



UNIVERSIDADE ESTADUAL DA PARAÍBA  
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA EM SAÚDE

**MODELOS MULTIVARIADOS DE ANÁLISE DISCRIMINANTE COMO  
FERRAMENTA DE PREVISÃO DE EPIDEMIAS DE DENGUE**

**DIOGO MEDEIROS NÓBREGA**

CAMPINA GRANDE - PB

2019

DIOGO MEDEIROS NÓBREGA

**MODELOS MULTIVARIADOS DE ANÁLISE DISCRIMINANTE COMO  
FERRAMENTA DE PREVISÃO DE EPIDEMIAS DE DENGUE**

Dissertação apresentada ao Programa de Pós-graduação em Ciência e Tecnologia em Saúde da Universidade Estadual da Paraíba como parte dos requisitos para obtenção do título de Mestre em Ciência e Tecnologia em Saúde.

**Orientador:** Dr. Wellington Candeia de Araújo

CAMPINA GRANDE - PB

2019

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

N337m Nóbrega, Diogo Medeiros.

Modelos multivariados de análise discriminante como ferramenta de previsão de epidemias de dengue [manuscrito] / Diogo Medeiros Nóbrega. - 2019.

47 p. : il. colorido.

Digitado.

Dissertação (Mestrado em Profissional em Ciência e Tecnologia em Saúde) - Universidade Estadual da Paraíba, Pró-Reitoria de Pós-Graduação e Pesquisa, 2019.

"Orientação : Prof. Dr. Wellington Candeia de Araújo, Departamento de Computação - CCT."

1. Análise discriminante. 2. Dengue. 3. Saneamento básico. 4. Software SPSS. I. Título

21. ed. CDD 600

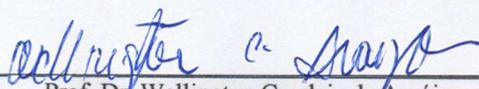
DIOGO MEDEIROS NÓBREGA

**MODELOS MULTIVARIADOS DE ANÁLISE DISCRIMINANTE COMO FERRAMENTA DE  
PREVISÃO DE EPIDEMIAS DE DENGUE**

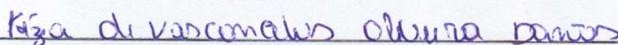
Dissertação apresentada ao Programa de Pós-graduação em  
Ciência e Tecnologia em Saúde da Universidade Estadual da  
Paraíba como parte dos requisitos para obtenção do título de  
Mestre em Ciência e Tecnologia em Saúde.

**Aprovado em 04 de junho de 2019.**

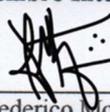
**BANCA EXAMINADORA**



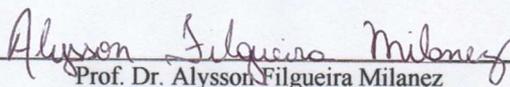
Prof. Dr. Wellington Candeia de Araújo  
Universidade Estadual da Paraíba (UEPB)  
**Orientador**



Prof. Dra. Kézia de Vasconcelos Oliveira Dantas  
Universidade Estadual da Paraíba (UEPB)  
**Membro Interno**



Prof. Dr. Frederico Moreira Bublitz  
Universidade Estadual da Paraíba (UEPB)  
**Membro Interno**



Prof. Dr. Alysson Filgueira Milanez  
Instituto Federal de Educação, Ciência e Tecnologia da Paraíba (IFPB)  
**Membro Externo**

## RESUMO

Esta pesquisa trata da utilização da Análise Discriminante e todos os seus pressupostos a partir de variáveis referentes a ações de saneamento básico, relacionando-as à incidência de casos de dengue em todos os Estados Brasileiros no período de 2013 a 2017 para modelar uma função discriminante de previsão capaz de classificar os Estados em grupos de baixo, médio e alto nível de incidência de dengue. Para as análises utilizou-se o software SPSS. Epidemias geram altos custos para os serviços de saúde, e a infraestrutura precária dos serviços públicos de saneamento básico podem contribuir diretamente para o crescimento nos casos de dengue. O uso de métodos estatísticos torna os processos de investigação mais diretos e apurados, o que possibilita perceber circunstâncias preocupantes e, dessa forma, se antecipar e agir para seu controle. Neste trabalho, são descritos o desenvolvimento teórico estatístico e as análises realizadas. As variáveis foram extraídas do boletim epidemiológico do Ministério da Saúde e da série histórica de ações de saneamento básico do Ministério das Cidades dos últimos 05 anos disponíveis. Identificou-se que as variáveis Índice de Coleta de Esgoto (ICE), Quantidade de Municípios Atendidos com Esgotamento Sanitário (QMAES), Despesa Total com Saneamento incluindo água e esgoto (DTS) e Quantidade de Municípios Atendidos com Abastecimento de Água (QMAAA) foram as mais estatisticamente significativas. Ao fim constatou-se que a separação entre os grupos de baixa, média e alta incidência é estatisticamente significativa de acordo com os testes, tendo a função discriminante alcançado um nível de acerto de 86,9% dos casos originais de Alta Incidência nos seus respectivos grupos.

**Palavras chave:** Análise Discriminante, Saneamento Básico, Dengue, Classificação, Epidemia.

## ABSTRACT

This research deals with the use of Discriminant Analysis and all its assumptions from variables related to basic sanitation actions, linking them to the incidence of dengue cases in the Brazilian States from 2013 to 2017 to model a discriminant prediction function capable of classifying the states into low-, medium- and high-level groups of dengue incidence. For the analyses, SPSS software was used. Epidemics generate high costs for health services, and the precarious infrastructure of public sanitation may directly contribute to growth in dengue cases. The use of statistical methods makes the investigation processes more direct and accurate, which makes it possible to perceive worrying circumstances and, thus, to anticipate and act for its control. In this work, the theoretical statistical development and the analyses are described. The features were extracted from epidemiological bulletin of the Ministry of Health and the time series of basic sanitation actions of the Ministry of Cities of the last five years available. It was identified that the variables Sewage Collection Index (ICE), Quantity of Municipalities Served with Sanitary Sewage (QMAES), Total Expenditure with Sanitation including water and sewage (DTS) and Quantity of Municipalities Served with Water Supply (QMAAA) were the most statistically significant. Finally, it was found that the separation between the low, medium and high incidence groups was statistically significant according to the tests, and the discriminant function reached a level of accuracy of 86.9% of the original High Incidence cases in their respective groups.

**keywords:** Discriminant Analysis, Basic Sanitation, Dengue, Classification, Epidemic.

# SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>07</b>
<b>2 OBJETIVOS.....</b>	<b>09</b>
2.1 OBJETIVO GERAL.....	09
2.2 OBJETIVOS ESPECÍFICOS.....	09
<b>3 REVISÃO DA LITERATURA.....</b>	<b>10</b>
3.1 A DENGUE NO BRASIL .....	10
3.2 ANÁLISE DISCRIMINANTE.....	11
3.2.1 Modelagem da Análise Discriminante.....	14
3.2.2 Método Stepwise.....	15
3.2.3 Função Discriminante.....	15
3.2.4 Estatísticas Associadas à Análise Discriminante.....	16
3.2.5 Lambda de Wilks.....	18
3.2.6 Teste Box's M (M de Box) .....	18
3.2.7 Teste de Qui-quadrado.....	19
3.3 ANÁLISE DISCRIMINANTE RELACIONADA AOS CASOS DE DENGUE.....	19
<b>4 MATERIAL E MÉTODO.....</b>	<b>24</b>
4.1 LOCAL DO ESTUDO.....	24
4.2 DELINEAMENTO DA PESQUISA.....	24
4.3 COLETA DE DADOS.....	25
4.4 VARIÁVEIS UTILIZADAS.....	26
<b>5 RESULTADOS.....</b>	<b>27</b>
5.1 VERIFICAÇÃO DE NORMALIDADE DOS DADOS.....	27
5.2 SIGNIFICÂNCIA DAS VARIÁVEIS.....	31
5.3 IGUALDADE DAS MATRIZES DE COVARIÂNCIA DOS GRUPOS.....	33
5.4 MÉTODO STEPWISE.....	33
5.5 AUTOVALORES.....	35
5.6 FUNÇÃO DISCRIMINANTE.....	37
5.7 ALOCAÇÃO DE NOVAS OBSERVAÇÕES.....	37
<b>6 CONCLUSÃO.....</b>	<b>40</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>42</b>

## 1. INTRODUÇÃO

Entender e explicar a natureza dos fenômenos epidemiológicos e suas relações de causa é uma efetiva ferramenta de suporte na tomada de decisão, no desenvolvimento de políticas de combate aos surtos das doenças transmitidas pelo *Aedes aegypti* ou na revisão de programas já em execução para que estes se tornem ainda mais eficientes (VANDENBROUCKE 2016).

A infraestrutura precária dos serviços públicos estruturais como fornecimento de água, sistemas de esgoto e coleta de resíduos sólidos, podem contribuir diretamente para o crescimento nos casos de doenças relacionadas ao mosquito *Aedes aegypti* nos centros urbanos (RIBEIRO, 2010).

Segundo a Associação Brasileira de Engenharia Sanitária e Ambiental (ABES, 2016), apenas através de infraestrutura sanitária básica será possível alterar o quadro epidemiológico de avanço das doenças transmitidas pelo *Aedes aegypti*, cuja falta pode se tornar um fator crucial para sua proliferação, uma vez que o abastecimento intermitente de água leva a população a acumular água em recipientes muitas vezes mal tampados, além dos esgotos a céu aberto e o acúmulo de resíduos que possam reter água das chuvas.

Muitos problemas de pesquisas em epidemiologia exigem a análise e previsão de classificação: prever se uma cidade deve ou não ser classificada em um grupo com mais chances de registrar um surto de dengue nos próximos anos, se residir em zona urbana é fator de risco para a infecção por determinado vírus ou não, e assim por diante (COSTA 2003).

O uso de métodos estatísticos torna os processos de investigação mais diretos e apurados, o que possibilita perceber circunstâncias preocupantes e, dessa forma, agir para seu controle, exercendo papel fundamental no estudo do histórico, bem como da incidência de diversas doenças (IVES, 2010).

Os métodos estatísticos de “análise multivariada” e “análise multinomial” são frequentemente usados alternadamente na literatura (BAGLEY, BRANCO, & GOLOMB, 2001). Em sentido estrito, a análise multivariada busca prever simultaneamente múltiplos resultados enquanto a análise multinomial utiliza múltiplas variáveis de prever um resultado único (KATZ, 1999).

As técnicas multivariadas exploram a relação entre várias variáveis independentes (previsão) e uma variável dependente (resposta), gerando um modelo que descreve a relação expressa do valor predito da variável de saída como uma soma de produtos, cada um dos produtos formados pela multiplicação do valor e coeficiente da variável independente (COX, D.R; SNELL, E.J 1989). São obtidos coeficientes com o melhor ajuste matemático para o

modelo especificado. Os coeficientes indicam o impacto de todas as variáveis independentes sobre a variável resposta.

O modelo pode inferir o valor da variável dependente para novos números das variáveis independentes, como também, pode ajudar a mensurar a proporcional contribuição de cada uma das variáveis independentes para com a variável dependente, controlando-se mutuamente a influência das demais variáveis independentes no modelo. Os métodos multivariados utilizados com mais frequência nas ciências da saúde são a regressão linear, regressão logística, análise discriminante e regressão de risco proporcional (KATZ 1999).

Nesse trabalho a técnica utilizada é a análise discriminante, por permitir fazer associação de múltiplas variáveis distintas para gerar uma função de previsão de classificação capaz de alocar indivíduos estudados em grupos pré-estabelecidos conforme sejam observados novos valores das variáveis que compõem o modelo. Constatou-se que a separação entre os grupos de baixa, média e alta incidência é estatisticamente significativa de acordo com os testes, tendo a função discriminante alcançado um nível de acerto de 86,9% dos casos originais de Alta Incidência nos seus respectivos grupos.

## **2. OBJETIVOS**

Nesta Seção será apresentado os objetivos geral e específico dos quais tratam a pesquisa.

### **2.1. OBJETIVO GERAL**

O objetivo geral do estudo é construir uma Função Discriminante Multivariada efetiva para a previsão de classificação dos Estados Brasileiros em grupos de baixa, média ou alta incidência de casos de Dengue.

### **2.2. OBJETIVOS ESPECÍFICOS**

- Verificar a literatura existente e tomar como base trabalhos já desenvolvidos na comunidade científico-acadêmica;
- Fazer o levantamento dos dados que possam ser utilizados de forma adequada nas análises;
- Verificar os pressupostos da análise discriminante e identificar as variáveis com maior poder de discriminação;
- Modelar uma função discriminante que permita realizar a previsão de classificação dos Estados nos grupos de interesse;
- Avaliar o poder de acerto da função em relação aos dados originais.

### **3. REVISÃO DA LITERATURA**

Neste capítulo será abordada a questão da dengue no Brasil, bem como a fundamentação teórica da análise discriminante e trabalhos que abordem a técnica relacionada a casos de dengue.

