

UNIVERSIDADE ABERTA



Modelos de Análise e Visualização de dados de COVID-19 em Moçambique

Francisco Alberto Arune

Mestrado em Bioestatística e Biometria 2019/2021

2025

UNIVERSIDADE ABERTA



Modelos de Análise e Visualização de dados de COVID-19 em Moçambique

Francisco Alberto Arune

Mestrado em Bioestatística e Biometria na área de especialização de Estatística aplicada às Ciências da Vida e da Saúde, utilizando recursos computacionais.

Dissertação orientada pela Prof.^a Doutora Filomena Teodoro e Coorientada pela Prof.^a Doutora Teresa Paula Costa Azinheira Oliveira

2025

Resumo

Na China, país originário da doença, até dia 23 de Fevereiro de 2020 foram notificados mais de 77.000 casos de infeção pela doença, e da análise feita 60% dos casos confirmados foram reportados na cidade de Wuhan. Moçambique declarou estado de emergência e diferentes medidas de prevenção estão sendo implementadas para controlar e dar resposta atempada à pandemia, incluindo o diagnóstico precoce de casos da doença. O presente trabalho tem como tema “*Modelos de Análise e visualização de dados de COVID-19 em Moçambique*” com objetivo central de descrever os modelos de análise e visualização de dados de COVID-19 em Moçambique. O tema enquadra-se na área de Estatística com propósito de trazer evidências que expliquem o atual estágio (desde a notificação do primeiro caso de COVID-19 em Moçambique (22 de Março de 2020) até o dia 31 de Maio de 2022) do país no concernente a evolução de casos de COVID-19, com enfoque para as províncias de Maputo, Nampula, Cabo Delgado e Niassa. O trabalho utilizará tipo de dados qualitativos e quantitativos para permitir a tomada de decisões na área de saúde sobre as medidas de prevenção da pandemia e a tendência dos casos e óbitos pela doença.

Palavras-chave: *Modelos de análise, visualização de dados e COVID-19.*

Abstract

In China, the country of origin of the disease, until February 23, 2020, more than 77,000 cases of infection with the disease were reported, and from the analysis carried out, 60% of confirmed cases were reported in the city of Wuhan. Mozambique has declared a state of emergency and different prevention measures are being implemented to control and respond in a timely manner to the pandemic, including the early diagnosis of cases of the disease. The present work has as its theme "Models of Analysis and Visualization of COVID-19 Data in Mozambique" with the central objective of describing the models of analysis and visualization of COVID-19 data in Mozambique. The topic falls within the area of Statistics with the purpose of providing evidence that explains the current stage (from the notification of the first case of COVID-19 in Mozambique (March 22, 2020) until May 31, 2022) of the country regarding the evolution of COVID-19 cases, with a focus on the provinces of Maputo, Nampula, Cabo Delgado and Niassa. The work will use qualitative and quantitative data to allow decision-making in the health area on measures to prevent the pandemic and the trend of cases and deaths from the disease.

Keywords: *Analysis models, data visualization and COVID-19.*

Dedicatória

A meus Pais

Alberto Arune e Luísa Domingos, pois a minha existência na terra é da vossa inteira vontade e pela vossa educação “serena, célere e rígida” que hoje fez de mim um homem novo. Muito obrigado,

A minha Esposa

Paneshia Silvestre João Arune, agradeço a sua viva coragem e dedicação em me confortar nos momentos de fraqueza. Obrigado,

A meus irmãos

Agradeço a vossa amizade e maior força que sempre dedicaram à minha vida. Obrigado,

A meus amigos

Aboo Camacho e Inácio Retene. Vocês tornaram-se a minha família.

Agradecimentos

A Universidade Aberta e a todos os docentes, em particular e de forma carinhosa as Professoras Filomena Teodoro e Teresa Paula Costa Azinheira Oliveira, orientadora e coorientadora respetivamente, pela paciência e carinho para dar seu suporte e apoio incondicional durante o desenvolvimento do trabalho.

A Direção do Serviço Provincial de Saúde de Nampula – SPSN e Ministério da Saúde - MISAU, por proporcionarem condições para a obtenção de dados que permitiram a efetivação desta dissertação.

A minha esposa (Paneshia), meu muito obrigado.

Ao meu irmão, Carlitos Arune, aos amigos, especialmente, Aboo Camacho pela preocupação, sugestões nos momentos de descontração.

Endereço a todos, o meu singelo agradecimento!

Obrigado,

Francisco Alberto Arune

Índice

[1].	Introdução.....	12
1.1.	Objetivos.....	15
[2].	Metodologia.....	16
2.1.	<i>Variáveis de base</i>	16
2.2.	<i>Amostra</i>	17
2.3.	<i>Descrição da Base de Dados</i>	17
[3].	Revisão da Literatura.....	19
3.1.	<i>Implicações epidemiológicas dos padrões de contacto social</i>	19
3.2.	Modelos de Análise de dados da COVID-19 em Moçambique.....	20
3.2.1.	<i>Modelos Estatísticos de Análise de dados</i>	20
3.2.1.1.	<i>Visão Geral dos Modelos Lineares</i>	26
[a]	Modelos de Regressão.....	26
[b]	Regressão Linear Simples.....	27
[c]	Regressão Linear Múltipla.....	29
[d]	Regressão Logística.....	32
C.1.	<i>Regressão logística univariável</i>	33
C.2.	<i>Regressão logística multivariável</i>	33
C.3.	<i>Odds Ratio</i>	34
C.4.	<i>Função Logística</i>	34
[e]	Modelos Autorregressivos.....	34
[f]	Modelos Médias - Móveis.....	34
3.2.2.	<i>Análise de dados da COVID-19 em Moçambique</i>	35
[a]	Modelos matemáticos de Análise de dados da COVID-19 em Moçambique.....	36
[4].	Caso de estudo – Aplicação.....	40
4.1.	Modelos de Visualização de dados.....	40
4.1.1.	<i>Visualização</i>	40
4.1.2.	<i>Visualização de Dados</i>	40
4.1.2.1.	<i>Infografia</i>	43
4.1.3.	Técnicas de Visualização de Dados.....	45

4.1.3.1.	Dados espaciais	50
[a]	<i>Software SIG</i>	51
[b]	<i>Dados Espaciais</i>	51
[c]	<i>Visualização de dados espaciais</i>	52
4.2.	Resultados	54
4.2.1.	Modelos de Análise de dados do COVID-19 em Moçambique	54
4.2.2.	Visualização de dados do COVID-19 em Moçambique	66
4.3.	Conclusões	88
4.4.	Perspetivas	89
4.5.	Considerações para o futuro	90
[5].	Referências Bibliográficas	92

Índice de gráficos/quadros/tabelas

<i>Gráfico 01: Exemplo dum pirâmide Etária</i>	49
<i>Gráfico 02: Exemplo dum mapa Coroplético</i>	52
<i>Gráfico 03: Exemplo dum mapa de calor de densidade de caso</i>	53
<i>Gráfico 04: Exemplo dum mapa de símbolos proporcionais</i>	54
<i>Gráfico 05: Análise gráfica dos pressupostos do modelo linear</i>	59
<i>Gráfico 06: Análise da multicolinearidade através do gráfico</i>	60
<i>Gráfico 07: Dados Cumulativos de Casos Confirmados Positivos</i>	64
<i>Gráfico 08: Série Temporal de Casos Positivos de COVID-19</i>	65
<i>Gráfico 09: Distribuição de Casos Positivos Confirmados de COVID-19 por Província</i>	65
<i>Gráfico 10: Distribuição de Casos Positivos de COVID-19 por Área Geográfica</i>	66
<i>Gráfico 11: Distribuição de Casos Positivos Confirmados de COVID-19 por Género</i>	66
<i>Gráfico 12: Dados Cumulativos de Casos Confirmados Positivos</i>	71
<i>Gráfico 13: Série Temporal de Casos Positivos de COVID-19</i>	72
<i>Gráfico 14: Distribuição de Casos Positivos Confirmados de COVID-19 por Província</i>	73
<i>Gráfico 15: Distribuição de Casos Positivos de COVID-19 por Área Geográfica</i>	74
<i>Gráfico 16: Distribuição de Casos Positivos Confirmados de COVID-19 por Género</i>	75
<i>Tabela 01: Dados da pandemia da COVID-19</i>	62
<i>Tabela 2: Distribuição de Casos por Província</i>	69
<i>Tabela 3: Distribuição de Casos Positivos Vs Peso da positividade por província</i>	69
<i>Tabela 04: Distribuição de Casos por área geográfica (Província) e Género</i>	70

Índice de figuras

<i>Figura 01: Esquema do modelo compartimental SIR</i>	63
<i>Figura 02: Apresentação da Situação Epidemiológica das Províncias de Maputo, Nampula e Cabo Delgado - COVID-19</i>	68
<i>Figura 03: Áreas de Alto Risco de Transmissão do COVID-19</i>	76
<i>Figura 04: Áreas de baixo e/ou medio Risco de Transmissão da COVID-19</i>	77
<i>Figura 05: Distribuição Geográfica de Casos e Níveis de Risco de transmissão</i>	78
<i>Figura 06: Distribuição Geográfica de Casos de COVID-19 no País – Moçambique</i>	79
<i>Figura 07: Distribuição Geográfica de Casos de COVID-19 – Densidade da População Infetada</i>	80
<i>Figura 08: Distribuição Geográfica de Casos de COVID-19 – Concentração de Casos em Coroplético</i>	81
<i>Figura 09: Distribuição Geográfica de Casos e Níveis de Risco de transmissão</i>	82
<i>Figura 10: Ilustração gráfica da Pirâmide Etária dos dados cumulativos das 04 províncias em análise</i>	83
<i>Figura 11: Ilustração gráfica da Pirâmide Etária de casos de COVID-19 na província de Cabo Delgado</i>	84
<i>Figura 12: Ilustração gráfica da Pirâmide Etária de casos de COVID-19 na província do Niassa</i>	85
<i>Figura 13: Ilustração gráfica da Pirâmide Etária de casos de COVID-19 na província de Maputo</i>	85
<i>Figura 14: Ilustração gráfica da Pirâmide Etária de casos de COVID-19 na província de Nampula</i>	86

Lista de abreviaturas, siglas e acrónimos

INS	Instituto Nacional de Saúde
OMS	Organização Mundial da Saúde
OMR	Observatório do Meio Rural
COVID-19	Coronavírus - 2019
VIF	Fator de Inflação de Variância
MISAU	Ministério da Saúde
INGD	Instituto Nacional de Gestão e Redução do Risco de Desastres
SIR	Suscetíveis, Infetados e Recuperados
SEIR	Suscetíveis, Expostos, Infetados e Recuperados

[1]. Introdução

Os coronavírus são uma família de vírus conhecidos desde os meados dos anos 1960. Estes vírus causam infecções respiratórias em seres humanos e em animais. Geralmente as infecções por coronavírus causam doenças leves às moderadas - resfriado comum (Instituto Nacional de Saúde - INS, 2020).

A Organização Mundial da Saúde (OMS) foi informada a 31 de Dezembro de 2019 sobre a existência de casos de pneumonia de origem não identificada na China, Província de Hubei, concretamente na cidade de Wuhan. Nas semanas subsequentes o agente etiológico foi identificado como um novo coronavírus designado por SARS-CoV-2 e a doença causada por esse agente denominada por COVID-19. Nos primeiros meses do ano 2020, casos de COVID-19 foram identificados em vários países do mundo e no dia 11 de Março de 2020 o surto desta doença foi declarada uma pandemia pela OMS (MF Teodoro et al., 2023).

O surto da doença de coronavírus 2019 (COVID-19) no mundo foi pela primeira vez diagnosticado na cidade de Wuhan, província de Hubei, na China e se espalhou de forma galopante para todas as províncias da China, incluindo 28 outros países do mundo. Na China, país originário da doença, até dia 23 de Fevereiro de 2020 foram notificados mais de 77.000 casos de infeção pela doença, e da análise feita 60% dos casos confirmados foram reportados na cidade de Wuhan.

Quando a doença de coronavírus abreviadamente chamada por COVID-19 foi declarada como pandemia pela OMS, em todo o mundo várias estratégias foram desenhadas e em particular Moçambique com vista à prevenção da doença, avançou com o confinamento, uso de máscaras, lavagem das mãos com água e sabão bem como o uso de álcool em gel para a descontaminação das mãos como medidas de combate à pandemia.

No sentido de garantir o controlo da pandemia e redução da propagação do vírus corona, vários países incluindo Moçambique declararam estado de emergência e diferentes medidas de prevenção estão sendo implementadas para controlar e dar resposta atempada à pandemia, incluindo o diagnóstico precoce de casos da doença.

Em Moçambique caso particular, há que destacar várias medidas de prevenção como a alocação de baldes com água e sabão para a lavagem das mãos em todas as portas de espaços públicos, uso de máscaras de proteção, distanciamento social, desinfeção das

mãos com álcool em gel para controlar a propagação da pandemia e consequente redução das mortes pela doença (MF Teodoro et al., 2024).

O primeiro caso notificado da doença em Moçambique, foi registado na cidade de Maputo, capital do país, no dia 22 de Março de 2020, num indivíduo do género masculino, com mais de 75 anos, de nacionalidade moçambicana, que retornou de uma viagem ao Reino Unido em meados de Março.

Trata-se, portanto, de um caso importado de infeção pelo novo Coronavírus, isto é, diagnosticado em pessoas que entraram no país vindo do estrangeiro (países de alta endemicidade) e por conta da fraca vigilância, falta de meio de diagnóstico e reduzidos mecanismos de proteção, proporcionou até então o alastramento da doença em todo o país com maior incidência para as cidades de Pemba, Nampula, Matola e Maputo Cidade.

Na sequência da dispersão acelerada de casos de COVID-19, o país registou o primeiro óbito pela doença no dia 25 de Maio 2020, em Nampula, na província da zona Norte de Moçambique. Foi uma criança de 13 anos, cuja amostra foi colhida no dia 20 de Maio naquela província (MF Teodoro et al., 2024).

A intensificação do terrorismo (insurgências) em Cabo Delgado gerou um desastre humanitário traduzido em centenas de milhares de famílias deslocadas, que procuraram, em função das suas condições, locais mais seguros e de integração social para se acomodarem. A cidade de Pemba foi o segundo local em Moçambique a ser considerado como um sítio com transmissão comunitária do novo coronavírus, em 21 de Junho, depois da cidade de Nampula, ambas províncias da zona Norte de Moçambique (OMR, 2022).

A situação de terrorismo em Cabo delgado provocou movimentos migratórios da população nos distritos de Palma, Mocímboa da Praia, Macomia e Quissanga, que por consequência, gerou mudanças imprevisíveis no quadro epidemiológico da doença nos distritos e províncias de reassentamento - comportamento anormal, mudanças extremas do fluxo da população, etc. (OMR, 2022).

A riqueza dos dados de deslocados dos conflitos terrorista em Cabo Delgado, é de permitir comparar a situação em diferentes pontos (província/distritos) com diferentes níveis de risco (comportamento normal e extremo) proporcionando informação útil sobre o estágio epidemiológico de COVID-19 na área de estudo.

As organizações humanitárias que apoiam os deslocados de conflitos terroristas em Cabo Delgado efetuaram levantamento de dados para ajuda humanitária que revelaram a existência de aproximadamente 660.190 deslocados, sendo na cidade de Pemba (151.553 indivíduos), Metuge-Sede (90.274 indivíduos) e Mueda-Sede (49.805 indivíduos) os locais com maior concentração de deslocados, acolhendo, no seu conjunto, 44,1% da população deslocada por este conflito (OMR, 2022).

Segundo o Instituto Nacional de Gestão e Redução do Risco de Desastres (INGD), a população deslocada por este conflito distribui-se pela província de Nampula – particularmente nos distritos de Meconta (20.229 indivíduos) e Cidade de Nampula (19.478 indivíduos). A forte concentração da população em áreas municipais (Pemba, Mueda, Nampula, Montepuez ou Ibo) traduz-se na densificação populacional, saturação de infraestruturas urbanas e serviços públicos (educação e saúde), no desemprego e informalização da economia (INGD, 2021 in OMR, 2022).

