



FACULDADE DE CIÊNCIAS  
Departamento de Matemática e Informática

---

Trabalho de Licenciatura em  
Estatística

**Análise dos Factores de Risco Associados a Ocorrência de Hepatite B  
nos Profissionais de Saúde do Hospital Central de Maputo.**

**Autor: Castro Rone Edmundo Maússe**



FACULDADE DE CIÊNCIAS  
Departamento de Matemática e Informática

Trabalho de Licenciatura em  
Estatística

**Análise dos Factores de Risco Associados a Ocorrência de Hepatite B  
nos Profissionais de Saúde do Hospital Central de Maputo.**

**Autor:** Castro Rone Edmundo Maússe

**Supervisor:** Paulo Ziaveia, MSc, UEM

Maputo, Julho 2024

## **Dedicatória**

*Este trabalho é dedicado aos meus pais, que tanto batalharam  
para que este dia se tornasse possível.*

Edmundo Fabião Maússe e

Rute João Mahanjane

## **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, Julho 2024

---

Castro Rone Edmundo Maússe

## **Agradecimentos**

Agradeço primeiramente a **DEUS**, pelo dom da vida e por ter me iluminado até aqui.

Ao meu supervisor, **Paulo Ziaveia**, MSc, UEM, pela sua disponibilidade, por ter aceitado e acreditado neste projecto, pelo conhecimento, profissionalismo, paciência, dedicação, esforço, pois várias foram as dificuldades para que este projecto se tornasse possível.

Ao **Dr. Jahit Sacarlal**, por ter aceitado, acreditado e disponibilizado os dados para a elaboração deste projecto de pesquisa, e pelo seu apoio durante o desenvolvimento do mesmo.

Aos meus pais **Edmundo Fabião Maússe e Rute João Mahanjane**, por terem me mostrado desde cedo a importância da educação e aos meus irmãos **Fabião Edmundo Maússe e Delfa Edmundo Maússe**, pelo amor, carinho, apoio, compreensão que sempre me concederam.

A minha família, minha esposa **Idalina Virgílio Macamo** e meu filho **Eddy de Castro Maússe**, por estarem presente e me motivarem diariamente para superar os obstáculos da vida.

A dra. **Lágrima Maússe** por ter me direccionado a este mundo de dados, no curso Técnico Médio de Estatística Sanitária e conseqüentemente o curso de Licenciatura em Estatística.

A toda equipe de digitação do C.S. 1 de Junho, em especial ao meu superior hierárquico **Anibal Massicane**, Oficial de Monitoria e Avaliação do Centro de Colaboração em Saúde, por ter dado apoio incondicional desde o início do projecto.

A todo o corpo docente do DMI, em particular os docentes do curso de Estatística, pelo conhecimento transmitido durante o curso.

A todos os colegas do curso de Estatística, em especial ao **Alfredo Baúle Júnior e Euflásio Francisco**, pela amizade, paciência e colaboração durante todo o curso.

A todos aqueles que não foram mencionados, e que directa ou indirectamente contribuíram para o término do curso e a realização deste trabalho.

## **Resumo**

A infecção pelo vírus de Hepatite B é uma preocupação em saúde pública contemporânea mundial, principalmente para os profissionais de saúde, por estarem expostos a sangue e outros fluidos ou tecidos orgânicos potencialmente contaminados, e apresentarem potencial risco de infecção e transmissão por várias doenças. Neste contexto, o presente estudo visa analisar os factores de risco associados à ocorrência do Vírus de Hepatite B nos profissionais de saúde. Para tal foram usados dados fornecidos pela Faculdade de Medicina da Universidade Eduardo Mondlane, do último estudo realizado sobre o vírus de Hepatite B nos profissionais de saúde, no Hospital Central de Maputo no ano 2015, cujo estudo é de coorte transversal. A população do estudo foi composta por profissionais de saúde do Hospital Central de Maputo, e usou-se a amostragem estratificada para se ter o número de participantes de cada departamento que fará parte da amostra, em que número de profissionais de saúde em análise foi de 587. A análise estatística foi feita mediante a aplicação de teste F de Thomas e Rao para independência e modelos de regressão logística múltipla. Com recurso ao modelo de regressão logística, verificou-se que a partilha de escovas e máquinas de barbear e os acidentes ocupacionais são os factores risco associados à ocorrência da hepatite B nos Profissionais de Saúde do Hospital Central de Maputo.

**Palavras-chave:** Hepatite B, Modelos Lineares Generalizados, Regressão Logística.

## **Abstract**

Hepatitis B virus infection is a contemporary global public health concern, especially for healthcare workers, as they are exposed to blood and other potentially contaminated organic fluids or tissues, and are potentially at risk of infection and transmission of various diseases. In this context, this study aims to analyse the risk factors associated with the occurrence of the Hepatitis B virus in health professionals. To this end, we used data provided by the Faculty of Medicine of the Eduardo Mondlane University, from the last study carried out on the Hepatitis B virus in health professionals, at Maputo Central Hospital in 2015, which is a cross-sectional cohort study. The study population was made up of health professionals from Maputo Central Hospital, and stratified sampling was used to determine the number of participants from each department that would be part of the sample, in which the number of health professionals analysed was 587. The statistical analysis was carried out using Thomas and Rao's F-test for independence and multiple logistic regression models. Using the logistic regression model, it was found that the sharing of brushes and razors and occupational accidents are the risk factors associated with the occurrence of hepatitis B among health professionals at Maputo Central Hospital.

Keywords: Hepatitis B, Generalised Linear Models, Logistic regression models.

## **Abreviaturas**

<b>DNA</b>	Ácido Desoxirribonucleico
<b>EPI</b>	Equipamentos de Proteção Individual
<b>FGEPA</b>	Factor Generalizado do Efeito do Plano Amostral
<b>HIV</b>	Vírus da Imunodeficiência Humana
<b>IC</b>	Intervalo de Confiança
<b>MISAU</b>	Ministério de Saúde
<b>OMS</b>	Organização Mundial de Saúde
<b>PS</b>	Profissional de Saúde
<b>SI</b>	Sistema de Informação
<b>TIC</b>	Tecnologias de Informação e Computação
<b>VHB</b>	Vírus de Hepatite B
<b>VHC</b>	Vírus de Hepatite C

## **Índice**

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.....	4
1.4. Relevância do Estudo.....	4
1.5. Estrutura do Trabalho.....	5
2 Revisão de Literatura.....	6
2.1 Conceitos Sobre Hepatite B.....	6
2.2 Hepatite B em Moçambique.....	7
2.3. Factores associados ao VHB.....	8
2.4. Prevenção e tratamento do Vírus de Hepatite B.....	8
2.5. Modelos Lineares Generalizados.....	9
2.5.1. Família Exponencial.....	9
2.5.2. Características do Modelo Linear Generalizado.....	10
2.5.3. Estimação de Parâmetros.....	12
2.5.4. Testes de Hipóteses.....	14
2.6. Análise de Regressão.....	14
2.6.1. Regressão Logística.....	15
2.6.2. Regressão Logística Simples.....	18
2.6.3. Regressão Logística Múltipla.....	18
2.6.4. Selecção das Variáveis e do Modelo.....	18
2.6.5. Análise de dados amostrais complexos.....	20

# Análise dos Factores de Risco Associados a Ocorrência de Hepatite B nos Profissionais de Saúde do Hospital Central de Maputo

---

3	Material e Métodos .....	23
3.1	Material .....	23
3.2	Métodos.....	24
3.2.1.	Medidas de associação.....	24
3.2.2.	Teste de Independência de Qui-Quadrado.....	24
3.2.3.	Teste Exacto de Fisher.....	25
3.2.4.	Estimação dos Parâmetros na Regressão Logística Binária.....	26
3.2.5.	Teste de Significância dos Parâmetros .....	26
3.2.6.	Qualidade de Ajuste na Regressão Logística Binária e Significância do Modelo .....	29
3.2.7.	Interpretação dos Parâmetros do Modelo .....	30
3.2.8.	Estratégias de Análise.....	33
3.2.9.	Especificação do Modelo.....	34
4	Resultados e Discussão .....	36
4.1	Análise descritiva dos dados e associação bivariada entre as variáveis em estudo. ....	36
4.2	Modelo de Regressão Logística Múltipla .....	39
4.2.1.	Modelo Logístico Final .....	40
4.2.2.	Interpretação das estimativas para a razão de chances do modelo de regressão logística final .....	41
4.3	Discussão dos Resultados.....	43
5	Conclusões e Recomendações .....	44
5.1.	Conclusões.....	44
5.2.	Recomendações .....	45
5.3.	Limitações .....	45
	Referências bibliográficas.....	46
	Anexos.....	50

## **Lista de Figuras**

Figura 4.1.1 Resultados laboratorias para o exame de presença do VHB.....	36
---	----

## **Lista de Tabelas**

<b>Tabela 2.5.1: Funções de ligação canônica mais utilizadas nos Modelos Lineares Generalizados</b>	<b>11</b>
<b>Tabela 3.1.1: Descrição das variáveis em estudo</b>	<b>24</b>
<b>Tabela 3.2.1: Distribuição de probabilidades</b>	<b>31</b>
<b>Tabela 3.2.2: Categoria das variáveis policotómicas</b>	<b>32</b>
<b>Tabela 4.1.1: Distribuição das frequências absolutas e relativas; e a relação entre os factores associados com a presença do VHB, com o resultado do exame do VHB nos profissionais de saúde</b>	<b>37</b>
<b>Tabela 4.2.1: Teste de razão de verosimilhança de Rao-Scott - Modelo Saturado</b>	<b>40</b>
<b>Tabela 4.2.2: Estimativa dos parâmetros e ajuste do Modelo Final</b>	<b>40</b>
<b>Tabela 4.2.3: Estimativa de razão de chances do Modelo Final</b>	<b>41</b>

## Introdução

Este capítulo apresenta a contextualização do tema do trabalho que foi desenvolvido, a problematização e a pergunta de pesquisa, os objectivos, a relevância deste trabalho, e também, como está estruturado o trabalho.

### 1.1. Contextualização

A exposição a agentes patogénicos constituiu, desde sempre, um sério risco para os profissionais de saúde. Este tema tem a sua importância, em virtude do melhor conhecimento das vias de transmissão dos agentes infecciosos, do desenvolvimento de medidas preventivas eficazes, de uma maior consciencialização/educação dos profissionais de saúde para os riscos de infecção no âmbito do exercício da sua actividade laboral e do desenvolvimento de medidas profiláticas eficazes que possibilitarão uma redução do número de acidentes ocupacionais e, conseqüentemente dos casos de doença infecciosa profissional (Beltrami *et al.*, 2000).

De acordo com Martins *et al.* (2013), anualmente, contabilizam-se mais de 500.000 acidentes ocupacionais potencialmente infectantes. Embora alguns outros microrganismos possam estar envolvidos, são sobretudo três os que maiores preocupações condicionam em termos de protecção/profilaxia da infecção dos referidos profissionais: o vírus da hepatite B (VHB), o vírus da hepatite C (VHC) e o vírus da imunodeficiência humana (HIV).

As hepatites virais são doenças provocadas por diferentes agentes etiológicos, com movimento primário pelo fígado, que apresentam características epidemiológicas, clínicas e laboratoriais distintas. A distribuição das hepatites virais é universal, sendo que a magnitude varia de região para região, de acordo com os diferentes agentes etiológicos (MISAU, 2007).

As hepatites virais são um sério problema de saúde pública afectando milhões de pessoas no mundo e causando cerca de um milhão de mortes por ano devido a doença hepática crónica (OMS Directoria Executiva, 2009). Estima-se que 57% dos casos de cirrose hepática e 78% dos tumores primários do fígado resultem da infecção pelos vírus das Hepatites B e C (Perz *et al.*, 2006).

De acordo com a Organização Mundial de Saúde (OMS, 2004), a infecção pelo Vírus da Hepatite B (VHB) está presente em mais de 2 biliões de pessoas a nível mundial, contando com aproximadamente 240 milhões de portadores crónicos e com 500.000 – 700.000 mortes por ano. As maiores prevalências da infecção pelo VHB são encontradas nos países em desenvolvimento, como

os países da África Subsaariana com 5 á 7%, pelas limitações presentes na área de saúde (OMS, 2015).

Nos países em desenvolvimento o risco destas infecções é elevado devido a alguns costumes e práticas inseguras, como a administração de injeções desnecessárias e precária esterilização do material hospitalar (Ozsoy *et al.*, 2003).

Trabalhar na área da saúde, principalmente em um ambiente hospitalar, é considerado insalubre, devido ao agrupamento de pacientes portadores de diversas enfermidades infectocontagiosas e o risco de acidentes existente em muitos procedimentos realizados (Nishide *et al.*, 2004).

Existem muito poucos estudos sobre as hepatites B no nosso país. O estudo feito pelo Cunha *et al.*, (2004) com uma amostra populacional de 1578 repositores de sangue, do banco de sangue do HCM, onde a prevalência do VHB encontrada foi de 10.6% para o sexo masculino e 4.5% para o sexo feminino.

Nas ciências sociais, normalmente estamos interessados em descobrir algo sobre um fenômeno que acreditamos que realmente exista (um fenômeno do mundo real). Esses fenômenos reais podem ser qualquer coisa desde o comportamento da incidência de doenças até comportamento sazonal de pandemias.

Neste contexto, emerge a possibilidade do uso de modelos estatísticos para o auxílio na identificação e avaliação dos factores de risco associados à ocorrência da Hepatite B nos profissionais de saúde, usando técnicas estatísticas multivariadas.

### **1.2. Definição do problema**

A hepatite B é uma enfermidade que ocorre em todo o mundo. Dados estatísticos apontam que 257 milhões de pessoas viviam com infecção crônica pelo VHB no ano 2015, e a prevalência da infecção global pelo VHB varia de 4.8 a 11.1% em trabalhadores da saúde, podendo ser, até três vezes maior que na população geral. O VHB causa epidemias principalmente no continente africano e no Pacífico Ocidental (OMS, 2015).

De acordo com Paiva *et al.* (2005) e Sanches *et al.* (2006), os resultados dos seus estudos realizados em Brasil, sugerem impacto positivo da vacinação. A eficiência da imunização e utilização dos equipamentos de segurança. Os mesmos autores demonstraram em seus estudos que o acidente ocupacional mais frequente foi o ferimento por agulha, mesmo assim devido às medidas de

profilaxia durante os acidentes e as vacinas, não foram encontrados resultados positivos para a Hepatite B.

A infecção pelo VHB ainda é uma preocupação em saúde pública, principalmente para os profissionais de saúde, por estarem expostos a sangue e outros fluidos ou tecidos orgânicos potencialmente contaminados, e apresentarem potencial risco de infecção e transmissão por várias doenças. O risco de contrair Hepatite B por profissional de saúde é quatro vezes maior, quando comparada a população geral ou aqueles que não trabalham em instituições de saúde (Feleke, 2016).