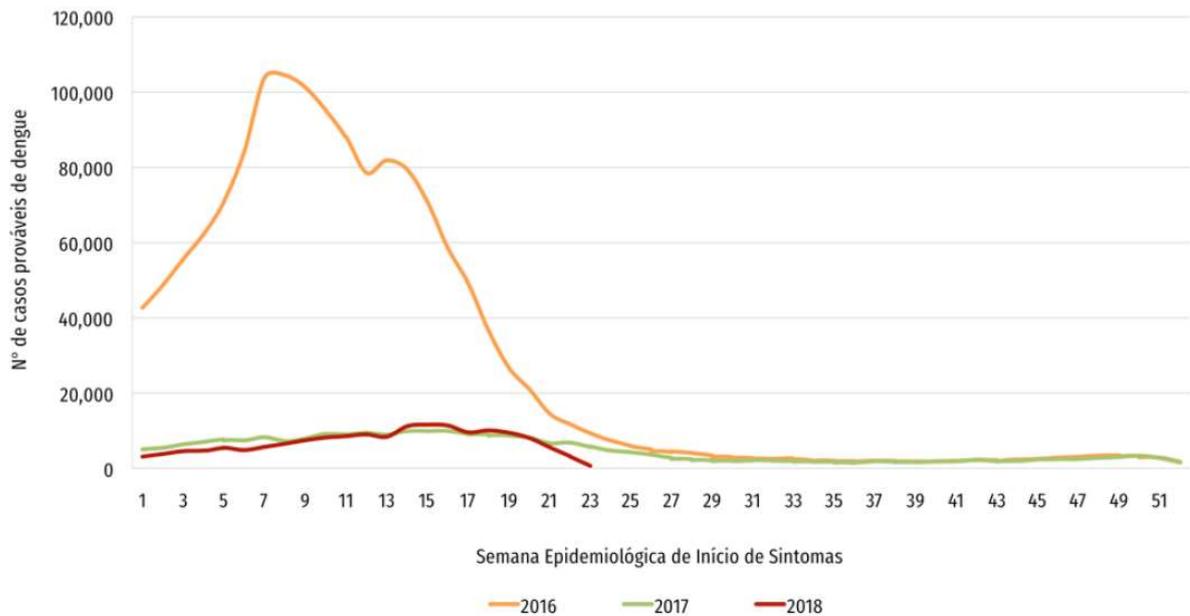
#### **3.1. A DENGUE NO BRASIL**

A dengue é uma das graves doenças infecciosas presentes no Brasil, representando um delicado problema de Saúde Pública no país e no mundo. Segundo a Organização Mundial da Saúde (OMS), atualmente a dengue continua sendo um problema de saúde pública nas Américas apesar dos esforços dos Estados Membros para deter e mitigar o impacto das epidemias. A dengue é considerada uma doença infecciosa sistêmica dinâmica (SILVEIRA, 2016).

A dengue tornou-se um problema de saúde pública, destacando-se no cenário epidemiológico mundial. Tanto países desenvolvidos quanto os em desenvolvimento (entre os quais encontra-se o Brasil) possuem índices de casos registrados da doença (TANNOUS, 2018).

Em queda no país desde o último ano, os casos notificados de dengue voltaram a crescer de acordo com boletins epidemiológicos do Ministério da Saúde. Em 2017, entre a semana 1 e a 52, foram registrados 239.389 casos prováveis de dengue, e em 2016, 1.483.623. Em 2018, até a semana 30 (31/12/2017 a 28/07/2018), foram registrados 187.830 casos prováveis de dengue no país, com uma incidência de 90,5 casos/100 mil habitantes, o que corresponde a 78,5% do número de casos de 2017. A Figura 01 apresenta a série histórica de casos de dengue entre 2016 e 2018.

Figura 01: Casos de Dengue ao longo das semanas (2016 à 2018)



Fonte: SINAN online

Dados do LIRAA (Levantamento Rápido de Índices de Infestação pelo *Aedes aegypti*) mostram que 1.153 municípios brasileiros (22%) apresentam alto índice de infestação do mosquito, com risco de surto para dengue, zika e chikungunya. O Ministério da Saúde adverte sobre a urgência de consolidar ações de combate ao *Aedes aegypti* em todo o país durante o ano inteiro, e afirma que a continuidade das ações é fundamental para manter baixos os índices dessas doenças no país.

Os impactos das epidemias de dengue são percebidos diretamente na assistência aos doentes e na vigilância epidemiológica, trazendo à tona sua vulnerabilidade e causando sérios efeitos de ordem socioeconômica, política e psicológica. Além disso, as epidemias de dengue geram altos custos com hospitalização, assistência médica e medidas de prevenção e controle do vetor, gerando uma sobrecarga para os serviços de saúde (OLIVEIRA, 2018).

### 3.2. ANÁLISE DISCRIMINANTE

A análise discriminante múltipla abrange um conjunto de variáveis preditoras e uma variável de critério (resposta), sendo a última uma variável de agrupamento com dois ou mais níveis, ou seja, dois ou mais grupos, dando origem a um conjunto de regras que consiste em combinações lineares tanto das variáveis preditoras quanto das categorias ou grupos (HUBERTY, 1989). A função discriminante formulada usa as pontuações (scores) nas variáveis preditoras para prever a categoria à qual o indivíduo pertence.

Para o caso de dois grupos, por exemplo, assim que os pesos preditores (scores) para duas combinações lineares (um associado a cada grupo) são determinados, duas probabilidades de associação ao grupo podem ser calculadas para observações subseqüentes com base nas duas combinações lineares, e essa nova observação é atribuída ao grupo com a maior pontuação de combinação linear. O foco da análise discriminante é a previsão e a precisão das taxas de acerto. Como Huberty (1989) destaca, a análise discriminante está basicamente interessada em determinar uma regra de classificação e avaliar sua precisão.

Os pré-requisitos básicos são que existam dois ou mais grupos e que supostamente suas variáveis difiram entre si, e que estas possam ser medidas no intervalo ou nível de razão. A análise discriminante permite analisar as diferenças entre os grupos e nos fornece um meio de atribuir (classificar) qualquer caso ao grupo em que mais se assemelhe.

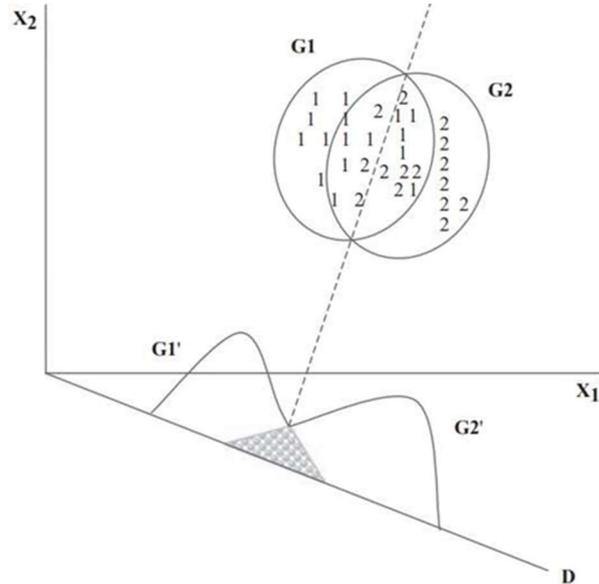
A análise discriminante é aplicável quando a variável dependente ou o critério é categórico e as variáveis preditivas ou independentes são de natureza intervalar, ou seja, contínuas. Os objetivos da análise discriminante são os seguintes:

1. Desenvolver funções discriminantes ou combinações lineares de variáveis preditivas ou independentes, que fazem uma melhor diferenciação entre as categorias das variáveis dependentes ou critérios (grupos). O número de funções equivale ao número de grupos menos 1 (ex.: 3 grupos = 2 funções);
2. Examinar se existem diferenças significativas entre os grupos, em termos das variáveis preditivas;
3. Determinar quais variáveis preditivas contribuem mais para as diferenças entre os grupos;
4. Classificar os casos em um dos grupos com base nos valores das variáveis preditivas;
5. Calcular o percentual de acerto da função.

As técnicas de análise discriminante são nomeadas de acordo com o número de categorias que a variável resposta possui. No caso de apenas duas categorias, a técnica é conhecida como análise discriminante simples. Quando existem três ou mais categorias, a técnica é conhecida como análise discriminante múltipla. A principal diferença é que, no caso de dois grupos, só pode ser modelada uma única função discriminante; enquanto na análise múltipla discriminante, o número de funções é dado pela quantidade de grupos menos um. Se há 3 grupos pode-se modelar duas funções discriminantes e avaliar qual delas serve melhor para o problema através de testes de significância. A figura 02 traz dois grupos, G1 e G2, onde cada

membro foi medido na escala de duas variáveis  $X_1$  e  $X_2$  e apresenta ainda um diagrama de dispersão, onde  $X_1$  e  $X_2$  são os dois eixos.

Figura 02: Interpretação geométrica da análise discriminante



Fonte: Naresh (2008)

O número 1 identifica as amostras do  $G_1$  e o número 2 identifica as amostras do  $G_2$ . As elipses resultantes abrangem uma quantidade específica de observações. Uma linha reta é traçada entre os dois grupos, e a partir daí é projetada para um novo eixo  $D$ . A sobreposição entre as distribuições de probabilidade univariadas  $G_1'$  e  $G_2'$  é representada pela área sombreada na figura (NARESH 2008).

A análise discriminante é um método estatístico que pode ser utilizado para inferir a classificação de unidades amostrais ou populacionais. Para que esta técnica seja aplicada, é fundamental que os grupos nos quais cada observação amostral pode ser classificada sejam definidos a priori, considerando-se suas particularidades. Isto dá o subsídio necessário para a estruturação de um modelo matemático para tomada de decisão chamado regra de classificação ou função discriminante, que é aplicada na classificação de novas observações nos grupos já existentes.

Uma vez que todo instrumento de tomada de decisões vem acompanhado de um provável erro de decisão, o propósito, portanto, é o de compor uma regra que minimize as classificações equivocadas, ou seja, o erro de afirmar que uma amostra pertence a uma população quando, na verdade, ele cabe à outra. Além disso, é importante construir uma regra que proporcionalmente minimize a classificação incorreta.

A regra de decisão da análise discriminante é aplicada através de testes de hipóteses, onde observa-se uma amostra e, a partir dos resultados obtidos, calcula-se o valor da estatística do teste, que é uma regra de diferenciação entre hipótese nula e alternativa. A depender do valor resultante da estatística de teste, decide-se pela aceitação, ou não, da hipótese nula. A constante delimitadora da região de não aceitação da hipótese nula é dada pela determinação do nível de significância do teste, isto é, a probabilidade de não se aceitar a hipótese nula em detrimento da alternativa, uma vez que a hipótese nula é verdadeira. Busca-se o teste que minimize a chance do chamado erro do tipo 2 (HOLM, 1979), que significa não se rejeitar a hipótese nula quando esta deveria ser de fato rejeitada.

### 3.2.1 Modelagem da Análise Discriminante

Antes de elaborar uma função discriminante, é oportuno que sejam verificados os pressupostos associados a este método. Existem dois pressupostos fundamentais, que se referem à presença de normalidade multivariada e a homogeneidade das matrizes de variância e covariância. Sobre a primeira suposição, a combinação linear das variáveis explicativas tem distribuição normal e, caso esse pressuposto não seja satisfeito, a avaliação do pesquisador corre o risco de ser enviesada. No entanto, se ela acontecer exclusivamente pela presença de assimetria da distribuição, a decisão sobre o uso do método não será alterada (SHARMA, 1996).

O segundo pressuposto diz respeito à homogeneidade das matrizes de variância e covariância. Esse requisito é examinado através da estatística M de Box (Box's M), que pode ser afetado pelo não atendimento do pressuposto da distribuição normal multivariada. Por felicidade, a análise discriminante é um método muito robusto quanto à violação destes pressupostos, uma vez que o tamanho do grupo seja maior que o número de variáveis em análise e que as médias dos grupos não sejam proporcionais às variâncias, isto é, caso a homogeneidade das matrizes não seja atendida, acarretará em aumento da probabilidade de classificação das amostras no grupo com maior grau de dispersão.

Além desses pressupostos é relevante destacar que a ausência de *outliers*, a existência de linearidade das relações e a falta de problemas conexos à multicolinearidade das variáveis explicativas também são consideradas pressupostos inerentes à análise discriminante.

É fundamental que se estabeleça corretamente a amostra a ser analisada, visto que esta técnica é sensível à proporção amostral relacionada ao número de variáveis preditoras e, dito isto, deve-se evitar grandes variabilidades de dimensões entre os grupos (HAIR, ANDERSON, TATHAM E BLACK, 2005).

Passados os pressupostos, pode-se partir para a composição propriamente dita das funções discriminantes, uma vez que isso caracteriza um dos objetivos primordiais da análise discriminante, através das quais as amostras serão discriminadas. Antes de iniciar essa etapa, é importante que o pesquisador já tenha delimitado o problema central da sua pesquisa e tenha bem definido pelo menos dois grupos de interesse para classificação.

Esta etapa consiste basicamente na escolha da variável categórica (dependente) e das variáveis numéricas (independentes), ou seja, a seleção de  $n$  variáveis discriminantes (explicativas) é realizada partindo de um conjunto maior de  $p$  possíveis variáveis.

A análise discriminante proporciona informações sobre as variáveis que mais se sobressaem na diferenciação dos grupos. Para tal, diversos resultados são gerados partindo de estatísticas e testes como o lambda de Wilks, a correlação canônica, o Qui-quadrado ou o autovalor.