A maior concentração populacional em locais de reassentamento dos deslocados de ataques dos terroristas traz consigo grandes desafios para os sectores sociais, criando uma pressão sobre os serviços de saneamento, educação e saúde para garantir uma vida plena aos imigrantes, sem se esquecendo que estamos a viver um período de novo normal (período de fraca mobilidade de pessoas – recolha obrigatória) devido à pandemia da COVID-19, colocando a saúde das pessoas em alto risco e vulnerabilidade da doença.

Para uma robusta apresentação espacial dos dados do problema em estudo, para além de diferentes ferramentas que permitem a manipulação de dados, é fundamental o uso dos SIG pela capacidade de reunir uma abundância de dados convencionais de expressão espacial, proporcionando a estruturação e integração adequada para a manipulação das informações geográficas.

Na pesquisa bibliográfica realizada distinguiram-se alguns trabalhos na área da epidemiologia, como são MF Teodoro et al. (2023), MF Teodoro et al. (2024), Taero (2023).

1.1.Objetivos

Objetivo Geral:

- Identificar e selecionar modelos de análise e mecanismo de visualização de dados de COVID-19 em Moçambique.

Objetivos Específicos:

- Explorar modelos teóricos e matemáticos que permitam a análise de dados de COVID-19;
- Aplicar diferentes mecanismos de visualização de dados de COVID-19 em Moçambique;
- Identificar um modelo matemático que permita analisar a tendência de casos da pandemia de uma forma otimizada.

[2]. Metodologia

Referente aos procedimentos técnicos foi efetuada uma pesquisa bibliográfica uma vez que serão consultadas várias obras que abordaram conteúdos sobre o tema em estudo e a pesquisa documental onde se fará consultas de vários relatórios sobre a pandemia de COVID-19 no mundo, no continente africano e em particular em Moçambique.

As pesquisas baseiam-se em fenómenos qualitativos e quantitativos. Na essência, a pesquisa quantitativa utiliza parâmetros estatísticos para a análise dos dados, onde a informação no seu todo é transformada em números; enquanto a pesquisa qualitativa usa a subjetividade que não pode ser traduzida em números, pois é mais descritiva.

A pesquisa qualitativa preocupa-se com aspetos da realidade que não pode ser quantificada, centrando-se na compreensão e explicação da dinâmica das relações sociais. A pesquisa qualitativa centra-se no universo de significados, motivos, aspirações, crenças, valores e atitudes, o que corresponde a um espaço mais profundo das relações, dos processos e dos fenómenos que dificilmente podem ser reduzidos à operacionalização de variáveis.

Os dados de COVID-19 reportados e publicados pelo Ministério da Saúde de Moçambique foram utilizados como fonte primária de dados e mecanismos de obtenção de resultados sobre a pandemia.

2.1. Variáveis de base

Para entender corretamente a disseminação de COVID-19, identificar tendências, avaliar o impacto da doença nas comunidades e orientar a tomada de decisões em saúde pública, no presente trabalho foram analisadas algumas variáveis de base como:

- Idade
- Sexo
- Nacionalidade
- Localização (zona urbana ou rural)
- Caso (Positivo ou Negativo)
- Medidas de prevenção (disponível ou não)

2.2. Amostra

Para a dissertação foi tomada como amostra quatro das onze províncias do país, sendo Niassa, Cabo-Delgado, Nampula e Província de Maputo, para permitir a análise e comparação do perfil epidemiológico dos casos da doença em Moçambique. O critério de inclusão e exclusão das outras províncias enquadra-se no raciocínio do perfil epidemiológico com transmissão comunitária e na evolução de casos da COVID-19. A situação de Cabo-Delgado poderá apresentar um padrão diferente devido às insurgências.

As quatro províncias selecionadas como amostra para a dissertação apresentaram maior número de casos e óbitos pela doença e um perfil com tendência crescente até um determinado período e logo a seguir a alteração do perfil apresentando uma curva de desaceleração de casos. Por outro lado, as condições climatéricas de cada província permitirão uma análise sucinta do problema em estudo. O presente trabalho baseou-se em dados reportados desde a notificação do primeiro caso de COVID-19 em Moçambique (22 de Março de 2020) até o dia 31 de Maio de 2022, considerando este período como de coorte para o estudo.

Com o registo da nova vaga da doença reportado em meados de mês de Janeiro de 2021, as províncias em referência registaram um aumento de novas infeções da COVID-19 e aumento galopante de casos e óbitos pela doença nos dois primeiros meses de 2021, tomando um novo perfil epidemiológico no que diz respeito a registos de novos casos de pessoas infetadas pela COVID-19.

Nestes termos, o estudo tomou como base a análise de dados da pandemia do COVID-19 desde o registo do primeiro caso da doença (2020) até Abril de 2021 (período que iniciou a vacinação contra pandemia).

2.3. Descrição da Base de Dados

Base de Dados: *Base_Dados_FINAL _ Dissertação (2020 - 2022)*.

A " *Base_Dados_FINAL _ Dissertação (2020 - 2022)*." é uma base de dados compilada e normalizada pelo autor do trabalho com base nos dados obtidos pelos Serviços Provinciais de Saúde das 04 províncias em estudo, com o objetivo de identificar e selecionar modelos de análise e mecanismo de visualização do objeto em estudo. Este

conjunto de dados abrange um período de dois anos, de 2020 a 2022, e oferece uma visão abrangente de várias variáveis relacionadas à saúde da população.

Características Principais:

A base de dados inclui informações demográficas, condição de saúde e outros fatores relevantes para a compreensão do panorama geral do COVID-19. As variáveis disponíveis abrangem uma ampla gama de tópicos, desde idade e gênero até a localização.

Metodologia de Coleta:

Os dados foram obtidos por meio de registos diários em livros de sistemas de informação de saúde em uso no país, bem como entrevistas domiciliares, conduzidas por profissionais de saúde de linha de frente (destacados para o seguimento de casos), utilizando fichas de notificação de casos padronizadas.

Formato dos Dados:

A base de dados está disponível em formatos acessíveis (Excel e arquivo CSV) facilitando sua utilização em diversas plataformas e ambientes de análise.

Acesso aos Dados:

Os dados podem ser acessados diretamente no site oficial do Ministério da Saúde, bem como pelos Departamentos de Planificação e Cooperação das instituições tuteladas pela área de saúde nas províncias. No site, os usuários encontrarão informações sobre a documentação técnica e instruções claras sobre como baixar os dados para análise e relatórios da pandemia de COVID-19.

Esta base de dados serve como uma valiosa fonte de informações para pesquisadores, profissionais de saúde e acadêmicos interessados em explorar e compreender os diversos aspetos relacionados à saúde da população no contexto da pandemia durante o período considerado.

[3]. **Revisão da Literatura**

A pandemia de COVID-19 é agora uma grande ameaça à saúde pública no global, com 332.930 casos e 14.510 mortes confirmadas em todo o mundo em 23 de Março de 2020. Desde a identificação inicial na China, a disseminação da doença foi rápida, com 182 dos 202 países relatando pelo menos um caso. A experiência nos países até ao momento enfatizou a intensa pressão que a pandemia de COVID-19 coloca nos sistemas nacionais de saúde, com demanda por leitos de terapia intensiva e ventiladores mecânicos (Imperial College COVID-19 Response Team, 2020).

Vários governos de países fustigados pela pandemia da COVID-19 foram introduzindo políticas e estratégias que visam combater e controlar o novo normal da doença, partindo dum modelo de análise do comportamento de casos da pandemia com pressupostos de distanciamento social, isolamento, bem como seguimento de casos positivos e seus contactos (Taero, E et al., 2025).

Não obstante, não é suficiente analisar a problemática da pandemia de COVID-19 olhando apenas as atitudes e comportamento da população suscetível e exposta à doença, mas também é fundamental analisar o problema em estudo na ótica ambiental, isto é, os modelos de análise devem incorporar como variáveis a região (espaço físico) e a temperatura, entre outras.

Muitas incertezas existem sobre os determinantes subjacentes da gravidade da infeção da COVID-19. No entanto, os fatores de risco muito claros sobre a doença incluem idade, com idosos mais propensos a necessitar de hospitalização e subsequentemente morrer como resultado de infeção e comorbidades subjacentes, incluindo hipertensão, diabete e doenças coronárias que permitem exacerbar os sintomas (Imperial College COVID-19 Response Team, 2020).

3.1. Implicações epidemiológicas dos padrões de contacto social

As consequências epidemiológicas dos padrões estimados de mistura são avaliadas simulando a distribuição etária dos indivíduos infetados durante a fase inicial de uma epidemia em uma população totalmente suscetível. Em particular, essa distribuição foi obtida computando o vetor próprio da matriz da próxima geração associada aos contatos, considerada para qualquer idade possível do caso-índice na comunidade, e assumindo

uma taxa de transmissão independente da idade por contacto, sob o chamado "Hipótese do contacto social (Imperial College COVID-19 Response Team, 2020)".

Segundo a mesma fonte, vale ressaltar que a distribuição etária dos casos durante a fase inicial da epidemia não depende nem da escolha da duração do período de infecciosidade, nem do valor considerado do número reprodutivo básico. Isso significa que o momento e a gravidade associados a uma infeção são geralmente específicos da doença, embora o impacto dos padrões de mistura na distribuição etária dos casos seja causado principalmente pelo tipo de contactos relevantes para a infeção, transmissão de infeções e estrutura sociodemográfica da população considerada.

3.2. Modelos de Análise de dados da COVID-19 em Moçambique

3.2.1. Modelos Estatísticos de Análise de dados

De diversas formas podem ser analisados os dados. De uma forma pormenorizada podemos os caracterizar conforme a característica temporal (estática ou dinâmica), distribuição, dimensão, sua natureza (hidráulico, geológico, contabilístico, linguístico, etc.), localização geográfica, entre outras formas.

Atualmente, existe uma gama de técnicas que permitem a modelação e visualização de dados utilizadas para mapear a informação. Na modelagem de dados, é prioritário antes da escolha das técnicas a serem utilizadas para a análise de dados a identificação do modelo matemático adequado. No entanto, o modelo deve “traduzir” de forma conveniente a relação existente entre as variáveis e por conseguinte, do mundo real e os respetivos dados.

É fundamental na modelação estatística (científica) a eliminação dos eventuais ruídos (viés nos dados com características anormais quando analisados ao longo do tempo) através da filtração, limpeza e normalização da base de dados.

Em problemas onde o objetivo é analisar e estudar a relação existente entre as variáveis (preditiva e preditora), são utilizados modelos de regressão. Existem várias metodologias de análise estatística que permitem explicar ou descrever a relação entre uma variável de interesse (variável resposta) e uma ou mais variáveis (variáveis explicativas).

Para entender e prever tendência dos casos de COVID-19, existem vários modelos de análise de dados que podem ser aplicados. Abaixo ir-se-á apresentar alguns dos modelos mais comuns e eficazes que podem ser utilizados para analisar os dados de COVID-19.

Modelos Epidemiológicos SIR e Variantes (SEIR, SIRD, etc.): esses modelos são usados para entender a propagação de doenças infecciosas em uma população. Eles dividem a população em compartimentos (Suscetíveis, Expostos, Infetados, Recuperados) e usam equações diferenciais para modelar as taxas de transmissão, incubação, recuperação, etc. Esses modelos são usados para prever a evolução da doença, entender o impacto de medidas de controle (como isolamento) e estimar a taxa de reprodução básica (Anderson & May, 1992).

Equações básicas diferenciais para o modelo SIR (*Suscetíveis – Infetados – Recuperados*):

$$1. \frac{dS}{dt} = -\beta \frac{SI}{N} \quad \mathbf{1}$$

$$2. \frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I \quad \mathbf{2}$$

$$3. \frac{dR}{dt} = \gamma I \quad \mathbf{3}$$

Onde:

- S – É o número de indivíduos suscetíveis
- I - É o número de indivíduos infetados
- R - É o número de indivíduos recuperados
- N – É a população total ($N=S+I+R$)
- β – é a taxa de contacto (quantas pessoas suscetíveis um infetado entra em contacto por unidade de tempo)
- γ - é a taxa de recuperação

Modelos de Crescimento Exponencial e Logístico: esses modelos são frequentemente usados para prever a disseminação inicial de uma doença, como a COVID-19. Eles se baseiam em uma taxa de crescimento exponencial que eventualmente atinge um limite (modelo logístico) à medida que a população se torna suscetível (Malthus, 1798).

Modelos de Crescimento Exponencial

O crescimento exponencial é modelado pela equação diferencial:

$$\frac{dN}{dt} = \gamma N$$

4

Onde:

- N – É a quantidade (população, valores, etc.)
- t – É o tempo
- γ – É a taxa de crescimento

Modelos de Crescimento Logístico

$$\frac{dN}{dt} = \gamma N \left(1 - \frac{N}{K}\right)$$

5

Onde:

- N – É a quantidade (população, valores, etc.)
- t – É o tempo
- γ – É a taxa de crescimento
- K – É a capacidade de suporte (limite máximo que o sistema pode sustentar)

Modelos de Séries Temporais: esses modelos analisam dados ao longo do tempo para identificar tendências e padrões sazonais. Eles podem ser usados para prever números futuros de casos, óbitos ou outros indicadores relacionados à COVID-19 bem

como para entender e prever padrões em dados temporais, como a evolução de uma variável ao longo do tempo. (Durbin & Koopman, 2012).

A seguir, ir-se-á apresentar de forma geral alguns modelos comuns e suas equações associadas que são usados em análises de séries temporais.

Médias Móveis (MA): Valores observados em períodos anteriores.

$$X_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \dots + \theta_q\epsilon_{t-q}$$

6

Onde:

- X_t – É a observação no tempo t ,
- μ – É a média da série,
- ϵ_t – São erros brancos (independentes e identicamente distribuídos),
- $\theta_1, \theta_2, \dots, \theta_q$ – São os coeficientes dos termos de erro anteriores.

Autorregressivos (AR): Observações atuais como uma combinação linear de observações anteriores.

$$X_t = \phi_1X_{t-1} + \phi_2X_{t-2} + \dots + \phi_pX_{t-p} + \epsilon_t$$

7

Onde:

- $\phi_1, \phi_2, \dots, \phi_p$ – São os coeficientes autorregressivos
- ϵ_t – São erros brancos.

ARIMA (Médias Móveis Integradas Autorregressivas): Combinação de modelos para séries temporais estacionárias.

$$(1 - \phi_1B - \phi_2B^2 - \dots - \phi_pB^p)(1 - B)^dX_t = (1 + \theta_1B + \theta_2B^2 + \dots + \theta_qB^q)\epsilon_t$$

Onde:

- B – É o operador de diferenciação, e d é a ordem de diferenciação.

Modelos de Redes: esses modelos consideram a estrutura da rede social e de contatos para entender como a doença se espalha em comunidades interconectadas. Eles podem ser úteis para analisar a eficácia de medidas de distanciamento social (Newman, 2010).

A modelação matemática deste é dada pela equação seguinte:

$$P(G) = \prod_{(i,j) \in E^p} \prod_{(i,j) \notin E^{(1-p)}}$$

9

Onde:

- $P(G)$ – é a probabilidade de uma configuração particular da rede G ,
- E – É o conjunto de todas as arestas possíveis.

Modelos de Análise de Risco: esses modelos avaliam o risco de disseminação da doença com base em fatores demográficos, comportamentais, geográficos, entre outros (Ave & Vinnem, 2007).

$$\frac{dR}{dt} = -\gamma I$$

10

Onde:

- R - É o número de indivíduos recuperados,
- I - É o número de indivíduos infectados,
- γ - é a taxa de recuperação.

Modelos de Análise Espacial: usando técnicas de reprocessamento e SIG (Sistemas de Informação Geográfica), esses modelos ajudam a entender como a COVID-19 se espalha geograficamente e como fatores locais podem influenciar a disseminação (Diggle & Ribeiro, 2007).