Considerando a dificuldade na obtenção de dados epidemiológicos, não existem muitos estudos neste campo. Segundo a Divisão de Assistência à Saúde do Estudante (2023), até à corrente data sabe-se que Moçambique tem uma prevalência das infecções pelos vírus de hepatites B estimada de 8.46%. A Hepatite B ainda é considerada uma doença de prevalência muito elevada neste país e a sua redução ao longo dos últimos 10 anos tem sido inferior ao que se esperava.

Apesar de ser possível minimizar o risco de infecção pelo VHB graças às medidas profiláticas implementadas, constitui uma realidade e condição diária na vida dos profissionais de saúde, por isso devem ser alertados sobre a sua existência e informados sobre os factores que podem influenciar na sua ocorrência e as formas de prevenção. O modelo de regressão logística que será encontrado neste trabalho, será de grande utilidade no sector de saúde pois permitirá descrever a relação entre ocorrência da Hepatite B e um conjunto de variáveis aleatória predictoras, para fins de previsão de ocorrência da Hepatite B nos profissionais de saúde quando expostos a determinadas realidades.

Daí que surge a seguinte pergunta de pesquisa:” **Quais são factores de risco associados a ocorrência da Hepatite B nos Profissionais de Saúde do Hospital Central de Maputo?**”.

### **1.3. Objectivos**

Nesta subsecção serão apresentados os objectivos que nortearão a elaboração do trabalho de fim de curso aqui proposto.

#### **1.3.1. Objectivo Geral**

Analisar os factores de risco associados a ocorrência da Hepatite B nos Profissionais de Saúde do Hospital Central de Maputo.

### 1.3.2. Objectivos Específicos

- Descrever o perfil sócio-demográfico dos profissionais de saúde do Hospital Central de Maputo.
- Verificar a associação bivariada entre a ocorrência da Hepatite B e as demais variáveis em estudo;
- Identificar os factores de risco que melhor explicam a ocorrência da Hepatite B nos profissionais de saúde do Hospital Central de Maputo.

### 1.4. Relevância do Estudo

O exercício da actividade profissional no ambiente hospitalar constitui factor de risco de infecção ocupacional pelo vírus da hepatite B, devido ao agrupamento de pacientes portadores de diversas enfermidades infectocontagiosas e o risco de acidentes existente em muitos procedimentos realizados (Silva e Alves, 2013). Segundo Pereira (2010), a prevenção das hepatites virais é uma prioridade em saúde pública especialmente para aqueles grupos que estão em maior risco de se tornarem portadores crónicos. Assim sendo, o estudo sobre a ocorrência do VHB é observado a sua pertinência uma vez que os profissionais de saúde fazem parte dos grupos de risco com maior exposição ao vírus da Hepatite B.

O uso de modelos de regressão na saúde tornou-se uma necessidade, numa área em que mais do que as técnicas de estimação e previsão utilizados, importa realçar os resultados que por eles são facultados. Dessa forma, conhecer os factores de riscos associados a ocorrência do VHB nos PS é de extrema importância para melhoria nos cuidados de saúde. No âmbito académico, este tema é explorado e traz a problemática da ocorrência do VHB e as suas complicações usando estatísticas multivariadas.

O conhecimento da prevalência da infecção pelo VHB nestes grupos, bem como os factores que influenciam na sua ocorrência, vai permitir que se criem estratégias de promoção de saúde e prevenção adequadas a realidade, criando melhores condições de segurança ocupacional e consciencializando a estes da existência destas infecções e das suas complicações a longo prazo.

Deste modo o presente estudo é elaborado para entender quais os factores associados a ocorrência do VHB nos profissionais de saúde.

### 1.5. Estrutura do Trabalho

O presente trabalho é constituído por cinco capítulos, descrito como:

- **Capítulo 1** - Apresenta a introdução e a contextualização do tema, a definição do problema de estudo, a questão de pesquisa, os objectivos do estudo e a relevância do mesmo.
- **Capítulo 2** - Apresenta alguns conceitos e fundamentação teórica relacionada aos factores que influenciam na ocorrência do VHB nos profissionais de saúde, e conceitos sobre as técnicas estatísticas aplicadas, como teste de Qui-quadrado de Pearson, teste de Fisher, Modelos Lineares Generalizados, especificamente o Modelo de Regressão Logística Múltipla.
- **Capítulo 3** - Material e Métodos: É apresentado o material usado para a realização do trabalho e os métodos usados para poder obter os resultados.
- **Capítulo 4** - Apresenta os resultados obtidos da análise estatística dos dados e discussão dos mesmos de forma comparativa com alguns estudos disponíveis.
- **Capítulo 5** - Apresenta as conclusões obtidas do trabalho e algumas recomendações.

Ainda também, consta na parte final as referências utilizadas no desenvolvimento do trabalho e materiais suplementares – Anexos.

## 2 Revisão de Literatura

Este capítulo é dedicado a apresentar os conceitos relacionados ao tema do trabalho, e também conceitos teóricos das principais técnicas estatísticas usadas para a concretização do trabalho.

### 2.1 Conceitos Sobre Hepatite B

A exposição a agentes patogénicos constituiu, desde sempre, um sério risco para os profissionais de saúde, contabilizando-se anualmente mais de 500.000 acidentes ocupacionais potencialmente infectantes (Martins *et al.*, 2013).

A infecção pelo VHB é um dos mais sérios problemas de saúde pública contemporânea mundial, dado o número elevado de pessoas infectadas por este agente etiológico e as consequências negativas para o indivíduo, família, colectividade e sistema público, e na área da saúde, a depender de condições organizacionais do ambiente de trabalho, recursos humanos e materiais, políticas institucionais, dentre outras. Os profissionais de enfermagem são os trabalhadores mais susceptíveis à exposição ao sangue e outros fluidos biológicos capazes de transmitir infecções, a exemplo do VHB (Morais *et al.*, 2016).

Segundo Silva (2019), as infecções pelos VHB, são um problema de saúde pública mundial crescente, e com grande relevância epidemiológica, em virtude das peculiaridades clínicas e alta morbidade e mortalidade dessas doenças.

De acordo com Filho (2016) Hepatite é caracterizada por um conjunto de lesões de necrose e inflamatórias que acometem o fígado de modo difuso, são causadas, principalmente, por vírus e ocasionalmente por drogas, distúrbios autoimunes e transtornos metabólicos. O causador da hepatite B é um vírus de ácido desoxirribonucleico (DNA), ou seja, após a infecção, o vírus se liga às células do fígado, onde o seu DNA fará o hepatócito produzir novos vírus (Silva *et al.*, 2013).

A Hepatite B é uma enfermidade que ocorre em todo o mundo. Estima-se que 257 milhões de pessoas viviam com infecção crônica pelo VHB no ano 2015. O VHB causa epidemias principalmente no continente africano e no Pacífico Ocidental. A Organização Mundial da Saúde (OMS) aponta que as hepatites virais consistem em um problema de saúde pública com necessidade de resposta urgente (OMS, 2015).

### ***Sintomas da Hepatite B***

A maioria dos casos da hepatite B não apresentam sintomas na fase aguda, mas os mais frequentes são cansaço, tontura, enjojo e/ou vômitos, febre, dor abdominal, pele e olhos amarelados, urina escura e fezes claras. Esses sinais costumam aparecer de um a seis meses após a infecção. A Hepatite B apresenta fases semelhantes que podem se desenvolver de duas formas, aguda e crônica (Vallezi e Júnior, 2011).

Os sinais e sintomas que aparecem durante a infecção aguda estão associados a idade do indivíduo infectado. Quando ocorre aparecimento de sinais e sintomas em pacientes sintomáticos estes são semelhantes aos sintomas de um resfriado, como febre, mal-estar, dores no corpo, fadiga e náuseas. Os recém-nascidos e as crianças na maioria dos casos são portadores assintomáticos (Pondé, 2016).

A fase aguda é quando a infecção tem curta duração. Os profissionais de saúde consideram a forma crônica quando a doença dura mais de seis meses. O risco de a doença tornar-se crônica depende da idade, neste caso as crianças e idosos são mais acometidos pelo avanço da doença quando ocorre a infecção e da resposta imunológica do infectado, onde existem indivíduos que são mais resistentes à infecção (Chavez *et al.*, 2003). Uma particularidade da Hepatite B é que a maioria dos pacientes elimina o vírus e evolui para a cura definitiva. Em menos de 5% dos casos, porém, o VHB persiste no organismo e a doença torna-se crônica sem apresentar sintomas que chamem a atenção durante muitos anos. E a doença pode evoluir dependendo de factores como a replicação do vírus; a resposta imunológica; o consumo de álcool e até infecção por outros vírus.

## **2.2 Hepatite B em Moçambique**

De acordo com Kiire (1996), considerando a dificuldade na obtenção de dados epidemiológicos, não existem muitos estudos neste campo. Até à corrente data sabe-se que Moçambique tem uma prevalência das infecções pelo vírus de hepatite B de 7.5%.

Existem muito poucos estudos sobre as hepatites B e C no nosso país. Um estudo feito, foi o estudo de Cunha *et al.*, (2004) com uma amostra populacional de 1578 repositores de sangue, do banco de sangue do HCM, onde a prevalência do VHB encontrada foi de 10.6% para o sexo masculino e 4.5% para o sexo feminino, e para o VHC as prevalências encontradas para o sexo masculino e feminino foram de 1.2% e 1.0% respectivamente.

### **2.3. Factores associados ao VHB**

O VHB pode ser transmitido por (i) solução de continuidade (pele e mucosas), (ii) relações sexuais desprotegidas, (iii) por via parenteral (compartilhamento de agulhas e seringas, tatuagens, piercings, procedimentos odontológicos ou cirúrgicos), (iv) da mãe portadora do VHB para o feto (transmissão vertical) e (v) no compartilhamento de instrumentos de manicure, escovas de dente, lâminas de barbear ou de depilar, canudo de cocaína e cachimbo (Dias *et al.*, 2014).

Segundo Dias *et al.* (2014), o VHB pode ser transmitido por meio de contacto com sangue ou outros fluidos corpóreos como a esperma, leite materno, secreção cervical e/ou vaginal, saliva de pessoas infectadas pelo vírus (fase aguda da doença). Esses fluidos podem ser transmitidos de pessoa para pessoa por meio de relações sexuais desprotegidas, uso de drogas injectáveis, transfusões sanguíneas, amamentação ou transmissão vertical e por meio de acidentes com material perfurantes e cortantes. Pode ocorrer também por meio da exposição percutânea sendo elas, piercing, tatuagens, e o compartilhamento de materiais contaminados, como navalhas, alicates de unha, lâminas de depilação, tesoura, entre outros.

### **2.4. Prevenção e tratamento do Vírus de Hepatite B**

Para prevenir ou eliminar os riscos de contaminação ocupacionais pelo VHB, os profissionais da saúde devem usar os equipamentos de protecção individual (EPI), entre eles luvas, máscaras, óculos e aventais durante a assistência directa ao paciente. Além destas medidas de prevenção, estes profissionais também devem ser imunizados com a vacina contra VHB, consequentemente o risco de infecção ocupacional pode ser significativamente reduzido (Filipe *et al.*, 2019).

De acordo com Morais *et al.* (2016), a imunização contra Hepatite B, por meio da administração de três doses sequenciais da vacina, é a medida de prevenção da infecção mais eficiente.

MISAU (2007) aponta que os pacientes devem ser orientados em relação ao não compartilhamento de objectos de uso pessoal como lâmina de barbear, escova de dente e alicates de unha. Deve-se utilizar preservativo nas relações sexuais e não compartilhar utensílios e materiais para colocação de piercing e tatuagem. Pessoas com passado de hepatite viral não são candidatas para doação de sangue. Os usuários de drogas injectáveis e inaláveis devem também ser orientados em relação ao não compartilhamento de agulhas, seringas, canudos e cachimbos para uso de drogas. Quanto aos filhos de mães com o VHB, para prevenir contra o vírus são necessário 3 doses de vacinas contra a hepatite, sendo a primeira dose da vacina nas primeiras 12 horas de vida, e a segunda e terceira

doses da vacina devem seguir o calendário vacinal normal, isto é, aos trinta dias e aos seis meses de idade, respectivamente. A amamentação não traz riscos adicionais para os seus recém-nascidos, desde que tenham recebido a primeira dose da vacina e imunoglobulina nas primeiras 12 horas de vida.

## **2.5. Modelos Lineares Generalizados**

Em estudos estatísticos, quer sejam de natureza experimental ou observacional, o objectivo principal é de estudar a relação entre variáveis, ou mais particularmente, analisar a influência que uma ou mais variáveis (explicativas), medidas em indivíduos ou objectos, têm sobre uma variável de interesse a que damos o nome de variável resposta.

O modelo linear é talvez o mais utilizado para modelar esta relação. Este modelo assume, entre outras, que o valor esperado da variável resposta é uma combinação linear das variáveis explicativas e que a variável resposta segue a distribuição Normal.

Esta teoria da modelação estatística é limitada, pois não pode ser utilizada se a distribuição da variável resposta é diferente da Normal. Para a resolução deste problema, o modelo linear generalizado foi introduzido em 1972, por Nelder e Wedderburn. São exemplos de modelos lineares generalizados, Regressão Logística, Poisson e Binomial Negativa.

O modelo linear generalizado é uma extensão do modelo linear, onde a distribuição da variável resposta não tem de ser Normal, mas sim uma outra distribuição da família exponencial (Binomial, Bernoulli, Poisson entre outras) e a função que relaciona o valor esperado e o vector de variáveis explicativas pode ser qualquer função diferenciável (Alvarenga, 2015).

### **2.5.1. Família Exponencial**

De acordo com Turkman e Silva (2000), uma variável aleatória  $Y$  tem distribuição pertencente à família exponencial de dispersão (ou simplesmente família exponencial) se a sua função densidade de probabilidade (f.d.p.) ou função massa de probabilidade (f.m.p.) se puder escrever na forma:

$$f(y|\theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\} \quad (2.1)$$

Onde  $y$  é a variável de interesse,  $\theta$  é o parâmetro de localização,  $\phi$  é o parâmetro de dispersão associado à variância,  $a(\cdot)$ ,  $b(\cdot)$  e  $c(\cdot)$  são funções reais conhecidas ou também denominadas funções específicas.

Simplesmente calculando-se as derivadas de primeira e segunda ordem da função  $b(\theta)$ , pode-se obter a média e a variância da variável  $y$ , como demonstrado na equação 2.2 e 2.3

$$E(y) = \frac{\partial b(\theta)}{\partial \theta} = \mu \quad (2.2)$$

$$var(y) = \frac{\partial^2 b(\theta)}{\partial \theta^2} a(\phi) \quad (2.3)$$

Como se pode observar, a variância de  $y$  é, na verdade, um produto que depende de sua média e de  $a(\phi)$ . A parcela correspondente à segunda derivada de  $b(\theta)$  é conhecida como função da variância,  $V(\mu)$ . O que faz sentido, dado que em algumas distribuições a variância dos dados muda conforme sua média. É interessante comparar, por exemplo, as distribuições normais  $(\mu, \sigma)$  e Poisson  $(\lambda)$ . A variância da distribuição Poisson é dada por  $\lambda-1$ , ou seja, depende diretamente de sua média  $\lambda$ , enquanto a variância da normal, dada por  $\sigma$ , não depende de sua média, mas sim de uma função de variância, que neste caso é igual a 1, como pode ser deduzido a partir da equação 2.4

$$V(\mu) = \frac{var(y)}{a(\phi)} = \frac{\partial^2 b(\theta)}{\partial \theta^2} = \frac{\partial \mu}{\partial \theta} \quad (2.4)$$

### 2.5.2. Características do Modelo Linear Generalizado

De acordo com Alvarenga (2015), os modelos lineares generalizados são caracterizados pela seguinte forma: Componente aleatória, Componente sistemática e Função de ligação.