### 3.2.2 Método Stepwise

Um pesquisador deve tomar decisões sobre quais variáveis estarão envolvidas em uma análise. Comumente, o pesquisador pode querer selecionar um subconjunto de variáveis da população de origem; ou mensurar a relativa importância do conjunto de variáveis, ainda que nenhuma variável seja eliminada. Alguns pesquisadores acreditam que métodos graduais podem ser usados para realizar qualquer uma dessas tarefas (HUBERTY, 1989).

Hawkins (1976) propôs que uma variável só seja incluída se esta for significativa ao nível  $\alpha / (k - p)$ , onde  $\alpha$  é o nível desejado de significância,  $p$  é o número de variáveis já inseridas e  $(k - p)$  é o número de variáveis disponíveis para serem incluídas.

### 3.2.3 Função Discriminante

O modelo de análise discriminante envolve um conjunto de combinações lineares e sua função pode ser escrita conforme a equação 1:

$$Z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (1.0)$$

Onde:

Z: variável dependente;

$\alpha$ : constante (intercepto);

$X_i$ : variáveis explicativas;

$\beta_i$ : coeficientes discriminantes para cada variável explicativa.

Segundo Naresh (2008), os coeficientes discriminantes são calculados de modo a fazer com que o grupo difira ao máximo nos valores da função discriminante. Isso ocorre quando a razão de soma de quadrados está no máximo entre grupos e intragrupos nos pesos discriminantes. Qualquer outra combinação linear dos preditores resultará em uma proporção menor de diferenciação.

Existem dois métodos gerais disponíveis para fins de cálculo dos coeficientes da função discriminante. O método direto envolve o cálculo da função discriminante, de forma que todas as variáveis independentes sejam incluídas ao mesmo tempo; e o método Stepwise (passo a passo), onde as variáveis preditivas são introduzidas em sequência com base no seu poder de discriminação entre os grupos (NARESH 2008).

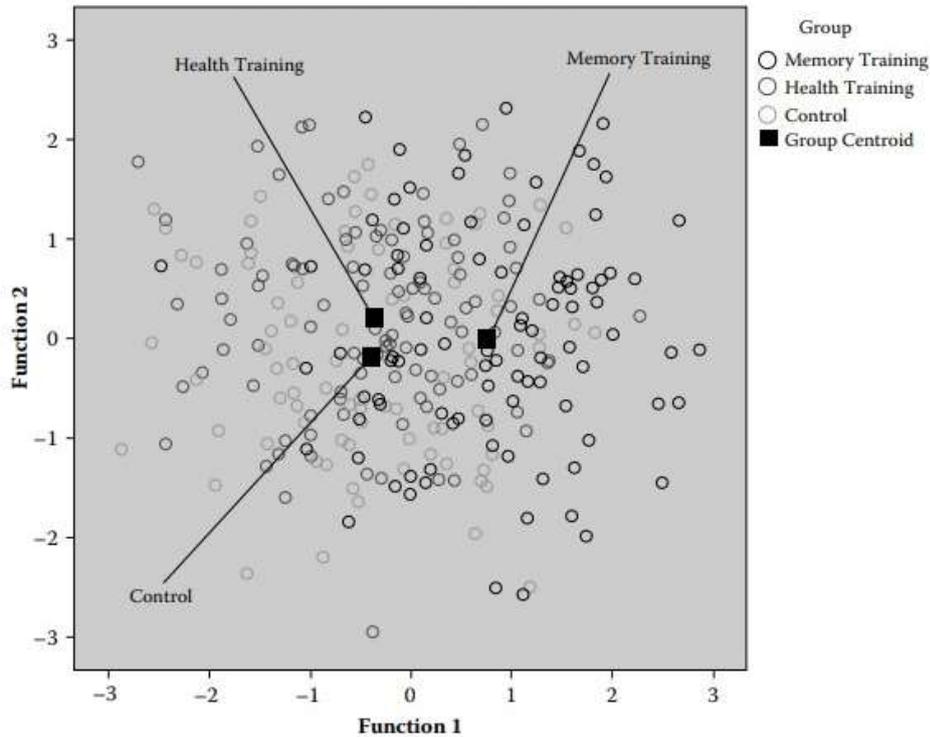
### 3.2.4 Estatísticas Associadas à Análise Discriminante

Naresh (2008) define as principais estatísticas associadas à análise discriminante como sendo:

**Correlação Canônica:** corresponde à razão entre a variação entre os grupos e a variação total e mede a associação entre a função discriminante e as variáveis dependentes. O quadrado do coeficiente de correlação canônica mostra a proporção de variância que consegue ser explicada pelas variáveis;

**Centroide:** o centroide é a média dos scores discriminantes de um grupo específico, sendo um por grupo. Os centroides são os pontos médios de cada grupo e servem para comparação com o valor resultante da função discriminante para fins de classificação à posteriori. A figura 03 ilustra os centroides em relação aos escores discriminantes:

Figura 03: Centroides dos grupos (quadrados) e escores da função discriminante (círculos)



Fonte: Stevens (2002)

**Matriz de classificação:** contém o número de casos cuja classificação foi correta e incorreta. Casos classificados corretamente aparecem na diagonal porque os grupos real e previsto são os mesmos. Os elementos fora da diagonal representam casos cuja classificação foi falha. A soma dos elementos da diagonal, dividida pelo número total de casos, representa a proporção de respostas corretas;

**Coefficientes da função discriminante:** os coeficientes da função discriminante são os multiplicadores das variáveis;

**Autovalores:** para cada função discriminante, o autovalor é a razão entre a soma dos quadrados entre grupos e intragrupos. Grandes autovalores supõem funções com maior poder discriminante;

**Médias e desvios padrão:** são calculados para cada variável em cada grupo;

**Matriz de correlações intragrupo:** para calcular a matriz de correlações agrupadas intragrupo, as médias das matrizes de covariância separadas de todos os grupos são calculadas;

**Matriz de correlação total:** os casos são tratados como se pertencessem a uma única amostra e as correlações são calculadas, obtendo-se uma matriz de correlação total;

Procedimentos inferenciais como testes de hipóteses, significância de discriminação, homogeneidade de grupos e significância da correlação canônica também fazem parte da

aplicação prática da análise discriminante (ALHASSAN, 2010). Os principais testes são listados na sequência.

### 3.2.5 Lambda de Wilks

Na análise discriminante, o lambda de Wilk testa quão bem cada variável independente contribui para o modelo. Seu valor flutua entre 0 e 1, onde 0 significa discriminação total e 1 significa não discriminação (BROWN, 2000). Cada variável independente é testada colocando-a e retirando-a do modelo, gerando uma estatística  $\lambda$ . A significância desta estatística é medida através de um teste F, e uma vez que o valor F for maior que o valor crítico, a variável será usada para compor a função discriminante, pois é estatisticamente significativa para fins de previsão de classificação. Essa estatística é definida conforme escrito na equação 2:

$$\lambda = \frac{|W|}{|T|} \quad (2.0)$$

Onde, W é a soma de quadrados intragrupos e T a soma dos quadrados totais.

### 3.2.6 Teste Box's M (M de Box)

Apresentado por Box (1949), este é um teste estatístico multivariado usado para verificar a igualdade de matrizes de variância-covariância múltiplas, sendo comumente usado para testar a suposição de homogeneidade de variâncias e covariâncias. As hipóteses a serem investigadas nesse teste são:

$$\begin{cases} H_0 = \Sigma_1 = \Sigma_2 = \dots = \Sigma_k \\ H_1 = \Sigma_1 \neq \Sigma_2 \neq \dots \neq \Sigma_k \end{cases}$$

Sua estatística de teste é calculada a partir da equação 3:

$$M = (n - g) \ln|S| - \sum_{k=1}^g v_k \ln|S_k| \quad (3.0)$$

Sendo  $n$  a dimensão total da amostra,  $v_k = n_k - 1$  os graus de liberdade associados a cada grupo,  $S_k$  a matriz de variâncias e covariâncias do grupo  $k$ ,  $k = 1, 2, \dots, g$ , e  $S = \frac{W}{n-g}$  a

matriz de variâncias e covariâncias geral, onde  $g$  é o número de populações, e  $W$  é a matriz de soma de quadrados e produtos cruzados totais do resíduo.

Box sugere ainda que no teste de homogeneidade de matrizes pode ser usado tanto a aproximação à distribuição de Qui-quadrado, como à distribuição  $F$  para fins de teste das hipóteses.

### 3.2.7 Teste de Qui-quadrado

Desenvolvido por Bartlett (1947), esse teste é usado na análise discriminante para testar a significância das correlações canônicas e sua estatística de referência é definida pela equação 4 para o caso de 2 graus de liberdade:

$$\chi^2 = -[n - 0,5(p_1 + p_2 + 1)] \log_e \lambda \quad (4.0)$$

O resultado da estatística  $\chi^2$  então será comparado ao respectivo valor crítico ( $\alpha$ ) para que se decida pela aceitação ou não da hipótese nula.

### 3.3. ANÁLISE DISCRIMINANTE RELACIONADA AOS CASOS DE DENGUE

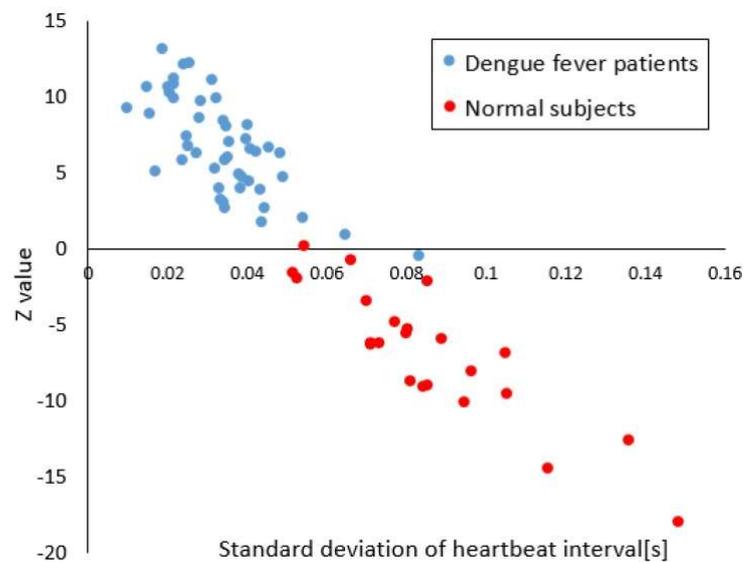
A compreensão dos fenômenos epidemiológicos requer a consideração simultânea das interações de múltiplas variáveis; e a extrema complexidade a ela atrelada, se torna inviável sem o auxílio da análise estatística multivariada (IVES, 2010). As estatísticas univariadas que examinam os efeitos de uma variável de cada vez são de pouca ajuda nessa empreitada, uma vez que fornecem um panorama limitado das interações dinâmicas que estão acontecendo simultaneamente.

A capacidade de entender procedimentos estatísticos e matemáticos de modelagem multivariada e aplicá-los efetivamente em problemas práticos, se mostra como sendo uma habilidade essencial no campo das ciências (BROWN 2000). O objetivo matemático da análise discriminante multivariada é ponderar e combinar linearmente as informações de um conjunto de variáveis dependentes de uma maneira que força os grupos a serem tão distintos quanto possível.

Yang et. al (2018) utilizou a análise discriminante linear para distinguir os pacientes infectados com dengue dos indivíduos não infectados. Neste estudo, foram utilizadas três

variáveis (HR, RR, e SDHI) para detectar se o indivíduo foi infectado ou não. A fim de verificar se estes três parâmetros são verdadeiramente significativos para diferenciar os indivíduos, foi realizado um teste-t. A partir disto, uma função discriminante de previsão de classificação foi formulada usando as três variáveis observadas a partir de 47 pacientes infectados com dengue e 23 indivíduos normais. A precisão de classificação da função discriminante foi de 97,6%. A Figura 04 mostra que 46 dos 47 casos de dengue tiveram diagnósticos positivos e 22 dos 23 indivíduos saudáveis tiveram seu diagnóstico como negativo para os parâmetros de teste.