Geralmente, a análise espacial de eventos epidemiológicos envolve o estudo de padrões e processos em relação à sua localização no espaço geográfico. Existem várias abordagens e modelos em análise espacial, e a equação específica dependerá do tipo de

modelo que está sendo utilizado para evidenciar o efeito da doença, com por exemplo, o método de interpolação espacial – Kriging (Oliveira et al, 2008).

O Kriging é um método de interpolação espacial que estima valores em locais não amostrados com base nas relações espaciais entre os pontos amostrados (Cressie, 1993).

Equação de Kriging:

$$Z(u) = \sum_{i=1}^n \lambda_i Z(x_i)$$

11

Onde:

- $Z(u)$ – é a estimativa do valor em um local não amostrado u ,
- $Z(x_i)$ – são os valores conhecidos nos locais de amostragem x_i ,
- λ_i – São os pesos atribuídos a cada ponto de amostra, determinados pela variância espacial e a covariância entre os pontos.

Em um modelo de regressão espacial, a relação entre a variável dependente e as variáveis independentes é ajustada levando em consideração a dependência espacial (Cressie, 1993).

Equação do Modelo de Regressão Espacial:

$$Y = X\beta + \rho WY + \varepsilon$$

12

Onde:

- Y – É o vetor de variáveis dependentes,
- X – É a matriz de variáveis independentes,
- β – é o vetor de coeficientes,
- W – É a matriz de pesos espaciais que captura a dependência espacial,
- ρ – é o parâmetro de auto-correlação espacial,
- ε – é o termo de erro.

No entanto, a escolha do melhor modelo depende dos objetivos da análise, dos dados disponíveis, da precisão desejada e do contexto específico da situação da COVID-19 na área de estudo. O uso holístico e combinado de vários modelos pode ajudar a entender perfeitamente o comportamento da doença, pois, nenhum modelo é perfeito.

3.2.1.1. Visão Geral dos Modelos Lineares

Uma análise de regressão linear produz uma equação matemática que chamamos de *modelo linear*, que estima uma variável dependente y a partir de um conjunto de variáveis preditoras ou repressoras x .

[a] Modelos de Regressão

Antes de aprofundar os modelos de regressão, é importante entender o termo abordado.

Regressão - é uma técnica estatística que busca identificar relações entre variáveis, permitindo a previsão ou a compreensão de como uma variável dependente é influenciada por uma ou mais variáveis independentes.

Na análise dos dados de COVID-19, os modelos de regressão podem ser aplicados de várias maneiras, como é o caso na previsão de casos, análise de fatores de risco, avaliação das intervenções desenvolvidas para o controlo da pandemia, análise do impacto das intervenções e entre outros aspetos que permitam a compreensão do fenómeno em estudo.

Abaixo irei caracterizar o contexto de cada abordagem:

- **Previsão de Casos:** os modelos de regressão podem ser usados para prever o número futuro de casos, óbitos ou outros indicadores relacionados à COVID-19 com base nos dados históricos. Essas previsões podem auxiliar na alocação de recursos, planeamento de medidas de saúde pública e tomada de decisões.
- **Análise de fatores de risco:** os modelos de regressão podem ajudar a identificar quais fatores (como idade, sexo, comorbidades, densidade populacional, medidas de distanciamento social) estão associados a um maior risco de infeção ou de

complicações da COVID-19. Isso é fundamental para proteger grupos mais vulneráveis.

- **Avaliação de intervenções:** os modelos de regressão podem ser usados para avaliar a eficácia de medidas de controle, como lockdowns, uso de máscaras ou campanhas de vacinação. Eles podem ajudar a entender como essas intervenções afetam a disseminação da doença.
- **Análise de impacto:** os modelos de regressão podem ser usados para avaliar o impacto de diferentes cenários ou variáveis no comportamento da doença. Isso é especialmente útil para antecipar o impacto de mudanças nas políticas de saúde.

No entanto, é importante ter em mente que os modelos de regressão têm limitações. Eles assumem relações lineares entre as variáveis, podem ser sensíveis a Outliers. Porém, os modelos de regressão é uma ferramenta valiosa para compreender os padrões da COVID-19, permitindo efetuar previsões que permitam a tomada de decisões baseada em evidências.

[b] Regressão Linear Simples

O modelo linear é o mais utilizado para modelar a relação entre as variáveis. Este modelo assume, de entre outras, o valor esperado da variável resposta é uma combinação linear das variáveis explicativas e que a variável resposta se reserva de margem de erro para explicar a influência que exerce sobre a outra, observando o pressuposto de que a outra variável segue a distribuição normal.

O modelo de regressão linear simples, pressupõe a existência de duas variáveis sendo uma independente (a considerada explicativa) e a outra dependente (a explicada), e é por isso que escrevemos $y = f(x)$. Quando são conhecidas as variáveis deve-se:

- *Escolher o atributo fatorial “de causa” e de resultado, baseando-se na experiência profissional ou numa análise qualitativa dos fenómenos ou variáveis que estão em estudo;*
- *Apresentar o diagrama de dispersão e escolher a partir deste, a forma da equação de regressão apropriada.*

O modelo geral da equação de regressão linear entre duas variáveis tem a seguinte forma:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

13

Onde:

y - É a variável dependente

β_0 - É o interceto ou a constante

β_1 - É a inclinação da reta

x - É o regressor ou variável independente na equação

ε - é o termo de erro e capta todos os fatores não observados no modelo

Na regressão linear simples exploram-se os conceitos e técnicas para se analisar e utilizar a relação linear entre duas variáveis. Esta análise conduz a uma equação que pode ser utilizada para se “predizerem” valores de uma variável dependente (a variável resposta) dados valores de uma variável independente associada (o regressor). A intuição deixa adivinhar que, geralmente, se pode melhorar esta “predição” se incluirmos novas variáveis independentes ao modelo (à equação de regressão).

Método dos Mínimos Quadrados

O método dos mínimos quadrados é uma técnica estatística utilizada para encontrar a melhor estimativa dos parâmetros de um modelo matemático, minimizando a soma dos quadrados das diferenças entre os valores observados e os valores previstos pelo modelo (Montgomery, Peck & Vining, 2012).

O objetivo é encontrar os valores de β_0 e β_1 que minimizam a soma dos quadrados dos resíduos (diferenças entre os valores observados e os valores preditos).

Equação matemática:

$$\text{Minimizar } \sum_{i=1}^n (Y_i - (\beta_0 + \beta_1 X_i))^2$$

14

Onde:

- n – É o número de observações,
- Y_i – São os valores observados,
- X_i – São valores da variável independente.

Para a solução analítica:

$$\beta_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

15

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$

16

Onde:

- \bar{X} – É a média de X,
- \bar{Y} – É a média de Y.

[c] Regressão Linear Múltipla

A regressão linear múltipla tem como objetivo o de estudar a relação entre duas ou mais variáveis explicativas que se apresentam na forma linear, e uma variável dependente métrica.

A regressão múltipla é adequada quando o investigador procura entender como múltiplas variáveis independentes (fatores) estão relacionadas a uma variável dependente contínua. Para o caso particular do tema em estudo, pode-se usar a regressão múltipla para analisar como a idade, o sexo, a comorbidades e medidas de distanciamento social estão relacionados com o número de caso de COVID-19 no país e em partículas nas 04 províncias em análise.

Pressupostos do modelo de regressão linear múltipla

1. A variável dependente é uma função linear de um conjunto específico de variáveis e do erro.

2. O valor esperado do termo de erro é zero.
3. O erro tem distribuição normal e não apresenta auto-correlação ou correlação com qualquer variável x .
4. As observações das variáveis explicativas podem ser consideradas fixas em amostras repetidas.
5. Não existe relação linear exata entre as variáveis explicativas e existem mais observações do que variáveis explicativas.

Na regressão linear múltipla assume-se que existe uma relação linear entre uma variável y (a variável dependente) e k variáveis independentes, $x_j (j = 1, \dots, k)$. As variáveis independentes são também chamadas variáveis explicatórias ou regressores, uma vez que são utilizadas para explicarem a variação de y . Muitas vezes são também chamadas variáveis de predição, devido à sua utilização para se predizer y .

As condições subjacentes à regressão linear múltipla são análogas à da regressão linear simples, resumidamente:

1. *As variáveis independentes x_j são não aleatórias (fixas);*
2. *Para cada conjunto de valores de x_j há uma subpopulação de valores de y . Para a construção dos intervalos de confiança e dos testes de hipóteses deve poder-se assumir que estas subpopulações seguem a distribuição normal;*
3. *As variâncias das subpopulações de y são iguais.*
4. *Os valores de y são estatisticamente independentes. Por outras palavras, quando se extrai a amostra, assume-se que os valores de y obtidos para um determinado conjunto de valores de x_j são independentes dos valores de y obtidos para outro qualquer conjunto de valores de x_j .*

Tal modelo linear em sua forma mais geral é expressa de seguinte forma:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

Os erros $\varepsilon_1, \dots, \varepsilon_n$ são independentes e distribuídos de forma idêntica com

$$E(\varepsilon_1) = 0 \quad \text{Var}(\varepsilon_1) = \sigma^2$$

18

A função linear estimada

$$\hat{f}(x_1, \dots, x_k) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k$$

19

Cada regressor em um modelo linear recebe um peso numérico, o chamado de coeficiente de regressão, inclinação de regressão ou simplesmente peso de regressão, que determina quanto a equação usa os valores dessa variável para produzir uma estimativa de y . Esses pesos de regressão são derivados por um algoritmo que produz uma equação matemática ou modelo para y que melhor se ajusta aos dados, usando algum tipo de critério para melhor definição.

A teoria da modelação estatística linear apresentação limitações pelo fato de assumir o pressuposto de normalidade, pois não pode ser utilizada se a distribuição da variável resposta é diferente da Normal, pelo que para a resolução de problemas com esta distribuição diferente da normal, é necessário utilizar o modelo linear generalizado.

Os modelos lineares generalizados é uma extensão do modelo linear, onde a distribuição da variável resposta não tem de ser normal, mas sim outra distribuição da família exponencial e a função que relaciona o valor esperado e o vetor de variáveis explicativas pode ser qualquer função diferenciável para uma função densidade de probabilidade escrita de forma geral:

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

20

Onde:

θ - É o parâmetro de localização,

φ - É o parâmetro de dispersão,

$a(\cdot)$, $b(\cdot)$ e $c(\cdot)$ - são funções reais conhecidas.

O Poder Explicativo do Modelo de Regressão - R^2

A fração da variância da amostra de Y_i que é explicada (prevista) pelas regressões é designada por R^2 . Da mesma forma, considera R^2 “a proporção da variação da amostra da variável dependente explicada pelas variáveis explicativas, e é utilizado como uma medida do grau de ajuste”.

O coeficiente de explicação é obtido da seguinte forma:

$$R^2 = \frac{SQR}{SQT}$$

21

Onde:

R^2 - *É o coeficiente de ajuste ou de explicação,*

SQR - *É a soma dos quadrados de regressão e,*

SQT - *é a soma dos quadrados totais.*

O coeficiente de ajuste ou de explicação pode variar de 0 e 1 (*0 a 100%*), porém é praticamente impossível a obtenção de um coeficiente de explicação igual a 1, uma vez que dificilmente todos os pontos cairão em cima de uma reta. Em outras palavras, se o coeficiente de explicação for 1, não haverá resíduos para cada observação da amostra em estudo. No entanto, se as variáveis explicativas não forem adequadas para explicar o comportamento de y , o coeficiente de explicação ficara próximo de 0. Desta forma, a medida que os erros de regressão forem minimizados, o coeficiente de explicação aumentará (Cordeiro, 2010).

[d] Regressão Logística

Em regressão logística, a variável resposta y é aleatória contínua ou dummy ou ainda uma variável dicotômica ou binária. Em outros casos a variável dependente y pode ser discreta, admitindo dois ou mais valores que representem uma categoria. As covariáveis do objeto de estudo dividem-se em variáveis ordinais e nominais (Agresti, 2013).

A regressão logística é apropriada quando o investigador trabalha com uma variável dependente binária, isto é, duas categorias que a resposta seja *infetado ou não infetado*

e/ou ordinal (categorias ordenadas). O pesquisador pode usar a regressão logística para, por exemplo na análise de casos de COVID-19, procurar entender como variáveis independentes como a idade, o sexo, a comorbidade influenciam a probabilidade de uma pessoa ser infectada pelo vírus corona (COVID-19).

Em estudo similares, poder-se-ia utilizar a regressão logística binária, onde a variável resposta tomará apenas os valores 0 e 1. Poderia-se-ia também utilizar outros modelos para analisar os dados do trabalho em alusão, mas existem razões fundamentais para usar o modelo de regressão logística: a primeira razão é que a regressão logística tem uma interpretação com significado clínico-bioestatístico e a segunda é que a liberdade dos pressupostos da regressão que permitem avaliar a significância de cada uma das variáveis independentes do modelo (Agresti, 2013).

É importante perceber que o modelo de regressão logística é amplamente utilizado quando a variável dependente é dicotômica, ou seja, tem apenas duas categorias. Essa técnica é especialmente útil quando o pesquisador concentra as suas atenções e está interessado em prever a probabilidade de um evento ocorrer, ou seja, quando deseja entender como diferentes variáveis independentes influenciam a probabilidade de uma resposta binária (Kleinbaum, 2010).

C.1. Regressão logística univariável

Considere-se uma variável independente X e uma variável aleatória real Y binária, tal que P seja a probabilidade:

- $P(Y = 1) = \pi(x)$, probabilidade de ter sucesso
- $P(Y = 0) = 1 - \pi(x)$, probabilidade de ter insucesso

C.2. Regressão logística multivariável

O modelo de regressão logística multivariável é dado pela expressão que define a probabilidade de que o acontecimento de interesse ocorra, que será dada por:

$$\pi(x) = P(Y = 1) = \frac{\exp(\beta_0 + \sum_{i=1}^p \beta_i x_i)}{1 + \exp(\beta_0 + \sum_{i=1}^p \beta_i x_i)}$$

C.3. Odds Ratio

Odds Ratio (OR) é uma medida de associação usualmente utilizada na regressão logística univariável para complementar o teste de significância da covariável (x). O facto de existir uma relação entre os parâmetros do modelo logístico e o odds ratio utilização desta medida constitui uma vantagem desde que se assuma que a covariável é binária.

C.4. Função Logística

Para os modelos de regressão logística, a probabilidade de ocorrência do acontecimento de interesse pode ser estimada diretamente. Se a variável dependente Y assumir apenas dois valores inteiros possíveis (0 ou 1) com um conjunto de n variáveis independentes X_1, X_2, \dots, X_p , o modelo de regressão logística será dado por:

$$P(Y = 1) = \frac{1}{1 + e^{-g(x)}}$$

23

Onde,

$$g(x) = B_0 + B_1X_1 + \dots + B_nX_n$$

24

[e] Modelos Autorregressivos

A classe de modelos conhecida como modelos autorregressivos - AR (no caso, de ordem p) é um processo iterativo em que há a identificação da ordem p através da função de auto-correlação; a partir daí, faz-se a estimativa de um modelo de previsão bem como a análise dos resíduos para a avaliação da existência de vieses e/ou grandes erros de estimativas (Hoffmann, 2017).

A dificuldade desta técnica é a identificação do modelo, pois é possível que pessoas diferentes identifiquem modelos de ordem diferentes para a mesma série temporal.

[f] Modelos Médias - Móveis

O modelo de média móvel de ordem q - $MA(q)$ é usado quando há autocorrelação entre os resíduos, ou seja, existe uma relação de dependência entre o conjunto de erro em

períodos passados. Este modelo em série temporal utiliza, como previsão para uma determinada observação no futuro, a média das observações passadas. As médias móveis podem ser simples, centradas ou ponderadas (Hoffmann, 2017).

O termo Média-Móvel é utilizado porque, à medida que a próxima observação está disponível, a média das observações é recalculada, incluindo esta observação no conjunto de observações e desprezando a observação mais antiga (Hoffmann, 2017).

