

Zona residencial	Viajou			χ^2 de Pearson	P-valor
	Não	Sim	Total		
Urbano	286	53	339	0,993	0,557
Rural	119	22	141		
Totais	405	75	480		

Realizado o teste de independência, onde verificou se que o p-valor associado ao teste χ^2 de Pearson é superior que o nível de significância ($\alpha = 5\% = 0.05$). Deste modo, existem evidências para não rejeitar a hipótese nula de que as variáveis em análise são independentes.

O mesmo teste foi realizado para verificar-se a situação de empregabilidade dos inquiridos está associada a prática do turismo doméstico, desta feita, de acordo com a tabela (4.3), o p-valor associado ao teste é superior ao $\alpha = 0.05$, o que levou a não rejeição da hipótese nula, afirmando que as variáveis emprego e viajou são independentes.

Tabela 4.3: Contingência entre Emprego*Viajou e a significância

	Viajou			χ^2 de Pearson	P-valor
Emprego	Não	Sim	Total	0,387	0,234
Sim	271	54	325		
Não	134	21	155		
Totais	405	75	480		

De seguida foi analisada a variável Sexo e Viajou, cerca de 271 do sexo masculino não viajou e 53 praticaram o turismo doméstico, sendo que, cerca de 159 do sexo feminino não chegou de viajar, e apenas 22 inquiridas viajaram. O teste cuja estatística de teste χ^2 de Pearson = 0.103 e o p-valor associado ao teste (0.065) que foi tabulado em (4.4), levaram a não rejeitar a hipótese nula, demonstrando que as variáveis em análise são independentes. Já que elas são independentes, então, pode se dizer que as variáveis sexo e viajou não estão associadas.

Tabela 4.4: Contingência entre Sexo*Viajou e a significância

	Viajou			χ^2 de Pearson	P-valor
Sexo	Não	Sim	Total	0,103	0,065
Masculino	246	53	299		
Feminino	159	22	181		
Totais	405	75	480		

De acordo com a tabela (4.5), pode-se verificar que os que vivem maritalmente destacaram se em maior número dos viajantes (com cerca de 245 inquiridos) e não viajantes (46 inquiridos). O χ^2 de Pearson foi de 4.389 e o p-valor associado ao teste (0.356) é superior a $\alpha = 0.05$, condicionando assim, as evidências para a não rejeição da hipótese nula, deste modo conclui-se que, o estado civil e o turismo doméstico são independentes.

Tabela 4.5: Contingência entre Civil*Viajou e a significância

	Viajou			χ^2 de Pearson	P-valor
Civil	Não	Sim	Total	4,389	0,356
Solteiro(a)	56	5	61		
Casado(a)	36	10	46		
União marital	245	46	291		
Divorciado(a)	48	11	59		
Viúvo(a)	20	3	23		
Totais	405	75	480		

Em análise das idades dos inquiridos e a prática do turismo doméstico, pode se verificar na tabela (4.6) que, o p-valor associado ao teste χ^2 de Pearson (0.356) é superior a $\alpha = 0.05$, assim sendo, não rejeita-se a hipótese nula e conclui-se que as variáveis são independentes.

Tabela 4.6: Contingência entre Idade*Viajou e a significância

	Viajou			χ^2 de Pearson	P-valor
Idade	Não	Sim	Total	4,602	0,331
[18; 28[95	23	118		
[28; 38[137	26	163		
[38; 48[98	19	117		
[48; 58[37	4	41		
[58; 68[38	3	41		
Totais	405	75	480		

Por conseguinte, verificou-se a existência de associação entre o estado civil e a variável viajou (Determinante da prática do turismo doméstico), a tabela (4.7) revelou que, o 2º grau e o nível médio de escolaridade têm o maior número dos que não viajaram com observações dos 116 inquiridos, do mesmo modo verificou-se que o menor número de observações é dos que fizeram doutoramento. Verificou-se ainda a associação entre as variáveis estado Civil e Viajou, foi assim constatado, pelo facto do p-valor associado ao teste χ^2 de Pearson (0.001) ser inferior ao nível de significância ($\alpha = 0.05$).

Tabela 4.7: Contingência entre Nível*Viajou e a significância

	Viajou			χ^2 de Pearson	P-valor
Idade	Não	Sim	Total	9,666	0,022**
Primário	184	28	212		
Nível básico	69	10	79		
Nível médio	121	23	144		
Nível superior	31	14	45		
Totais	405	75	480		

O alojamento dos inquiridos e a variável Viajou também foram analisadas com objectivo de verificar a relação entre elas. Pelos resultados tabulados em (4.8), onde foram ilustradas as observações de forma cruzada entre as duas variáveis, os valores de χ^2 de Pearson = 9,568 e o p-valor = 0,297 > $\alpha = 0,05$ levaram a não rejeição da hipótese nula, deste modo, as duas variáveis são consideradas independentes.