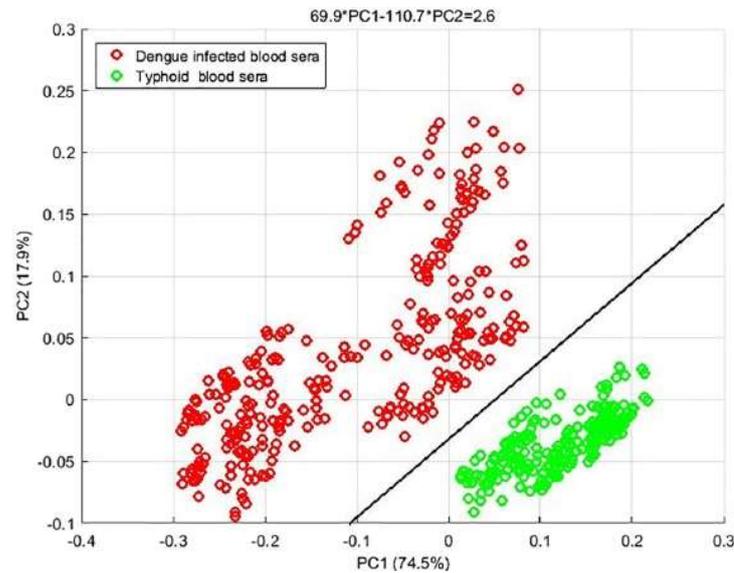
Figura 04: Relação entre o desvio padrão do intervalo de pulsação e o valor Z



Fonte: <https://ieeexplore.ieee.org/document/8512355>

Naseer et. al (2018) aplicou a técnica de análise discriminante para diferenciar grupos de pacientes com febre tifoide dos grupos de pacientes com dengue. Ao todo foram coletadas 20 amostras positivas de febre tifoide e 22 de infectados por dengue, no período de outubro de 2016 a novembro do respectivo ano, todas coletadas em pacientes de ambos os sexos e idades semelhantes. O autor utilizou a análise discriminante em conjunto com a técnica de análise de componentes principais para escrever o modelo de previsão de classificação para análise dos dados de espectro de Raman para conseguir uma alta precisão na separação dos dois grupos, conforme mostra a Figura 05:

Figura 05: Resultado da separação dos grupos de doenças estudadas



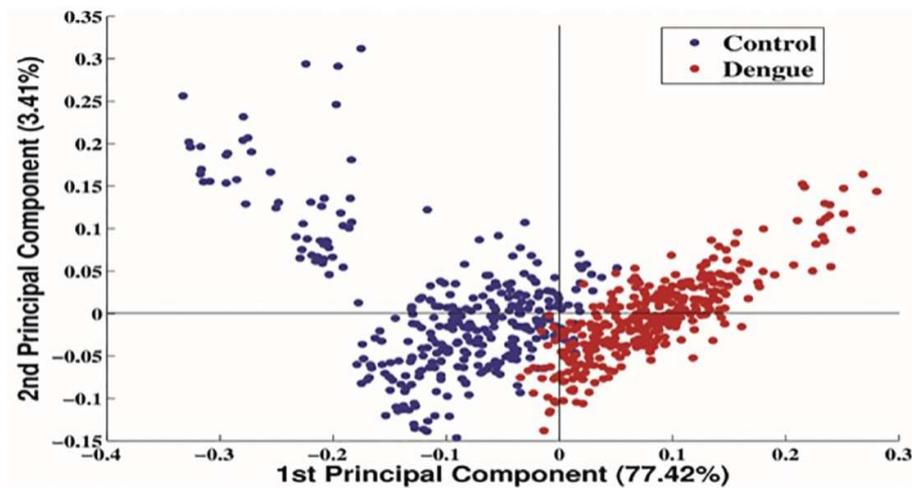
Fonte: <https://www.ncbi.nlm.nih.gov/pubmed/30114579>

Banerjee et. al (2015) fez uso da análise discriminante para estudar se existem diferenças significativas na produtividade larval dos resíduos domiciliares com base nos materiais (porcelana e plástico), e se o contexto urbano-rural afeta os resultados da produção dessas larvas. Os coeficientes da função discriminante foram obtidos a partir da soma dos quadrados e matrizes produto transversais das variáveis explicativas. Os resultados da análise discriminante indicaram diferenças na abundância do *Aedes aegypti* para as áreas urbanas e rurais. Os resíduos porcelana e plástico foram mais produtivos em áreas urbanas em comparação com as áreas rurais, indicando uma possível diferença na geração de lixo doméstico. O autor destaca que a ligação entre resíduos domésticos com a produtividade de larvas do mosquito deverá aumentar o risco de epidemias de dengue na área, e recomenda estratégias alternativas como a substituição destes materiais estudados por outros com baixo tempo de vida, como forma de amenizar o problema.

Rogers et. al (2014) usou dados remotos de satélite, dados meteorológicos interpolados, mapas de distribuição preditos das duas principais espécies de vetores da dengue, *Aedes aegypti* e *Aedes albopictus*, dados de densidade digital de superfície e densidade populacional humana como potenciais variáveis preditoras em um estudo não-linear de modelagem de análise discriminante. O estudo destacou uma coincidência entre áreas de alta densidade populacional humana e áreas com alto risco de dengue, mas não identificou a variável densidade humana como preditora para o modelo de classificação.

Mahmood et. al (2018), realizou um levantamento prévio das variáveis proteínas não estruturais (NS1), imunoglobulina M (IgM) e imunoglobulina G (IgG) em amostras de sangue humano para diagnóstico de pacientes saudáveis e infectados com dengue e utilizou a análise discriminante para quantificar o potencial de diagnóstico da espectroscopia Raman na caracterização desses parâmetros. Os resultados obtidos indicam que a análise discriminante em conjunto com a análise de componentes principais atingiu valores de 97.95% e 95.40% de sensibilidade e especificidade respectivamente, para a identificação de infecção por dengue nos indivíduos. A figura 06 mostra a precisão na separação entre os grupos controle e dengue.

Figura 06: Separação dos grupos de caso e controle através da função discriminante



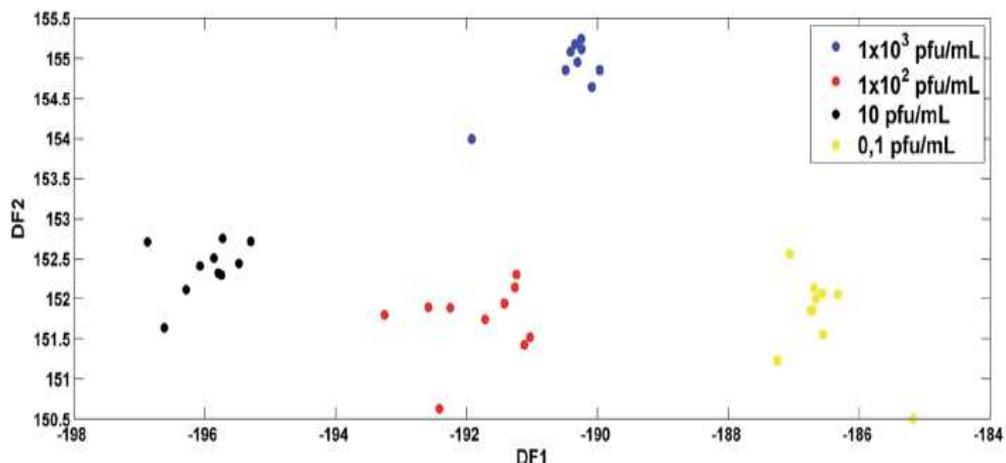
Fonte: <https://www.sciencedirect.com/science/article/pii/S1386142518303172>

Resti et. al (2017), aplicou a abordagem de análise discriminante no mapeamento do vírus da dengue em 5 áreas de Palimbão (cidade da Indonésia localizada ao sul da ilha de Samatra), sendo elas: Central, Sul, Norte, Oeste e Leste. Os resultados apontam que o mapeamento da incidência da dengue em cada área é significativamente afetado pelos fatores estudados, com exceção das variáveis idade do membro da família diagnosticado com dengue, comportamento higiênico do entrevistado, idade e ocupação do chefe da família e número de quartos existentes na casa. As variáveis gênero, grupo sanguíneo do paciente, idade, grau de escolaridade, ocupação e renda do entrevistado, ocupação e renda do chefe da família, tipo de residência, número de banheiros com bacia sanitária na casa e fonte de água limpa diferem estatisticamente entre si, de acordo com cada área, tomando como base os resultados do lambda de Wilks e do teste F. O índice global de acerto do mapeamento é de 66,7%. Essa porcentagem, que está acima de 50%, indica que o modelo proposto é suficiente para mapear a incidência da dengue nas cinco áreas estudadas na cidade.

Santos et. al (2017) fez uso da análise discriminante em conjunto com a análise de componentes principais, algoritmo de projeção sucessiva e algoritmo genético com o propósito de identificar e discriminar o sorotipo 3 da dengue (DENV-3) diluído em diferentes concentrações em amostras de soro e sangue, com o objetivo de desenvolver um metodologia rápida e não destrutiva para uma análise quantitativa do vírus da dengue. Ao todo, foram coletadas quarenta amostras (40 para soro e 40 para sangue) infectadas com DENV-3 em diferentes concentrações (dez amostras para cada concentração) e analisadas por espectroscopia de infravermelho.

Os resultados deste estudo sugerem que a espectroscopia de infravermelho aliada a técnicas de classificação multivariada pode ser aplicada na perspectiva clínica. Comparando as amostras de DENV-3 no soro e no DENV-3 no sangue, os modelos de classificação implementados obtiveram melhores resultados para amostras de sangue, mostrando um bom grau de separação entre os grupos, conforme mostra a figura 07:

Figura 07: Separação dos grupos através do uso de análise discriminante



Fonte: <https://pubs.rsc.org/en/content/articlepdf/2017/ra/c7ra03361c>

Por fim, o autor sugere que é necessário realizar estudos com maior número de amostras para maior confiabilidade e propõe que mais pesquisas merecem ser realizadas nessa perspectiva.

## **4. MATERIAL E MÉTODO**

A princípio é apresentado o local de estudo da pesquisa, suas denominações e características. Em seguida, é apresentado o delineamento da pesquisa, explicando o tipo e natureza da pesquisa, bem como, a coleta de dados da pesquisa. Por fim, serão apresentadas as variáveis adotadas para a modelagem e execução da pesquisa.

### **4.1. LOCAL DO ESTUDO**

A pesquisa foi desenvolvida no Núcleo de Tecnologias Estratégicas em Saúde (NUTES), situado no CAMPUS I da Universidade Estadual da Paraíba (UEPB). Por se tratar de um estudo de análise exploratória com auxílio de software, o local de pesquisa se limitou ao ambiente do laboratório NUTES, pois não foi necessária nenhuma intervenção externa para implementar etapas de execução do projeto.

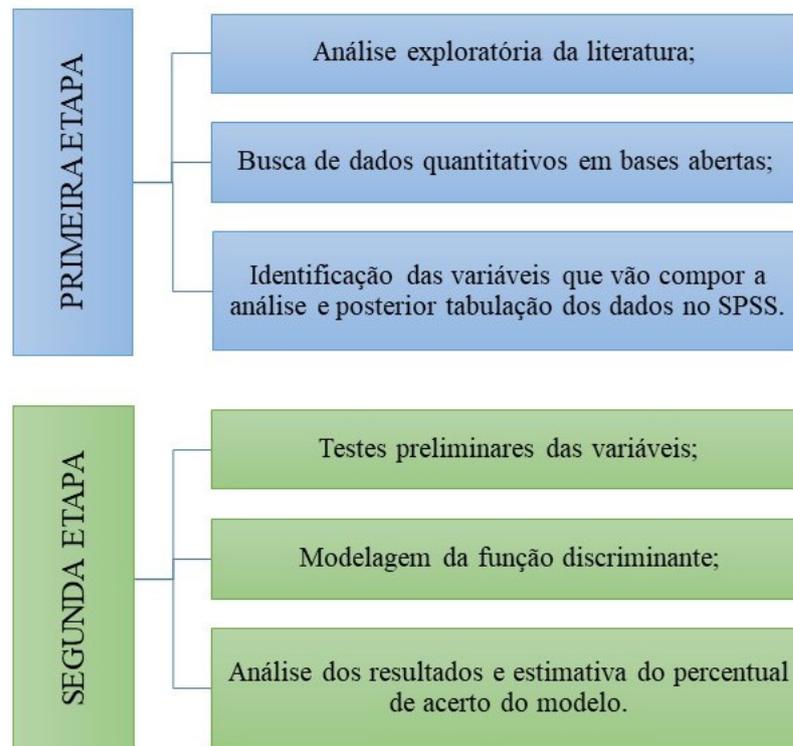
### **4.2. DELINEAMENTO DA PESQUISA**

Este trabalho fundamenta-se na análise exploratória de dados abertos do Ministério da Saúde e Ministério das Cidades, trazendo os aspectos relacionados aos casos de dengue nos Estados Brasileiros, utilizando a análise discriminante como forma de extrair e contextualizar o máximo de informações que possam servir de base para previsões e decisões estratégicas no combate aos altos índices de casos dessa doença.

O estudo dessa técnica traz em sua principal característica tentar explicar e entender os fenômenos e características das variáveis estudadas, bem como o nível de significância de cada uma delas no quadro epidemiológico de casos de dengue no país.

Em seguida, será apresentado o desenho da pesquisa dividida em duas etapas, dispostas na Figura 08:

Figura 08: Desenho da Pesquisa



Fonte: do autor

### 4.3. COLETA DE DADOS

Foram utilizados dados do período de 2013 à 2017 (últimos 5 anos disponíveis) do Boletim Epidemiológico do Ministério da Saúde para o levantamento dos números acumulados da incidência de casos de dengue que serão usados como base de classificação dos grupos da variável dependente. Dados do Sistema Nacional de Informações sobre Saneamento (SNIS) do Ministério das Cidades referentes ao tratamento de água, esgoto e manejo de resíduos sólidos foram selecionadas como as variáveis dependentes que serão cruzadas para fins de previsão de classificação e modelagem da função discriminante com base nas variáveis identificadas como estatisticamente significativas para tal finalidade através de testes específicos.