3.2.2. Análise de dados da COVID-19 em Moçambique

O número de casos da pandemia da COVID-19 em Moçambique tem tido oscilações. A análise deste fenómeno deve ser acompanhada ao longo tempo (série temporal) permitindo assim observar o fator de crescimento de infetados a cada dia que passa, avaliando a sua evolução e tendências (eventualmente sazonais e/ou fatores ambientais).

Tratando-se de dados obtidos no mesmo intervalo de tempo e com atualizações a cada hora cujo armazenamento é sucessivamente em função do tempo, a análise temporal do fenómeno é viável.

Em epidemiologia, a COVID-19 é uma doença viral altamente contagiosa, sem nenhuma medida conhecida de controlo efetivo. Como o fator de crescimento é elevado, a aproximação pela curva exponencial parece adequada.

Moçambique e China, no cumprimento dos acordos de cooperação socioeconómica entre os dois governos e no âmbito do combate à pandemia da COVID-19 recebeu uma doação de 200 mil doses de vacina VeroCell contra a COVID-19 no intuito de prevenir a contaminação pela doença que numa primeira fase o lote da vacina foi administrada aos profissionais de saúde que constituem o grupo com maior risco de contrair a infeção pelo novo Coronavírus (SARS-COV2) em seguimento às orientações da OMS.

Moçambique recebeu as 200 mil doses da Vacina VeroCell, produzida pela Farmacêutica chinesa, Sinopharm (Beijing Institute of Biological Products) a 24 de Fevereiro de 2021 e espera receber vários lotes de vacinas no âmbito das iniciativas criadas para a aquisição e disponibilização de vacinas. O Plano Nacional de Vacinação contra a COVID-19 preconiza a cobertura de cerca de 20% da população em maior risco

(trabalhadores de saúde, população idosa, pessoas com comorbilidade e outros grupos de maior risco de contaminação).

[a] Modelos matemáticos de Análise de dados da COVID-19 em Moçambique

A cada dia que passa o número de casos da pandemia da COVID-19 tende a aumentar em Moçambique (até ao período da coorte do estudo – 31 de Maio de 2022). A análise deste fenómeno prende-nos no pensamento de que este pode ser analisado ao longo tempo (série temporal) desde o seu início no mundo e até no caso particular do nosso país, fazendo assim se perceber um fator de crescimento de infetados a cada dia que passa a uma evolução exponencial de casos.

Em epidemiologia, a COVID-19 é uma doença viral, é altamente contagiosa, sem nenhuma medida de restrição, o fator de crescimento é alto, podendo ser muito bem aproximado pela curva exponencial.

Os dados cumulativos de casos confirmados da COVID-19 em Moçambique mostram uma taxa de positividade crescente. A partir dos dados reais e preditivos foi possível fazer uma análise de regressão e se encontrar a curva exponencial e o modelo matemático que a evolução dos casos.

i. Modelos de Regressão

Modelo de Regressão Linear Múltipla:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

25

Modelo Matemático Exponencial:

$$NC = E \times \rho \times N_d$$

26

Onde:

NC – Novos Casos; *E* – N° de Contactos;

ρ - Probabilidade de se infetar e *N_d* – Total de Casos Reportados

Analisando o propósito do trabalho e para compreender o comportamento dos casos de COVID-19 nas províncias identificadas para o presente trabalho, foi aplicado o

Modelo de Regressão Linear Múltipla para explicar se a variação do número de casos da doença tem relação com variáveis “*área geográfica, Idade e Género*”.

ii. Modelos Epidemiológicos

Dentre vários modelos epidemiológicos que podem ser usados para analisar o perfil epidemiológico duma determinada doença, para o presente trabalho de culminação do curso ir-se-á abordar sobre os chamados modelos compartimentais (modelo SIS) de grande utilidade para simular epidemias em larga escala populacional. Os modelos SIR são também designados por modelos compartimentais porque se baseiam na divisão da população em compartimentos, isto é, levando em conta as características físicas e epidemiológicas que refletem o estado em que os indivíduos se encontram no desenvolvimento da doença.

O modelo SIR ou compartimental é um modelo epidemiológico clássico utilizado para descrever a dinâmica da transmissibilidade e/ou propagação de doenças infecciosas em uma população. O modelo SIR representa três compartimentos principais do modelo, a saber:

1. **Suscetíveis (S):** Indivíduos que ainda não foram infetados pela doença e são suscetíveis a contrair a infeção.

- *Taxa de mudança de suscetíveis*

$$\frac{dS}{dt} = -\beta \frac{S \cdot I}{N}$$

2. **Infetados (I):** Indivíduos que estão atualmente infetados e podem transmitir a doença para indivíduos suscetíveis.

- *Taxa de mudança de infetados*

$$\frac{dI}{dt} = \beta \frac{S \cdot I}{N} - \gamma I$$

3. **Recuperados (R):** Indivíduos que se recuperaram da doença e, em muitos modelos, assumem que são imunes e não podem mais ser reinfectados.

○ *Taxa de mudança de recuperados*

$$\frac{dR}{dt} = -\gamma I$$

Para melhor compreender os modelos matemáticos com linhagem epidemiológica (modelos compartimental), é necessário falar sobre os principais termos epidemiológicos mais usuais que nortearão o presente trabalho, abaixo descrito:

- **Indivíduos Suscetíveis:** caracterizados por indivíduos saudáveis, que passam a ser vulneráveis a contrair a doença fazendo algum tipo de contato potencialmente infeccioso. Indivíduos caracterizados suscetíveis podem ou não desenvolver a doença, isto é, não são indivíduos infetados.
- **Indivíduos infetados:** caracterizado por indivíduos que contraíram o agente patogénico (a doença). Quando o indivíduo está infetado, automaticamente, começamos a chamá-lo de infeccioso, e este indivíduo passa a ser o principal meio de propagação da doença.
- **Indivíduos removidos:** caracterizados por indivíduos que passam do quadro de infetados para removidos, ou seja, quando o indivíduo é totalmente curado da doença ou morre.
- **Incidência:** pode-se caracterizar a incidência como a proporção entre o número de indivíduos que adoecem durante um intervalo de tempo pelo total da população. Na maioria das vezes, a incidência é determinada pelo número de casos já confirmados de infecção, os quais subestimam a verdadeira incidência, pois são ignorados os casos suspeitos.
- **Mortalidade induzida pela doença:** é a proporção do número de pessoas que morreram da doença em uma unidade de tempo pelo total da população.
- **Taxa de contato:** medida de frequência de encontro entre indivíduos suscetíveis e infetados.

- **Transmissão:** caracterizado pelo processo no qual um vírus passa de uma fonte de infecção para um novo hospedeiro.

Os modelos epidemiológicos permitem entender a propagação da doença, bem como identificar as medidas que podem ser adotadas para controlar a disseminação da doença. Os modelos epidemiológicos partem de duas hipóteses:

- **Compartimentalização** – considera indivíduos suscetíveis, infectados e recuperados, que podem se deslocar entre os compartimentos, isto é, quando um indivíduo suscetível entra em contato com um infectado, pode se tornar infectado; caso combata o agente patógeno, se torna recuperado e fica imune.
- **Mistura homogênea** – assume que cada indivíduo tem a mesma chance de entrar em contato com um indivíduo infectado, considerando que a rede de contatos é aleatória.

Para uma análise mais profunda, robusta e detalhada da propagação da doença em estudo (COVID-19) obedeceu uma combinação do modelo SIR com os modelos de regressão linear múltipla que permitiu modelar e prever a propagação da doença, bem como analisar a influência de fatores externos que podem influenciar a situação epidemiológica da doença na província.

O modelo SIR entra na combinação com os modelos de regressão para:

Previsão da Propagação: estimar como a doença pode se espalhar ao longo do tempo.

Análise de Intervenções: avaliar o impacto de medidas de controle, como a vacinação, recolha obrigatória, distanciamento social e a quarentena.

Compreensão das dinâmicas da doença: estudar como a taxa de transmissão e recuperação pode afetar a propagação da doença.

[4]. Caso de estudo – Aplicação

Nesta secção do trabalho serão descritos os resultados esperados e discussão destes com as opiniões dos autores e investigadores que relataram sobre o tema em estudo.

Durante a recolha e análise de dados do presente trabalho espera-se que existam fundamentos (fatores associados ao perfil epidemiológico da doença bem como os modelos matemáticos) que explicam a variação numérica dos casos e óbitos por COVID-19 em Moçambique, em particular para as províncias seleccionadas.

4.1. Modelos de Visualização de dados

4.1.1. Visualização

A razão para coletar dados é para permitir a tomada de decisões mais bem informadas com base nos dados, sendo importante não apenas coletar dados sem maneira útil de acessá-lo. E com dados de gráfico, isso normalmente significa desenhar o gráfico.

Entenda-se de **visualização** qualquer método de usar imagens para transmitir um ponto ou conhecimento. Ao trabalhar com gráficos de computador, geralmente significa encontrar uma maneira de mostrar abundância de dados em uma única exibição através da criação de imagens e gráficos para mostrar dados ou informação a um padrão de conhecimento e/ou entendimento da realidade.

Em Saúde Pública, os dados são essenciais na tomada de decisões, porque, por meio deles, os profissionais de saúde e quaisquer outros utilizadores de dados podem compreender informações a respeito do alastramento e magnitude de doenças, da situação clínica de seus pacientes, de estoques de medicamentos, entre outras situações. Acreditando que a Visualização de Dados pode fornecer importantes contribuições à Saúde, o presente trabalho de pesquisa tem a sua essência em mapear e gerar diferentes possibilidades de visualizações de dados.

4.1.2. Visualização de Dados

Antes de mais, é fundamental entender os diferentes conceitos sobre a visualização e saber diferenciar a visualização de dados e de informação (a visualização). Comumente

pode-se entender como o mesmo o conceito de visualização de dados e a visualização de informação.

*O termo **Visualização de dados** refere-se a todas representações visuais que suportam a exploração, análise e comunicação de dados, enquanto a **Visualização de Informações** refere-se ao uso de representações visuais interativas para dados relacionados, facilitando a sua análise e compreensão com o propósito de auxiliar a tomada de decisão com combinação de aspetos de computação gráfica, interação humano computador e mineração de dados.*

As técnicas de visualização de dados consistem na sua representação gráfica, tornando-se essencial a partir do momento em que a informação nos chega como dados de diferentes estruturas, tal como vídeos, fotografias, mapas, etc. O processo de conversão dos dados, que na sua origem são abstratos e desprovidos de qualquer forma geométrica em formas gráficas, é conhecido como visualização, apoiando-se na habilidade natural humana de compreender a informação de maneira visual.

No quotidiano dos indivíduos, é notória a familiaridade com dados, que de forma significativa intensifica a comunicação em sociedade, mediada por dispositivos de processamento de dados digitais. Lidar de maneira direta com grandes volumes de dados constitui uma tarefa ingrata e desgastante, para a qual têm sido desenvolvidos instrumentos de mineração de dados e outros, adequados para agregar sentido interpretativo a esses dados e extrair informação útil.

Uma forma acessível de visualizar (ver) e entender exceções, tendências e padrões nos dados é utilizando elementos visuais, em modelos tais como: dashboards, tabelas, diagramas, gráficos e mapas.

A visualização de dados resulta em informações que são mais facilmente compreensíveis, pois traduz a complexidade do relacionamento entre as variáveis em modelos visuais. Pode-se compreender dados numéricos apresentados em forma de tabela como uma tarefa possível para um ser humano, mas quando ilustrados em forma de gráficos a partir da seleção de determinado conjunto de registos em uma base de dados, a interpretação da informação fica muito mais fácil, clara e evidente para a sua percepção. As variações e os padrões de precipitação (grandes concentrações de massas de dados,

variações cíclicas e acíclicas, desníveis longitudinais dos padrões e a suavização da curva) se revelam quando são traduzidos em formas geométricas e/ou em mapas.

Os princípios da exploração das ferramentas de visualização de dados giram em torno de evidenciar os relacionamentos presentes num universo de dados. É fundamental entender que o objetivo básico da visualização de dados está focado em perceber o todo ou parte de um conjunto de dados de “*n*” dimensões para encontrar os padrões, as relações, as limitações, as falhas e os *clusters* do objeto de estudo.

Os artistas da visualização de dados transformam o caos informacional de pacotes de dados que se locomovem através da rede em formas claras e ordenadas. (...) A visualização de dados nos permite enxergar padrões e estruturas por detrás do vasto e aparente fortuito conjunto de dados. (...). Os dados quantitativos são reduzidos a seus padrões e estruturas, os quais, a seguir, explodem em inúmeras imagens visuais ricas e concretas. (Manovich, 2004:157)

A visualização de dados resulta em informações que são facilmente compreensíveis por utilizadores e/ou consumidores, pois traduz a complexidade de relacionamento entre as variáveis em modelos visuais. Compreender dados numéricos registados e reportados em forma de tabela é uma tarefa possível para um ser humano, mas quando apresentado em gráficos a partir da seleção de um conjunto de dados, a interpretação da informação fica muito mais clara e evidente.

Em apresentações gráficas e em mapas, as variações e os padrões de repetição se revelam de forma minuciosa e localizada virtualmente que conduza a uma compreensão do fenómeno em alusão. Esse é o princípio explorado pelas ferramentas de visualização: evidenciar relacionamentos presentes num universo de dados, mas cujas interpretações ainda se encontram latentes, isto é, a visualização se constitui como instrumento fundamental para revelar sentidos ocultos, invisíveis numa observação restrita aos dados em si.

4.1.2.1. Infografia

Pode-se conceituar **infografia** uma representação visual que combina elementos de texto, imagens e gráficos para transmitir informações de maneira clara e eficaz sobre um determinado assunto e/ou tópico. A infografia visa de simplificar conceitos complexos, destacar padrões, contar histórias e fornecer uma visão geral rápida de um determinado tópico, muitas vezes usando recursos visuais para tornar a informação mais compreensível e memorável ao “consumidor” dos dados.

É importante recordar que os infográficos (gráficos informativos) são as imagens frequentemente usadas em apresentações em simpósios, reuniões formais com base em tabelas e gráficos estilizados usados em jornais e revistas. Os gráficos de informação para fins operacionais ganharam campo e até aos nossos dias são usados diariamente por milhares de pessoas para melhorar sua eficiência e eficácia bem como a melhoria da qualidade interativa dos dados; resolver problemas relacionados com a apresentação de dados, planejar, ensinar, treinar, monitorar processos, estudar a distribuição geográfica de dados, apresentar tendências e relacionamentos, revisando o status de projetos, desenvolvendo ideias, redigindo relatórios, analisando dados de censos, estudando resultados de vendas e acompanhando as finanças domésticas.

Com a necessidade de lidar com o aumento da quantidade de dados e, ao mesmo tempo, aumentar a qualidade e a produtividade no processamento de dados, as tabelas, os gráficos e os mapas estão sendo cada vez mais utilizados em apresentações de dados em situações operacionais. Com o avanço no desenvolvimento de equipamentos informático e softwares de apresentação e análise de dados, a maioria das tabelas e gráficos populares usados diariamente podem ser gerados de forma rápida, fácil e em tempo real.

A infografia no contexto de visualização de dados tornou-se uma ferramenta poderosa de transmissão de conhecimentos devido a sua característica de síntese visual, compreensão rápida, comparação, memorabilidade e apelo visual.

Uma informação complexa pode ser insuficiente e difícil de se entender quando apresentada apenas com textos corridos, mas quando é simplificada e apresentada em imagens traduzidas em mapas e gráficos, destacando os pontos-chave, torna-se mais fácil para o público entender rapidamente a mensagem.

Por outro lado, a combinação de texto e elementos visuais, como gráficos, ilustrações – imagens e ícones, ajudam a sintetizar informações, tornando-as mais

acessíveis e atraentes aos interessados. Esta combinação é denominada por Síntese Visual. Os gráficos, mapas e diagramas podem ser incorporados em infográficos para representar dados numéricos e relacionamentos entre variáveis, permitindo que o público identifique padrões e tendências do problema em estudos (casos de COVID-19 em Moçambique).