1. **Componente aleatória:** Dado o vector de covariáveis  $x_i$  as variáveis  $Y_i$  são (condicionalmente) independentes com distribuição pertencente à família exponencial na forma canônica, com o seu valor médio dado por  $E(Y_i/x_i) = \mu_i = b'\theta_i$  para  $i = 1, \dots, n$  e, possivelmente, um parâmetro de dispersão  $\phi$  não dependente de  $i$ .
  
2. **Componente Sistemática:** Define-se o preditor linear  $\eta_i$  como combinação linear das variáveis explicativas, ou seja, a componente sistemática é formada pelas variáveis explicativas, que entram sob forma de uma soma linear de seus efeitos, dando origem a um vector de preditores lineares, representado pela expressão dado por:

$$\mu_i = h(\eta_i) = h(z_i^T \beta), \text{ sendo que, } \eta_i = g(\mu) \quad (2.5)$$

**Onde:**

$h$  é uma função monótona e diferenciável;

$g = h^{-1}$  é a função de ligação;

$\beta$  é um vector de parâmetros de dimensão  $p$ ;

$Z_i$  é um vector de especificação de dimensão  $p$ , função do vector de covariáveis  $x_i$

**3. Função de ligação:** Representa uma função que relaciona a componente aleatório ao componente sistemático, relacionando a média ao predictor linear, isto é,

$$\eta_i = g(.) \quad (2.6)$$

Onde  $g(.)$  é função real, monótona e diferencial.

Turkman e Silva (2000), apontam que os casos em que o predictor linear coincide com o parâmetro canônico  $\theta_i$ , ou seja,  $\theta_i = \eta_i$ , implica que  $\theta_i = z_i^T \beta$ , a função de ligação chama-se função de ligação canônica.

As funções de ligação canônica mais utilizadas, nos Modelos Lineares Generalizados, são descritas na Tabela 2.5.1.

Tabela 2.5.1: **Funções de Ligação Canônica mais utilizadas nos Modelos Lineares Generalizados**

Distribuição	Função de Ligação Canônica
Normal	Identidade: $\eta = \mu$
Binomial	Logística: $\eta = \log\left(\frac{\pi}{1-\pi}\right) = \log\left(\frac{\mu}{m-\mu}\right)$
Poisson	Logarítmica : $h = \log(m)$
Gama	Recíproca: $\eta = \frac{1}{\mu}$
Normal Inversa	Recíproca do quadrado: $\eta = \frac{1}{\mu^2}$

Fonte: Madson e Thyregod (2010).

### 2.5.3. Estimação de Parâmetros

Num modelo linear generalizado o parâmetro de interesse é  $\beta$ , sendo a sua estimação feita pelo método da máxima verossimilhança. O parâmetro de dispersão  $\phi$ , quando existe, é considerado parâmetro perturbador e é estimado pelo método dos momentos. A verossimilhança também é considerada como base fundamental no processo inferencial, no caso dos testes de hipóteses sobre os coeficientes estimados e da qualidade do ajustamento.

Sabendo que  $\theta_i$  é a função de  $\mu_i$ , sendo  $b'(\theta_i) = \mu_i = h(\eta_i)$  onde  $h(\cdot)$  é a função inversa da função de ligação  $g(\cdot)$  e que  $var(Y_i) = \frac{\phi}{w_i} b''(\theta_i)$ .

A função de verossimilhança, como função de  $\beta$ , é:

$$\begin{aligned} l(\beta) &= \prod_{i=1}^n f(y_i|\theta_i, \phi) \\ &= \prod_{i=1}^n \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\} \\ &= \exp \left\{ \sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + \sum_{i=1}^n c(y_i, \phi) \right\} \end{aligned} \quad (2.7)$$

O logaritmo da verossimilhança apresentando, chamado também por log-verossimilhança é dado por

$$\begin{aligned} \ln(l(\beta)) &= l(\beta) \\ &= \sum_{i=1}^n \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\} \\ &= \sum_{i=1}^n l_i(\beta) \end{aligned} \quad (2.8)$$

Onde  $l_i(\beta) = \frac{\omega_i(y_i \theta_i - b(\theta_i))}{\phi} + c(y_i, \phi, \omega_i)$

é a contribuição de cada observação  $y_i$  para a verossimilhança. Os estimadores de máxima verossimilhança para  $\beta$  são obtidos como solução do sistema de equações de verossimilhança.

As equações de verosimilhança são dadas por:

$$\frac{\partial l\beta}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l_i\beta}{\partial \beta_j} = 0, j = 1, \dots, p \quad (2.9)$$

Para se obter estas equações escreve-se da seguinte forma (Turkman e Silva, 2000),

$$\frac{\partial l_i\beta}{\partial \beta_j} = \frac{\partial l_i(\theta_i)}{\partial \theta_i} \frac{\partial \theta_i(\mu_i)}{\partial \mu_i} \frac{\partial \mu_i(\eta_i)}{\partial \eta_i} \frac{\partial \eta_i(\beta)}{\partial \beta_i} \quad (2.10)$$

Onde

$$\begin{aligned} \frac{\partial l_i(\theta_i)}{\partial \theta_i} &= \frac{w_i(y_i - b'(\theta_i))}{\phi} = \frac{w_i(y_i - \mu_i)}{\phi} \\ \frac{\partial \mu_i}{\partial \theta_i} &= b''(\theta_i) = \frac{w_i \text{var}(Y_i)}{\phi} \\ \frac{\partial \eta_i(\beta)}{\partial \beta_i} &= x_{ij} \end{aligned} \quad (2.11)$$

Assim

$$\frac{\partial \eta_i(\beta)}{\partial \beta_i} = \frac{w_i(y_i - \mu_i)}{\phi} \frac{\phi}{w_i \text{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} \quad (2.12)$$

e as equações de verosimilhança para  $\beta$  são

$$\sum_{i=1}^n \frac{(y_i - \mu_i)x_{ij}}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0, j = 1, \dots, p \quad (2.13)$$

A primeira derivada da função log-verosimilhança em ordem a  $\beta$  é dominada por função Score, definida como vector  $p$ -dimensional, e é dada por:

$$s(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^n s_i(\beta) \quad (2.14)$$

Onde  $s_i(\beta)$  é o vector de componentes  $\frac{\partial l_i(\beta)}{\partial \beta}$  obtido em 2.6.

A matriz de covariância da função score  $I(\beta)$ , é designada por matriz de informação de Fisher e é dada por:

$$I(\beta) = E \left[ - \frac{\partial s(\beta)}{\partial \beta} \right] \quad (2.15)$$

Para obter a matriz de informação de Fisher tem-se:

$$\begin{aligned}
 I(\beta) &= -E\left(\frac{\partial^2 l_i}{\partial \beta_j \partial \beta_k}\right) = E\left(\frac{\partial l_i}{\partial \beta_j} \frac{\partial l_i}{\partial \beta_k}\right) \\
 &= E\left[\left(\frac{(Y_i - \mu_i)x_{ij}}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i}\right) \left(\frac{(Y_i - \mu_i)x_{ik}}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i}\right)\right] \\
 &= E\left[\frac{(Y_i - \mu_i)^2 x_{ij} x_{ik}}{(\text{Var}(Y_i))^2} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\right] \\
 &= \frac{x_{ij} x_{ik}}{\text{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2
 \end{aligned} \tag{2.16}$$

e, portanto o elemento genérico de ordem  $(j, k)$  da matriz de informação de Fisher é

$$I(\beta) = -\sum_{i=1}^n E\left(\frac{\partial^2 l_i}{\partial \beta_j \partial \beta_k}\right) = \sum_{i=1}^n \frac{x_{ij} x_{ik}}{\text{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \tag{2.17}$$

Na forma matricial tem-se:

$$I(\beta) = X^T W X \tag{2.18}$$

onde  $W$  é a matriz diagonal de ordem  $n$  cujo  $i$ -ésimo elemento é

$$w_i = \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \times \frac{1}{\text{Var}(Y_i)} \tag{2.19}$$

#### 2.5.4. Testes de Hipóteses

Depois de obter as estimativas para os coeficientes da regressão é necessário avaliar a sua significância, isto é, determinar se as variáveis independentes introduzidas no modelo estão significativamente associadas à variável dependente (Hosmer e Lemeshow, 2013).

Segundo Williams e Iyer (1958), os testes de hipóteses são utilizados para determinar se os resultados de um estudo científico podem levar a rejeição da hipótese nula  $H_0$ , a um nível de significância pré-estabelecido. O estudo da teoria das probabilidades e a determinação da estatística de teste correcta são fundamentais para a coerência de um teste de hipótese.

#### 2.6. Análise de Regressão

Sempre é interessante conhecer os efeitos que algumas variáveis exercem, ou que parecem exercer, sobre outras. Mesmo que não exista relação causal entre as variáveis podemos relacioná-las por

meio de uma expressão matemática, que pode ser útil para se estimar o valor de uma das variáveis quando conhecemos os valores das outras sob determinadas condições.

Segundo Gujarati e Porter (2011), o termo regressão foi criado por Francis Galton. Em um artigo famoso, Galton verificou que, embora existisse uma tendência de que pais altos tivessem filhos altos e pais baixos tivessem filhos baixos, a estatura média das crianças nascidas de pais com uma dada altura tendia a mover-se ou "regredir" à altura média da população como um todo.

Gujarati e Porter (2011) apontam ainda que a análise de regressão diz respeito ao estudo de técnicas estatísticas de dependência de uma variável dependente, em relação a uma ou mais variáveis explicativas, visando estimar e/ou prever o valor de uma ou mais variáveis dependentes em termos dos valores conhecidos ou fixados das variáveis independentes. Este relacionamento é representado por um modelo matemático que associa a variável dependente com as variáveis independentes.

### 2.6.1. Regressão Logística

Segundo Mesquita (2014), a técnica de regressão logística, desenvolvida no século XIX, obteve maior visibilidade após 1950 ficando então, mais conhecida. Conseguiu um grande avanço a partir dos trabalhos de Cox & Snell (1989) e Hosmer e Lemeshow (2000). Caracteriza-se por descrever a relação entre uma variável dependente qualitativa binária, associada a um conjunto de variáveis independentes qualitativas ou métricas.

De acordo com Fávero *et al.* (2009), a regressão logística é uma técnica estatística utilizada para descrever o comportamento entre uma variável dependente binária e um conjunto de variáveis independentes métricas ou não métricas. Ou seja, destina-se a investigar o efeito das variáveis pelas quais os objectos ou indivíduos estão expostos sobre a probabilidade de ocorrência de determinado evento de interesse.

O modelo de regressão Logística é um dos casos mais importante do modelo linear generalizado, quando se pretende modelar uma variável resposta categórica com dois valores possíveis, dado um conjunto de variáveis explicativas (Alvarenga, 2015).

Segundo Mestre (2009), a regressão logística pode ser considerada como sendo uma extensão da regressão linear, pois assim como na regressão linear, ela estuda relações existentes entre variáveis, identificando as variáveis que podem influenciar de alguma forma em uma variável resposta, sendo que na regressão logística essa variável resposta deve ser categórica dicotômica. Enquanto a regressão linear dá uma resposta métrica, a regressão logística fornece a resposta em termos de

chance de ocorrência do fenómeno em estudo. Hosmer & Lemeshow (2013), concordando com este autor acrescenta ainda que a principal diferença entre modelos de regressão logística e regressão linear é a distribuição da variável resposta, a regressão logística substitui a distribuição normal da variável resposta pela distribuição de Bernoulli, cuja função de probabilidade é dada por:

$$f(y_i|p_i) = p_i^{y_i}(1 - p_i)^{1-y_i}, y_i = 0,1 \quad i = 1, \dots, n \quad (2.20)$$

Onde  $p_i$  é o parâmetro desconhecido, que significa a probabilidade do sucesso  $P(y_i = 1)$  e a probabilidade do insucesso  $P(y_i = 0) = 1 - p_i$ .

Com base nos conceitos acima citados, o objectivo fundamental da regressão logística é prever uma variável dependente dicotómica a partir de um conjunto de variáveis independentes, mediante um modelo que apresenta um bom ajuste.

#### 2.6.1.2. *Pressupostos da regressão logística*

De acordo com Fávero *et al.* (2009) é recomendável atender certas condições para uma aplicação apropriada de muitas técnicas estatísticas multivariadas. a regressão logística assume os seguintes pressupostos: (1) a relação linear entre o vector das variáveis explicativas  $X$  e a variável dependente  $Y$ ; (2) o valor esperado dos resíduos é igual a zero; (3) ausência de heterocedasticidade; e (4) ausência de multicolinearidade.

Diferentemente da análise de regressão múltipla e da análise discriminante, a regressão logística não pressupõe normalidade dos resíduos. Isso acaba representando, quando da aplicação das outras técnicas, uma grande limitação, uma vez que, na presença de várias variáveis dicotómicas, este pressuposto acaba sendo violado. Além disso, a regressão logística também não pressupõe homogeneidade de variância e a redução de número de pressupostos torna-se preferível em muitas situações práticas.

Segundo Sharma (1996), quando se tem em um mesmo modelo, variáveis explicativas com escalas de mensuração qualitativa e quantitativa, o pressuposto da normalidade multivariada não será atendido na análise discriminante. Nestes casos, o pesquisador pode optar pelo uso da regressão logística, uma vez que esta não faz nenhuma consideração sobre a distribuição das variáveis explicativas.

#### 2.6.2.2. *Modelos de regressão logística*

De acordo Hair Jr. *et al.*, (2006) O objectivo do modelo de regressão logística é encontrar um modelo adequado e parcimonioso que permite descrever a relação entre uma variável aleatória

binária  $Y$  e um conjunto de variáveis aleatórias predictoras  $X_1, X_2, X_3, \dots, X_p$ . A ideia não é obter um modelo que determine de forma eficaz e total os dois resultados possíveis de  $Y$ , mas sim que indique com precisão a probabilidade de ocorrência do sucesso ou do insucesso.