Para a definição dos parâmetros qualitativos da variável resposta será utilizado o critério de classificação para os grupos definidos pelo Programa Nacional de Controle da Dengue através do Sistema de Informação de Agravos de Notificação (SINAN):

- **Baixa Incidência:** Até 100 casos de dengue para cada 100 mil habitantes;
- **Média Incidência:** De 101 até 299 casos de dengue para cada 100 mil habitantes ou;
- **Alta Incidência:** 300 ou mais casos de dengue para cada 100 mil habitantes.

#### 4.4. VARIÁVEIS UTILIZADAS

Para realização das análises, testes e elaboração do modelo discriminante, serão utilizadas variáveis referentes à resíduos sólidos, água e esgoto extraídas dos relatórios do Sistema Nacional de Informações sobre Saneamento (SNIS) e cruzadas com as variáveis ligadas aos casos de Dengue retiradas dos boletins epidemiológicos. Foram selecionadas as variáveis que guardassem relação direta com ações de controle do aumento nos casos da doença de acordo com as Diretrizes Nacionais para a Prevenção e Controle de Epidemias de Dengue do Ministério da Saúde, e cujos dados estavam consolidados no período levantado, tendo sido todos preenchidos em sua totalidade, sem lacunas e sem erros de digitação. As siglas são apresentadas no Quadro 01:

Quadro 01: Variáveis que serão utilizadas nas análises

<b>CASOS</b>		<i>Incidência de casos (por 100 mil Habitantes)</i>
<b>RESÍDUOS SÓLIDOS</b>	<b>CRDO</b>	<i>Taxa (%) de cobertura da coleta Resíduos Domiciliares em relação à população total.</i>
	<b>LIXÕES</b>	<i>Número de lixões no Estado.</i>
	<b>QRJL</b>	<i>Quantidade de resíduos jogados nos lixões (em Toneladas).</i>
<b>ÁGUA E ESGOTO</b>	<b>QMAA</b>	<i>Quantidade de municípios atendidos com abastecimento de água.</i>
	<b>IMAA</b>	<i>Índice de municípios atendidos com água.</i>
	<b>QMAES</b>	<i>Quantidade de municípios atendidos com esgotamento sanitário.</i>
	<b>IMAE</b>	<i>Índice de municípios atendidos com esgoto.</i>
	<b>ICE</b>	<i>Índice de coleta de esgoto.</i>
	<b>DTS</b>	<i>Despesa total com os serviços de água e esgoto.</i>
	<b>PTAAA</b>	<i>População total atendida com abastecimento de água.</i>
	<b>PTAES</b>	<i>População total atendida com esgotamento sanitário.</i>
	<b>PSA</b>	<i>Paralisações em sistemas de água. (Número de paralisações no ano)</i>
	<b>DPSA</b>	<i>Duração total das paralisações em sistemas de água no ano. (Em horas/ano)</i>
	<b>DMP</b>	<i>Duração média das paralisações nos sistemas de água. (Em horas)</i>
	<b>CMAP</b>	<i>Consumo médio de água per capita. (litros/hab/dia)</i>

Fonte: SNIS

## **5. RESULTADOS**

Nesta sessão apresentam-se os resultados das análises dos dados através da aplicação da técnica de análise discriminante com o auxílio do software SPSS. As análises realizadas estão descritas abaixo e serão aplicadas conforme a seguinte sequência:

1. Verificação de normalidade dos dados através do teste de Kolmogorov-Smirnov e geração de gráficos QQ-plot;
2. Verificação de significância das variáveis através do teste de igualdade de médias nos grupos;
3. Verificação da igualdade das matrizes de covariância dos grupos através do teste Box's M;
4. Seleção das variáveis que garantirão os melhores resultados preditivos do modelo discriminante através do método Stepwise;
5. Cálculo do coeficiente de correlação canônica e aplicação dos testes Lambda de Wilks e Qui-quadrado para testar a significância da função discriminante;
6. Cálculo dos Coeficientes Discriminantes das variáveis explicativas e construção do modelo discriminante de classificação;
7. Definição do valor de corte que servirá como referência para a alocação de novas observações em um dos grupos estudado através do cálculo do ponto médio dos centroides;
8. Resultado da classificação e índice global de acerto.

### **5.1. VERIFICAÇÃO DE NORMALIDADE DOS DADOS**

Um dos pré-requisitos básicos para que se possa utilizar a técnica de Análise Discriminante é a suposição de normalidade. A mesma é entendida como sendo o pressuposto de que as variáveis de interesse são distribuídas normalmente, ou aproximadamente normais. Intuitivamente, a normalidade pode ser entendida como o resultado da soma de um grande número de eventos aleatórios independentes.

Para que os dados se adequassem à condição de normalidade foi aplicado uma transformação logarítmica, que estende os valores da variável para a reta real.

De acordo com Pino (2014), a transformação logarítmica é bem vista e tem sido largamente aplicada com sucesso há muito tempo, sendo bastante utilizada para fins de análise de regressão e na econometria.

A avaliação de normalidade das variáveis se deu através do teste de Shapiro Wilks, que segundo Lopes (2013), aparenta ser o melhor teste de aderência à normalidade, sendo eficiente para diferentes distribuições e tamanhos de amostras se comparado a outros testes. Seus resultados estão expressos na Tabela 01:

Tabela 01: Resultados do teste de Shapiro Wilks para aderência à normalidade

	<b>Estatística</b>	<b>Sig.</b>
CRDO	,987	,903
LIXOES	,912	,003
QRJL	,890	,001
QMAA	,962	,169
IMAA	,941	,028
QMAES	,975	,477
IMAE	,962	,167
ICE	,960	,137
DTS	,972	,369
PTAAA	,959	,131
PTAES	,988	,923
PSA	,973	,398
DPSA	,986	,856
DMP	,816	,000
CMAP	,968	,265

Fonte: do autor

A Tabela acima mostra os resultados do nível de significância do teste de Shapiro-Wilks, utilizado para verificar se os dados seguem distribuição normal ou aproximadamente normal. A regra de decisão para o teste de hipóteses é a seguinte:

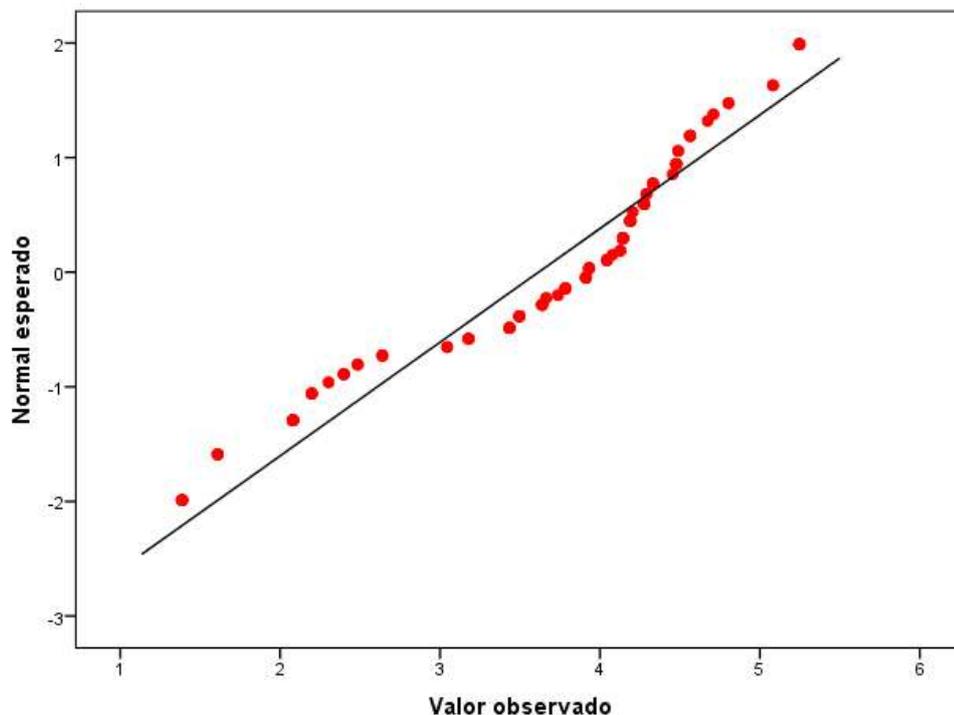
$H_0$ : Aceitamos que os dados têm distribuição normal (se o p-valor (Sig) > 0,05), e;

$H_1$ : Assumimos que os dados não seguem distribuição normal (se o p-valor (Sig) < 0,05)

As variáveis marcadas na Tabela acima não satisfazem a condição de normalidade, de acordo com o teste de Shapiro Wilks, pois seus níveis de significância calculados estão abaixo do valor crítico de 0,05.

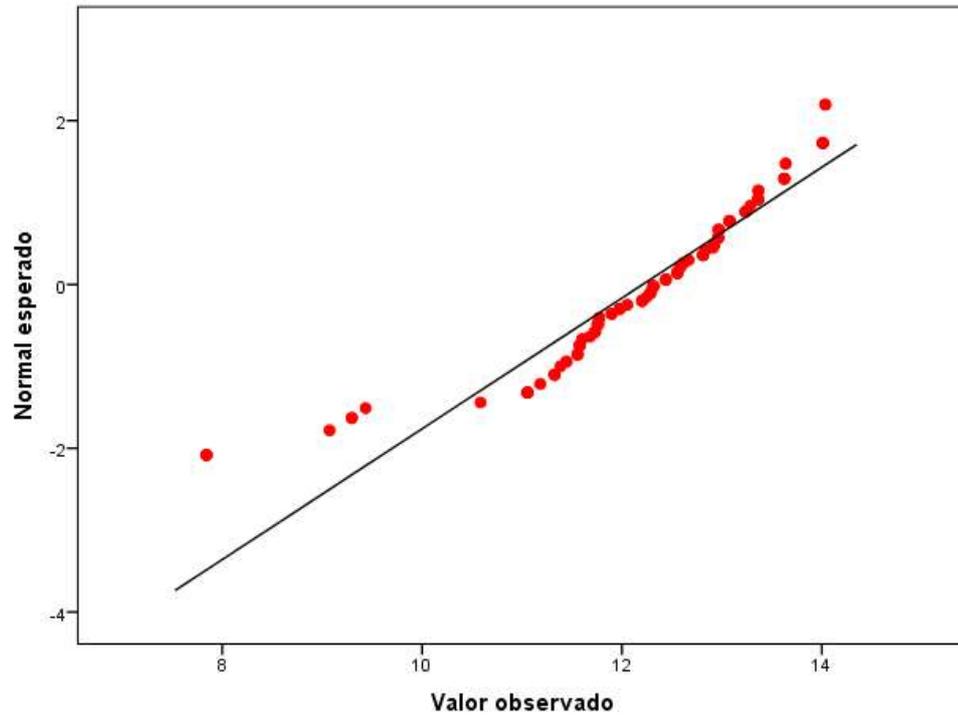
Para reforçar o resultado de não normalidade das variáveis, foram gerados os gráficos QQ-plot (Quantil-Quantil), que é uma técnica gráfica que dá indícios de que dois conjuntos de dados provêm de populações com uma distribuição comum ou não. Se estes vierem de uma população normal, os pontos do cruzamento entre os valores observados e os valores normais esperados deverão se ajustar ao longo da reta de referência. Quanto maior o afastamento dessa linha, maiores as evidências que os dados não têm distribuição normal. Os Gráficos QQ-plot das variáveis tidas como não normais pelo teste de Shapiro Wilks estão listados a seguir:

Gráfico 01: QQ-plot da variável Número de Lixões (LIXOES)



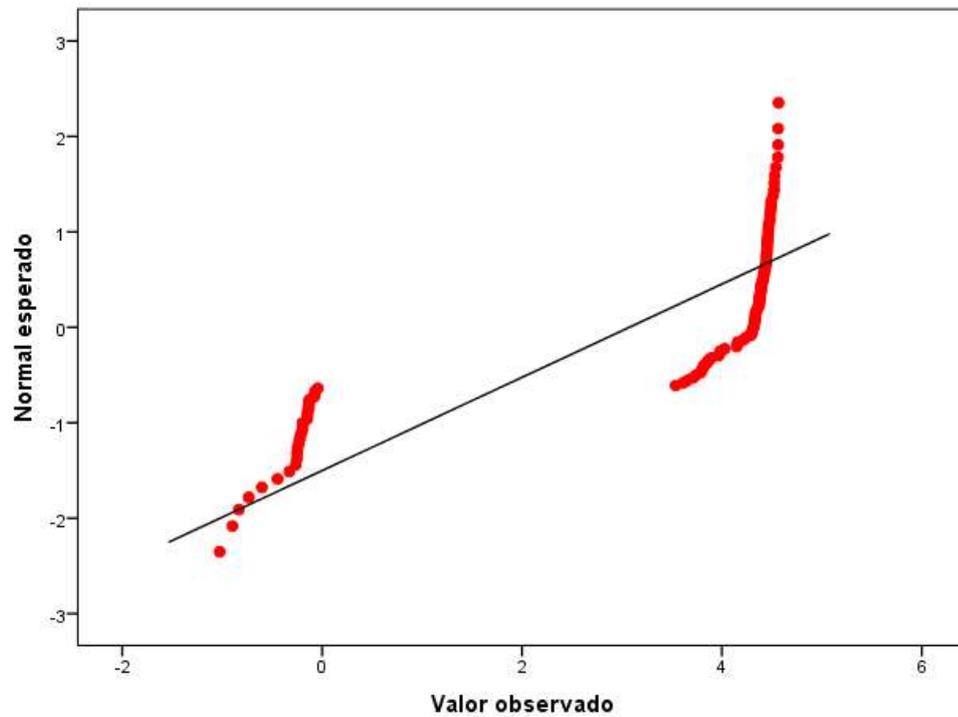
Fonte: do autor

Gráfico 02: QQ-plot da variável Quantidade de Resíduos Jogados nos Lixões (QRJL)