No domínio de visualização de dados, os infográficos também permitem fazer comparação de informações, tornando mais fácil para o público identificar diferenças e semelhanças entre vários dados apresentados. Ela (infografia), pode ser usada para destacar progressões cronológicas, narrativas ou fluxos de informação de maneira visualmente atraente orientada para uma história de acontecimentos do passado.

Todas informações apresentadas visualmente são frequentemente mais memoráveis do que textos puros. Nos dias de hoje, o maior desafio das entidades é fazer compreender os acontecimentos de forma rápida e adequada para que as pessoas tendem a lembrar melhor de imagens e gráficos, proporcionando assim os infográficos mais eficazes na comunicação de conceitos importantes.

A infografia é uma ferramenta versátil que ajuda na visualização de dados de diferentes domínios, tornando informações complexas em mais fáceis, simples e acessíveis, envolventes e memoráveis para o público consumidor dos dados. Para além do domínio de análise de dados, ela é amplamente usada em jornalismo, marketing, educação e em muitas outras áreas para comunicar informações de forma diversificada e eficaz.

Existem muitas técnicas de visualização de dados que, cada uma delas, permite analisar os dados com metáforas visuais diferentes. No entanto, ao aplicar-se uma determinada técnica de visualização, diversos fatores devem ser tomados em consideração, tais como: A classificação dos dados e das tarefas, a característica cognitiva do potencial utilizador ou o tipo de ênfase visual desejada ao resultado (conteúdo versus quantidade).

Abaixo, de forma muito resumida, será descrita cada modelo e/ou técnica de visualização de dados considerados pelo autor como os potenciais e comuns para expressar o significado de dados.

4.1.3. Técnicas de Visualização de Dados

[a] Visualização de Dados em Dashboard

Os *Dashboards* são painéis que mostram métricas e indicadores importantes de forma visual, facilitando a compreensão das informações geradas para o consumo público.

O objetivo dos dashboards é de possibilitar o monitoramento dos resultados de uma organização ou programa, distribuídos em diversos indicadores procurando responder perguntas essenciais para ter as respostas desejadas com a apresentação da informação neste fórum.

[b] Visualização em Tabelas Resumidas

As tabelas resumidas agrupam registros de valores exclusivos e um ou mais campos-chave efetuando uma contagem para o número de registros conforme a correspondência.

i. Tabela de Frequências

Este tipo de tabela exhibe a frequência de valores entre cada combinação exclusiva de categorias. O seu uso é frequente para medir a ocorrência de valores ou categorias específicas em um conjunto de dados.

As tabelas de frequência são comumente usadas em estatística para análises descritivas.

ii. Tabela Cruzada ou de Contingência

A tabela cruzada ou de contingência resume a relação entre duas ou mais variáveis categóricas, descrevendo o número de vezes que cada uma das possíveis combinações de categorias ocorreu nos dados. É especialmente utilizada quando se deseja investigar como duas variáveis estão relacionadas e se existe alguma associação entre elas.

Este tipo de tabela é amplamente utilizado em estatística e pesquisa de mercado para identificar associações e padrões entre variáveis categóricas. Para o caso particular do presente trabalho, as tabelas de contingência ou cruzada são úteis para responder

perguntas como: *existe uma relação entre o número de casos de COVID-19 e a faixa etária da população suscetível ou infectada pela doença?*".

[c] Gráficos

Os gráficos são úteis para comparar categorias de dados, onde o eixo x representa a categoria e o eixo y representa o valor a comparar. Pode-se citar diferentes tipos de gráficos a considerar: Gráfico de linha ou linear, de barras, de colunas, histograma, circular ou de pizza, teia de aranha, etc. todos que exibem a distribuição dos dados.

Para os dados em análise, o sector da saúde utiliza os gráficos de linha, de coluna e de barras para visualizar os dados da pandemia da COVID-19 em Moçambique.

i. Gráfico de Linha

Os gráficos de linhas mostram tendências ou alterações ao longo do tempo exibindo uma série de pontos de dados conectados por segmentos de linha reta. Foi utilizado para monitorar mudanças ao longo do tempo para ajudar na análise de dados preditiva.

ii. Gráfico de Barras

Um *gráfico de barras* é composto por barras com formato retangular, onde o comprimento de cada barra de forma proporcional representa o valor do evento correspondente. Geralmente é utilizado para comparar duas ou mais variáveis.

iii. Gráfico Circular ou de Pizza

Os gráficos de pizza mostram a proporção/percentagem total das categorias. Este tipo de gráfico procura traduzir a composição de dados categóricos com cada segmento proporcional à quantidade que representa.

[d] Mapas

Na atualidade e com os avanços tecnológicos, assiste-se à representação da realidade em ambiente virtual/digital. A complexidade dos sistemas tornou-se mais

simples, sobretudo no que diz respeito ao tempo a despendido para visualização de grandes volumes de dados. Os mapas têm uma história antiga e continuam a ser uma ferramenta essencial nos dias de hoje, quer que seja para servir de guia durante viagens, para analisarmos a distância entre lugares, estudo de questões topográficas, incidências epidemiológicas, entre outros.

Nos momentos atuais, os mapas tornaram-se referência na visualização de dados, possibilitando o homem a ver dados de localização representados em um mapa e incluídos em visualizações que permitam o entendimento de fenómenos e casos de estudo.

Diversas formas de apresentação de dados podemos destacar no contexto de infografia, a destacar:

a. Gráficos Abstratos

Estes tipos de gráficos, também conhecidos como diagrama, resumem-se em gráficos que normalmente não tem escalas, marcas ou linhas de grade e é usado principalmente para investigações, ilustrações e transmissão ideias e conceitos em vez do que valores específicos.

Outrossim, os gráficos abstratos são representações visuais de dados que podem incluir elementos simplificados, formas geométricas, linhas, cores e padrões, sem necessariamente representar um objeto real ou uma escala precisa. Este tipo de gráfico é utilizado quando se deseja destacar padrões gerais sem enfatizar os valores exatos dos dados do estudo ou do pensamento.

b. Mapas Abstratos

São mapas que transmitem uma visão focada e limitada a quantidade de informações específicas, como localização, direção, relacionamentos, tempo ou informações por área. Os mapas abstratos geralmente transmitem um ou no máximo alguns conjuntos de fatos, ideias, conceitos, etc.

Também, pode-se entender os mapas abstratos como sendo representações artísticas ou simplificadas de áreas geográficas que podem não ser geograficamente

precisas, mas que são projetados para transmitir conceitos, padrões ou informações de maneira mais visual, interativa e criativa.

Duma forma geral, os gráficos e mapas abstratos são formas de representação visual que utilizam elementos simplificados e estilizados para visualizar informações de maneira conceitual ou artística, muitas vezes sem tentar representar detalhes precisos ou geograficamente precisos.

Estes são usados para criar uma visualização mais estilizada, criativa ou conceitual dos dados, permitindo que o observador entenda padrões e tendências de maneira intuitiva, sem se prender a detalhes realistas.

c. Pirâmide Etária

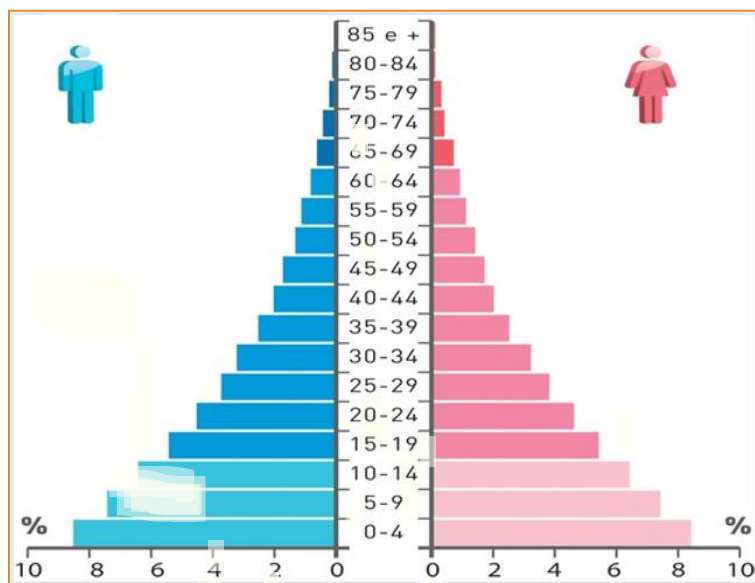
Pirâmide etária, também chamado de pirâmide populacional aplica-se especificamente em estudar um determinado fenômeno que envolva a população e que a sua análise se baseia no gênero e faixa etária (população por idade e sexo).

A pirâmide populacional é uma ferramenta poderosa para visualizar e entender a estrutura demográfica duma população, destacando a variabilidade das variáveis de estudo dentro duma população buscando compreender as características de pessoas em diferentes grupos de idade e como essa distribuição varia entre homens e mulheres.

Para além de estudo caracterizados pela análise de dados focados a idade e sexo, a pirâmide etária também pode ser usada para fazer previsões sobre tendências demográficas futuras com base nas mudanças das variáveis bases (sexo e faixas etárias).

Ao aplicarmos a pirâmide etária como uma ferramenta robusta de visualização de dados, permite também compreender a composição demográfica de uma população, identificar padrões, tomar decisões de políticas públicas e muito mais.

Gráfico 01: Exemplo duma pirâmide etária



A pirâmide etária para análise de dados de COVID-19 permite uma representação gráfica que mostra como a infecção pode se distribuir por faixas etárias. É útil para fazer entender como a pandemia afeta diferentes grupos de idade por sexo permitindo desta forma a formulação de políticas de saúde que visem uma intervenção pontual e acertada.

d. Mapa Auxiliar

Mapa auxiliar é um modelo de mapa simplificado e estilizado usualmente utilizado como plano de fundo para ajudar a contextualizar os dados que estão sendo visualizados. Não é um mapa geográfico preciso, mas fornece informações de localização aproximadas bem como características geográficas gerais que permitem enquadrar os dados de forma mais compreensível.

e. Mapa base

Refere-se mapa base a camada de fundo sobre a qual os dados são projetados ou representados em um Shapefile. Pode ser um mapa geográfico real, como um mapa de ruas, um mapa topográfico ou um mapa de satélite. O mapa base fornece informações detalhadas sobre características geográficas, como estradas, rios, fronteiras, relevos e outros elementos.

Na visualização de dados em uma infografia no contexto geográfico, pode-se utilizar um mapa base para fornecer referências visuais precisas. Os dados são então sobrepostos ao mapa base para mostrar sua relação com o ambiente geográfico real.

No contexto de visualização de dados que dentre vários, existem vários mecanismos e técnicas que podem ser aplicados para criar infografias eficazes, que passo a citar:

- a. **Hierarquia Visual:** colocar os elementos mais importantes (títulos, subtítulos, ícones ou gráficos) em destaque, para direcionar a atenção para os pontos-chave.
- b. **Gráficos Claros:** utilização de legendas, rótulos e escala apropriada nos eixos para garantir que os dados transmitam a ideia central do problema e que sejam compreendidos corretamente.
- c. **Consistência Visual:** manter o padrão visual e consistente no que tange a cores, fontes, estilos de ícones e layouts.

4.1.3.1. Dados espaciais

O conhecimento da distribuição espacial dum determinado evento (foco para um problema em saúde pública) é de extrema importância no desenvolvimento das atividades humanas que visem a solução e/ou a mitigação do mesmo bem como a interação do homem no espaço.

Para investigar e efetuar o seguimento de um surto em saúde pública, deve se conhecer a localização geográfica do fenómeno, se estabelecer relações com outros elementos da área geográfica de ocorrência do evento, de identificar padrões geo-espaciais para permitir uma tomada de decisões com base nas suas características geográficas.

Um dos grandes desafios em saúde pública é a possibilidade de representar, analisar e conhecer a dimensão espacial associada a ocorrência de determinados fenómenos, colocando o investigador do surto a recorrer os Sistemas de Informação Geográfica para visualizar os dados e mapear toda informação relacionada com o problema que se pretende resolver.

Sistema de Informação Geográfica (SIG) - é uma estrutura ou ambiente aplicado para coletar, gerenciar, analisar e visualizar dados espaciais.

[a] Software SIG

Para visualizar e analisar dados através do Sistemas de Informação Geográfica recorre-se a alguns softwares SIG populares que permitem interação e desenvolvimento de mapas e análise espacial, que abaixo passo a citar:

- **ArcGIS** - Um software SIG comercial desenvolvido pela empresa ESRI, que é muito popular; e
- **QGIS** - Um software SIG de código aberto gratuito.

[b] Dados Espaciais

As duas formas principais de dados espaciais usados em SIG são dados vetoriais e raster:

Dados vetoriais: os dados vetoriais são compostos de recursos geométricos de vértices e caminhos. Estes dados podem ser divididos em três tipos amplamente usados:

- *Pontos* - Um ponto consiste em um par de coordenadas (x, y) representando uma localização específica em um sistema de coordenadas. Esta é a forma mais básica de dados espaciais e podem ser usados para denotar um caso (um evento de interesse público) ou um local (que pode ser um hospital, uma escola, etc.) em um mapa.
- *Linhas* - Uma linha é composta por dois pontos conectados. As linhas têm um comprimento e podem ser usadas para denotar estradas e/ou rios.
- *Polígonos* - Um polígono é composto por pelo menos três segmentos de linha conectados por pontos. As características do polígono têm um comprimento (perímetro da área) bem como uma medida de área. Os polígonos podem ser usados para indicar uma área (uma aldeia em zona de expansão) ou uma estrutura (a área real de um estabelecimento comercial).

Dados Raster – é um formato alternativo para dados espaciais caracterizada por uma matriz de células (geralmente pixels) com cada célula contendo informações como altura, temperatura, inclinação, cobertura florestal, etc. Geralmente são fotografias aéreas, imagens de satélite, etc. Os rasters também podem ser usados como “mapas base” abaixo dos dados vetoriais.

[c] Visualização de dados espaciais

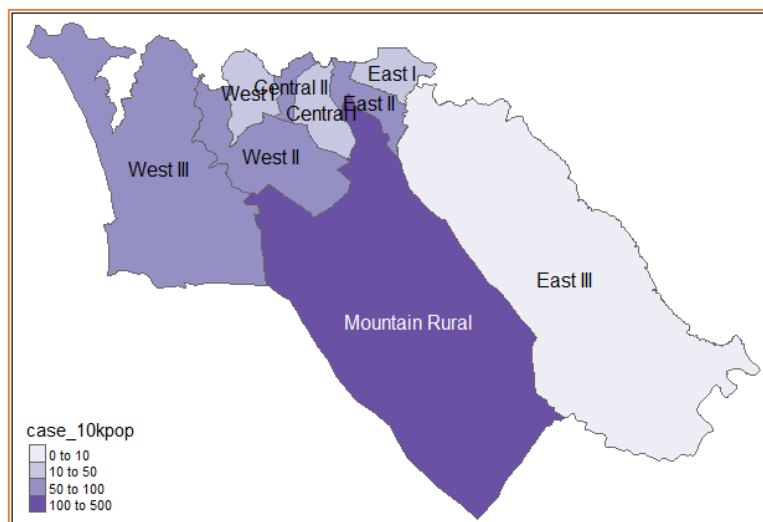
Para visualizar os dados espaciais em um mapa, o software SIG exige informações suficientes sobre a localização dos diferentes recursos, um em relação ao outro. Essas informações normalmente são armazenadas em um shapefile.

Shapefiles é um formato de dados comum para armazenar dados espaciais “vetoriais” que consistem em linhas, pontos ou polígonos. Um único shapefile é uma coleção de pelo menos três arquivos - .shp, .shx e .dbf.