Tabela 4.8: Contingência entre Alojamento*Viajou e a significância

	Viajou			χ^2 de Pearson	P-valor
Alojamento	Não	Sim	Total	9,568	0,297
Hotel	6	2	8		
Pensão	34	3	37		
Acampamento	7	0	7		
Outro	331	67	398		
Totais	378	72	450		

De acordo com a tabela (4.9), verificou-se que cerca de 301 dos inquiridos que fizeram parte da amostra dos que usam autocarro, não viajou. Pode-se verificar ainda que, o χ^2 de Pearson = 10,502 e o p-valor associado ao teste (0,105), levam a concluir que o transporte usado pelos inquiridos e a variável viajou são independentes.

Tabela 4.9: Contingência entre Transporte*Viajou e a significância

	Viajou			χ^2 de Pearson	P-valor
Transporte	Não	Sim	Total	8,137	0,087
Autocarro	359	67	426		
Avião	8	5	13		
Barco	4	1	5		
Comboio	10	1	11		
Outro	24	1	25		
Totais	405	75	480		

Para as variáveis Cheias e viajou, verificou-se que, os que não sofreram com as cheias e não viajaram foram 367 e 67 deles viajaram, apenas 8 dos inquiridos sofreram de cheias mas mesmo assim viajaram (como ilustrado em 4.10). Realizado o teste de independência, verificou-se que as variáveis são independentes, pois, o p-valor associado ao teste (0,729) é superior ao $\alpha = 0.05$.

Tabela 4.10: Contingência entre Cheias*Viajou e a significância

	Viajou			χ^2 de Pearson	P-valor
Cheias	Não	Sim	Total	0,120	0,729
Não	367	67	434		
Sim	38	8	46		
Totais	405	75	480		

A variável referente as mudanças climáticas envolvendo ventos fortes (Ciclones) também foi analisada com a variável viajou, para certificar-se da sua independência, onde o p-valor associado ao teste é inferior ao nível de significância (p-valor = 0,006 < $\alpha = 0.05$), com essas evidências, rejeita a H_0 pode-se concluir que as variáveis em análise são dependentes.

Tabela 4.11: Contingência entre Ciclone*Viajou e a significância

	Viajou			χ^2 de Pearson	P-valor
Cheias	Não	Sim	Total	7,645	0,006**
Não	375	62	437		
Sim	30	13	43		
Totais	405	75	480		

Realizados os testes de independência entre as variáveis categóricas e a variável Viajou, verificou-se que, as únicas associadas a prática do turismo doméstico foram o nível acadêmico e o ciclone.

4.3 Regressão logística binária

Para verificar a influência das variáveis independentes à probabilidade dos inquiridos praticarem o turismo doméstico foi inicialmente verificada a multicolinearidade. Segundo a tabela (4.12), todos os valores da tolerância são superiores a 0.1 e os valores de VIF não superam 10, isso significa que, todas as variáveis não apresentam alguma colinearidade, assim sendo, pode se prosseguir com a análise de regressão logística.

Tabela 4.12: Estatísticas de Multicolinearidade

Variáveis em análise	Estatísticas	
	Tolerância	VIF
Zona	0,863	1,158
Emprego	0,765	1,307
Sexo	0,878	1,139
Civil	0,825	1,212
Idade	0,851	1,176
Escolaridade	0,852	1,174
Riqueza	0,985	1,015
Transpor	0,979	1,022
Alojamen	0,980	1,021
Cheias	0,968	1,033
Ciclone	0,988	1,012

As outras suposições dessa regressão, foram verificadas depois de estimar o modelo de regressão logística que se ajusta aos dados, através da análise residual, descrita no processo do diagnóstico do modelo ajustado. Assim sendo, prossegue se com a análise da regressão logística, tendo se começado por codificar a variável dependente prática de turismo descrita por "viajou".

Tabela 4.13: Codificação de variável Viajou

Valor original	Valor interno
Não viajou	0
viajou	1

4.3.1 Análise do modelo nulo

Em análise do modelo nulo, que contém apenas a constante (β_0), pode se verificar na tabela (4.14) que, a constante revelou se significativa, pois, o p-valor = 0,000 e é inferior ao $\alpha = 0,05$, e rejeita hipótese de que $\beta_0 = 0$.

Tabela 4.14: Equação do modelo nulo

	β_0	S.E.	Wald	df	Sig.	Exp(B)
Constante	-1,609	0,122	172,686	1	0,000**	0,200

A tabela de classificação inicial (4.15), foi levado em consideração as observações com maior frequência sobre a variável viajou, onde a categoria "Não" esteve em destaque e explicam cerca de 83.3% dos casos correctamente classificados.

Tabela 4.15: Tabela de classificação inicial

Categorias	Viagens previstas		Percentagem	
	Não	Sim	Correcto	
Viagens observadas	Não	400	0	100,0
	Sim	80	0	0,0
Percentagem global				83,3

Por conseguinte, passa-se a analisar o modelo ajustado, com vista a certificar se da existência de alguma influência das variáveis independentes e também compara-lo com o modelo nulo.