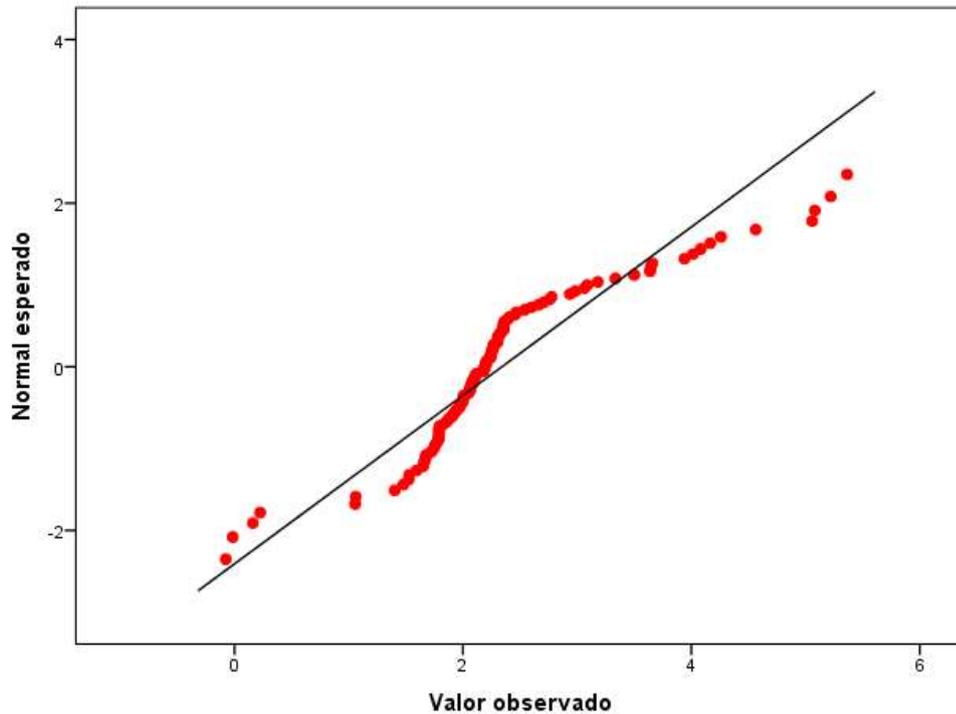
Fonte: do autor

Gráfico 03: QQ-plot da variável Índice de Municípios Atendidos com Abastecimento de Água (IMAAA)



Fonte: do autor

Gráfico 04: QQ-plot da variável Duração Média das Paralisações nos Sistemas de Água (DMP)



Fonte: do autor

De acordo com os resultados calculados, percebemos que apesar da transformação logarítmica dos dados, algumas variáveis ainda violam a condição de normalidade (com  $\text{sig} < 0,05$ ). A análise dos Gráficos 01, 02, 03 e 04 reforça que estas não seguem distribuição normal, pois os dados não se ajustam à reta. Sendo assim, aceitamos a suposição de normalidade (ou normalidade aproximada) das demais variáveis, e excluindo-se as não normais, pode-se dar sequência às próximas etapas das análises.

## 5.2. SIGNIFICÂNCIA DAS VARIÁVEIS

Nesta etapa foi calculado quais variáveis disponíveis na amostra são significativas para a construção do modelo de previsão de classificação através do Lambda de Wilks. A Tabela 02 traz os resultados:

Tabela 02: Resultados dos testes de igualdade de médias de grupo

Variável	Lambda de Wilks	F	Sig.
QMAA	,929	3,958	,022
QMAES	,861	8,342	,000
IMAE	,925	4,189	,018
ICE	,904	5,452	,006
DTS	,909	5,126	,008
PTAAA	,932	3,738	,027
PTAES	,885	6,677	,002
PSA	,996	,223	,800
DPSA	,999	,033	,968
CMAF	,999	,071	,932
CRDO	,938	3,421	,036

Fonte: do autor

No teste de Igualdade de médias entre grupos a regra de decisão para o teste de hipóteses é a seguinte:

$H_0$ : Não há diferença entre as médias (se o  $p$ -valor (Sig)  $> 0,05$ ), e;

$H_1$ : Há diferença estatística entre as médias (se o  $p$ -valor (Sig)  $< 0,05$ ).

Para que uma variável seja considerada adequada para se fazer inferência, sua média deve ser estatisticamente diferente. A Tabela acima mostra que as variáveis marcadas não atenderam ao pré-requisito e, com isso, as variáveis que mostraram grau de significância maior que 0,05 não servem para compor a função discriminante.

### 5.3. IGUALDADE DAS MATRIZES DE COVARIÂNCIA DOS GRUPOS

Mais um pressuposto no uso da análise discriminante é o da igualdade das matrizes de covariância entre os grupos, que é testado através do teste Box's M, conforme os resultados estão expressos na Tabela 03.

Tabela 03: Resultados do teste Box's M

M de Box		9,691
F	Aprox.	1,567
	<b>Sig.</b>	<b>,152</b>

Fonte: do autor

Para esse teste, é necessário que se verifique as seguintes hipóteses:

$H_0$ : *As matrizes de covariância são iguais (se o p-valor > 0,05) e;*

$H_1$ : *As matrizes de covariância são diferentes (se o p-valor < 0,05).*

É interessante que as matrizes de covariância sejam iguais, para que se tenha o máximo de homogeneidade entre os dados. Como a significância encontrada satisfaz a condição de ser maior que o valor crítico de 0,05, as evidências estatísticas mostram que há igualdade nas matrizes de covariância, permitindo assim o prosseguimento das análises dos dados.

### 5.4. MÉTODO STEPWISE

Uma maneira de construir modelos cuja variabilidade seja melhor explicada, ao mesmo tempo em que pode maximizar a diferença entre os grupos, é o método stepwise (por etapas). Esse método busca estimar a função discriminante selecionando variáveis para serem acrescentadas ou eliminadas do modelo até que alguma medida de critério geral seja alcançada.

Este tem por objetivo selecionar variáveis que maximizam a previsão com o menor número de variáveis empregadas. Com isso, pode-se identificar as variáveis estatisticamente significativas para compor a função discriminante, uma vez que somente variáveis que aumentem significativamente a capacidade discriminativa do modelo são adicionadas.

Sendo assim, as variáveis selecionadas para compor o modelo e que garantirão os melhores resultados preditivos estão listadas na Tabela 04.

Tabela 04: Variáveis selecionadas pelo método Stepwise

Lambda de Wilks			
Passo	Inseridas	Estatística	Sig.
1	ICE	,886	,001
2	QMAES	,760	,000
3	DTS	,672	,000
4	QMAAA	,624	,000

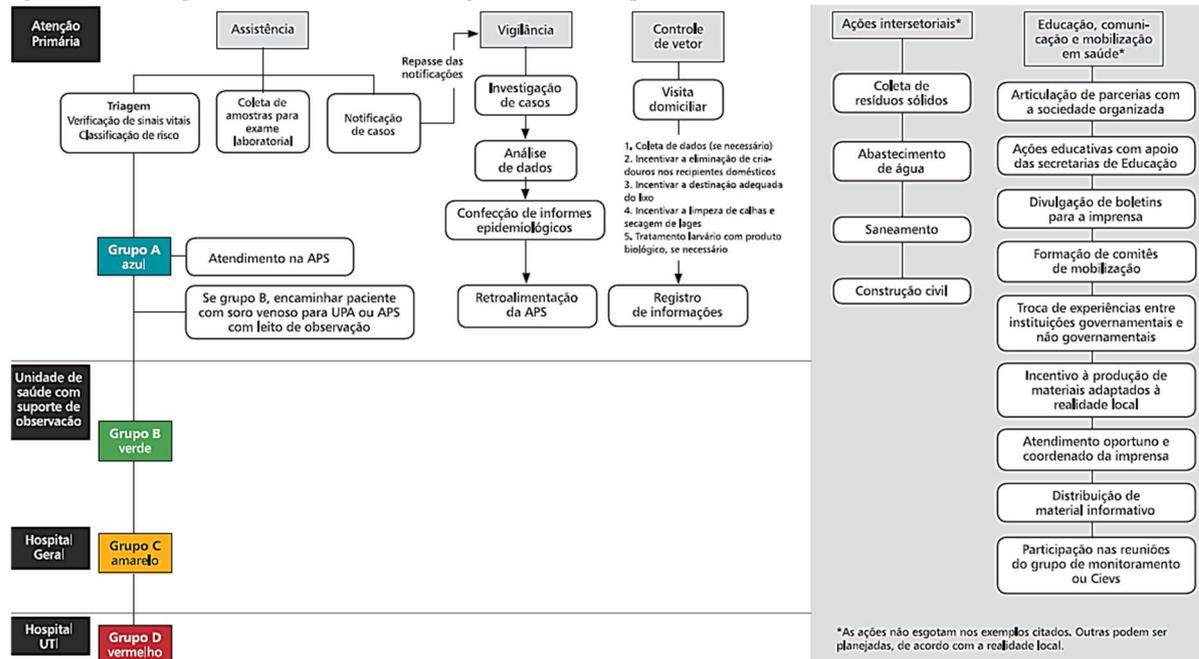
Fonte: do autor

As variáveis Índice de Coleta de Esgoto (ICE), Quantidade de Municípios Atendidos com Esgotamento Sanitário (QMAES), Despesa Total com Saneamento -incluindo água e esgoto - (DTS) e Quantidade de Municípios Atendidos com Abastecimento de Água (QMAAA) são as mais significativas para classificar os Estados nos grupos de baixo, médio e alto índice de casos de dengue.

O Ministério da Saúde destaca em seu Plano de Contingência Nacional para Epidemias de Dengue o abastecimento de água e saneamento como fatores condicionantes da expansão da Dengue no Brasil, sendo classificadas como lacunas resultantes do crescimento desordenado dos centros urbanos.

Abastecimento de água e Saneamento são ações intersetoriais propostas pelo Ministério da Saúde nas Diretrizes Nacionais para a Prevenção e Controle de Epidemias de Dengue, conforme a figura 09.

Figura 09: Fluxograma do Controle de Dengue nos Municípios



Fonte: Diretrizes Nacionais para a Prevenção e Controle de Epidemias de Dengue

## 5.5. AUTOVALORES

Os autovalores são uma medida relativa de quão diferentes são os grupos na função discriminante, ou seja, quanto mais próximos de 1 for a correlação canônica dos autovalores, menores serão as variações entre os grupos explicados pela função discriminante. Os resultados são apresentados na Tabela 05.

Tabela 05: Autovalor e Correlação Canônica

Função	Autovalor	% de Variância	Correlação canônica
1	,456 <sup>a</sup>	81,8	,560
2	,101 <sup>a</sup>	18,2	,303

Fonte: do autor

O coeficiente de correlação canônica calculado dá ideia do grau de associação dos grupos, que nesse caso, traz a primeira função mais próxima de 1, o que demonstra que esta possui maior correlação e menor variação que a segunda função, dando indícios de que a primeira serve melhor para o propósito ao qual se pretende usá-la.

A Função 1 é a que apresenta o maior % de variância com 81,8%, demonstrando que essa é a que mais contribui para explicar a diferença entre os grupos, uma vez que a função 2 apresenta apenas 18,2% de poder discriminante.

Foi aplicado em seguida os testes Lambda de Wilks e Qui-quadrado, calculados a partir do determinante da matriz da soma dos quadrados e produtos cruzados dentro dos grupos e do determinante da matriz da soma dos quadrados e produtos cruzados totais para testar a significância da função discriminante. Tem-se então as seguintes hipóteses:

$H_0$ : A média populacional dos dois grupos é igual (se o p-valor > 0,05) e;

$H_1$ : A média populacional dos dois grupos é diferente (se o p-valor < 0,05).

Os valores de significância das funções para comparação com os respectivos valores críticos estão listados na Tabela 06.

Tabela 06: Teste Lambda de Wilks e Qui-quadrado

Teste de funções	Lambda de		
	Wilks	Qui-quadrado	Sig.
1	,624	55,951	,000
2	,726	5,452	,065

Fonte: do autor

Com p-valor menor que 0,05, rejeitamos  $H_0$  para a função 1 e assumimos que as médias são diferentes e, portanto, a função é significativa e tem poder de discriminação.

Como a função 2 mostrou p-valor maior que 0,05, concluímos que as médias são iguais e, portanto, esta não possui poder de discriminação e não pode ser usada para fins de previsão de classificação.

Como o Lambda de Wilks é menor na primeira função, conclui-se também que esta tem melhor poder de discriminação que a segunda. O valor mais alto do Qui-quadrado da primeira função é mais adequada para a separação das observações de cada grupo que a função 2. Portanto, opta-se por trabalhar com a função 1 para fins de previsão de classificação.