○ **Tipos de mapas**

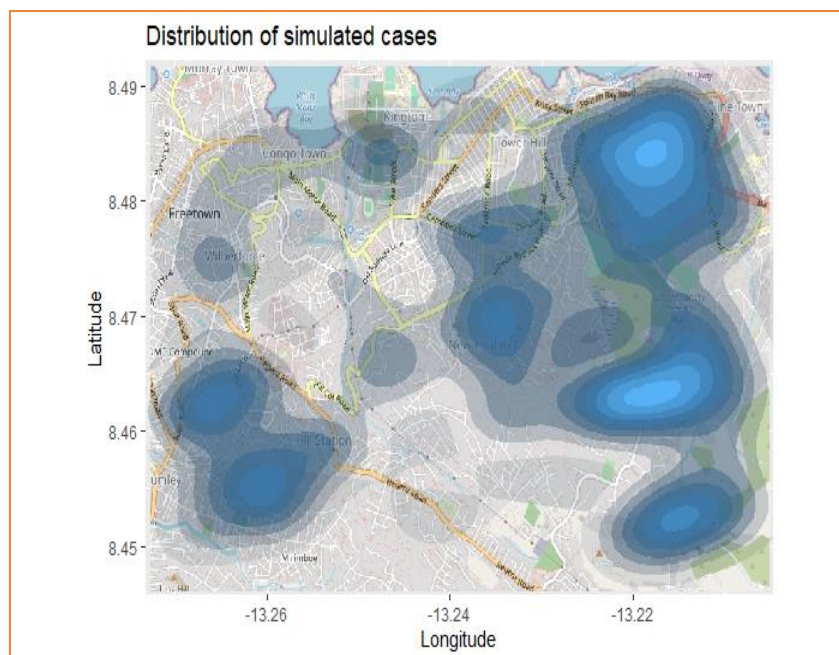
Mapa coroplético - um tipo de mapa temático em que cores, sombreamentos ou padrões são usados para representar regiões geográficas em relação ao valor de um atributo. Por exemplo, um valor maior pode ser indicado por uma cor mais escura do que um valor menor. Esse tipo de mapa é particularmente útil ao visualizar uma variável e como ela muda em regiões definidas ou áreas geopolíticas.

Gráfico 02: Exemplo dum mapa Coroplético



Mapa de calor de densidade de caso - um tipo de mapa temático onde as cores são usadas para representar a intensidade de um valor, porém, não usa regiões definidas ou limites geopolíticos para agrupar dados. Este tipo de mapa é normalmente usado para mostrar 'pontos quentes' ou áreas com alta densidade ou concentração de pontos.

Gráfico 03: Exemplo dum mapa de calor de densidade de caso

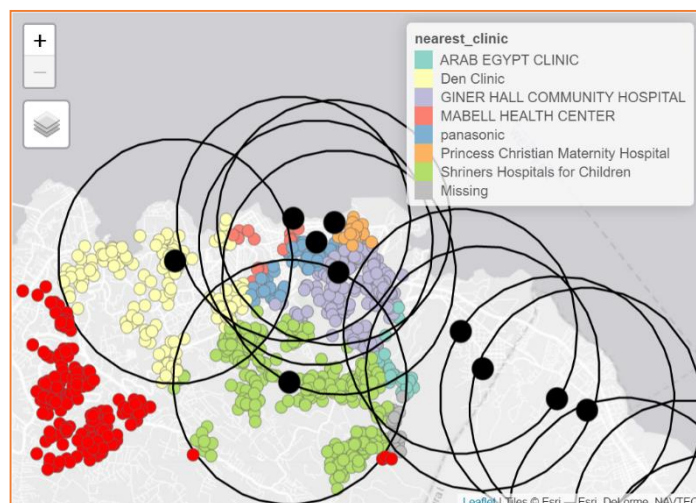


Mapa de densidade de pontos - um tipo de mapa temático que usa pontos para representar valores de atributo em seus dados. Esse tipo de mapa é melhor usado para visualizar a dispersão de seus dados e verificar visualmente os clusters.

Mapa de símbolos proporcionais (mapa de símbolos graduados) - um mapa temático semelhante a um mapa coroplético, mas em vez de usar cores para indicar o valor de um atributo, usa um símbolo (geralmente um círculo) em relação ao valor.

Por exemplo, um valor maior pode ser indicado por um símbolo maior do que um valor menor. Esse tipo de mapa é melhor usado quando você deseja visualizar o tamanho ou a quantidade de seus dados em regiões geográficas.

Gráfico 04: Exemplo dum mapa de símbolos proporcionais



Os casos (pontos) no mapa acima são coloridos de acordo com a concentração de unidade sanitárias mais próximas (conforme a legenda apresentada no mapa). Os grandes círculos vermelhos mostram as áreas de captação das unidades de saúde de um determinado raio, e os pontos vermelhos brilhantes aqueles que estavam fora de qualquer faixa de captação.

4.2. Resultados

4.2.1. Modelos de Análise de dados do COVID-19 em Moçambique

[a] Modelos de Regressão Linear Múltipla

Feita uma análise temporal dos casos notificados no país e em particular nas 4 províncias em análise, os dados apresentam uma tendência decrescente. Porém, os modelos a seguir ilustrarão a real situação do problema em estudo.

A Regressão é uma técnica utilizada na investigação da relação entre variáveis que surgem em problemas das mais variadas áreas da ciência. Dum modo geral, o pesquisador procura entender e aferir a influência de uma variável explicativa X sobre o valor esperado de uma variável de resposta denominada por Y.

Os modelos de regressão são vários e são definidos consoante o número e distribuição das variáveis explicativas, cujos efeitos na variável Y, que se pretendem

estudar. Se o investigador está interessado na relação de apenas uma variável explicativa com a variável resposta então aplica-se a Regressão Linear Simples. Mas se pretende relacionar a variável resposta com mais do que uma variável explicativa, então um modelo de Regressão Linear Múltipla será o mais apropriado.

Para caso particular deste relatório, foi investigada a influência de certas variáveis que contribuíram para explicar a variabilidade dos casos de COVID-19 no país e em particular para as 4 províncias destacadas como amostra para a análise da pandemia. A regressão múltipla fornece análises diagnósticas que podem determinar se tais efeitos existem com base em argumento empírico ou teórico. Indicações de um grau elevado de interrelações (multicolinearidade) entre as variáveis independentes podem sugerir o uso de escalas múltiplas.

Uma relação estatística é caracterizada por dois elementos:

- *Quando múltiplas observações são coletadas, mais de um valor da medida dependente geralmente será observado para qualquer valor de uma variável independente.*
- *Com base no uso de uma amostra aleatória, o erro na previsão da variável dependente também é considerado aleatório, e para uma dada variável independente, podemos apenas esperar estimar o valor médio da variável dependente associada a ela.*

Porém, antes de iniciar-se com a análise das variáveis de dependência que influenciaram na variação do número de casos da doença, vale referenciar que as províncias tomadas como amostra apresentaram ao longo do tempo a transmissão comunitária, o que proporcionou a aceleração no aumento de número de casos de COVID-19. Além dos fatores endógenos (a renda do agregado familiar, a insurgência em Cabo Delgado – ataques armados, o incumprimento de normas de segurança – distanciamento social e a recolha obrigatória) os fatores sócios económicos e políticos influenciaram negativamente no perfil epidemiológico do COVID-19.

Para analisar o perfil epidemiológico do COVID-19 em moçambique e em particular nas províncias em análise, é importante que se olhe os múltiplos fatores que influenciaram no comportamento da transmissibilidade da doença em diferentes áreas geográficas do pelouro do índico (Moçambique), que, o seu pano de fundo será utilizado

a Regressão Linear Múltipla para permitir a determinação do valor da variável resposta Y com base nas múltiplas variáveis de influência (preditoras).

Um modelo de regressão múltipla é expresso como:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

27

Para compreender a variabilidade do número de casos do COVID-19 em Moçambique e em particular nas províncias do Niassa, Cabo Delgado, Nampula e Maputo Província foram tomados como variáveis de influências para o número de casos e que condicionam o perfil epidemiológico da pandemia, as seguintes:

1. *Idade da população suscetível (Infantil, Jovem, Adulta ou Idosa)*
2. *Género da população suscetível (Masculino, Feminino)*
3. *Província - Área geográfica (Urbana ou Rural)*

As condições de vida deploráveis, o fraco envolvimento comunitário nas atividades de saúde, os fatores climáticos/ambientais (temperaturas), socioeconómicos (pobreza) e incumprimento de normas de prevenção (distanciamento social e recolha obrigatória) podem ser analisadas de forma isolada como também uma parte de fatores que influenciem a cadeia de transmissão da doença.

Para analisar a correlação entre as variáveis (fatores influenciadores) em estudo, foi tomada a medida estatística que mostra o grau de relacionamento ou associação entre as variáveis, onde estas, uma é dependente e a outras independentes (variáveis preditoras).

Para tal, considerou-se:

- a) **Casos de COVID-19** como variável dependente ou explicada,
- b) **Idade, Género e área geográfica** com variáveis independentes ou explicativas, sem descartar a possibilidade de outros fatores externos de influência como é o caso de distanciamento social.

Da análise feita das variáveis de influência, obteve-se os seguintes coeficientes:

$$\begin{array}{ll} \beta_0 = 3474.32 & \beta_2 = -708.91 \\ \beta_1 = -409.53 & \beta_3 = -9.46 \end{array}$$

Sendo: $x_1 = \text{Área geográfica (Província)}$
 $x_2 = \text{Género da população suscetível}$
 $x_3 = \text{Idade da população suscetível}$

Assim, teremos o seguinte modelo matemático que explique o perfil epidemiológico da pandemia de COVID-19 em Moçambique:

$$Y = 3\,474.32 - 409.53 * \text{Area}_g - 708.91 * \text{Genero} - 9.46 * \text{Idade}$$

Interpretação dos coeficientes

β_0 - Número esperado de casos de COVID-19 no grupo de referência (masculino, urbano, idade) quando as variáveis de influência são nulas (igual a Zero).

β_1 - Diferença média no número de casos entre mulheres e homens, mantendo as outras variáveis constantes.

β_2 - Diferença média no número de casos de COVID-19 entre áreas rurais e urbanas, ajustando para sexo e idade.

β_3 - Variação média no número de casos de COVID-19 para cada ano adicional de idade que entra no grupo de infetados pela doença.

O modelo sugere que, os casos de COVID-19 são mais comuns em moradores das cidades, em homens e pessoas mais jovens, isto é, ser da área urbana, homem e de idade ativa têm maior probabilidade de influenciar no aumento significativo do número de casos do que ser da zona rural, mulher e criança. Esse padrão de doença faz sentido olhando nos contextos de que as áreas urbanas têm maior densidade populacional e exposição devido aos contactos com pessoas infetadas, os homens estão mais expostos à procura de

sustento familiar e as crianças e idosos têm baixa imunidade e maior vulnerabilidade necessitando sempre dum isolamento.

O modelo prevê aproximadamente a 3 475 o número de casos de COVID-19 em Moçambique (inferido) até ao momento da coorte quando a área geográfica, o género e a idade forem iguais a 0 (zero).

Conforme o modelo, os mais jovens têm maior número de casos de COVID-19 do que os mais velhos. A cada aumento de 1 ano, mantendo constantes as variáveis área geográfica e género, o número estimado de casos diminui em cerca de 9,5 unidades.

Para $\beta_1 = -409.53$ (área geográfica) significa que, mantendo género e idade constantes, viver em área rural está associado a uma média de 409,5 casos a menos de COVID-19 do que viver em área urbana.

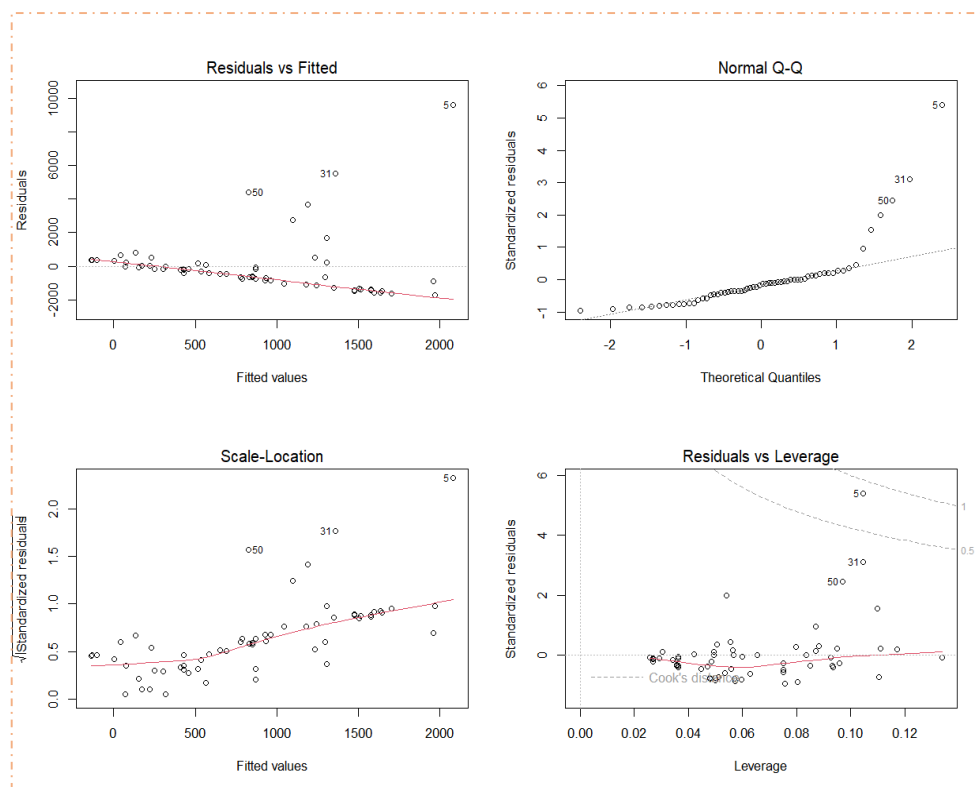
A necessidade da utilização da variável “área geográfica – província” e sua inclusão no modelo é de procurar explicar se a variação do número de caso de COVID-19 no país e em particular nas províncias em análise é influenciada pela concentração de massa população nas zonas urbanas ou o fraco acesso de meios de prevenção a tempo real bem como a inobservância do cumprimento de políticas de controlo da pandemia.

Análise gráfica do Modelo

A análise de resíduos, seja com gráficos de resíduos ou testes estatísticos, fornece um conjunto simples, mas poderoso, de ferramentas analíticas para o exame da adequação de nosso modelo de regressão. No entanto, muito frequentemente essas análises não são feitas e as violações de suposições são mantidas intactas.

O ajustamento do modelo da série original, os resíduos das observações mostram uma determinação sobre a tendência dos casos do COVID-19 em Moçambique. Apesar do variável tempo não influenciar diretamente no comportamento do fenómeno em estudo, os fatores externos como climáticos, ambientais e socioeconómicos têm uma influência direta na tendência dos casos da pandemia.

Gráfico 05: Análise gráfica dos pressupostos do modelo linear



O gráfico acima permite analisar a Linearidade, a normalidade dos resíduos e a homocedasticidade. O primeiro quadro apresenta o estágio dos resíduos de valores ajustados para analisar a linearidade dos resíduos do modelo de regressão linear dos fatores impulsionadores do problema em estudo. No entanto, os resíduos do modelo não são lineares.

O segundo quadro, evidencia os resíduos padronizados versus resíduos teóricos para compreender a normalidade dos resíduos. Analisando a variabilidade dos resíduos em torno da reta de regressão, a concentração dos pontos mostra que os resíduos estão normalmente distribuídos, isto é, apresentam uma distribuição normal.

No que concerne a análise da homocedasticidade, os resíduos permitem efetuar a descrição de dados para os quais a variância dos termos de erro (ϵ) aparece constante no intervalo de valores de uma variável independente. Ora vejamos, os resíduos não apresentam um padrão triangular em torno da linha de regressão, o que permite afirmar que os resíduos não apresentam homocedasticidade.

Em relação a análise dos dados que fogem da normalidade e se diferenciam drasticamente de todos os outros dados (Outliers), pode-se afirmar que existe Outliers nos

resíduos e podem causar anomalias nos resultados obtidos por meio de algoritmos e sistemas de análise do modelo gerado a partir das variáveis de influência.