4.3.2 Análise do modelo ajustado

Para fazer análise do modelo ajustado foi inicialmente feito teste de razão de verossimilhança, também denominado teste de significância dos coeficientes. De acordo com a tabela (4.16), pode se verificar que o p-valor associado ao teste (0,000) é inferior ao nível de significância $\alpha = 0,05$, o que leva a rejeitar a hipótese nula, concluindo se assim que, no modelo ajustado existe pelo menos um coeficiente estimado estatisticamente diferente de zero.

Tabela 4.16: Teste de significância dos coeficientes (Omnibus)

	Chi-square	df	Sig.
Step	73,706	36	0,000
Block	73,706	36	,000
Model	73,706	36	,000

A tabela (4.17), apresenta o resumo do modelo ajustado, que contém informações também usadas para avaliar o modelo, onde pode se ver que, cerca de 32,8% da probabilidade da prática do turismo é determinada pelas variáveis independente significativas.

Tabela 4.17: Resumo do modelo ajustado

Step	-2 Log Likelihood	R ² de Cox & Snell	R ² de Nagelkerke
1	314,810	0,190	0,328

Segundo os resultados descritos em (4.18), o p-valor (0.674) associado ao teste de Hosmer & Lemeshow, não rejeita se a hipótese nula de que o modelo ajustado se ajusta aos dados.

Tabela 4.18: Teste Hosmer & Lemeshow

Step	χ^2 - Quadrado	df	Sig.
1	5,758	8	0,674

4.3.3 Avaliação de ajuste do modelo

Depois de certificar-se da existência de pelo menos um parâmetro significativo para a previsão da prática do turismo, deste modo, passa-se para a verificação da capacidade do modelo logístico classificar correctamente os casos.

Tabela 4.19: Tabela de classificação final

Categorias		Viagens previstas		Percentagem
		Não	Sim	Correcto
Viagens observadas	Não	397	8	98,0
	Sim	56	19	55,3
Percentagem global				89,7

De acordo com a tabela (4.19), fez-se a classificação dos sujeitos observados e os previstos pelo modelo, onde verificou-se que a sensibilidade do modelo é de 55.3%, e a especificidade como indicador de classificação correcta dos que foram previsto como não tendo praticado o turismo doméstico, quando de facto não o praticaram com um percentual de 98.0%, por isso, pode-se dizer que cerca de 89.4% dos casos, foram correctamente classificados pelo modelo ajustado.

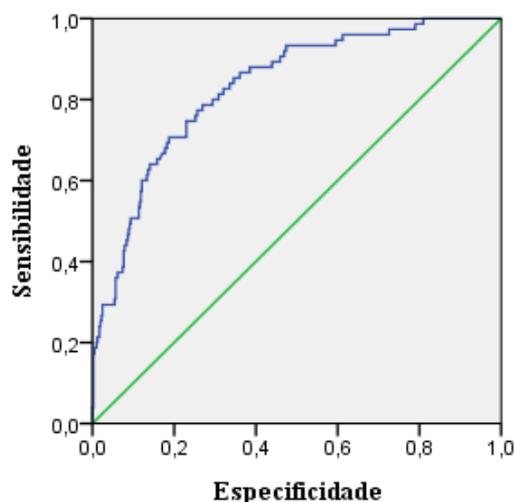


Figura 4.6: Representação da curva de ROC

Para complementar as classificações correctas dos casos, fez-se a análise da curva de ROC, em que de acordo com os resultados obtidos, pode-se verificar que a área abaixo da curva ROC, é maior do que 50%, pois, está acima da recta que faz o **Cut-Off** pela metade (0.5), como ilustra a figura (4.6) acima. Como forma de confirmar a representação gráfica descrita em (4.6), verificou-se que, sub a suposição da variável dependente não ser paramétrica, a Tabela (4.20) apresenta o teste formal.

Tabela 4.20: Área sob a curva

			Intervalo de confiança	
			Assintótico a 95%	
Área	Erro padrão ^a	Sig. assintótico ^b	inferior	superior
0,831	0,025	0,000	0,783	0,880

a. Sub a suposição da variável não paramétrica

b. Hipótese nula: Área da curva do ROC = 0.5

Como pode se verificar na tabela (4.20), o valor da área abaixo da curva de ROC que é igual a $0.831 > 0.8$ (80%) superando o valor do corte 0.5, e o p-valor = $0.000 < 0.05$ leva a concluir que o modelo ajustado têm um bom poder discriminativo.

4.3.4 Significância individual dos coeficientes

A significância de cada coeficiente foi feita através do teste de wald que foi apresentado nas tabelas (4.21) e (4.22), onde as variáveis significativas foram destacadas e sinalizadas com o sinal (**) como um simbolo de variável ou categoria estatisticamente significativa.