Considera-se o seguinte modelo:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (2.21)$$

Onde  $Y_i$  assume os valores de 0 (insucesso) e 1 (sucesso). Desta relação pode-se escrever:

$$E(Y_i = 1|X_i) = \beta_0 + \beta_1 x_i \quad (2.22)$$

Onde  $E(Y_i|X_i)$  representa a média condicional, e pode-se interpretar como valor esperado de  $x$ , dado a distribuição logística usada, a fim de simplificar a notação pode ser usada  $E(Y_i|X_i) = \pi_i$ , assume-se que:

$$\pi_i = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (2.23)$$

Onde o modelo de regressão logística é dado em 2.24, dessa forma, a relação não é linear nos parâmetros, daí que a interpretação dos parâmetros de um modelo de regressão logística é obtida comparando a probabilidade de sucesso com a probabilidade de fracasso, usando a função odds ratio - OR (razão de chances), que é estimada da seguinte fórmula:

$$g(x) = \frac{\pi(x)}{[1 - \pi(x)]} = \frac{\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}}{1 - \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}} = \frac{\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}}{\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}}} = e^{\beta_0 + \beta_1 x_i} \quad (2.24)$$

Aplicando o logaritmo natural na expressão em 2.25, obtém-se a expressão em 2.26.

$$\ln \left[ \frac{\pi(x)}{[1 - \pi(x)]} \right] = \beta_0 + \beta_1 x_i \quad (2.25)$$

Assim sendo, a razão de chances de ocorrência do sucesso quando  $x = 1$  e quando  $x = 0$  é dado da forma seguinte:

$$OR = \frac{\frac{\pi(1)}{1 - \pi(1)}}{\frac{\pi(0)}{1 - \pi(0)}} \quad (2.26)$$

### 2.6.2. Regressão Logística Simples

Segundo Gonzalez (2018), a regressão logística simples, representa os casos de regressão logística em que a variável dependente  $Y$  é dicotômica ou binária, ou seja, tem duas categorias e tem apenas uma variável independente.

A média condicional (probabilidade condicional) para modelo de regressão logística simples é dada por:

$$E [Y_i] = \pi_i = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (2.27)$$

A equação acima é chamada de equação de regressão estimada, e é, essencialmente, a função que representa o objectivo do modelo de regressão logística, pois  $\pi_i$  é a probabilidade estimada para quaisquer valores de coeficientes e variáveis que são colocadas na equação.

### 2.6.3. Regressão Logística Múltipla

Segundo Hosmer e Lemeshow (2000), a regressão logística múltipla representa o contexto de regressão logística, em que a variável dependente  $Y$  é binária ou dicotômica, ou seja, tem duas categorias e que há mais de uma variável independente. Pode-se considerar a regressão logística múltipla como uma generalização da regressão logística binária.

Realizando uma transformação, o modelo logit é dado por:

$$g(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (2.28)$$

O modelo logístico é dado pela expressão em 2.30:

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad (2.29)$$

Onde  $g(x)$  apresenta parâmetros lineares, que podem ser contínuos e podem variar  $-\infty$  a  $+\infty$  e  $\beta$  determina a taxa de crescimento ou decréscimo da função.

### 2.6.4. Selecção das Variáveis e do Modelo

De acordo com Braga (1994), o modelo de regressão deve ser parcimonioso, isto é, deve envolver o mínimo de parâmetros possíveis a serem estimados e que explique bem o comportamento da variável resposta. Geralmente deve-se escolher o modelo mais simples. O problema da selecção do modelo corresponde à procura do “melhor modelo”, no sentido de ser um modelo que atinge um

bom equilíbrio entre os três factores “bom ajustamento”, “parcimónia” e “interpretação”. A inclusão ou a exclusão de uma variável no modelo, pode variar conforme o problema a considerar ou até mesmo a área científica em análise.

Segundo Alvarenga (2015), na prática há geralmente um elevado número de variáveis que podem ser potencialmente importantes para explicar a variabilidade da variável resposta. Isto implica a existência de vários modelos com combinações diferentes das variáveis explicativas para explicar o fenómeno em causa, o que torna o processo da selecção mais difícil e mais moroso. Para facilitar o processo da selecção vários investigadores utilizaram o método de selecção stepwise.

O método stepwise é um procedimento automático de selecção das variáveis em direcção backward, forward e both.

- A **direção forward** inicia-se a partir de um modelo nulo e adiciona uma variável de cada vez, que podem ser significativas para explicar a variabilidade da variável resposta. O modelo nulo é um modelo simples com apenas um parâmetro que representa o mesmo valor médio  $\mu$  para todas as observações  $Y_i$ . Uma característica primordial desse método de selecção é que, uma vez que a variável foi seleccionada e incluída por ser significativa, ela não deve mais ser excluída do modelo.
- O caso da **direção backward** é ao contrário da direcção forward, inicia-se a partir de um modelo completo e verifica a cada passo se uma variável pode ser ou não eliminada do modelo. O modelo completo ou saturado é o maior modelo que temos a possibilidade de considerar. Dada uma amostra com  $n$  observações, o número máximo de parâmetros para esse modelo é igual a  $n$ , isto é, um parâmetro para cada observação.
- O método **both stepwise** é uma combinação dos dois métodos (forward e backward). A fase de incluir ou excluir a variável do modelo é a fase de avaliação da significância das variáveis ou comparação dos modelos. Para isso utilizam-se medidas estatísticas adequadas para a sua avaliação.

De acordo com Turkman e Silva (2000), dado que no processo de selecção do modelo, há uma série de modelos em consideração, convém descrever vários que são comumente referidos durante o processo.

### Modelo nulo

O modelo mais simples que se pode imaginar é o com um único parâmetro. Corresponde a assumir que todas as variáveis  $Y_i$  têm o mesmo valor médio  $\mu$ . É um modelo de interpretação sem dúvida simples, mas que raramente captura a estrutura inerente aos dados.

### **Modelo saturado**

O modelo saturado é o que contém o maior número de parâmetros, e, portanto, o mais complexo que se está preparado a considerar.

### **Modelo minimal**

Contrariamente ao modelo saturado, o modelo minimal é o mais simples, com o menor número de parâmetros, que ainda se ajusta adequadamente aos dados. Este modelo embora adaptando-se aos dados e podendo até ser adequado para réplicas do estudo, pode esconder características ainda importantes dos dados.

### **Modelo ajustado**

Em geral trabalha-se com modelos encaixados, isto é, passagem do modelo saturado para o minimal por exclusão de termos de desvio. O modelo corrente, é qualquer modelo com  $q$  parâmetros linearmente independentes situado entre o modelo saturado e o minimal, e que está sujeito à investigação.

#### **2.6.5. Análise de dados amostrais complexos**

Diversos métodos estatísticos são utilizados para analisar dados provenientes de pesquisas amostrais. Tais métodos frequentemente pressupõem que os dados são provenientes de uma amostra aleatória simples, que é uma técnica de amostragem probabilística em que se assume que as observações são independentes e cada elemento na amostra tem a mesma probabilidade de ser seleccionado (Casela e Berger, 2001).

Na prática, raramente a selecção de amostras para pesquisas é realizada por amostragem aleatória simples. Com frequência são utilizados esquemas mais complexos de amostragem, tais como amostragem estratificada e amostragem por conglomerados, por exemplo (Pereira, 2017).

Os motivos para se realizar pesquisas por meio de planos amostrais complexos são, entre outros, limitações financeiras dos órgãos financiadores, facilitação logística da aplicação das provas e a possibilidade da produção de estimativas com níveis de precisão bastante satisfatórios. Além disso, a limitação dos recursos disponíveis para a realização das pesquisas poderia aumentar a

possibilidade de ocorrência de erros não amostrais caso todas as pesquisas fossem realizadas de forma censitária (Vicente *et al.*, 2001).

Segundo Heeringa *et al.* (2010), tem crescido o uso de dados oriundos de amostras complexas, que são definidas como sendo amostras envolvendo múltiplos estágios de seleção, sorteio sistemático, amostragem com probabilidade proporcional a uma medida de tamanho, estratificação, etc. Logo, é necessário poder incorporar estas características do plano amostral no uso descritivo ou analítico dos dados, informando os pesos das unidades da amostra e as informações estruturais da pesquisa, identificando o estrato e pelo menos as unidades primárias de amostragem.

De acordo com Morais (2008), as estimativas pontuais são influenciadas por pesos distintos associados às unidades da amostra, enquanto as estimativas de variância são influenciadas pela estratificação, conglomeração e pesos amostrais.

Segundo Pessoa e Silva (1998), as dificuldades na aplicação do método de máxima pseudoverosimilhança é a necessidade de conhecimento das características do plano amostral e o fato de que suas propriedades não são conhecidas para pequenas amostras, o que impossibilita algumas análises estatísticas. Para situações que os dados são provenientes de amostras complexas, o estimador de máxima pseudo-verosimilhança é obtido a partir de correções que levam em consideração os pesos amostrais. Para o estimador de máxima pseudo-verosimilhança de  $\beta$  é necessária uma ponderação. Assim tem-se:

$$\sum_{i \in s} w_i \left[ \sum_{i=1}^n \left[ y_i - \frac{1}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}} e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p} \right] \right] \quad (2.30)$$

$$\sum_{i \in s} w_i \left[ \sum_{i=1}^n \left[ y_i x_i - \frac{x_i}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}} e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p} \right] \right] \quad (2.31)$$

onde  $w_i$  é o peso da  $i$ -ésima observação amostral.

O tipo do estimador (linear ou não linear) e as características dos planos amostrais interferem diretamente no processo de estimação da variância. Entre os métodos mais usuais para a estimação da variância convém citar o método do Conglomerado Primário, os métodos de replicação (reamostragem): Jackknife e Bootstrap e Linearização em série de Taylor.

Quando o estimador é não linear, geralmente não existe uma expressão matemática exacta para sua variância. No entanto, se este estimador poder ser escrito como uma função não linear de estimadores de totais populacionais torna-se possível a construção da expressão analítica da variância aproximada de primeira ordem, através da Linearização em série de Taylor

## **3 Material e Métodos**

Neste capítulo, são apresentados de forma detalhada os materiais usados e métodos aplicados para a produção dos resultados que são apresentados na secção 4 desta pesquisa. Na secção 3.1, referente aos materiais, é apresentada a fonte de dados e a descrição das variáveis (tabela 3.1.1). Na secção 3.2, referente aos métodos, é apresentada a classificação metodológica da pesquisa e os testes usados para a realização do trabalho.

### **3.1 Material**

Para a realização deste trabalho, foram utilizados dados fornecidos pela Faculdade de Medicina da Universidade Eduardo Mondlane, obtidos através do último inquérito do estudo sobre o vírus de Hepatite B nos profissionais de saúde, realizado no Hospital Central de Maputo, no ano 2015, cujo estudo é de coorte transversal e representativo.

No que se refere no planeamento amostral deste estudo transversal, é aplicado um plano amostral complexo que envolve estratificação, com probabilidades desiguais de selecção, o que resultam em pesos amostrais distintos para as unidades amostrais. A população do estudo foi composta por profissionais de saúde do HCM, e a concepção da amostra foi composta por duas etapas. Na primeira, a amostra foi estratificada em departamentos, de modo a obter o número de participantes de cada departamento que fez parte da amostra. Na segunda etapa usou-se a amostragem sistemática para a selecção dos elementos que fizeram parte do estudo dentro dos estratos. A recolha de dados ocorreu em 2015.

O tamanho da amostra para o presente estudo é constituído por 66 profissionais do departamento de administração, 133 do departamento de cirurgia, 27 de clínica especial, 3 de farmácia, 20 de laboratório, 148 de medicina, 30 do departamento de obstetrícia, 45 de pediatria e 115 dos serviços de urgência, totalizando assim 587 profissionais de saúde.

Tabela 3.1.1: Descrição das variáveis em estudo

Variável	Descrição da Variável	Categoria
HEP00	<b>Hepatite B</b>	<b>0-Não, 1-Sim</b>
HEP01	Idade dos Participantes	
HEP02	Sexo dos Participantes	0- Masculino, 1-Feminino
HEP03	Estado Civil	0-Nenhum, 1-Casado, 2-Divorciado, 3-
HEP04	Consumo de Álcool	0-Não consome, 1-Consome
HEP05	Consumo de Drogas como Cannabis e Cigarros	0-Não consome, 1-Consome
HEP06	Consumo de Drogas por Via Endovenosa	0-Não consome, 1-Consome
HEP07	Uso Habitual de Protecção no Acto Sexual	0-Não usa, 1-Usa
HEP08	Vários Episódios de Sexo Ocasional	0-Não, 1-Sim
HEP09	Prática de Sexo em Grupo	0-Não pratica, 1-Pratica
HEP10	Antecedentes de ITS	0-Não tem antecedentes, 1-Tem
HEP11	Troca de Sexo por Bens e Favores	0-Nunca trocou, 1-Já trocou
HEP12	Transusão Sanguínea ou Derivados	0-Nunca recebeu, 1-Já recebeu
HEP13	Antecedentes de Cirurgias	0-Não tem antecedentes, 1-Tem
HEP14	Ter Piercing ou Tatuagens	0-Não tem, 1-Tem
HEP15	Partilhas de Escovas e Máquinas de Barbear	0-Não partilha, 1-Partilha
HEP16	Acidente Ocupacional	0-Nunca teve, 1-Ja teve
HEP17	Vacina Contra Hepatite B	0-Não, 1-Sim 1-Administracao, 2-Cirurgia, 3-Clinica
HEP18	Departamento	Especial, 4-Farmacia, 5-Laboratorio, 6-Medicina,7-Obstetricia, 8-Pediatria, SUR
HEP20	Hábito de Manter Relações Sexuais Desprotegidas	0-Não tem hábito, 1-Tem hábito

## 3.2 Métodos

### 3.2.1. Medidas de associação

Medidas de associação quantificam a relação entre uma dada exposição e uma consequência (Rodrigues, 2015).

### 3.2.2. Teste de Independência de Qui-Quadrado

De acordo com Rodrigues (2015), o teste  $\chi^2$  (qui-quadrado) proposto pelo estatístico Karl Person em 1900, serve para comprovar se existem diferenças significativas entre duas distribuições quaisquer. É um dos principais testes para associação, ele calcula o total de desvios entre o número de ocorrências observadas e o de esperadas, e observa sua probabilidade de ocorrência segundo uma distribuição  $\chi^2$  com número de graus de liberdade adquiridos da estrutura da tabela de contingência

da forma:  $gl = (C - 1)(L - 1)$ . Dessa maneira ele é adequado para testar a hipótese nula se há não relação entre as categorias. Então, as hipóteses são dadas por:

$H_0$ : Não existe associação entre as categorias

$H_1$ : Existe alguma associação entre as categorias.

A estatística do teste é dada por:

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

Onde,  $L$  é o número de linhas,  $C$  é o número de colunas,  $O_{ij}$  é a frequência observada na linha  $i$  da coluna  $j$  e  $E_{ij} = \frac{L_i \cdot C_j}{N}$  é a frequência esperada na linhas  $i$  das colunas  $j$ .

### 3.2.3. Teste Exacto de Fisher

O teste de Fisher faz uso de tabelas de contingências 2x2, para comparar 2 grupos. É indicado quando o tamanho das duas amostras é pequeno e consiste em determinar a probabilidade exacta de ocorrência de uma frequência observada, ou de valores mais extremos (Guimarães, 2011).