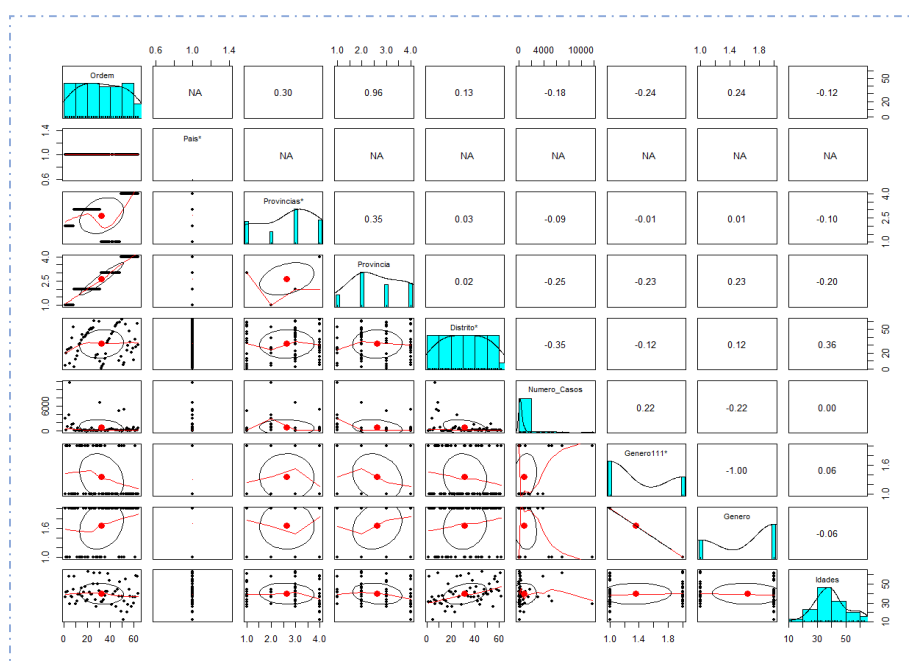
Multicolinearidade entre as variáveis

Uma questão-chave na interpretação da variável estatística de regressão é a correlação entre as variáveis independentes. Esse é um problema de dados, e não de especificação de modelo. A situação ideal para um pesquisador seria ter diversas variáveis independentes altamente correlacionadas com a variável dependente, mas com pouca correlação entre elas próprias.

O VIF traduz o valor de tolerância, o qual expressa diretamente o grau de multicolinearidade em um impacto sobre o processo de estimação. Quando o erro padrão aumenta, os intervalos de confiança em torno dos coeficientes estimados ficam maiores, dificultando a demonstração de que o coeficiente é significativamente diferente de zero.

Os valores de VIF (área geográfica = província: 1.095, género: 1.056 e idade: 1.041) mostram que não há uma forte associação entre as variáveis independentes, pois estes são iguais a 1.0, permitindo assim afirmar que não existe multicolinearidade entre as variáveis (fatores associados ao comportamento dos casos de COVID-19).

Gráfico 06: Análise da multicolinearidade através do gráfico



Análise Residual

Para analisar e/ou avaliar se um conjunto de observações apresentam uma distribuição normal utiliza-se o Teste de Shapiro-Wilk. A distribuição normal também pode ser chamada de gaussiana e sua forma assemelha-se a de um sino. Esse tipo de distribuição é muito importante, por ser frequentemente usada para modelar fenômenos naturais.

O pressuposto da normalidade dos resíduos é que o *p-value* que subescreve a hipótese nula seja maior de 0.05, para satisfazer a afirmação de aceitar não rejeitar a hipótese nula. O teste de normalidade (Shapiro-Wilk Test) mostra que os resíduos apresentam uma distribuição aproximadamente a normal pelo que aceita afirmar que os resíduos do modelo têm uma distribuição normal.

Análise de Outliers dos Resíduos

Os resíduos padronizados mostram que os resíduos do modelo não estão dentro do padrão, pois a mediana se localiza fora dos extremos mínimo e máximo dos resíduos.

Análise da Independência

O *p-value* reforça o valor do teste de Durbin-Watson que mostra que não existe resíduos dependentes ou autocorrelacionados.

Análise da Homocedasticidade

De acordo com os resultados do teste de Breusch-Pagan, o *p-value* situa-se abaixo do valor padrão, denunciando assim a não existência da homocedasticidade, pois o *p-value* é inferior a 0.05 o que explica a inexistência de uma variância constante dos resíduos. A variância ser constante equivale a supor que não existem observações incluídas na variável residual cuja influência seja mais intensa na variável dependente.

[b] Modelo Epidemiológico SIR

Para descrever a evolução de uma determinada doença contagiosa dentro duma população, pensa-se a priori em modelos compartimentais, com ênfase para SIR que divide a população em três classes disjuntas: suscetíveis (S), formada por indivíduos não infetados e que podem contrair a doença; infetados (I), indivíduos que têm a doença e podem transmiti-la e recuperados (R), indivíduos que contraíram a doença e foram curados e adquiriram imunidade em determinado período de tempo (em tempo contínuo ou discreto).

O modelo SIR, proposto por Kermack e McKendrick em 1927, é, até então considerado de um modelo de compartimentos que caracteriza o estado do indivíduo ao longo do desenvolvimento de uma doença, relacionando-os por meio de um sistema não linear de equações diferenciais.

Os dados apresentados para o presente trabalho mostram que, no início da epidemia a fração de indivíduos infetados tendem a um aumenta exponencialmente, sob pressuposto de que um indivíduo infetado encontra apenas indivíduos suscetíveis e permite que o patógeno se dissipe facilmente, e na medida em que o tempo passa, o indivíduo infetado deixa de encontrar pessoas suscetíveis colocando a curva da propagação da doença na tendência decrescente com maior probabilidade da epidemia terminar quando todos os suscetíveis forem infetados.

Tabela 01: Dados da pandemia da COVID-19

Província	População (N)	População Exposta (S)	Casos Positivos (I)	Altas (R)	Taxa de Transmissão (β)	Taxa de Recuperação (γ)
Maputo	2 131 384	1 811 676	23 783	22 928	0,35	96,40
Nampula	6 036 169	5 130 744	11 089	10 997	0,05	99,17
Cabo Delgado	2 455 072	2 086 811	8 177	8 169	0,04	99,90
Niassa	1 933 505	1 643 479	9 886	9 872	0,06	99,86
	12 556 130	10 672 711	52 935	51 966	0,05	98,17

Fonte: MISAU - Base de dados de COVID-19

Como pode-se entender sobre os modelos compartimentas, baseiam-se na divisão da população em compartimentos, isto é, tomando em consideração as características físicas

e epidemiológicas do estado em que se encontram os indivíduos na cadeia da doença (Suscetíveis, Infetados, Recuperados).

A tabela acima apresenta os dados de população Exposta (S= Suscetível), Casos Positivos (I=população Infetada), Altas (R=população Recuperado) que permitirão simular e modelar as epidemias e/ou pandemias em larga escala populacional com fundamentos teóricos e matemáticos.

Esquemáticamente, o modelo compartimental SIR pode-se representar da seguinte forma:

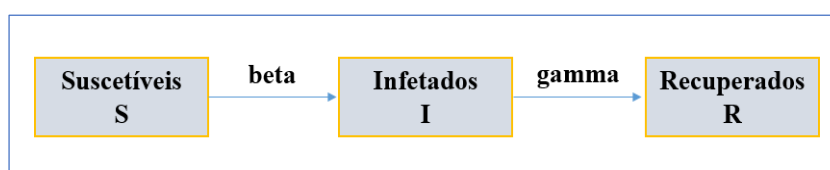


Figura 01: Esquema do modelo compartimental SIR

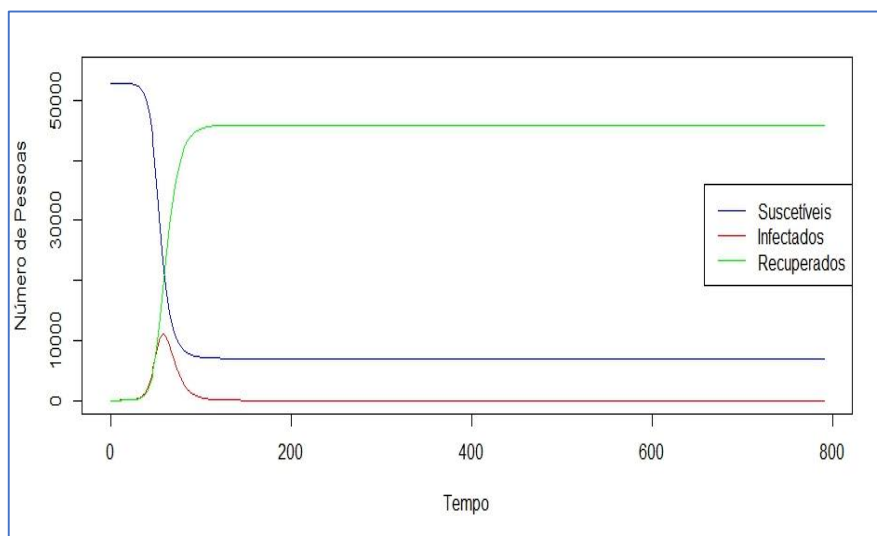
A partir do esquema do modelo compartimental apresentado na figura 02 acima, monta-se o modelo SIR com dinâmica vital em tempo discreto. Para realizar as simulações que se seguem utilizamos o software *RStudio* para plotar as curvas do modelo.

Consideramos como unidade de tempo o primeiro dia de registo do primeiro caso de COVID-19 em uma progressão do tempo de 1095 dias (correspondente a 3 anos) de análise de dados e como parâmetros, a taxa de transmissão e recuperação de 0.05 e 0.98 respetivamente para uma doença cujo tempo de recuperação é de 7 dias.

Vale referir que, devido à insuficiência de fontes fiáveis para a obtenção de dados sobre os parâmetros do modelo epidemiológico em análise, estes, juntamente com a taxa de infeção, foram obtidos através de estimativas manual do investigador. Os valores de β e γ foram estimados com base em conhecimento prévio do comportamento observado da doença, isto é, β (Taxa de Transmissão) foi estimado com base em informações sobre a probabilidade de transmissão da doença entre indivíduos e γ (Taxa de Recuperação) estimado com base no tempo médio de recuperação da doença dum indivíduo infetado pela COVID-19.

Abaixo, ilustramos o comportamento da doença de forma geral para as 4 províncias em estudo, mostrando as curvas de infeção e recuperação da doença.

Gráfico 07: Curvas Gerais do modelo SIR para COVID-19



Fonte: MISAU - Base de dados de COVID-19

O gráfico acima mostra como a doença se espalhou no período de transmissão massiva e como ela foi controlada à medida que a população atingia uma quantidade significativa de pessoas imunes.

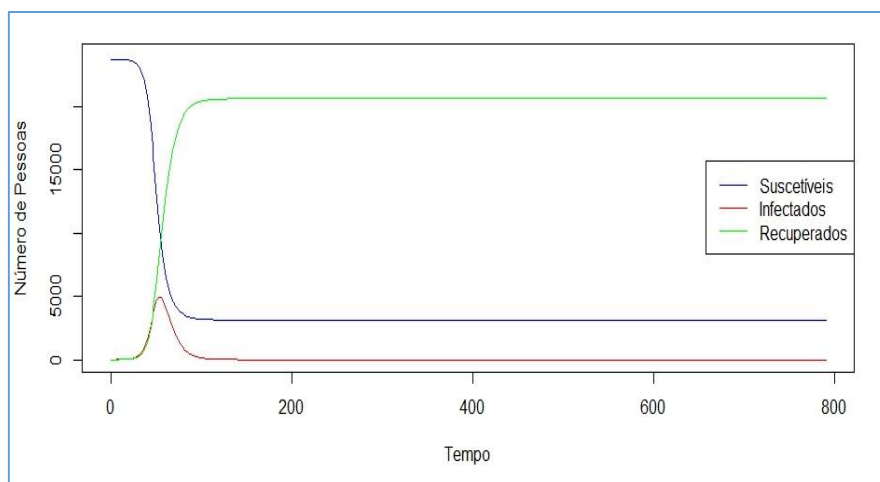
Como se pode observar, tem-se a explicar nos seguintes termos:

A *Curva S (Suscetíveis)*: no início da pandemia, assumir-se-ia que quase toda a população estava exposta ao fator de risco e que poderia contrair a doenças tomando o ponto mais alta e, diminui na medida em que mais pessoas se infetavam e se tornavam imunes (sem descorar a probabilidade de co infetar-se). A curva começa com mais de 50 mil pessoas suscetíveis e foi reduzindo até cerca de 8 mil.

A *Curva I (Infetados)*: começa baixa (com um número de pessoas infetadas inferior a 10 unidades estatística no princípio da pandemia), aumentando com o passar do tempo até ao pico (cerca de 53 mil pessoas infetadas) e depois diminuiu na medida em que os infetados se recuperam e há menos suscetíveis restantes.

A *Curva R (Recuperados)*: começa em zero até o 7º dia de seguimento dos infetados e a partir do 8º dia se registam os primeiros indivíduos recuperados da doença e, aumenta continuamente ao longo do tempo conforme as pessoas se recuperam.

Gráfico 08: Curvas do modelo SIR para COVID-19 _ Província de Maputo

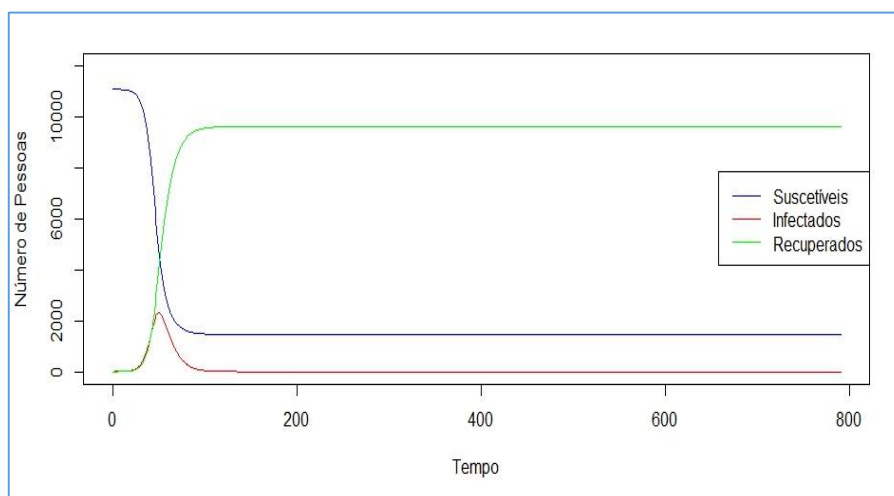


Fonte: MISAU - Base de dados de COVID-19

O gráfico 08 ilustra as curvas do modelo Epidemiológico SIR da província de Maputo. A Curva S (Suscetíveis) inicia alta no período crítico da transmissibilidade da doença e desce em direção ao valor mínimo reportado. No mesmo período, a curva I (Infetados) cresce rapidamente, até determinado período de tempo atinge o pico e depois começa a decrescer. A curva R (Recuperados) começa em zero e aumenta gradualmente na medida em que os infetados em seguimento médico a partir do 8º dia registam melhorias.

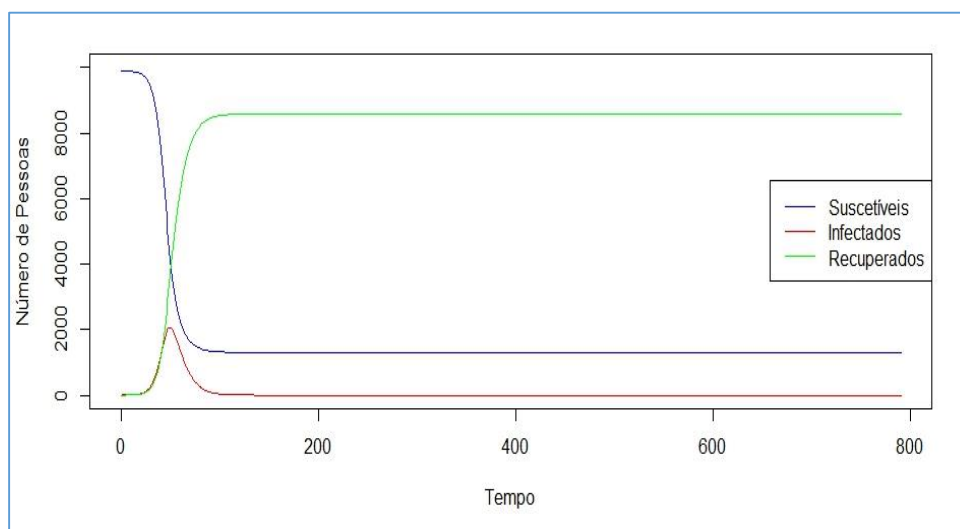
O mesmo cenário se regista em todos outros gráficos abaixo apresentados.