Tabela 4.21: Significância individual dos coeficientes

Variáveis	β	S.E.	Wald	Df	Sig.	Exp(β)
Província			24,677	10	0,006*	
Cabo delgado	0,668	0,637	1,101	1	0,294	1,950
Niassa	-0,9666	0,827	1,364	1	0,243	0,381
Nampula	-0,335	0,640	0,275	1	0,600	0,715
Zambézia	-1,222	0,717	2,908	1	0,088	0,295
Tete	-1,769	0,686	6,651	1	0,010*	0,171
Manica	-1,495	0,598	6,263	1	0,012*	0,224
Sofala	-0,580	0,654	0,785	1	0,376	0,560
Inhambane	0,120	0,628	0,037	1	0,848	1,127
Gaza	-1,181	0,634	3,462	1	0,063	0,307
Maputo Província	-0,312	0,581	0,289	1	0,591	0,732
Maputo Cidade (ref.)	—	—	—	—	—	—
Zona (Urbano)	0,165	0,360	0,210	1	0,647	1,179
Rural(ref.)	—	—	—	—	—	—
Emprego (Sim)	0,079	0,384	0,042	1	0,837	1,082
Não (ref.)	—	—	—	—	—	—

A tabela (4.22) apresenta a continuação da tabela acima, que contém os resultados do teste de significância individual dos coeficientes estimados (Wald).

Tabela 4.22: Significância individual dos coeficientes - Continuação

Variáveis	β	S.E.	Wald	Df	Sig.	Exp(β)
Sexo (Masculino)	0,849	0,371	5,232	1	0,022**	2,338
Feminino (ref.)	—	—	—	—	—	—
Civil			8,575	4	0,073	
Solteiro(a)	-1,646	0,959	2,947	1	0,086	0,193
Casado(a)	-0,297	0,885	0,113	1	0,737	0,743
Maritalmente	-0,857	0,829	1,071	1	0,301	0,424
Divorciado(a)	0,025	0,858	0,001	1	0,977	1,025
Viúvo(a) (ref.)	—	—	—	—	—	—
Idade			11,300	466	0,023*	
[18; 28[1,899	0,702	7,328	1	0,007*	6,680
[28; 38[1,150	0,696	2,731	1	0,038*	3,160
[38; 48[1,206	0,685	3,100	1	0,078	3,339
[48; 58[0,235	0,837	0,079	1	0,778	1,265
[58; 68[(ref.)	—	—	—	—	—	—
Escolaridade			10,079	4	0,039	
Primário	-4,001	1,504	7,078	1	0,008*	0,018
Nível básico	-4,124	1,537	7,199	1	0,007*	0,016
Nível médio	-3,985	1,507	6,997	1	0,008*	0,019
Licenciatura	-3,156	1,522	4,300	1	0,038*	0,043
Pós-graduação (ref.)	—	—	—	—	—	—
Riqueza	0,064	0,113	0,318	1	0,573	1,066
Transporte			3,516	4	0,475	
Autocarro	1,305	1,062	1,511	1	0,219	3,688
Avião	2,174	1,263	2,964	1	0,085	8,795
Barco	1,965	1,612	1,487	1	0,223	7,138
Comboio	1,707	1,346	1,608	1	0,205	5,514
Outro	—	—	—	—	—	—
Alojamento			2,318	4	0,677	
Hotel	-0,514	0,672	0,585	1	0,444	0,598
Pensão	0,186	0,887	0,044	1	0,833	1,205
Campismo	-0,867	0,663	1,708	1	0,191	0,420
Outros (ref)	—	—	—	—	—	—
Cheias (Sim)	0,128	0,491	0,068	1	0,794	1,137
Não (ref.)	—	—	—	—	—	—
Ciclone (Sim)	-1,008	0,439	5,275	1	0,022**	0,365
Não (ref.)	—	—	—	—	—	—
Constante	-38,975	4093	0,000	1	0,999	0,000

Os resultados acima tabulados revelam que as variáveis que influenciam de forma significativa a prática do turismo doméstico são: A província, sexo, idade, escolaridade e o ciclone.

4.3.5 Interpretação e estimação do modelo ajustado

Os coeficientes $\hat{\beta}_i$ que têm sinais positivos, influenciam no aumento da chance ou probabilidade dos inquiridos terem praticado turismo doméstico no período em análise, já os coeficientes negativos influenciam negativamente nas chances da ocorrência da prática do turismo. Assim sendo, a um nível de confiança de 95%, verificou-se que os resultados tabulados na tabelas (4.21) e (4.22), por conseguinte passa-se a interpretar os valores de $\mathbf{Exp}(\beta)$ nos seguintes termos:

- Nota-se uma redução na chance de um inquirido escolhido aleatoriamente praticar o turismo doméstico em cerca de 89,3% e 71,1%, nas províncias de Tete e Gaza respectivamente, quando comparados com os residentes da cidade de Maputo, considerando o resto das variáveis constantes. Para o sexo, verificou se que as chances de um inquirido praticar turismo doméstico é 2,338 vezes superior para os do sexo masculino, quando comparados aos do sexo feminino;
- Na variável escolaridade, constatou se uma redução das chances de praticar o turismo doméstico, em 98,2% para os com o ensino primário. As idades inferiores a 18 e dos inquiridos com as idades compreendidas dos 18 à 28 anos têm o aumento das chances em 6,680 e estas chances vão reduzindo quando comparadas com os mais idosos;
- Em relação aos ciclones, verifica-se que há uma redução na chance dum inquirido selecionado ter praticado turismo doméstico, na ordem de 63,5% dos que já sofreram de ciclone, quando comparado aos que não.