Os critérios para o uso do teste é que as amostras sejam aleatórias e independentes, constituam duas classes mutuamente exclusivas e variáveis qualitativas ao menos. As frequências observadas têm distribuição hipergeométrica. A probabilidade de ocorrência será dada por:

$$P(x) = \frac{\prod_{i=1}^l L_i! \prod_{j=1}^c C_j!}{N! \prod_{i=1}^l \prod_{j=1}^c O_{ij}!} \sim \text{Hipergeométrica}(L_i, C_j, N) \quad (3.2)$$

Se o resultado for inferior ao nível de significância  $\alpha$ (alfa), devemos rejeitar a hipótese nula de independência.

Rao e Scott (1984) propuseram um ajuste do teste Qui-Quadrado, de modo a incorporar as características do plano amostral, que consiste em dividir o valor da sua estatística por um factor generalizado do efeito do plano amostral (FGEPA),  $\chi^2_{R-S} = \frac{\chi^2}{FGEPA}$  que representa o efeito do plano amostral. O FGEPA é tido como a média dos autovalores da matriz  $D = V_{Design} V_{SRD}^{-1}$ , em que

$V_{Design}$  é uma matriz de variância-covariância baseada no plano amostral para o vector de proporções usadas na construção da estatística Qui-Quadrado e  $V_{SRD}^{-1}$  é a matriz de variância-covariância para as proporções estimadas dada uma amostra aleatória simples.

Foi proposta ainda uma correcção de segunda ordem que consiste em dividir a estatística de Rao e Scott pela quantidade  $(1 + a^2)$  de modo a incorporar a variabilidade nos autovalores da matriz D, em que a representa o coeficiente de variação dos autovalores. Thomas e Rao (1984), propuseram uma versão F-ajustado da correcção de segunda ordem, que melhor controla o erro do tipo I, dada pela expressão em:

$$F_{T-R} = \frac{\chi_{R-S}^2}{(L-1)(C-1)} \quad (3.3)$$

#### **3.2.4. Estimação dos Parâmetros na Regressão Logística Binária**

De acordo com Rodrigues (2015), o método de estimação dos parâmetros de regressão logística binária é a máxima verossimilhança, que consiste em maximizar a função de verossimilhança da amostra. Em linhas gerais, o método da máxima verossimilhança fornece estimativas para os parâmetros que maximizam a probabilidade de obter o conjunto observado de dados. Para aplicar tal método deve-se, em primeiro lugar, construir a função chamada função de verossimilhança. Os estimadores de máxima verossimilhança destes parâmetros são escolhidos entre aqueles que maximizam esta função.

Segundo Pessoa e Silva (1998), no caso de amostras complexas, é usado o método de máxima pseudo-verossimilhança, método este que consiste em encontrar valores que maximizam o valor da função de pseudo-máxima verossimilhança, pois, produz estimativas levando em consideração as características do plano amostral. Mesmo que o estimador encontrado na máxima pseudo-verossimilhança possa ser em algumas situações igual ao de máxima verossimilhança, estimativa da variância pelo máxima pseudo-verossimilhança é preferível em relação ao método que não leva em consideração o plano amostral. Este procedimento fornece estimativas mais robustas para a realização de inferência para a população (Pessoa e Silva, 1998).

#### **3.2.5. Teste de Significância dos Parâmetros**

Após a estimação dos coeficientes, tem-se interesse em verificar a significância das variáveis no modelo. Isto geralmente envolve formulação e teste de uma hipótese estatística para determinar se a variável independente no modelo é significativamente relacionada com a variável resposta. Para

isso, existem os testes de hipóteses. Neste trabalho, serão utilizados os testes de Razão de Verossimilhança de Rao-Scott e Wald.

### Teste de Razão de Verossimilhança

Segundo Marôco (2011), o teste do razão de verossimilhanças consiste em comparar as funções de verossimilhança de dois modelos, um modelo só com a constante (modelo nulo) e um modelo com as covariáveis incluídas (modelo completo).

De acordo com Hosmer e Lemeshow (2000), Os métodos em regressão logística seguem o mesmo princípio que em regressão linear que é de comparar os valores observados com os valores preditos da variável resposta. A estatística  $D$  tem como objectivo, comparar o modelo em análise e o modelo saturado, ou seja, a comparação dos valores observados e dos valores esperados usando a função de verossimilhança, assim sendo, para avaliar a significância de uma variável explicativa, comparam-se os valores de  $D$  com ou sem a variável em questão e verifica-se se é significativo, testando as hipóteses:

$$\begin{cases} H_0: \beta_1 = \beta_2 = \dots \beta_p = 0 & \text{(isto é, o modelo não é estatisticamente significativo)} \\ H_1: \exists_i: \beta_j: \neq 0 \ (i = 1, \dots, p) & \text{(isto é, o modelo é estatisticamente significativo).} \end{cases} \quad (3.4)$$

A função de verossimilhança que é baseada na seguinte expressão:

$$D = -2 \ln \left[ \frac{\text{Verossilhança do modelo ajustado}}{\text{Verossimilhança do modelo saturado}} \right] \quad (3.5)$$

$$D = -2 \sum_{i=1}^n \left[ y_i \ln \left( \frac{\hat{\pi}_i}{y_i} \right) + (1 - y_i) \ln \left( \frac{1 - \hat{\pi}_i}{1 - y_i} \right) \right] \quad (3.6)$$

Onde : **modelo ajustado** corresponde ao modelo com apenas as variáveis desejadas para o estudo e o **modelo saturado** corresponde ao modelo com todas as variáveis e interações.

A estatística do teste de razão de verossimilhanças para testar a significância conjunta dos coeficientes, isto é, a significância do modelo, é dada pela equação em 3.6.

$$G^2 = -2 \ln \left[ \frac{L(R)}{L(F)} \right] \quad (3.7)$$

Onde:

$L(R)$  é a função de verossimilhança do modelo ajustado, isto é, do modelo contendo  $q$  parâmetros e  $L(F)$  é a verossimilhança do modelo completo.

$G^2$  segue aproximadamente uma distribuição  $\chi^2_{p-q}$ , sendo a sua regra de decisão procedida de forma análoga ao teste  $G$ .

Segundo Lumley (2018), para amostras complexas, o teste de razão de verossimilhanças é baseado em diferenças entre Deviance resultantes da ponderação das funções de pseudo-máxima verossimilhança.

### Teste de Wald

Segundo Silva (2011), o teste de Wald, descrito por Polit (1996) e Agresti (1990), é uma das possíveis formas de testar se o parâmetro associado toma o valor zero. Este teste é utilizado para avaliar se o parâmetro é estatisticamente significativo.

O teste de Wald é obtido por comparação entre a estimativa de máxima verossimilhança do parâmetro ( $\beta_0$ ) e a estimativa de seu erro-padrão. A razão resultante, sob as hipóteses em 3.4 tem distribuição normal padrão.

$$\begin{cases} H_0: & \beta_j = 0 \\ H_1: & \beta_j \neq 0 \end{cases} \quad (3.8)$$

A estatística de teste é dada por:

$$W_j = \frac{\hat{\beta}_j}{\text{var}(\hat{\beta}_j)} \cap N(0,1) \quad (3.9)$$

Onde  $\beta_i$  é vector de estimativas para  $\beta$  e  $\text{Var}(\hat{\beta}_j)$  é a matriz de variâncias.

De modo a testar a significância dos parâmetros do modelo é usado um teste de Wald ajustado, baseado no plano amostral, dado pela equação 3.9:

$$F = \frac{s-p+1}{sp} W \quad (3.10)$$

Onde  $s = (\sum_{k=1}^k m_k) - K$  é o número de conglomerados seleccionados menos o número de estratos e  $p$  é o número de variáveis explicativas e  $W$  é a versão multivariada do teste de Wald. A estatística  $F$  segue a distribuição  $F$  com  $p$  com  $(s - p + 1)$  graus de liberdade (Cassy *et al.*, 2016).

**3.2.6. Qualidade de Ajuste na Regressão Logística Binária e Significância do Modelo**

Segundo Alvarenga (2015), depois de obter as estimativas dos coeficientes da regressão, é necessário avaliar a qualidade do modelo ajustado. O teste de Hosmer e Lemeshow é utilizado para avaliar de uma forma geral a qualidade do ajuste de um modelo. O teste avalia se o modelo obtido pode explicar adequadamente os dados observados. Este teste baseia-se na divisão dos dados de acordo com as probabilidades previstas. As observações são separadas em  $g$  grupos de acordo com as probabilidades previstas.

Segundo Marôco (2018), o teste de Hosmer e Lemeshow consiste em comparar as frequências previstas e as frequências observadas de ocorrência de sucesso em cada grupo por meio de uma estatística Qui-Quadrado. A estatística do teste é dada pela equação:

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - e_k)^2}{e_k \left(1 - \frac{e_k}{n_k}\right)} \tag{3.11}$$

Onde:  $O_k = \sum_{j=1}^{n_k} yk_j$  representa o número de casos registados no  $k$ -ésimo decil, e  $e_k = \sum_{j=1}^{n_k} \hat{\pi}_{kj}$  representa o número esperado de casos no  $k$ -ésimo decil.

As hipóteses a testar são:

$$\begin{cases} H_0: & \text{O modelo se ajusta aos dados.} \\ H_1: & \text{O modelo não se ajusta aos dados.} \end{cases} \tag{3.12}$$

Sob  $H_0$ , o modelo estimado ajusta-se aos dados, isto é, não existem diferenças significativas entre os valores previstos e os observados, em que a regra de decisão consiste em:

- Se  $\hat{C} \leq \chi^2_{1-\alpha;g-2}$  ou  $P(\hat{C} > \chi^2_{1-\alpha;g-2}) \geq \alpha$ , não rejeitar  $H_0$ .
- Se  $\hat{C} > \chi^2_{1-\alpha;g-2}$  ou  $P(\hat{C} > \chi^2_{1-\alpha;g-2}) < \alpha$ , rejeitar  $H_0$ .

De acordo com Mesquita (2014) em modelos de regressão logística, medidas de qualidade do ajuste são funções dos resíduos definidos como a diferença entre o valor observado e o valor ajustado. Por se tratar de regressão logística, não existe uma estatística que seja equivalente ao  $R^2$ , utilizado em regressão linear.

A denominação de pseudo- $R^2$  deve-se ao facto de eles se parecem com um  $R^2$  do modelo de regressão linear, pois estão em uma escala similar, variando de 0 à 1, mas apesar dessa similaridade não podem ser interpretados da mesma forma como se interpreta em regressão linear.

A Razão de verossimilhança de teste fornece a estimativa da razão de verossimilhança entre o modelo logístico final (com as variáveis independentes incluídas) e o modelo nulo (sem as variáveis independentes). Para verificar a significância do modelo, é aplicado o teste de razão de verossimilhança, com hipóteses:

$$H_0: \beta_i = 0, i = 1, 2.$$

$$H_1: \beta_i \neq 0, i = 1, 2.$$

### 3.2.7. Interpretação dos Parâmetros do Modelo

De acordo com Alvarenga (2015), depois de ajuste do modelo, assumindo o pressuposto de que o modelo se ajuste bem aos dados e de que os coeficientes estimados são significativos, é necessário interpretar os valores associados aos coeficientes do modelo.

De acordo com Hair *et al.* (2006), uma das vantagens de regressão logística é que precisa-se saber apenas se um evento ocorreu para usar um valor dicotômico como variável dependente. A partir desse valor dicotômico, o procedimento prevê sua estimativa da probabilidade de que o evento ocorrerá ou não. O procedimento que calcula o coeficiente da regressão logística compara a probabilidade de ele não ocorrer. Essa razão de desigualdade pode ser expressa como:

$$\frac{\text{Prob (evento ocorrer)}}{\text{Prob (evento não ocorrer)}} = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k} \quad (3.13)$$

Os coeficientes estimados ( $\beta_0, \beta_1, \dots, \beta_k$ ) são medidas das variações das probabilidades, chamada de razão de desigualdade. Além disso, eles são apresentados em logaritmos, precisando ser transformados de forma que seu efeito relativo sobre as probabilidades seja avaliado de forma fácil. Um coeficiente positivo mostra que aumenta a probabilidade, ao passo que um valor negativo diminui a probabilidade prevista.

Segundo Alvarenga (2015), a interpretação dos coeficientes do modelo de regressão depende da natureza das variáveis explicativas que podem ser categóricas ou contínuas. No caso da variável explicativa categórica é necessário criar variáveis auxiliares para que possam ser introduzidas no modelo. Estas variáveis também são designadas de variáveis dummy.

### Variáveis Independentes Dicotômicas

Quando a variável explicativa,  $X$ , é categórica e pode assumir dois valores possíveis. Se  $X$  tomar os valores 0 e 1, podemos construir uma tabela de contingência com as probabilidades que se pretendem estimar, nomeadamente a probabilidade em que a variável resposta,  $Y$ , pode assumir o valor 0 e o valor 1, e tenha a seguinte distribuição de probabilidade:

Tabela 3.2.1: **Distribuição de probabilidades**

	X=0	X=1
Y=0	$1-\pi(0)$	$1-\pi(0)$
Y=1	$\pi(0)$	$\pi(1)$

Em que a expressão do cálculo de  $\pi(1)$  e  $\pi(0)$  é respectivamente dada por:

$$\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}; \pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}; 1 - \pi(0) = \frac{1}{1 + e^{\beta_0}}; 1 - \pi(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}} \quad (3.14)$$

Portanto, pode definir-se a chance (odds) da seguinte forma:

$$\frac{\pi_1}{1-\pi_1} = e^{\beta_1 + \beta_2} \text{ e } \frac{\pi_0}{1-\pi_0} = e^{\beta_1} \quad (3.15)$$

A razão  $\frac{\pi_1}{1-\pi_1}$  corresponde à chance da variável resposta assumir o valor 1 em relação ao valor 0, quando a variável explicativa é igual a 1,  $X = 1$ . Da mesma forma a razão  $\frac{\pi_0}{1-\pi_0}$  significa a chance da variável resposta assumir valor 1 em relação ao valor 0 quando a variável explicativa é igual a 0,  $X = 0$ .

Aplicando o logaritmo, ou seja a função logit, obtém-se:

$$g(1) = \ln \left[ \frac{\pi(1)}{1 - \pi(1)} \right]; g(0) = \ln \left[ \frac{\pi(0)}{1 - \pi(0)} \right] \quad (3.16)$$

A razão entre as chances é designada por razão da chance ou em termo inglês odds ratio, a sua expressão é dada por

$$OR = \frac{\frac{\pi(1)}{1 - \pi(1)}}{\frac{\pi(0)}{1 - \pi(0)}} \quad (3.17)$$

Substituindo os valores  $\frac{\pi_1}{1-\pi_1}$  e  $\frac{\pi_0}{1-\pi_0}$  de pelas expressões apresentadas anteriormente, vem

$$odds\ ratio = \frac{e^{\beta_1 + \beta_2}}{e^{\beta_1}} = e^{\beta_2}$$

Para amostras suficientemente grandes, o estimador para  $\beta_i$ ,  $\hat{\beta}_i$ , segue uma distribuição aproximadamente normal. Supondo válida essa aproximação à Normal dos  $\hat{\beta}_i$  encontra-se primeiro o IC para  $\beta_i$ .

$$\hat{\beta}_i \pm Z_{1-\frac{\alpha}{2}} * Ep(\hat{\beta}_i) \tag{3.18}$$

Onde  $Z_{1-\frac{\alpha}{2}}$  representa o quantil de probabilidade  $(1 - \frac{\alpha}{2})$  da distribuição Normal do valor médio nulo e variância unitária e seguidamente  $e^{\beta_i}$ .