## 5.6. FUNÇÃO DISCRIMINANTE

Através dos Coeficientes Discriminantes das variáveis explicativas, podemos escrever o modelo discriminante de classificação. Seus valores estão descritos na Tabela 07.

Tabela 07: Coeficientes de Função Discriminante

	Função	
	1	2
QMAA	1,041	-,406
QMAES	-1,858	,181
ICE	,396	,110
DTS	,987	,868
(Constante)	-18,538	-16,499

Fonte: do autor

A função discriminante  $Z$  é dada pela equação 5.0:

$$Z = -18,538 + (1,041 \times \ln(QMAA)) - (1,858 \times \ln(QMAES)) + (0,396 \times \ln(ICE)) + (0,987 \times \ln(DTS)) \quad (5.0)$$

## 5.7. ALOCAÇÃO DE NOVAS OBSERVAÇÕES

Através do ponto médio dos centroides da Tabela abaixo, podemos definir um valor de corte que servirá como referência para a alocação de novas observações em um dos grupos estudados. Os resultados estão expressos na Tabela 08:

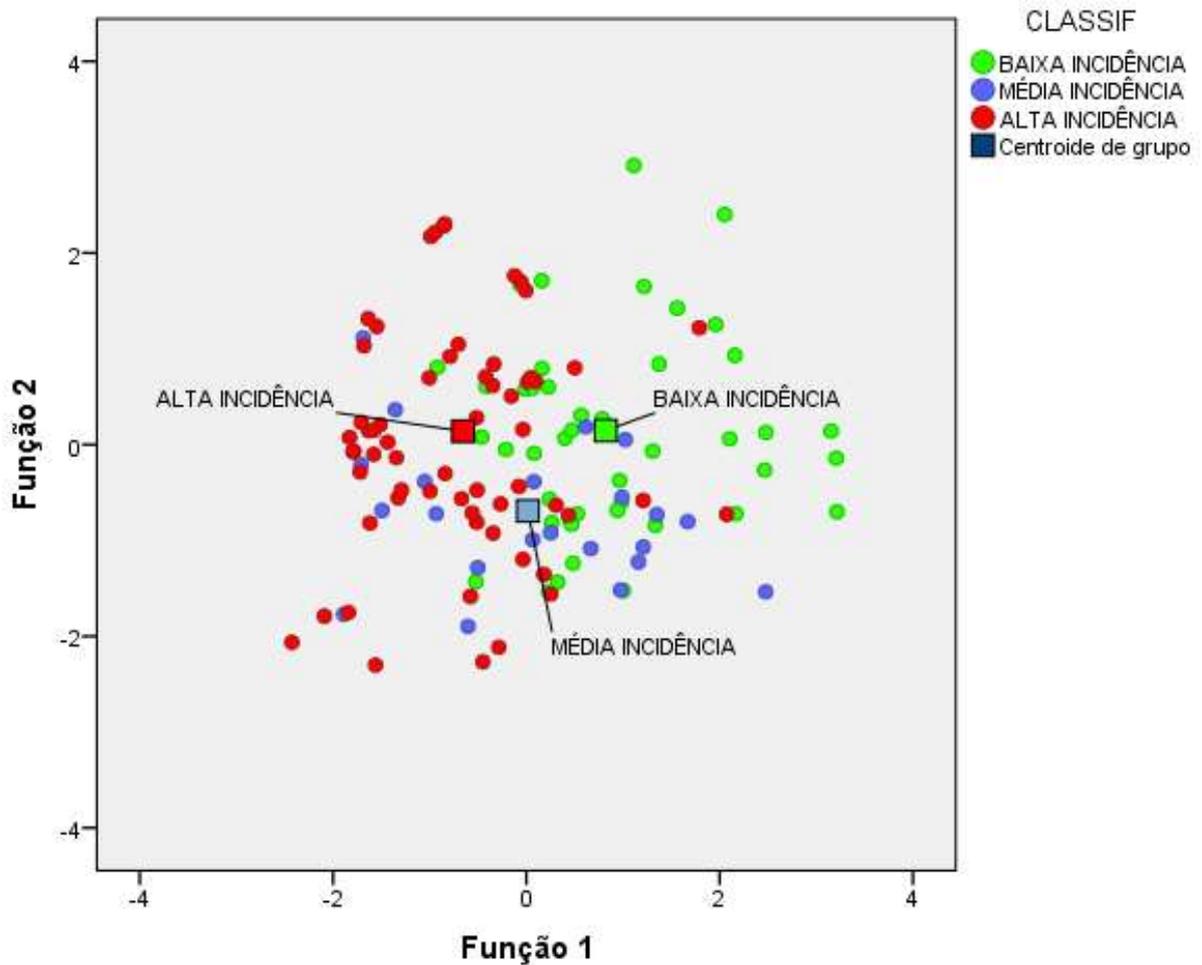
Tabela 08: Funções em Centroides de Grupos

CLASSIF	Função	
	1	2
Baixa Incidência	,821	,147
Média Incidência	,015	-,692
Alta Incidência	-,654	,139

Fonte: do autor

Visto isso, ao se atribuir novos valores observados na função discriminante, compara-se o valor obtido com os centroides dos grupos, classificando a nova observação onde a diferença entre o resultado e o centroide for menor, ou seja, no grupo onde o resultado da função for mais próximo do seu centroide. A Figura 10 ilustra essa relação:

Figura 10: Centroides em Relação aos Escores da Função Discriminante



Fonte: do autor

Substituindo-se os casos originais na função discriminante, comparando-os aos centroides dos grupos e classificando nos grupos obtém-se o índice global de acerto, cujo resultado dessa classificação está descrito na Tabela 09:

Tabela 09: Resultado da Classificação

		Associação ao grupo				
		Baixa	Média	Alta		
CLASSIF		Incidência	Incidência	Incidência	Total	
Original	Contagem	Baixa Incidência	30	2	14	46
		Média Incidência	10	1	11	22
		Alta Incidência	6	2	53	61
%		Baixa Incidência	65,2	4,3	30,4	100,0
		Média Incidência	45,5	4,5	50,0	100,0
		Alta Incidência	9,8	3,3	86,9	100,0

Fonte: do autor

A partir da Tabela 09 temos informações acerca do sucesso (ou não) do modelo discriminante. Podemos ver que a função tem bom poder de discriminação, tendo êxito em classificar 86,9% dos casos originais de Alta Incidência nos seus respectivos grupos, valor esse que é compatível com os alcançados em pesquisas citadas ao longo desse trabalho.

## 6. CONCLUSÃO

O objetivo desta dissertação foi aplicar a técnica de análise discriminante multivariada para modelar uma função discriminante de previsão capaz de classificar os Estados em grupos de baixo, médio e alto nível de incidência de dengue, utilizando o software SPSS para tabular variáveis extraídas do boletim epidemiológico do Ministério da Saúde e da série histórica de ações de saneamento público do Ministério das Cidades, ambos entre os anos de 2013 e 2017.

Assim sendo, foi realizada uma revisão bibliográfica acerca da atual situação da Dengue no Brasil os riscos de volta das epidemias, das teorias e pressupostos que envolvem a utilização da análise discriminante multivariada e da aplicação da técnica relacionada especificamente em estudos que tratavam de forma direta ou indireta de casos de dengue.

As principais conclusões deste trabalho são apresentadas a seguir.

- Os resultados da pesquisa trazem um detalhado passo a passo da aplicação da análise discriminante, podendo ser usado como base para outros trabalhos na área;
- Para que se pudesse aplicar técnicas multivariadas que necessitam da suposição de normalidade para funcionar sem viés, os dados precisaram passar por uma transformação logarítmica. A transformação foi satisfatória para que pudesse se prosseguir com as análises sem maiores problemas com a exclusão de algumas das variáveis que continuaram com indícios de não-normalidade;
- Através do teste Lambda de Wilks foi possível testar as hipóteses de interesse, mostrando quais variáveis disponíveis na amostra são significativas para a construção do modelo de previsão de classificação e quais não são;
- Através do teste de Box's M foi verificado mais um pressuposto da análise discriminante que é a igualdade das matrizes de covariância, sendo realizado o teste de hipóteses e posteriormente aceita a hipótese de igualdade;
- Com o uso do método Stepwise foi possível identificar as variáveis com maior poder de discriminação entre todas as analisadas, sendo Índice de Coleta de Esgoto (ICE), Quantidade de Municípios Atendidos com Esgotamento Sanitário (QMAES), Despesa Total com Saneamento -incluindo água e esgoto - (DTS) e Quantidade de Municípios Atendidos com Abastecimento de Água (QMAAA) as mais significativas estatisticamente para fins de classificação posterior nos grupos de interesse;
- O cálculo dos autovalores mostrou a função 1 com um bom nível percentual de variância na explicação do modelo com 81,8% e um grau de correlação canônica de 0,560, o que indica um grau de correlação moderada entre as variáveis;

- O teste de Lambda de Wilks avaliou as duas funções discriminantes geradas, chegando no resultado que de a função número 1 é a mais significativa para fins inferenciais;
- Através dos Coeficientes Discriminantes das variáveis explicativas foi gerado o modelo discriminante de classificação como sendo =  $-18,538 + (1,041 \times \ln(QMAA)) - (1,858 \times \ln(QMAES)) + (0,396 \times \ln(ICE)) + (0,987 \times \ln(DTS))$ . Devido à transformação logarítmica realizada nos dados, é importante que as novas observações a serem inseridas na função discriminante também estejam transformadas;
- Para a classificação propriamente dita, foram calculados os centroides dos grupos, que servirão de referência para alocação de novas observações nos grupos baixa, média e alta incidência de dengue;
- Os resultados obtidos com a análise discriminante foram satisfatórios a um nível de significância de 0,05 e foi possível constatar que a separação entre os grupos de baixa, média e alta incidência é estatisticamente significativa de acordo com o resultado dos testes estatísticos implementados no SPSS, tendo a função discriminante alcançado um nível de acerto de 86,9% dos casos originais de Alta Incidência nos seus respectivos grupos.

Todos os objetivos deste trabalho foram alcançados de forma satisfatória, podendo ser usados para aumentar a eficiência do combate estratégico da doença à níveis Nacionais, proporcionando economia de tempo e recursos financeiros. A seguir tem-se sugestões para continuidade deste trabalho.

- Utilizar como base a maneira como a análise discriminante foi aplicada nessa pesquisa para a aplicação em outras áreas;
- A implementação da função discriminante em algum aplicativo ou planilha que conforme seja alimentada com os dados, mostre o resultado da classificação nos grupos descritos no trabalho para servir como ferramenta de apoio à tomada de decisão;
- Aplicabilidade das análises aplicadas neste trabalho em outros estudos de casos para pesquisa e aprimoramento do método;
- Utilizar a metodologia aplicada para criar funções de previsão de outros tipos de doenças (Ex.: zika e chikungunya);
- Modificar a escala de classificação dos grupos a fim de aumentar o índice global de acerto da função discriminante.

## REFERÊNCIAS

- ANDERSON, T. W. **An Introduction to Multivariate Statistical Analysis**, New York: John Wiley. 2003.
- ARTUSO, Alysson; NETO, Anselmo Chaves. **Análise discriminante e regressão logística – reconhecimento de padrões para a seleção de portfólios no mercado acionário brasileiro**. UFPR, 2 abr. 2019. Disponível em: [https://www.researchgate.net/publication/265685925\\_Analise\\_discriminante\\_e\\_regressao\\_logistica\\_-\\_reconhecimento\\_de\\_padroes\\_para\\_a\\_selecao\\_de\\_portfolios\\_no\\_mercado\\_acionario\\_brasileiro](https://www.researchgate.net/publication/265685925_Analise_discriminante_e_regressao_logistica_-_reconhecimento_de_padroes_para_a_selecao_de_portfolios_no_mercado_acionario_brasileiro). Acesso em: 20 fev. 2019.
- ALAA, Tharwat, T. GABER, ABDELHAMEED, Ibrahim AND A. E. HASSANIEN. **Linear Discriminant Analysis: A Detailed Tutorial**. AI Communications, IOS press, 2017.
- ALHASSAN, Abukari. **Discriminant analysis as a tool for under five anthropometric classification in Ghana**. UNIVERSITY FOR DEVELOPMENT STUDIES, 2010. Disponível em: <http://udsspace.uds.edu.gh/handle/123456789/1607>. Acesso em: 20 fev. 2019.
- ASSOCIAÇÃO BRASILEIRA DE ENGENHARIA SANITÁRIA E AMBIENTAL. **Dengue e zika: ABES alerta sobre a importância do saneamento básico no combate às doenças**. Disponível em: <http://abes-dn.org.br/?p=923>. Acesso em: 20 fev. 2019.
- BAGLEY, S. C., White, H., & GOLOMB, B. A. (2001). **Logistic regression in the medical literature: Standards for use and reporting, with particular attention to one medical domain**. *Journal of Clinical Epidemiology*, 54(10), 979-985.
- BANERJEE, Soumyajit et al. **Household Wastes as Larval Habitats of Dengue Vectors: Comparison between Urban and Rural Areas of Kolkata, India**. 2015. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/26447690>. Acesso em: 20 fev. 2019.
- BARTLETT, M. S. **The general canonical correlation distribution**. Cambridge, 1947. Disponível em: <https://projecteuclid.org/euclid.aoms/1177730488>. Acesso em: 20 fev. 2019.
- BOX, G. E. P. **A general distribution theory for a class of likelihood criteria** *Biometrika*. 1949.