Usando as variáveis significativas, o modelo de regressão logística pode ser descrita substituindo as variáveis significativas na equação que se segue:

$$\hat{\pi}(x) = P(Y = 1|\mathbf{X}) = \frac{e^{(\hat{\beta}_0 + \sum_{i=1}^p \hat{\beta}_i x_i)}}{1 + e^{(\hat{\beta}_0 + \sum_{i=1}^p \hat{\beta}_i x_i)}}$$

Onde para determinar a probabilidade de um inquirido praticar o turismo doméstico, deve se substituir a parte da combinação linear das variáveis influentes como demonstrado de seguida:

$$\begin{aligned} \hat{\beta}_0 + \sum_{i=1}^p \hat{\beta}_i x_i &= \hat{\beta}_{Pro} * \mathbf{Província} + \hat{\beta}_{Sex} * \mathbf{Sexo} + \hat{\beta}_{Ida} * \mathbf{Idade} \\ &+ \hat{\beta}_{Esc} * \mathbf{Escolaridade} + \hat{\beta}_{Cic} * \mathbf{Ciclone} \end{aligned}$$

Onde:

β_0 é a constante; e os coeficientes estimado das variáveis significativas: $\hat{\beta}_{Pro}$ - estimativa da província, $\hat{\beta}_{Sex}$ - coeficiente do sexo, $\hat{\beta}_{Esc}$ - coeficiente da escolaridade, $\hat{\beta}_{Ida}$ - coeficiente da idade e $\hat{\beta}_{Cic}$ - coeficiente da variável ciclone. Cada coeficiente é substituído de acordo com a categoria da variável significativa.

4.3.6 Análise residual

O gráfico (4.7), sugere um bom ajuste do modelo aos dados, visto que, a linha central denominada de "Lowess", apresentam alguns pontos sobre a linha central, e alguns pontos atípicos são os que superam o resíduo " $\varepsilon_i = 2$ ".

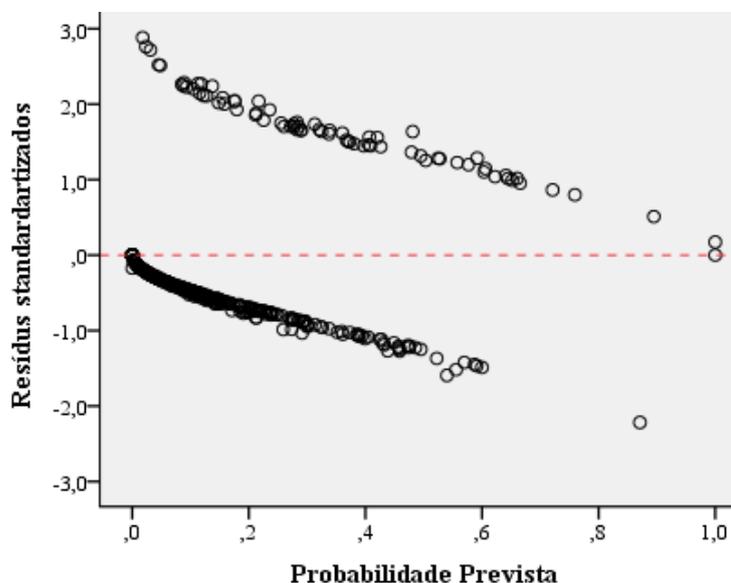


Figura 4.7: Representação gráfica dos resíduos estandardizados

Os tais pontos atípicos representados graficamente em (4.7), foram tabulados em (4.23). De acordo com essa tabela acima, cerca de 15 observações revelam ser atípicas ou *outliers*. De seguida, foram representadas graficamente as observações influentes.

Tabela 4.23: Casos de anomalias listados

Caso	Selecionado	Observado	Previsto	Grupo previsto	Variável temporária	
					Resid	ZResid
24	S	S**	0,093	N	0,907	3,127
34	S	S**	0,128	N	0,872	2,612
40	S	S**	0,217	N	0,783	1,901
68	S	S**	0,113	N	0,887	2,801
100	S	S**	0,024	N	0,976	6,440
131	S	S**	0,148	N	0,852	2,395
143	S	S**	0,176	N	0,824	2,165
164	S	S**	0,118	N	0,882	2,732
179	S	S**	0,155	N	0,845	2,334
184	S	S**	0,087	N	0,913	3,234
190	S	S**	0,089	N	0,911	3,198
236	S	S**	0,085	N	0,915	3,282
292	S	S**	0,031	N	0,969	5,574
328	S	N**	0,871	S	-0,871	-2,597
393	S	S**	0,045	N	0,955	4,584

De acordo com a representação do quadrado da variação dos resíduos ($\Delta\chi^2$) e a probabilidade prevista, onde verificou-se que, existem observações que superam 4 e são consideradas influentes no modelo de regressão estimado ou ajustado, mas para que essas observações sejam assim consideradas e removidas têm que se verificar a sua quantidade.