Assim, o intervalo de  $100(1-\alpha)\%$  de confiança para  $e^{\beta_i}$  é dado por:

$$e^{\left(\hat{\beta}_i \pm Z_{1-\frac{\alpha}{2}} * Ep(\hat{\beta}_i)\right)} \tag{3.19}$$

Se o IC contiver o valor 1, não existe relação significativa entre as variáveis em causa,  $X$  (variável explicativa) e  $Y$  (variável resposta).

**Variáveis Independentes Policotômicas**

Quando a variável explicativa,  $X$ , é categórica com mais de duas categorias distintas, então é necessário criar  $m-1$  variáveis dummy. Estas variáveis podem assumir apenas os valores 0 ou 1. Por conveniência as  $m$  categorias são numeradas de 0 a  $m-1$ , sendo a categoria 0 a classe de referência.

Os valores possíveis 0 ou 1 das variáveis dummy significam que se a característica de um objecto ou indivíduo pertence à classe  $i$ , com  $i = 1, \dots, m-1$ , a todas as variáveis dummy corresponderá o valor 0, com exceção da  $i$ -ésima classe, à qual corresponderá o valor 1. E no caso da classe referência, se a característica de um objecto pertence a esta classe, então a todas as  $m-1$  variáveis dummy corresponderá o valor 0.

Tabela 3.2.2: Categoria das variáveis Policotômicas

Categoria da variável X	Variáveis dummy			
	Categoria 1	Categoria 2	...	Categoria $m-1$
Categoria 1	0	0	...	0
Categoria 2	1	0	...	0
...	...	...	...	...
Categoria $m$	0	0	...	1

Assim, para cada categoria da variável explicativa pode estimar-se a probabilidade da variável resposta assumir o valor 1 e a chance desta variável assumir o valor 1 em relação ao valor 0. O cálculo e interpretação do valor da razão da chance é semelhante ao caso da variável dicotômica.

### Variáveis Independentes Contínuas

Quando um modelo de regressão logística contém uma variável independente contínua, a interpretação do coeficiente estimado será feita com base no pressuposto da linearidade entre a variável resposta e a variável independente, cuja relação linear é dada por  $\beta_1 + \beta_2 x$ .

Sendo assim a interpretação do coeficiente estimado é semelhante ao do modelo clássico de regressão. O coeficiente  $\beta_2$  representa a variação no logaritmo da chance por uma unidade de variação no valor da variável independente,  $x$ . Assim, é necessário saber de quantas unidades se deverá aumentar  $x$ , de modo que a interpretação seja considerada adequada.

No caso de a variável ser dicotômica vimos que existe uma diferença no logit ou logaritmo da chance quando a variável independente passa da classe referência para outra classe. Aqui o processo é semelhante, ao aumentar uma unidade no valor da variável,  $x$ , haverá uma diferença  $\beta_2$  no logaritmo da chance e se aumentarmos  $k$  unidades, haverá uma diferença de  $k\beta_2$ . Para estimar o valor de odds ratio é a exponencial de  $\beta_2$  ou  $k\beta_2$ . O intervalo de  $100(1-\alpha)\%$  de confiança para a estimativa e  $k\beta_2$  é dado por:

$$e^{\left(k\hat{\beta}_i \pm Z_{1-\frac{\alpha}{2}} * Ep(\hat{\beta}_i)\right)} \quad (3.20)$$

No caso de não se verificar a condição de linearidade no logit ou mesmo no caso em que a escolha de  $k$  seja difícil, poderá pôr-se a hipótese de categorizar a variável em estudo, em duas ou mais categorias.

#### 3.2.8. Estratégias de Análise

O processo de selecção iniciou com o teste F de Thomas e Rao para independência de cada variável explicativa em relação à variável dependente, a presença do vírus de hepatite B (ter ou não hepatite B), com a finalidade de verificar a associação à nível de significância inferior à 5% (p-valor menor que 0.05), no SPSS.

Fez-se uma avaliação das variáveis a serem incluídas no modelo de regressão logística múltiplo, e buscou-se uma forma de encontrar um modelo ajustado que incluía as variáveis mais importantes

para explicar a probabilidade de ter Hepatite B  $P(Y = 1/X)$ , sendo que  $Y$  toma dois valores possíveis, tem hepatite B (1) e não tem hepatite B (0).

A investigação dos factores associados a presença de Hepatite B foi feita utilizando a regressão logística clássica, através do software SPSS e R para confirmação. Baseando-se no método backward, foram eliminadas as variáveis segundo o grau de importância das mesmas, de acordo com um critério estatístico. O grau de importância foi avaliado pelo p-value do teste de razão de verossimilhança de Rao-Scott, com recurso ao R-studio, segundo os pressupostos de quanto maior for o valor do teste de razão de verossimilhança de Rao-Scott mais, importante será considerada a variável. Terminada a selecção do modelo final, será feita a interpretação da razão de chances das categorias estatisticamente significativas (p-value < 0.05).

Recorrendo ao R, avaliou-se a significância do modelo por meio do teste de razão de verossimilhança entre o modelo logístico final e o modelo nulo, de modo a testar a hipótese de que todos os parâmetros do modelo final sejam estatisticamente iguais a zero ao nível de significância de 5%, e recorrendo ao SPSS, avaliou-se a qualidade de ajuste através do teste de Hosmer-Lemeshow para o mesmo nível de significância.

### **3.2.9. Especificação do Modelo**

Para o presente estudo, o modelo adequado e parcimonioso proposto para explicar a relação entre a ocorrência do VHB e o conjunto de variáveis aleatórias explicativas descritas na tabela 3.1.1. é dado por

$$Y_i = g(x) = \text{logit}(\pi_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i \quad (3.21)$$

Onde  $Y_i$  é a probabilidade de um profissional de saúde do Hospital Central de Maputo adquirir (1) ou não adquirir (0) o Vírus de Hepatite B em razão dos valores assumidos pelas variáveis explicativas ( $x_i$ ). O intercepto ( $\beta_0$ ) representa a chance de um profissional adquirir o VHB quando todas as variáveis do modelo ( $x_i$ ) assumirem o valor 0.  $x_1, x_2, x_2 \dots x_i$  é o conjunto de variáveis predictoras. Os coeficientes  $\beta_i$  representam a variação observada da ocorrência ou não do VHB em razão dos valores assumidos pelas variáveis explicativas ( $x_i$ ).

De forma específica, o modelo estimado proposto é descrito da seguinte forma:

$$\begin{aligned} \text{VHB} = & \beta_0 + \beta_1.\text{HEP01} + \beta_2.\text{HEP02} + \beta_3.\text{HEP03} + \beta_4.\text{HEP04} + \beta_5.\text{HEP05} \\ & + \beta_6.\text{HEP06} + \beta_7.\text{HEP07} + \beta_8.\text{HEP08} + \beta_9.\text{HEP0} + \beta_{10}.\text{HEP10} \\ & + \beta_{11}.\text{HEP11} + \beta_{12}.\text{HEP12} + \beta_{13}.\text{HEP13} + \beta_{14}.\text{HEP14} + \beta_{15}.\text{HEP16} \\ & + \beta_{16}.\text{HEP17} + \beta_{17}.\text{HEP17} + \beta_{18}.\text{HEP18} + \beta_{18}.\text{HEP20} \end{aligned}$$

Onde a descrição destas variáveis bem como as suas categorias estão descritas no corrente capítulo, na tabela 3.1.1.

Um coeficiente positivo ( $\beta_i > 0$ ) mostra que aumenta a probabilidade de ocorrência ou de não ocorrência do VHB nos profissionais de saúde quando ( $x_i$ ) aumenta, ao passo que um valor negativo de ( $\beta_i < 0$ ) diminui a probabilidade prevista.

Como Dias *et al.* (2014) apontam que o VHB pode ser transmitido por meio de relações sexuais desprotegidas, uso de drogas injectáveis, transfusões sanguíneas, amamentação ou transmissão vertical e por meio de acidentes com material perfurantes e cortantes. Pode ocorrer também por meio da exposição percutânea sendo elas, piercing, tatuagens, e o compartilhamento de materiais contaminados, como navalhas, alicates de unha, lâminas de depilação, tesoura, entre outros, espera-se que os coeficientes  $\beta_i$  sejam maiores que 0 ( $\beta_i > 0$ ), de modo com que os PS expostos as condições descritas sejam mais propensos ou tenham mais chances de ter o VHB em relação aos PS não expostos as mesmas condições.

## 4 Resultados e Discussão

O presente capítulo é dedicado à apresentação dos resultados. No mesmo será apresentada, inicialmente, uma análise descritiva e os resultados do teste de independência entre as variáveis do estudo, e por fim, os resultados referentes aos modelos estimados.

### 4.1 Análise descritiva dos dados e associação bivariada entre as variáveis em estudo.

A amostra analisada no presente estudo é composta por 587 profissionais de saúde de diferentes categorias afectos ao Hospital Central de Maputo. No que diz respeito ao sexo dos participantes, 69.3% são do sexo feminino e 30.7% do sexo masculino. Relativamente ao estado civil, a maior parte dos participantes são solteiros (60.3%), 35.9% são casados, 2.2% são viúvos e 1.5% dos participantes são divorciados. No atinente a idade dos profissionais de saúde que fizeram parte do estudo, 80.4% dos profissionais tem idades compreendidas entre 30 a 59 anos, 16.9% com idades compreendidas entre 18 à 29 anos e 2.7% com 60 anos ou mais. A idade média dos profissionais de saúde que participaram no estudo foi de 38.76 anos, com idade mínima de 19 anos e idade máxima de 82 anos.

Através dos resultados da figura 4.1.1, ilustra-se que a taxa de positividade do vírus de hepatite B nos profissionais de saúde foi de 6.5%.

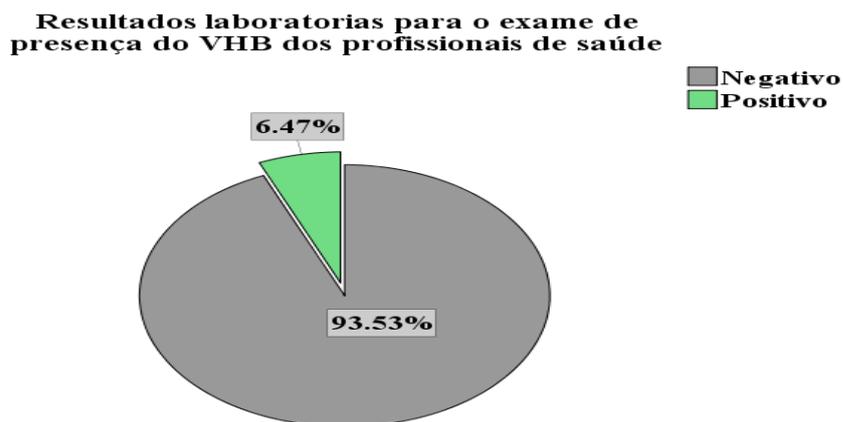


Figura 4.1.1: Resultados laboratorias para o exame de presença do VHB

Num total de 354 profissionais de saúde solteiros, 7.63% apresentaram resultado positivo no exame de presença do VHB. Quanto aos participantes do estudo casados, dentre 211 profissionais de saúde 4.74% tiveram resultado positivo para o exame de VHB. Os participantes com o estado civil viúvo, correspondentes a 13 profissionais de saúde, tiveram uma taxa de positividade do VHB de 7.69%. Por fim, dos 9 participantes divorciados do estudo, nenhum profissional teve o resultado positivo para o VHB, tendo assim uma taxa de positividade de 0.0%.

**Tabela 4.1.1: Distribuição das frequências absolutas e relativas; e a relação entre os factores associados com a presença do VHB, com o resultado do exame do VHB nos profissionais de saúde.**

Variável	Categoria	Tem Hepatite B				Total	P-value
		Nao	(%)	Sim	(%)		
Sexo dos Participantes	Masculino	164	91.11%	16	8.89%	180	0.114
	Feminino	385	94.59%	22	5.41%	407	
Estado Civil	Casado	201	95.26%	10	4.74%	211	0.467
	Divorciado	9	100.00%	0	0.00%	9	
	Solteiro	327	92.37%	27	7.63%	354	
	Viuvo	12	92.31%	1	7.69%	13	
Consumo de Álcool	Nao consome	300	94.64%	17	5.36%	317	0.236
	Consome	249	92.22%	21	7.78%	270	
Consumo de Drogas como Cannabis e Cigarros	Nao consome	520	93.36%	37	6.64%	557	0.713
	Consome	29	96.67%	1	3.33%	30	
Consumo de Drogas por Via Endovenosa	Nao consome	547	93.50%	38	6.50%	585	0.875
	Consome	2	100.00%	0	0.00%	2	
Uso Habitual de Protecção no Acto Sexual	Nao usa	287	92.58%	23	7.42%	310	0.325
	Usa	262	94.58%	15	5.42%	277	
Vários Episódios de Sexo Ocasional	Sim	498	93.79%	33	6.21%	531	0.394
	Nao	51	91.07%	5	8.93%	56	
Prática de Sexo em Grupo	Nao Pratica	540	93.43%	38	6.57%	578	0.545
	Pratica	9	100.00%	0	0.00%	9	
Antecedentes de ITS	Nao tem antecedentes	447	93.71%	30	6.29%	477	0.706
	Tem antecedentes	102	92.73%	8	7.27%	110	
Troca de Sexo por Bens e Favores	Nunca trocou	545	93.48%	38	6.52%	583	0.765
	Ja trocou	4	100.00%	0	0.00%	4	
Transusão Sanguínea ou Derivados	Nunca recebeu	490	93.87%	32	6.13%	522	0.295
	Ja recebeu	59	90.77%	6	9.23%	65	
Antecedentes de Cirurgias	Nao tem antecedentes	350	93.09%	26	6.91%	376	0.562
	Tem antecedentes	199	94.31%	12	5.69%	211	
Ter Piercing ou Tatuagens	Nao tem	499	93.27%	36	6.73%	535	0.564
	Tem	50	96.15%	2	3.85%	52	
Partilhas de Escovas e Máquinas de Barbear	Nao Partilha	451	94.95%	24	5.05%	475	<b>0.006</b>
	Partilha	98	87.50%	14	12.50%	112	
Acidente Ocupacional	Nunca teve	412	95.15%	21	4.85%	433	<b>0.009</b>
	Ja teve	137	88.96%	17	11.04%	154	
Vacina Contra Hepatite B	Nao	427	94.26%	26	5.74%	453	0.184
	Sim	122	91.04%	12	8.96%	134	
Hábito de Manter Relações Sexuais Desprotegidas	Nao tem habito	341	94.20%	21	5.80%	362	0.401
	Tem habito	208	92.44%	17	7.56%	225	
Idade dos Participantes	18-29 anos	92	92.93%	7	7.07%	99	0.556
	30-59 anos	441	93.43%	31	6.57%	472	
	60+	16	100.00%	0	0.00%	16	
Departamento	Administração	61	92.42%	5	7.58%	66	0.533
	Cirurgia	126	94.74%	7	5.26%	133	
	Clínica Especial	25	92.59%	2	7.41%	27	
	Farmácia	2	66.67%	1	33.33%	3	
	Laboratório	18	90.00%	2	10.00%	20	
	Medicina	136	91.89%	12	8.11%	148	
	Obstetrícia	29	96.67%	1	3.33%	30	
	Pediatria	42	93.33%	3	6.67%	45	
SUR	110	95.65%	5	4.35%	115		

Em relação ao sexo dos participantes, dentre os 180 profissionais de saúde do sexo masculino, 8.89% tiveram resultados positivos para o exame do VHB, e nos profissionais de saúde do sexo feminino, dentre 407 participantes, 5.41% tiveram resultado positivo para o VHB.