BROWN, Steven; TINSLEY, Howard. **Handbook of Applied Multivariate Statistics and Mathematical Modeling 1st Edition**. University Of Florida, Gainesville, U.S.A.: Academic Press, 2000. 721 p. Disponível em: <https://www.elsevier.com/books/handbook-of-applied-multivariate-statistics-and-mathematical-modeling/tinsley/978-0-12-691360-6>. Acesso em 20 fev. 2019.

CASELLA, G., BERGER, R. **Statistical Inference (2nd ed.)**, Paci. C Grove, CA: Duxbury Press. 2002.

COSTA, Maria Fernanda Lima; BARRETO, Sandhi Maria. **Tipos de estudos epidemiológicos: conceitos básicos e aplicações na área do envelhecimento**. Belo Horizonte-MG, 2003. Disponível em: [http://scielo.iec.gov.br/scielo.php?script=sci\\_arttext&pid=S1679-49742003000400003](http://scielo.iec.gov.br/scielo.php?script=sci_arttext&pid=S1679-49742003000400003). Acesso em: 20 fev. 2019.

COX, D.R; SNELL, E.J. **Analysis of Binary Data, Second Edition**. USA: CRC Press, 1989. 240 p.

CRUVINEL, Evelyn de Castro. **Discriminante para mistura de distribuições GEV**. Brasília, 2017. Disponível em: [http://repositorio.unb.br/bitstream/10482/24006/1/2017\\_EvelyndeCastroCruvinel.pdf](http://repositorio.unb.br/bitstream/10482/24006/1/2017_EvelyndeCastroCruvinel.pdf). Acesso em: 20 fev. 2019.

DOS REIS, Déborah Aparecida Souza. **Análise discriminante e otimização aplicadas ao controle de qualidade de ruído de compressores usando os softwares r e gurobi**. Uberlândia, 2017. Disponível em: <https://repositorio.ufu.br/bitstream/123456789/19801/1/AnaliseDiscriminanteOtimizacao.pdf>. Acesso em: 20 fev. 2019.

FÁVERO, Luiz Paulo. **Análise de dados: modelagem multivariada para tomada de decisões**. Rio de Janeiro: Elsevier, 2009.

HAIR, J. F., ANDERSON, R. E., TATHAM, & BLACK, W. C. (2005). **Análise multivariada de dados** (A. S. Sant'Anna & A. C. Neto, trad.). Porto Alegre: Bookman.

HAWKINS, Douglas M. **The Subset Problem in Multivariate Analysis of Variance**. University of the Witwatersrand South Africa, 1976. Disponível em:

[https://www.jstor.org/stable/2985017?seq=1#page\\_scan\\_tab\\_contents](https://www.jstor.org/stable/2985017?seq=1#page_scan_tab_contents). Acesso em: 20 fev. 2019.

HOLM, Sture. **A Simple Sequentially Rejective Multiple Test Procedure**. Goteborg, 1979. Disponível em: <https://www.ime.usp.br/~abe/lista/pdf4R8xPVzCnX.pdf>. Acesso em: 20 fev. 2019

HUBERTY, C. J. **An Introduction to Discriminant Analysis**. [S. l.], 1989. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/07481756.1989.12022925>. Acesso em: 20 fev. 2019.

IVES, A.R; GARLAND , T. Jr. **Phylogenetic logistic regression for binary dependent variables..** Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/20525617>. Acesso em: 20 fev. 2019.

JOBSON, J.D., **Applied Multivariate Data Analysis, Volume II: Categorical and Multivariate Methods**, Springer-Verlag, 1992.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. Fifth Edition. New Jersey: Prentice Hall, 2002.

KATZ, Mitchell H. **Multivariable Analysis: A Practical Guide for Clinicians**. Cambridge: Cambridge University Press, 1999. 208 p.

KLECKA, William R. **Discriminant analysis**. Londres: Sage Publications, 2011. 71 p.

LEITE, Marcos Esdras. **Análise da correlação entre dengue e indicadores sociais a partir do sig.** 2010. Disponível em: <http://www.seer.ufu.br/index.php/hygeia/article/view/16981/9367>. Acesso em: 20 fev. 2019.

LOPES, Manuela de Mesquita; CASTELO BRANCO, Verônica T. F.; SOARES, Jorge Barbosa. **Utilização dos testes estatísticos de Kolmogorov-Smirnov e Shapiro-Wilk para verificação da normalidade para materiais de pavimentação**. [S. l.], 2013. Disponível em: <https://www.revistatransportes.org.br/anpet/article/download/566/467>. Acesso em: 20 fev. 2019.

MAHMOOD, T. et al. **Raman spectral analysis for rapid screening of dengue infection**. 2018. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1386142518303172>. Acesso em: 20 fev. 2019.

MAROCO, J. **Análise Estatística com a utilização do SPSS**. 3ª Ed. Silabo. Lisboa. 2007 pp.

MIEKO, Daniela. **Avaliação do sexo por análise de função discriminante a partir de dimensões lineares do crânio**. Piracicaba, 2000. Disponível em: [http://repositorio.unicamp.br/bitstream/REPOSIP/289046/1/Abe\\_DanielaMieko\\_M.pdf](http://repositorio.unicamp.br/bitstream/REPOSIP/289046/1/Abe_DanielaMieko_M.pdf). Acesso em: 20 fev. 2019.

MINISTÉRIO DA SAÚDE, Agência Saúde. **Mais de mil cidades podem ter surto de dengue, zika e chikungunya**. 2018. Disponível em: <http://portalmms.saude.gov.br/noticias/agencia-saude/43454-brasil-pode-ter-aumento-de-casos-de-dengue-zika-e-chikungunya>. Acesso em: 20 fev. 2019.

MINISTÉRIO DA SAÚDE. **Diretrizes Nacionais para a Prevenção e Controle de Epidemias de Dengue**. Brasília, 2009. Disponível em: [http://bvsmms.saude.gov.br/bvs/publicacoes/diretrizes\\_nacionais\\_prevencao\\_controle\\_dengue.pdf](http://bvsmms.saude.gov.br/bvs/publicacoes/diretrizes_nacionais_prevencao_controle_dengue.pdf). Acesso em: 20 fev. 2019.

MINISTÉRIO DA SAÚDE. **Plano de Contingência Nacional para Epidemias de Dengue**. Brasília, 2015. Disponível em: <http://portalarquivos2.saude.gov.br/images/pdf/2015/janeiro/20/plano-contingencia-dengue-19jan15-web.pdf>. Acesso em: 20 fev. 2019.

MINISTÉRIO DA SAÚDE, Secretaria de Vigilância em Saúde. **Boletim Epidemiológico de número 34**. Volume 49. 2018. Disponível em: <http://portalarquivos2.saude.gov.br/images/pdf/2018/agosto/21/Publicacao-BE-2018-SE-30.pdf>. Acesso em: 20 fev. 2019.

NARESH K., Malhotra. **Investigación de mercados**. 5. ed. Cidade do México: Pearson, 2008. 920 p. Disponível em: <http://www.cars59.com/wp-content/uploads/2015/09/Investigacion-de-Mercados-Naresh-Malhotra.pdf>. Acesso em: 20 fev. 2019.

NASEER, Khulla et al. **Raman spectroscopy based differentiation of typhoid and dengue fever in infected human sera.** 2018. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/30114579>. Acesso em: 20 fev. 2019.

NIST/SEMATECH. **E-Handbook of Statistical Methods.** [S. l.], 2012. Disponível em: <https://www.itl.nist.gov/div898/handbook/eda/section3/eda33o.htm>. Acesso em: 20 fev. 2019.

NÓBREGA, Diogo Medeiros. **Análise Discriminante Utilizando o Software SPSS.** 2010. 56 p. Monografia (Bacharelado em Estatística)- UEPB, Campina Grande PB, 2010. Disponível em: <http://dspace.bc.uepb.edu.br/jspui/bitstream/123456789/498/1/PDF%20Diogo%20Medeiros%20N%C3%B3brega.PDF>. Acesso em: 20 fev. 2019.

OLIVEIRA, Rhaquel; ARAÚJO, Fernanda; CAVALCANTI, Luciano. **Aspectos entomológicos e epidemiológicos das epidemias de dengue em Fortaleza, Ceará, 2001-2012.** Disponível em: <http://www.scielo.br/pdf/ress/v27n1/2237-9622-ress-27e201704414.pdf>. Acesso em: 20 fev. 2019.

PIGNATTI, Marta Gislene. **Políticas ambientais e saúde: as práticas sanitárias para o controle do dengue no ambiente urbano.** Disponível em: [http://www.anppas.org.br/encontro\\_anual/encontro2/GT/GT12/gt12\\_marta\\_pignatti.pdf](http://www.anppas.org.br/encontro_anual/encontro2/GT/GT12/gt12_marta_pignatti.pdf). Acesso em: 20 fev. 2019.

PINO, Francisco Alberto. **A Questão da Não Normalidade: uma revisão.** São Paulo, 2014. Disponível em: <http://www.iea.sp.gov.br/ftp/iea/publicar/rea2014-2/rea2-22014.pdf>. Acesso em: 20 fev. 2019.

RÊGO, Nathielly Lima *et al.* **Aplicação do teste de igualdade de matrizes de variância/covariância (M de Box), aos dados da população Europeia.** Campina Grande, 2012. Disponível em: [http://editorarealize.com.br/revistas/enect/trabalhos/Comunicacao\\_758.pdf](http://editorarealize.com.br/revistas/enect/trabalhos/Comunicacao_758.pdf). Acesso em: 20 fev. 2019.

RESTI, Yulia et al. **Quadratic Discriminant Analysis of Dengue Viruses Disease Incidence in Palembang.** 2017. Disponível em: <https://thescipub.com/pdf/10.3844/ajassp.2017.578.582>. Acesso em: 20 fev. 2019.

RIBEIRO, Júlia Werneck; ROOKE, Juliana Maria Scoralick. **Saneamento básico e sua relação com o meio ambiente e a saúde pública**. Juiz de Fora - MG, 2010. Disponível em: <http://www.ufjf.br/analiseambiental/files/2009/11/TCC-SaneamentoeSa%25C3%25BAde.pdf>. Acesso em: 20 fev. 2019.

ROCHA, Hermano; CARVALHO, Eduardo. **Conceitos Básicos em Epidemiologia e Bioestatística**. Disponível em: <http://www.epidemiologia.ufc.br/files/ConceitosBasicosemEpidemiologiaeBioestatistica.pdf>. Acesso em: 20 fev. 2019.

ROGERS, David J. et al. **Using global maps to predict the risk of dengue in Europe**. 2017. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/23973561>. Acesso em: 20 fev. 2019.

SANTOS, Marfran C. D. et al. **ATR-FTIR spectroscopy coupled with multivariate analysis techniques for the identification of DENV-3 in different concentrations in blood and serum: a new approach**. 2017. Disponível em: <https://pubs.rsc.org/en/content/articlelanding/2017/ra/c7ra03361c#!divAbstract>. Acesso em: 20 fev. 2019.

SILVEIRA, Lucia T. C. da. **Revisão sistemática da eficácia da vacina da dengue**. Rio de Janeiro, 2016. Disponível em: <http://www.mestradoinc.com.br/wp-content/uploads/2019/01/REVIS%C3%83O-SISTEM%C3%81TICA-DA-EFIC%C3%81CIA-DA-VACINA-DA-DENGUE.pdf>. Acesso em: 20 fev. 2019.

SHARMA, S. 1996. **Proactive environmental responsiveness: Catalysts for changing organizational paradigms and organizational capabilities**. BPS, ONE, and SIM. Acade of Management Annual Meetings. 1996.

STEVENS, J.P. **Applied Multivariate Statistics for the Social Sciences**. Lawrence Erlbaum Associates, Mahwah, NJ. 2002.

TANNOUS, Isabele P. Perfil epidemiológico e geográfico da infecção pelo vírus da dengue em um município do sudoeste de goiás: um estudo transversal. Jataí - GO, 2018. Disponível em: [https://cienciasaplicasaude.jatai.ufg.br/up/374/o/Disserta%C3%A7%C3%A3o\\_Biblioteca\\_2018\\_%281%29.pdf](https://cienciasaplicasaude.jatai.ufg.br/up/374/o/Disserta%C3%A7%C3%A3o_Biblioteca_2018_%281%29.pdf). Acesso em: 20 fev. 2019.

VANDENBROUCKE, Jan P; BROADBENT, Alex; PEARCE, Neil. **Causality and causal inference in epidemiology: the need for a pluralistic approach**. Denmark, 22 jan. 2019. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5841832/>. Acesso em: 20 fev. 2019.

WHITAKER, Jean S. **Use of Stepwise Methodology in Discriminant Analysis**. Texas A&M University, 1997. Disponível em: <http://ericae.net/ft/tamu/STEPWIS.htm>. Acesso em: 20 fev. 2019.

YANG, Xiaofeng et al. **Dengue Fever Detecting System Using Peak-detection of Data from Contactless Doppler Radar**. 2018. Disponível em: <https://ieeexplore.ieee.org/document/8512355>. Acesso em: 20 fev. 2019.