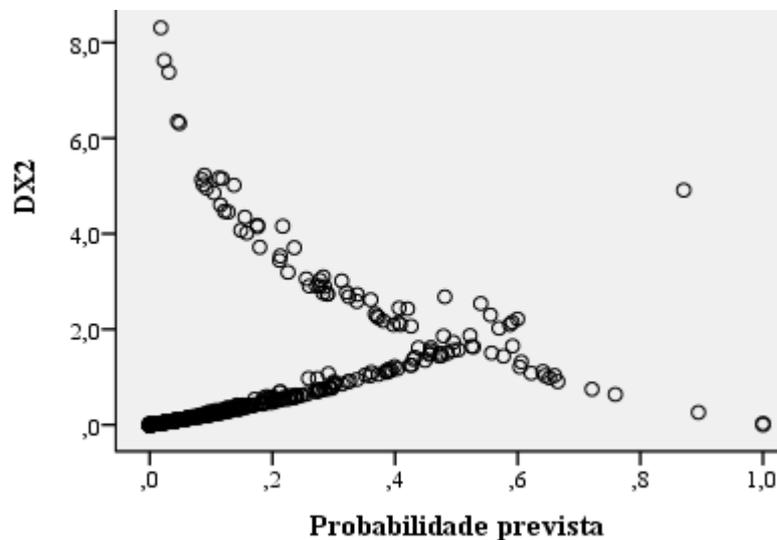


Figura 4.8: Representação gráfica dos resíduos estandardizados

Deste modo, através da figura (4.8), foi possível concluir que, embora as observações influentes, a percentagem destas não superou os 10% da amostra, então, a influência dessas observações não é significativa ou não afectam de forma significativa no modelo ajustado.

4.4 Discussão dos resultados

Diante dos estudos relacionados com o tema referente aos factores que influenciam para a prática do turismo doméstico em Moçambique, dentre os 3 estudos enunciados os estudos do Wache (2021) e Silva (2019), foram classificadas como pesquisas quantitativas, bibliográfica e documental, sendo que, os preços de produto e serviços, alojamento turístico, estabilidade política e segurança, infra-estruturas e acessibilidade, patrimônio natural, patrimônio histórico e cultural, hospitalidade para com os turistas, turismo de sol e mar foram descritas como influentes na prática do turismo pelo Wache (2021).

Entretanto, de acordo com Silva (2019), A variável localização geográfica representa alguma oportunidade, mas também um desafio para os praticantes do turismo, pois, por um lado a localização geográfica pode ser visto como lugares paradisíacos que propiciam o turismo, com desastres naturais excepcional e que impulsionem a prática do turismo, e por outro lado, dependendo da sua localização geográfica, os outros locais podem propiciar a vulnerabilidade aos ciclones, cheias e inundações, o que têm posto em causa os investimentos feitos em infra-estruturas em geral e nas facilidades turísticas em particular. Neste contexto, nesse estudo existe um ponto de concordância em relação a localização geográfica, pois, as variáveis que divergem da localização como as províncias, mas a variável ciclones também se revelou ser influente na prática do turismo doméstico.

O estudo levado a cabo pelo Gêmo (2017), foi uma pesquisa quantitativa e constatou-se que, as variáveis preços de produtos e serviços, alojamento turístico, estabilidade política e segurança, infra-estruturas e acessibilidade, patrimônio natural, patrimônio histórico e cultural, hospitalidade para com os turistas, turismo de sol e mar. O que revelou vários pontos de discordância com o presente estudo, mas acredita se que, se algumas das variáveis incluídas fizessem parte do estudo, de certeza algumas dessas variáveis seriam consideradas influentes na prática do turismo doméstico. Toda via, neste estudo, as variáveis província, Sexo, estado civil, idade e ciclones revelaram ser influentes para a prática do turismo doméstico pelos moçambicanos no período em análise.

Capítulo 5

Conclusões e Recomendações

Neste capítulo, foram apresentadas as principais conclusões do estudo, bem como as recomendações consideradas pertinentes pelo autor.

5.1 Conclusões

Com base nas análises feitas através das 480 observações, que foram filtradas da base de dados da IOF 2019/2020, tendo sido a variável viajou a chave para tal filtração, com o auxílio das técnicas da estatística multivariada, foi possível analisar os factores que influenciam na prática do turismo doméstico em moçambique no período de referência, deste modo passa-se a apresentar as principais conclusões:

- Os inquiridos nesse estudo foram caracterizados com um maior número de Homens (62.29%), em que destes cerca de 67,71% são assalariados. Em relação as idades, verificou-se que, os grupos com maior frequências são os que têm idades compreendidas dos [28; 38] e [18; 28] anos de idade com 33,96% e 24,58% respectivamente, a maioria dos inquiridos vivem em união marital (60,63%) e apenas têm o ensino primário concluído. Além dos aspectos descritos acima, verificou-se de forma cruzada que a maioria dos inquiridos residem em zonas urbanas com destaque a cidade de Maputo, em que todos fazem parte da zona urbana, seguido de Manica (78,4%) e Gaza (75%) ambas percentagens dos inqueridos em zonas urbanas.
- Foram realizados testes de independência entre as variáveis independentes (qualitativas categóricas) e a dependente denominada Viajou (Também categórica), onde constatou-se que, apenas as variáveis nível académico e o ciclone demonstraram estarem associados à prática do turismo doméstico.
- A estimação do modelo logístico foi possível através da verificação das variáveis estatisticamente significativas para a probabilidade do inquirido praticar o turismo doméstico, onde verificou-se que as variáveis província, sexo, idade, escolaridade e o ciclone demonstraram serem importantes para estimar a probabilidade de um inquerido ser praticante do turismo doméstico.

As duas províncias da região central (Tete e Manica) demonstraram a redução das chances deste praticarem o turismo doméstico quando comparados aos residentes da cidade de Maputo, os que são do sexo masculino demonstraram ter 2,338 vezes a chance de viajarem em comparação com as mulheres. Quanto mais novos ou jovens os inquiridos forem, estes têm mais chances de viajarem quando comparados aos mais velhos. Já em relação ao nível académico, verificou-se que os inquiridos com os níveis mais baixos que os pós-graduados tinham menores chances de viajarem, o mesmo refletiu para os que sofreram pela ocorrência de algum ciclone, cujas suas chances tinham uma redução de até 63,5% quando comparados aos inquiridos que não sofreram de nenhum desastre natural (ciclone).

5.2 Recomendações

Com as conclusões tiradas no presente estudo, delimitaram se algumas recomendações que acredita se ser importante para impulsionar o turismo doméstico em Moçambique, que são:

Ao governo moçambicano

As recomendações destinadas ao governo foram:

1. Moçambique precisa de promover mais instituições para a formação profissional qualificada a todos os níveis, ou seja, precisa de formar e consciencializar as pessoas, observando com atenção a importância das profissões de turismo ao nível das escolas;
2. Moçambique precisa de políticas próprias e estabilidade para desenvolver a área do turismo, pois, o mesmo têm um grande potencial por aproveitar;
3. Promover mais estratégias de divulgação do seu potencial turístico moçambicano, de procurar condicionar formas para criação de mais parceiros e expandir a indústria extractiva que é um dos factores impulsionadores para o turismo, uma vez que o país é cada vez mais procurado para negócios e investimentos;

A comunidade

1. Os poderes públicos locais e centrais têm que mudar na forma como vêm a cultura, porque só com cultura pode-se olhar para frente e tomar-se decisões sábias e consequentes;
2. Cooperação das comunidades com as entidades empregadores da área, cultivando um ambiente que melhoraria o turismo em Moçambique, bem como procurar condições para o estabelecimento de parcerias entre a comunidade e os investidores, públicas-privadas

para melhor criação de políticas em matérias de turismo regidas com vista a garantir a sua implementação.

Aos futuros investigadores

1. Pode-se fazer uma análise mais profundada das diversas regiões específicas do país, tendo em conta as enormes diversidades de produtos turísticos, desde praias com águas cristalinas, um património cultural bastante rico e zonas florestais de grande valor a nível mundial;
2. Pode-se enquadrar e aplicar-se o inquérito as agências de viagens moçambicanas que vendem pacotes tanto nacionais como internacionais, com vista a perceber qual a percentagem nacional e internacionais registadas nas diferentes agências, quais são os principais clientes e os pacotes turísticos;
3. Pode-se também fazer o mesmo inquérito aos turistas estrangeiros que já fizeram turismo em moçambique com vista a compreender quais são os principais motivos que levam os mesmos a fazerem turismo em Moçambique, bem como saber quais os principais motivos que podem contribuir para o aumento do turismo em Moçambique, entre outros.