Pode-se verificar que num universo de 475 profissionais de saúde que não partilham escovas ou máquinas de barbear, 94.95% dos profissionais tiveram resultado negativo para o exame do VHB e 4.85% tiveram resultado positivo. Num total de 112 profissionais de saúde que partilham escovas e máquinas de barbear, 87.50% tiveram resultado negativo para o exame do VHB, e 12.50% tiveram resultado positivo para o exame do VHB.

É possível verificar também que num total de 433 profissionais que não tiveram algum tipo de acidente ocupacional, 95.15% dos profissionais de saúde tiveram resultado negativo para o exame do VHB, e 4.85% tiveram resultado positivo para o mesmo exame. Dentre os 154 profissionais que tiveram algum tipo de acidente ocupacional, 88.96% dos profissionais de saúde tiveram resultado negativo para o exame do VHB, e 11.04% tiveram resultado positivo para o exame do VHB.

Dentre os 453 profissionais de saúde que não tiveram a vacina contra o vírus de Hepatite B, 94.26% dos profissionais tiveram resultado negativo para o exame do VHB, sendo que dos 134 profissionais de saúde que foram vacinados contra o vírus de Hepatite B, 91.04% tiveram resultado negativo para o exame do VHB e 8.96% tiveram resultado positivo para o mesmo exame.

A associação entre a variável de interesse e as demais variáveis em estudo procedeu-se usando o teste F de Thomas e Rao que é um ajuste ao teste Qui-Quadrado para amostras complexas. Todas as variáveis em análise, cujo p-value foi menor do que o nível de significância ( $p\text{-value} < 0.05$ ) tem um contributo estatisticamente significativo em relação ao resultado do exame do VHB. Serão usadas as variáveis da tabela 4.1.1 para a construção do modelo de regressão logístico múltiplo saturado e ajustado, aplicando respectiva técnica.

## **4.2 Modelo de Regressão Logística Múltipla**

Inicialmente, foi construído um modelo de regressão logístico saturado (tabela 5.1 em anexo), com todas as variáveis em estudo. Após a estimação do modelo saturado, pelo teste de razão de verosimilhança de Rao-Scott, foi avaliada a significância global de cada variável, excluindo as que não são estatisticamente significativas a um nível de significância de 5%.

Tabela 4.2.1: Teste de razão de verossimilhança de Rao-Scott - Modelo Saturado

Variáveis	p-valor
Partilhas de Escovas e Máquinas de Barbear	0.004
Acidente Ocupacional	0.007

Este processo culminou no modelo logístico final só com variáveis significativas a um nível de significância de 5%, modelo este que envolve o mínimo de parâmetros possíveis (modelo parcimonioso) e com grande poder predictivo e explicativo. Usando todas as variáveis do modelo saturado, e já tendo as variáveis estatisticamente significativas pelo teste de razão de verossimilhança de Rao-Scott, usou-se o Método Backward, descrito em 2.6.4, para flexibilizar o processo de selecção e confirmação de variáveis a reter.

#### 4.2.1. Modelo Logístico Final

Tabela 4.2.2: Estimativa dos parâmetros e ajuste do Modelo Final

Predictores	Coefficientes Estimados	Erro Padrão	Teste Wald	de	P-valor	IC (95%) dos coeficientes
Intercepto	-3.2375	0.2578	15.136		< 2e-16	[(-3.779) - (-2.764)]
Partilhas de Escovas e Máquinas de Barbear (Partilha)	0.984	0.357	7.584		0.006	[0.262- 1.674]
Acidente Ocupacional (Já teve)	0.887	0.344	6.649		0.010	[0.202 - 1.559]
Teste de Razão de máxima verossimilhança			$\chi^2_{(3)} =$	13.419	$p = 0.001$	
Teste de Hosmer e Lemeshow			$\chi^2_{(3)} =$	0.883	$p = 0.643$	

Para verificar o ajuste do modelo logístico foi aplicado o teste de Hosmer e Lemeshow. Os resultados obtidos através do teste de Hosmer e Lemeshow, com um p-value de 0.643, apontam que há evidências suficientes para não se rejeitar a hipótese nula comparado ao nível de significância de 5% , ajustando-se, deste modo, o modelo aos dados.

Testando assim a significância do modelo, por meio do teste de Razão de Verossimilhança entre o modelo nulo e o modelo final, rejeita-se a hipótese nula de que todos os parâmetros sejam estatisticamente iguais a zero ao nível de significância de 5%, sendo  $\chi^2_{(3)} = 13.419$  e  $p = 0.001$ . Deste modo, podemos concluir que pelo menos uma variável independente é estatisticamente significativa para explicar a probabilidade de os profissionais de saúde ter o VHB.

**4.2.2. Interpretação das estimativas para a razão de chances do modelo de regressão logística final**

Os resultados da razão de chances do modelo de regressão logística final são demonstrados na tabela 4.2.3.

Tabela 4.2.3: Estimativa de Razão de Chances do Modelo Final

Predictores	OR	Erro Padrão	Teste Wald	de P-valor	IC (95%) OR
Intercepto	0.039	0.2578	15.136	< 2e-16	[0.023 - 0.063]
Partilhas de Escovas e Máquinas de Barbear (Partilha)	2.676	0.3574	7.584	0.006	[1.299 - 5.332]
Acidente Ocupacional (Já teve)	2.427	0.3439	6.649	0.010	[1.223 - 4.755]

O modelo final reescrito em termo de probabilidade e a interpretação dos parâmetros estimados em termos de razão de chance:

$$\pi_i = \frac{e^{-3.238+0.984 \times PartilhasdeEscovaseMaquinasBarbear+0.887 \times Acidente\ Ocupacional}}{1-e^{-3.238+0.984 \times PartilhasdeEscovaseMaquinasBarbear+0.887 \times Acidente\ Ocupacional}} \quad (4.1)$$

E o logit na forma:

$$\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = -3.238 + 0.984 \times PartilhasdeEscovaseMaquinasBarbear + 0.887 \times Acidente\ Ocupacional \quad (4.2)$$

Em que:

**Intercepto:** O valor estimado -3.238 representa a log-odds (logit) de um profissional de saúde ter Hepatite B quando todas as variáveis do modelo forem constantes e iguais a 0, (ou seja, quando o profissional de saúde não partilha escovas e máquinas de barbear e nunca teve acidente ocupacional). Assim sendo, as chances de um profissional de saúde ter o vírus da hepatite B, quando o profissional de saúde não partilha escovas e máquinas de barbear e nunca antes teve um acidente ocupacional, reduz em cerca de 96.1%.

**Partilhas de Escovas e Máquinas de Barbear (Partilha):** Mantendo todas as outras variáveis independentes constantes, espera-se que o log-odds da categoria partilha (profissionais de saúde que partilham escovas e máquinas de barbear) seja 0.984 unidades maior do que o log-odds da categoria de referência (Pacientes que não partilham escovas e máquinas de barbear).

Como o coeficiente  $\beta_1 = 0.984 > 0$ , a probabilidade de ocorrência de VHB nos PS aumenta quando o PS partilha escovas e máquinas de barbear. A razão de chances correspondente é 2.676, o que significa que os profissionais de saúde que partilham escovas e máquinas de barbear tem 2.676 (167.6%) vezes mais chances de ter Hepatite B em relação aos profissionais de saúde que não partilham escovas e máquinas de barbear, mantendo as outras variáveis constantes.

**Acidente Ocupacional (Já teve):** Mantendo todas as outras variáveis independentes constantes, espera-se que o log-odds da categoria “Já teve” (profissional de saúde que já tiveram acidente ocupacional) seja 0.887 unidades maior do que o log-odds da categoria de referência (profissional de saúde que nunca tiveram acidente ocupacional).

Como o coeficiente  $\beta_2 = 0.887 > 0$ , a probabilidade de ocorrência de VHB nos PS aumenta quando o PS tem acidente ocupacional. A razão de chances correspondente é 2.427, o que significa que os profissionais de saúde que já tiveram acidente ocupacional tem 2.427 (142.7%) vezes mais chances de ter Hepatite B em relação aos profissionais de saúde que nunca tiveram acidente ocupacional, mantendo as outras variáveis constantes.

### 4.3 Discussão dos Resultados

Analisando os possíveis factores associados à ocorrência do vírus da hepatite B nos profissionais de saúde do Hospital Central de Maputo, constatou-se que apenas os factores partilhas de escovas e máquinas de barbear e acidente ocupacional têm um contributo estatisticamente significativo com a variável de interesse. Para tal efeito, foi necessário estimar modelos de regressão logística clássico que pudessem identificar tais factores. A variável de interesse é binária (representa a ocorrência do VHB), cujas categorias são: 0 para os profissionais que não tem o vírus e 1 para os profissionais que tem o vírus.

Quanto ao acidente ocupacional, Lopes *et al.* (2000) e Morais *et al.* (2016), ambos com coeficientes  $\beta_i > 0$  para os factores exposição ocupacional e não uso de equipamentos de proteção, demonstraram no seu estudo que quanto maior a exposição ocupacional e não uso de equipamentos de proteção, as chances de seropositividade ao VHB aumentam significativamente, ou seja, as chances de ter o vírus, aumentam para os profissionais que tem esses acidentes ocupacionais. O presente estudo sobre os factores de risco associados a ocorrência de hepatite B nos profissionais de saúde do Hospital Central de Maputo, provou que os profissionais que tem ou já tiveram acidentes ocupacionais tem mais chances de ter o vírus, com um  $\beta_2 > 0$ .

A partilha de escovas e máquinas de barbear foi considerado um factor de risco significativo para a ocorrência do VHB, contribuindo para um aumento de aproximadamente 3 vezes nas chances, corroborando assim com os achados de diversos estudos que apresentam também um  $\beta_i > 0$  para este factor, como de Luz *et al.* (2001), Braga *et al.* (2004), Silva *et al.* (2010), Paiva E. M. M. (2008), e Dias *et al.* (2014), que enfatizam ainda de forma generalizada, que o risco da partilha de escovas e máquinas de barbear, bem como outros objectos perfurantes e cortantes, vem da exposição percutânea ou de mucosas a sangue ou fluidos do corpo contaminados com o VHB.

## **5 Conclusões e Recomendações**

O presente capítulo é dedicado à apresentação das conclusões, olhando para os objectivos que foram propostos para este trabalho no primeiro capítulo e as recomendações.

### **5.1. Conclusões**

O trabalho teve como objectivo principal analisar os factores de risco associados à ocorrência da Hepatite B nos Profissionais de Saúde do Hospital Central de Maputo, e para alcançar este objectivo foram usados os dados do Faculdade de Medicina da Universidade Eduardo Mondlane, de um estudo realizado no Hospital Central de Maputo no ano 2015, cujo estudo é de coorte transversal e representativo. Sendo assim, pôde-se constatar que:

1. Cerca de 6.5% dos profissionais de saúde do Hospital Central de Maputo, tem o vírus de hepatite B, sendo a maioria do sexo feminino. Grande parte dos profissionais de saúde do HCM tem o estado civil solteiro e tem idades entre 30 a 59 anos.
2. A partilha de escovas e máquinas de barbear e acidente ocupacional são as variáveis explicativas associadas à ocorrência do VHB nos profissionais de saúde do Hospital Central de Maputo, tornando-se assim os mais importantes para analisar a ocorrência do vírus de hepatite B no Hospital Central de Maputo.
3. Os factores de risco que melhor explicam à ocorrência de hepatite B nos profissionais de saúde do Hospital Central de Maputo são a partilha de escovas e máquinas de barbear pelos profissionais de saúde e a ocorrência de acidentes ocupacionais. Os profissionais de saúde que partilham escovas e máquinas de barbear, e que já tiveram algum tipo de acidente ocupacional, tem 3 vezes mais chances de ter o vírus de hepatite B, relativamente aos profissionais que não partilham escovas e máquinas de barbear e não tiveram algum tipo de acidente ocupacional.

Conclui-se ainda que, o presente estudo possa contribuir para a melhoria das políticas de saúde nos nossos hospitais, ao identificar os factores de risco da ocorrência do vírus de hepatite B nos profissionais de saúde, alertando para a necessidade de melhorar o uso de equipamentos de protecção que reduzam as chances de transmissão de vírus através de acidentes ocupacionais, bem como alertar sobre o risco de partilha de objectos capazes de perfurar e cortar em simultâneo, entre os profissionais de saúde, bem como pela comunidade no geral, visto que

pelo sinal dos coeficientes encontrados no trabalho, as chaces de ocorrência do VHB aumentam quando os PS partilham escovas e máquinas de barbear e quando sofrem algum tipo de acidente ocupacional.

## **5.2. Recomendações**

Tomando em consideração as conclusões obtidas e observações feitas ao longo da realização do presente trabalho, recomenda-se que:

1. A realização de estudos similares por forma a identificar os factores associados a ocorrência do vírus de hepatite B nos profissionais de saúde, ao nível do país, ampliando assim a discussão e trazer novos esclarecimentos sobre este tema;
2. Haja mais publicação de informação relativa aos riscos associados a ocorrência do vírus de hepatite B;
3. Haja mais orientação sobre o vírus de hepatite B, quanto ao risco de transmissão do vírus por via parental, sexual, vertical e percutânea, bem como quanto aos métodos de prevenção do vírus, como por exemplo, o uso de equipamentos de protecção individual; o não compartilhamento de objectos perfurantes e cortantes, escova de dente; o uso de protecção nas relações sexuais; entre outros métodos.

## **5.3.Limitações**

A fraca disponibilidade de estudo similares publicados na área de saúde ao nível nacional com aplicação de modelos lineares generalizados, especificamente a regressão logística, foi a grande limitação para a elaboração deste trabalho.