Referências bibliográficas

1. Acerenza, M. A. (2002). Administração do turismo: conceituação e organização (Vol. 1). Bauru, SP: EDUSC;
2. Agranonik, M. (2005) . Técnicas de diagnóstico aplicadas ao Modelo de Regressão Logística. Monografia para obtenção de título de Bacharel em Estatística, Porto Alegre, Brasil;
3. Agresti, A. (2007) . An introduction to categorical data analysis . New Jersey, Florida, 2ª edição, Editora John Wiley and Sons;
4. Agresti, A. e Franklin, C. (2013). Statistics: the art and science of learning from data. 3ª ed;
5. Assis J. P.; Sousa, R. P. e Dias, C. T. S. (2020), Glossário de Estatística. Editora Universitária (Edufersa);
6. Beni, M. C. (2007). Análise estrutural do turismo (12ª ed.). São Paulo: Senac São Paulo;
7. Bittencourt, H. R. (2003). Regressão logística politômica: revisão teórica e aplicações: ACTA SCIENTIAE . 5(1), pp.77-86;
8. Carvalho F. C. C. e Pimentel T. D. (2012). A Influência dos Factores Externos sobre os Destinos Turísticos, IX Seminário da Associação Nacional Pesquisa e Pós-Graduação em Turismo - Universidade Anhembi Morumbi - São Paulo;
9. Corrar, L. J., Paulo, E. e Filho, J. M. (2007). Análise multivariada: para cursos de administração, ciências contábeis e economia. FIPECAFI – Fundação Instituto de Pesquisas Contábeis, Atuariais e Financeiras. São Paulo: Atlas;
10. Cox, D. e Snell, E. (1989). Analysis of Binary Data, Second Edition. [S.l.]: Taylor & Francis, (Chapman & Hall/CRC Monographs on Statistics & Applied Probability). ISBN 9780412306204;
11. Duhamel, P. H. e Sacareau, I. E. (1998). Tourisme dans le monde, Coll. Prépas, Paris: Armand Colin;
12. Gêmo R. C. (2017). Factores críticos do sucesso do turismo em Moçambique, Universidade Lusófona de Humanidades e Tecnologias Escola de Comunicação, Arquitectura, Artes e Tecnologias da Informação, Lisboa;
13. Guambe, J. J. J. et al. (2021). Covid-19, Transporte Aéreo e Turismo em Moçambique. Geo UERJ, n. 39, jul;

14. Hair, J. F., Anderson, R. E., Tatham, R. L. e Black, W. C. (2005) . Análise multivariada de dados . 5^a edição, Porto alegre, Bookman.
15. Hair, J. F., Anderson, R. E., Tatham, R. L. e Black, W. C. (2009). Análise multivariada de dados. 6^a Edição, Porto Alegre. Editora Bookman;
16. Hosmer, W. D. e Lemeshow, S. (1989) . Applied Logistic Regression. New York;
17. Hosmer, D. W. e Lemeshow, S. (2000). Applied Logistic Regression. 2 Ed. New York: John Wiley & Sons;
18. Maroco, J. e Bispo, R. (2003). Estatística Aplicada às Ciências Sociais e Humanas, Climepsi, Lisboa;
19. Maroco, J. (2007). Análise Estatística com utilização do SPSS. 3^a Edição. Edições Sílabo;
20. Ministério da Cultura e Turismo (2015). Plano Estratégico para o Desenvolvimento do Turismo em Moçambique 2016-2025. Maputo: MITUR;
21. Moçambique (2004). Plano Estratégico para o Desenvolvimento do Turismo em Moçambique 2004-2013. Maputo: MITUR;
22. Mota, K. C. N. (2001). Marketing Turístico: promovendo uma atividade Sazonal. São Paulo: Atlas;
23. Organização Mundial de Turismo (OMT, 2003). Introdução ao Turismo, Roca;
24. Palhares, G. L. (2006). Ameaças, crises e riscos no turismo globalizado e a suscetibilidade do sistema de transporte. In: AZAMBUJA, M. S. de. (Org.). O turismo e a ação das variáveis controláveis e incontroláveis(pp. 103-123). Porto Alegre: EDIPUCRS;
25. Paula, G. A. (2013). Modelos de Regressão com Apoio Computacional. São Paulo: Universidade de São Paulo;
26. Pearce, D. G. (2003) . Geografia do turismo: fluxos e regiões no mercado de viagens. São Paulo: Aleph;
27. Rabahy, W. (2003). Turismo e desenvolvimento: estudos econômicos e estatísticos no planejamento. Barueri, SP: Manole;
28. Salvian, M. (2016). Multicolinearidade. Piracicaba - São Paulo, Brazil;
29. Silva, J. J. (2019). Turismo em Moçambique: oportunidades, desafios e riscos, Universidade Pedagógica de Maputo, AbeÁfrica: Revista da Associação Brasileira de Estudos Africanos, v.03, n.03, p.58
30. Triola, M.F. (2005). Introdução à Estatística (9^a ed.) Ri ode Janeiro: LTC Editora;

31. Truett, J.; Cornfield, J. e Kannel, W. (1967) A multivariate analysis of the risk of coronary heart disease in Framingham. *Journal of Chronic Diseases*. v. 20, pp. 511-524;
32. Wache, T. A. (2021). *Factores Determinantes na Escolha de um Destino Turístico: Caso do Município de Inhambane*, Escola Superior de Hotelaria e Turismo de Inhambane, trabalho para obtenção do grau de Licenciatura em Informação Turística;

APÊNDICES

Tabela 5.1: Tabela de contingência para teste de Hosmer e Lemeshow

Grupos	Viajou = Não		Viajou = Não		Total
	Observado	Esperado	Observado	Esperado	
1	48	47,917	0	0,083	48
2	46	47,165	2	0,835	48
3	47	46,415	1	1,585	48
4	46	45,616	2	2,384	48
5	44	44,102	4	3,898	48
6	44	42,788	4	5,212	48
7	39	40,849	9	7,151	48
8	40	37,226	8	10,774	48
9	28	32,054	20	15,946	48
10	23	20,867	25	27,133	48