## Referências bibliográficas

1. Alvarenga, A. (2015). *Dissertação: Modelos lineares generalizados: aplicação a dados de acidentes rodoviários*. Lisboa, Portugal.
2. Beltrami E.M. *et al.* (2000). *Risk and Management of Blood-Borne Infections in Health Care Workers*. Clinical Microbiology Reviews;
3. Braga, A. (1994). *Acidente Vascular Cerebral e seus Factores de Risco. Estudo de ocorrência de quatro tipos de AVC - Universidade do Lisboa*. Lisboa, Portugal;
4. Casella, G., Berger, R. (2001). *Statistical Inference*. Duxbury Resource Center;
5. Chavez, J. H., Campana, S. G. e Haas, P. (2003). *Panorama da hepatite B no Brasil e no Estado de Santa Catarina*. Rev Panam Salud Publica, vol.14. Brasil.
6. Cox, D. R., e Snell, E. J. (1989). *Analysis of binary data*. Routledge.
7. Cunha L *et al.* (2004). *Use of replacement blood donors to study the epidemiology of major blood-borne viruses in the general population of Maputo, Mozambique*;
8. Dias at al. (2014). *Factores associados à infecção pelo vírus da hepatite B: um estudo caso-controlado no município de São Mateus, Espírito Santo*. Epidemiologia e Serviços de Saúde;
9. Divisão de Assistência à Saúde do Estudante. (2023). *Boletim epidemiológico. Hepatites virais*. UNILAB. Brasil;
10. Fávero, L. P., Belfiore, P., Silva, F. L. e Chan, B. L. (2009). *A Análise de Dados- Modelagem Multivariada para Tomada de Decisões*. Rio de Janeiro. Elsevier;
11. Feleke, B. E. (2016). *Low Coverage of Hepatitis B Vaccine and Determinants Among Health Professionals Working in Amhara Regional State Hospitals*. Ethiopia; J Public Health Afr.;
12. Filho B. e Bogliolo G. (2016). *Patologia*. 9. ed. Rio de Janeiro: Guanabara Koogan, Brasil.
13. Filipe, C. A., *et al.*, (2019). *Acidentes de trabalho e o uso de equipamentos de proteção individuais pelos profissionais de saúde em um Hospital Terciário*. Rev Med UFC, 2019;
14. Gonzalez, L.A. (2018). *Regressão Logística e Suas Aplicações*. Brasil, São Luis.
15. Guimarães, P.R.B. (2011). *Testes para comparação de k n amostras relacionadas*. Material didático;

16. Hair Jr., J. F., *et al.*, (2006). *Análise Multivariada de dados*. 5ª edição. Porto Alegre, Bookman.
17. Heeringa, S. G., West, B. T. e Berglund, P. A. (2010). *Applied Survey Data analysis*. Boca Raton, FL: Chapman & Hall CRC.
18. Hosmer, D. W., Lemeshow, S. (2000). *Applied Logistic Regression*. 2nd ed, New York: John Wiley Sons;
19. Hosmer, D. W., Lemeshow, S. (2013). *Applied Logistic Regression*. Wiley;
20. Lopes, C. L. R. *et al.*, (200). *Perfil soropidemiológico da infecção pelo vírus da hepatite B em profissionais das unidades de hemodiálise de Goiânia-Goiás, Brasil Central*. Rev. Soc. Bras. Med. Trop. vol.34;
21. Kiire C.F. (1996). *The Epidemiology and Prophylaxis of Hepatitis B in Sub-Saharan Africa: a view from tropical and subtropical Africa*, Gut, 38 (suppl 2), Zimbabwe;
22. Luz, J. A. *et al.* (2001). *Soroprevalência das infecções pelos vírus das hepatites b e c em profissionais de hemodiálise do tocantins*. Revista de Patologia Tropical.
23. Marôco, J. (2018). *Análise estatística com o spss statistics*.: 7ªedição, ReportNumber, Lda;
24. Marôco, J. (2011). *Análise estatística com o spss statistics*. 5ªPêro Pinheiro:ReportNumber;
25. Martins, A. C. P. *et al.*, (2013). *A Transmissão do VHB, do VHC e do VIH aos Profissionais de Saúde*. FMUC;
26. Mesquita, P. S. B. (2014). *Um Modelo de Regressão Logística para Avaliação dos Programas de Pós-Graduação no Brasil*. Rio de Janeiro. Brasil;
27. Mestre, P. (2009). *Aplicação dos Modelos Lineares Generalizados às Telecomunicações Móveis: Caracterização dos clientes que desactivam os seus serviços*. Lisboa, Portugal;
28. MISAU (2007). *Guia de vigilância epidemiológica*. 2ª ed, Brasília: MS;
29. MISAU. (2007). *Hepatites Virais*. Editora MS. Brasília;
30. Morais, J. (2008). *Efeito do plano amostral em modelo logístico ordinal: uma análise do estado de saúde autorreferido de adultos no Brasil usando a Pesquisa Nacional por Amostra de Domicílios de 2008*. Minas Gerais, Brasil;
31. Morais, L.Q. *et al.*, (2016). *Hepatite B em profissionais de enfermagem: prevalência e factores ocupacionais de risco*. UERJ. Rio de Janeiro, Brasil;
32. Nishide V. M. *et al.*, (2004). *Ocorrência de Acidente do Trabalho em uma Unidade de Terapia Intensiva*. Rev. Latino-am Enfermagem;

33. OMS. (2015). *Prevenção e controle de Infecções Virais Hepáticas: Quadro de Acção Global*, Geneva. Suíça;
34. OMS Directoria Executiva (2009). *Hepatites Virais. Relatório do Secretariado*. ED 126/15, [http://apps.OMS.int/gb/ebwha/pdf\\_files/EB126/B126\\_15-en.pdf](http://apps.OMS.int/gb/ebwha/pdf_files/EB126/B126_15-en.pdf);
35. OMS (2004). *Vacinas de Hepatite B.Registo Epidemiológico Semanal*.
36. Ozsoy M. F. et al., (2003). *A Seroprevalence of Hepatitis B and C Among Health Care Workers in Turkey*. J. Viral Hepatitis;
37. Paiva, E. M. M., et al., (2005). *Marcadores sorológicos e fatores de risco relacionados ao vírus da hepatite B em dentistas da região Centro Oeste do Brasil*. vol.39. Brazil;
38. Paiva, E. M. M. (2008). *Seroprevalência da infecção pelo vírus da hepatite B e avaliação da imunidade vacinal em cirurgiões-dentistas de Goiânia-GO*. Tese (Doutorado em Ciências da Saúde) - Convênio Rede Centro-Oeste, UnB/UFG/UFMS. Goiânia, Brasil.
39. Pereira, S. R. C. G. (2010). *Co-Infecção por Vírus da Hepatite B e Vírus da Imunodeficiência Humana em Países Africanos*, Tese de Mestrado, Lisboa;
40. Pereira, S.S. (2017). *Análise de Dados Amostrais Complexos Utilizando Redes Neurais*. Tese de Doutorado, Rio de Janeiro;
41. Pessoa, G., e Silva, P. (1998). *Análise de Dados Amostrais Complexos*.
42. Perz JF et al., (2006). *The Contributions of Hepatitis B Virus and Hepatitis C Virus Infections to Cirrhosis and Primary Liver Cancer Worldwide*. Journal of Hepatology;
43. Pondé, R. A. A. (2016). *Acute hepatitis B virus infection or acute exacerbation of chronic hepatitis B infection: the differential serological diagnosis*. European Journal of Clinical Microbiology & Infectious Diseases;
44. Rao, J. e Scott, A. (1984). *Chi-squared test for multiway contingency tables with cell proportions estimated from survey data*, Annals of Statistics;
45. Rodrigues, C. S. (2015). *Análise categórica para dados de HIV: Enfoque teórico e computacional no R-Project*. Brasil. São Cristovão;
46. Sanches, et al., (2006). *Caracterização soroepidemiológica da infecção pelo vírus da hepatite B em profissionais de saúde da atenção básica no Estado de Mato Grosso do Sul, Brasil*. Rev Panam Infectol;
47. Sharma, S. (1996). *Applied multivariate techniques*. New York: Wiley;

48. Silva, J.R., Alves F. (2013). *O Risco dos Profissionais de Saúde à Infecção por Hepatites B e C*. Brasil;
49. Silva, L.C.M., (2019). *Hepatites B e C e Infecção pelo HIV em Profissionais de Embelezamento, Estética e Tatuagem: Prevalência e Fatores de Risco*. PUC- Goiás. Brasil;
50. Silva, A., (2011). *Análise Estatística de Inquéritos online*. Rio Grande do Norte. Brasil.
51. Silva, A. C. L. G. et al., (2013). *Incidência e mortalidade por hepatite B, de 2001 a 2009: uma comparação entre o Brasil, Santa Catarina e Florianópolis*. Cadernos Saúde coletiva. Vol. 21. Rio de Janeiro, Brasil.
52. Silva, T. R. et al., (2010). *Acidente com material perfurocortante entre profissionais de enfermagem de um hospital universitário*. Rev. Gaúcha Enferm. vol.31.
53. Turkman, A. e Silva, G. (2000). *Modelos Lineares Generalizados - da teoria à prática*. Lisboa, Portugal;
54. Vallezi, D. R. R. e Zanusso, J. G., (2011). *Diagnóstico laboratorial da Hepatite B*. Rev. Uningá. vol.1;
55. Vicente, P. et al., (2001). *Sondagens. A Amostragem como Factor Decisivo de Qualidade*. 2 ed. rev. e corrigida. Lisboa: Sílabo;
56. Williams, J. S. e Iyer, H. K. (1985). *Scientific Theories, Observational Studies and Statistical Methods: A Rejoinder to AS Heath*. Department of Statistics, Colorado State University.

## Anexos

Tabela 5.1: Modelo Logístico incluindo todas as variáveis

Variáveis	Categorias	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Consumo de Álcool (Não)	Não consome (ref.)								
	Consome	0.203	0.382	0.282	1.000	0.595	1.225	0.579	2.592
Consumo de Drogas como	Não consome (ref.)								
	Cannabis e Cigarros	-0.823	1.066	0.595	1.000	0.440	0.439	0.054	3.550
Consumo de Drogas por Via	Não consome (ref.)								
	Endovenosa	-18.066	27121.642	0.000	1.000	0.999	0.000	0.000	
Hábito de Manter Relações	Não tem hábito (ref.)								
	Sexuais Desprotegidas	0.030	0.422	0.005	1.000	0.943	1.031	0.450	2.360
Uso Habitual de Protecção no	Não usa (ref.)								
	Acto Sexual	-0.348	0.418	0.694	1.000	0.405	0.706	0.312	1.601
Vários Episódios de Sexo	Não (ref.)								
	Ocasional	-0.168	0.579	0.084	1.000	0.771	0.845	0.272	2.629
Prática de Sexo em Grupo	Não pratica (ref.)								
	Antecedentes de ITS	-18.267	13189.503	0.000	1.000	0.999	0.000	0.000	
Troca de Sexo por Bens e	Não tem antecedentes	-0.038	0.453	0.007	1.000	0.933	0.963	0.396	2.341
	Favores	-18.737	19536.806	0.000	1.000	0.999	0.000	0.000	
Transfusão Sanguínea ou	Nunca recebeu (ref.)								
	Derivados	0.431	0.542	0.632	1.000	0.426	1.539	0.532	4.453
Antecedentes de Cirurgias	(ref.)Não tem antecedentes								
	Tem antecedentes	-0.056	0.422	0.017	1.000	0.895	0.946	0.413	2.163
Ter Piercing ou Tatuagens	Não tem (ref.)								
	Tem	-0.962	0.783	1.510	1.000	0.219	0.382	0.082	1.772
Partilhas de Escovas e	Não partilha (ref.)								
	Máquinas de Barbear	0.809	0.428	3.571	1.000	0.059	2.245	0.970	5.196
Acidente Ocupacional	Nunca teve (ref.)								
	Vacina Contra Hepatite B	0.941	0.375	6.305	1.000	0.012	2.561	1.229	5.337
Idade dos Participantes	Não (ref.)								
	Sim	0.389	0.414	0.883	1.000	0.347	1.476	0.655	3.326
Departamento	18-29 anos (ref.)								
	30-59 anos	0.219	0.506	0.187	1.000	0.911	1.244	0.462	3.355
Sexo dos Participantes	60+	-18.433	9673.812	0.000	1.000	0.998	0.000	0.000	
	Administracção (ref.)								
Estado Civil	Cirurgia	-0.373	0.644	0.335	1.000	0.822	0.689	0.195	2.435
	Clínica Especial	0.026	0.914	0.001	1.000	0.563	1.026	0.171	6.153
Sexo dos Participantes	Farmácia	1.949	1.538	1.607	1.000	0.977	7.023	0.345	143.076
	Laboratório	0.278	0.963	0.084	1.000	0.205	1.321	0.200	8.721
Sexo dos Participantes	Medicina	0.055	0.605	0.008	1.000	0.772	1.057	0.323	3.457
	Obstetrícia	-0.713	1.178	0.366	1.000	0.927	1.057	0.490	4.936
Sexo dos Participantes	Pediatria	0.029	0.828	0.001	1.000	0.545	1.030	0.203	5.219
	SUR	-0.572	0.690	0.686	1.000	0.972	1.030	0.203	5.219
Sexo dos Participantes	Masculino (ref.)								
	Feminino	-0.264	0.431	0.375	1.000	0.407	0.564	0.146	2.184
Sexo dos Participantes	Solteiro(a) (ref.)								
	Casado(a)	-0.572	0.414	2.223	3.000	0.540	0.768	0.330	1.787
Estado Civil	Divorciado (a)	-18.284	13274.924	0.000	1.000	0.528	0.565	0.251	1.272
	Viúvo (a)	0.448	1.167	0.147	1.000	0.168	1.565	0.159	15.413
	(Intercepto)	-2.804	0.847	10.965	1.000	0.001	0.061		

**Tabela 5.2: Teste de razão de verossimilhança de Rao-Scott - Modelo saturado**

<b>Variáveis</b>	<b>p-valor</b>
Idade dos Participantes	0.486
Sexo dos Participantes	0.550
Estado Civil	0.457
Consumo de Álcool	0.575
Consumo de Drogas como Cannabis e Cigarros	0.425
Consumo de Drogas por Via Endovenosa	0.697
Uso Habitual de Protecção no Acto Sexual	0.448
Vários Episódios de Sexo Ocasional	0.844
Prática de Sexo em Grupo	0.398
Antecedentes de ITS	0.947
Troca de Sexo por Bens e Favores	0.577
Transfusão Sanguínea ou Derivados	0.293
Antecedentes de Cirurgias	0.719
Ter Piercing ou Tatuagens	0.218
Partilhas de Escovas e Máquinas de Barbear	0.004
Acidente Ocupacional	0.013
Vacina Contra Hepatite B	0.184
Departamento	0.767
Hábito de Manter Relações Sexuais Desprotegidas	0.857

$P < 0.05$