Gráfico 09: Curvas do modelo SIR para COVID-19 _ Província de Nampula



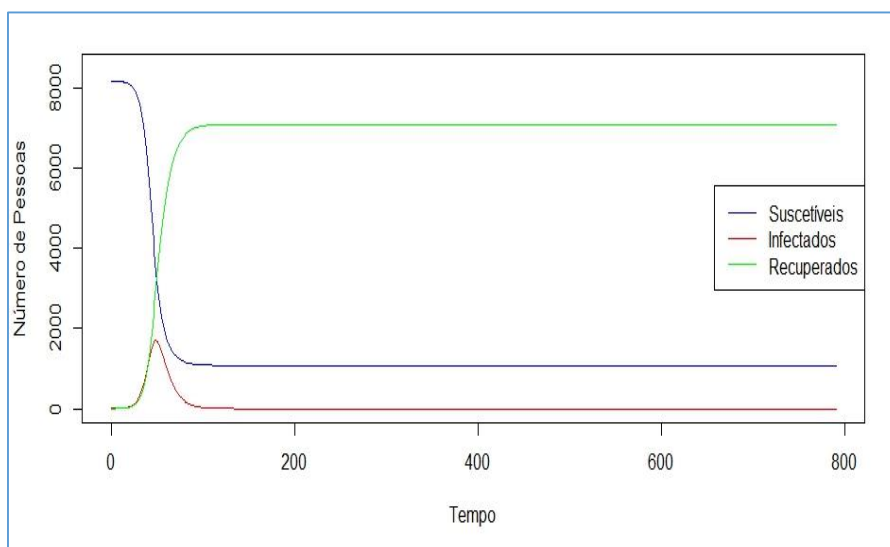
Fonte: MISAU - Base de dados de COVID-19

Gráfico 10: Curvas do modelo SIR para COVID-19 _ Província de Cabo Delgado



Fonte: MISAU - Base de dados de COVID-19

Gráfico 11: Curvas do modelo SIR para COVID-19 _ Província do Niassa



Fonte: MISAU - Base de dados de COVID-19

4.2.2. Visualização de dados do COVID-19 em Moçambique

Existem diversos tipos de geração de informação, mas o formato visual ganha muitos pontos quando se trata de entender como estão os resultados, possibilitando que

qualquer pessoa consiga consumir a informação. A visualização dinâmica de dados é uma das formas culturais genuinamente novas que se tornaram possíveis graças à computação.

Com os computadores podemos visualizar conjuntos diferenciado de dados muito mais amplos, criar visualizações dinâmicas, alimentar dados em tempo real, basear as representações gráficas de dados em sua análise matemática, usando vários métodos, da estatística clássica à prospeção de dados, mapear um tipo de representação em outro (imagens em sons, sons em espaços tridimensionais, etc.).

A visualização de dados resulta em informações que são mais facilmente compreensíveis, pois traduz a complexidade do relacionamento entre as variáveis em modelos visuais.

[a] Visualização de Dados em Dashboard

Os *Dashboards* são painéis que mostram métricas e indicadores importantes de forma visual, facilitando a compreensão das informações geradas para o consumo público.

O objetivo dos dashboards é de possibilitar o monitoramento dos resultados de uma organização ou programa, distribuídos em diversos indicadores, procurando responder perguntas essenciais para ter as respostas desejadas com a apresentação da informação neste fórum.

Qualquer tipo de informação pode ser configurado para exibição em um dashboard, desde que tenha valor para o negócio ou gestão. Alguns exemplos: taxa de ROI sobre projetos, informações sobre compras em sua loja virtual, desempenho de unidades filiais da empresa, detalhes sobre a disponibilidade de equipamentos, dados sobre atendimento aos clientes (ex: service desk), funcionamento de sistemas críticos, perfil epidemiológico e entre diversos outros parâmetros.

Figura 02: Apresentação da Situação Epidemiológica das Províncias de Maputo, Nampula e Cabo Delgado - COVID-19



Fonte: MISAU - Base de dados de COVID-19

A figura 02 acima ilustra um dashboard operacional. Este tipo de dashboard permitiu monitorar os níveis de operação de estratégias levadas a cabo para o controlo da pandemia da COVID-19 em Moçambique e em particular nas províncias em análise (Província de Maputo, Nampula, Cabo Delgado e Niassa).

Conforme a classificação dos dashboards, os operacionais normalmente mostraram as métricas que devem ser acompanhadas por provedores de cuidados de saúde, pesquisadores e até analistas para realizarem otimizações em seus trabalhos e assim permitir maior agilidade em correções de estratégias e ações implementadas para controlar a pandemia da COVID-19. Durante a vigilância dos casos de COVID-19 nas 4 províncias em análise, foram operacionalizadas diferentes estratégias combinadas para o controlo da propagação e das consequências da doença com a criação de Clusters (Vigilância, Logística, Manejo de Casos e Wash).

A partir do dashboard foi possível diariamente fazer o acompanhamento dos progressos das intervenções realizadas para o controlo da transmissão e o estágio da pandemia no concernente aos novos casos notificados de forma geral (nacional) e particularizada, focada para as províncias em estudo. É possível visualizar no mapa os pontos epicentros onde eclodiu a pandemia (cidades capitais provinciais) e ao longo do tempo foi-se alastrando para as vilas municipalizadas e vilas distritais.

[b] Visualização em Tabelas Resumidas

As tabelas resumidas agrupam registros de valores exclusivos e um ou mais campos-chave efetuando uma contagem para o número de registros conforme a correspondência.

Tabela 02: Distribuição de Casos por Província

PROVÍNCIAS	NÚMERO DE CASOS
Cabo Delgado	8 177
Maputo	23 783
Nampula	11 089
Niassa	9 886
Total Geral	52 935

Fonte: MISAU - Base de dados de COVID-19

De forma global, até ao momento da coorte dos casos reportados pelas províncias em estudo, a província de Maputo lidera com o número de casos positivos reportados ao longo do tempo, seguido da província de Nampula com uma diferença de 24% do total de casos reportados pela província de Maputo. No período em análise, das 4 províncias em análise, a província de Cabo Delgado é a que notificou menos casos positivos de casos de COVID-19.

○ *Tabela de Frequências*

Este tipo de tabelas exhibe a frequência de valores entre cada combinação exclusiva de categorias.

Tabela 03: Distribuição de Casos Positivos Vs Peso da positividade por província

PROVINCIAS	CASOS POSITIVOS	PERCENTAGEM
Maputo	23 783	45
Nampula	11 089	21
Cabo Delgado	8 177	15
Niassa	9 886	19
TOTAL	52 935	100

Fonte: MISAU - Base de dados de COVID-19

A tabela acima elucida que a concentração de pessoas diagnosticadas positiva para a COVID-19 está na província de Maputo (Cidade de Matola), seguida da maior concentração de casos positivos na província de Nampula (Cidade de Nampula) com percentagens cumulativas desde do início da pandemia até ao período da coorte com 45 e 21% respetivamente.

○ **Tabela Cruzada**

Este tipo, resume a relação entre duas ou mais variáveis de categoria, descrevendo o número de vezes que cada uma das possíveis combinações de categorias ocorreu num conjunto de dados em análise.

Tabela 04: Distribuição de Casos por área geográfica (Província) e Género

	GÉNERO		TOTAL
	Masculino	Feminino	
Maputo	12 153	11 630	23 783
Nampula	5 627	5 462	11 089
Cabo Delgado	4 832	3 345	8 177
Niassa	4 955	4 931	9 886
TOTAL	27 567	25 368	52 935

Fonte: MISAU - Base de dados de COVID-19

Os dados apresentados na tabela 4 ilustram a informação dos doentes dos COVID-19 confirmados positivos discriminados por Género e localização (província de ocorrência), onde podemos destacar o Género masculino como o mais afetado pela doença e a maior concentração de casos positivos masculinos na província de Maputo. As características socioeconómicas do país e em particular nas províncias vincadas no trabalho, contribuíram significativamente na concentração de casos positivos em indivíduos de Género masculino quando comparado com o Género feminino em todas áreas geográficas.

A mobilidade de pessoas e bens bem como a “luta” pela sobrevivência e autossustento das famílias pressupõem-se como o pano de fundo que coloca os indivíduos de Género masculino mais suscetíveis a alto fator de risco de se infetar por COVID-19.

[c] Gráficos

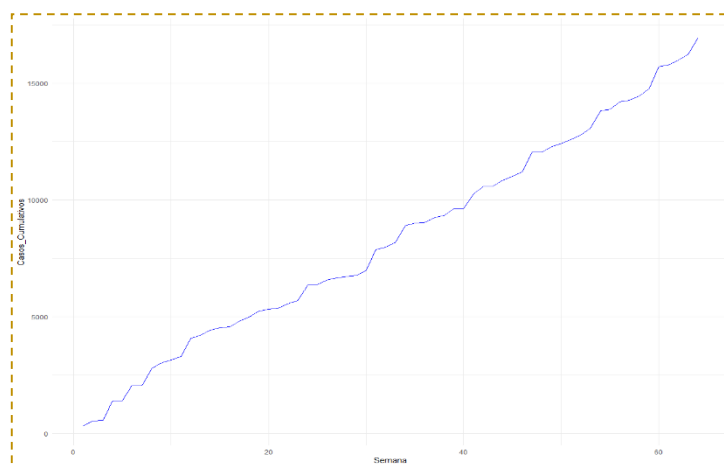
Os gráficos são úteis para comparar categorias de dados, onde o eixo *x* representa a categoria e o eixo *y* representa o valor a comparar. Pode-se citar diferentes tipos de gráficos a considerar: Gráfico de linha ou linear, de barras, de colunas, histograma, circular ou de pizza, teia de aranha, etc. todos que exibem a distribuição dos dados.

Para os dados em análise, o sector da saúde utiliza os gráficos de linha, de coluna e de barras para visualizar os dados da pandemia da COVID-19 em moçambique.

○ *Gráfico de Linha*

Os gráficos de linhas mostram tendências ou alterações ao longo do tempo exibindo uma série de pontos de dados conectados por segmentos de linha reta. Foi utilizado para monitorar mudanças ao longo do tempo para ajudar na análise de dados preditiva.

Gráfico 12: Dados Cumulativos de Casos Confirmados Positivos

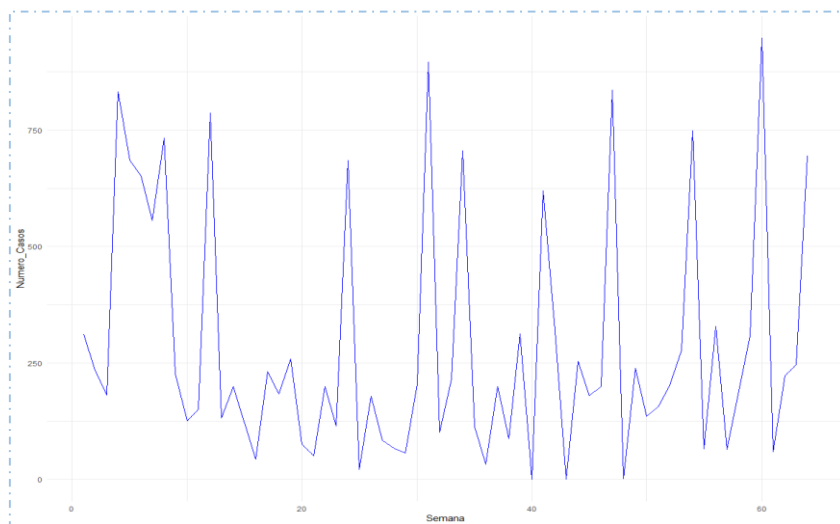


Fonte: MISAU - Base de dados de COVID-19

O gráfico 12 ilustra o número cumulativo de casos positivos de COVID-19 em Moçambique e de forma particular nas 4 províncias amostrais em análise desde a notificação do primeiro caso da doença até o período da coorte.

Foi utilizado para monitorar mudanças ao longo do tempo para ajudar na análise de dados preditivos e cumulativamente sobre o progresso na notificação, seguimento e reporte dos dados.

Gráfico 13: Série Temporal de Casos Positivos de COVID-19



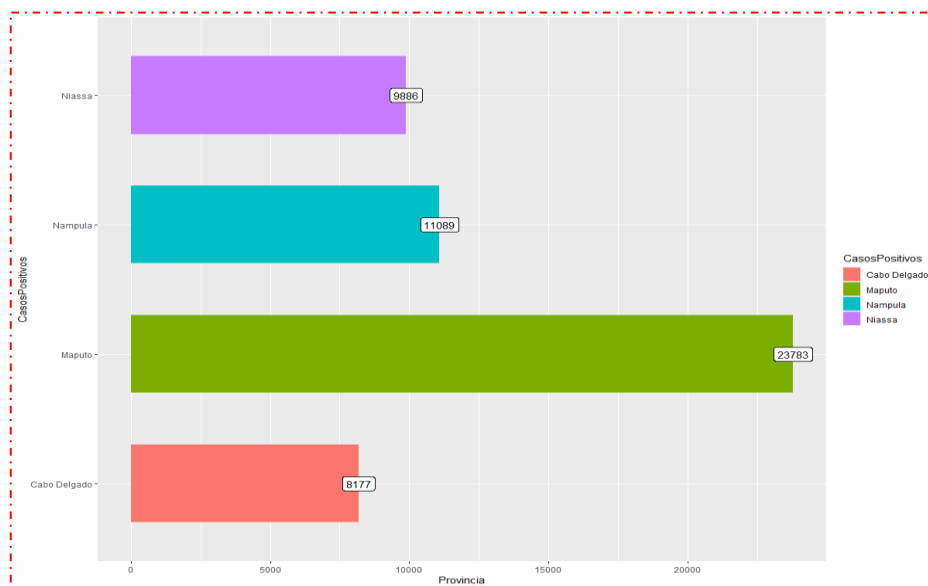
Fonte: MISAU - Base de dados de COVID-19

O gráfico 13 acima mostra o comportamento (a tendência) do número semanal de casos confirmados positivos de COVID-19 em 4 províncias do país. Feita a análise ao longo tempo, desde a notificação do primeiro caso da pandemia no país, devido à fraca vigilância e controlo de casos, em algumas semanas de seguimento dos contactos de casos positivo da doença caracterizou-se por envio de amostras espezas para a análise laboratorial devido à redução da oferta de meios, reduzindo a taxa de positividade de casos reportados em aproximadamente 3% até a data da extração de dados em todas as províncias envolvidas no estudo.

○ *Gráfico de Barras*

Um *gráfico de barras* é composto por barras com formato retangulares, onde o comprimento de cada barra de forma proporcional representa o valor que do evento correspondente. Geralmente é utilizado para comparar uma ou duas variáveis usando um único valor.

Gráfico 14: Distribuição de Casos Positivos Confirmados de COVID-19 por Província



Fonte: MISAU - Base de dados de COVID-19

O gráfico 14 ilustra-nos a distribuição de pessoas infetada pelo vírus corona (COVID-19) nas 4 províncias mais afetadas do país e que durante a vigilância e seguimento de casos estas apresentaram um perfil epidemiológico com transmissão comunitária e evolução galopante e acelerado de casos da COVID-19 em curto tempo (nos princípios da eclosão da doença).

A situação de Cabo-Delgado incitou grandemente na mudança do perfil epidemiológico da província de Nampula ao apresentar um padrão diferente devido a insurgências.

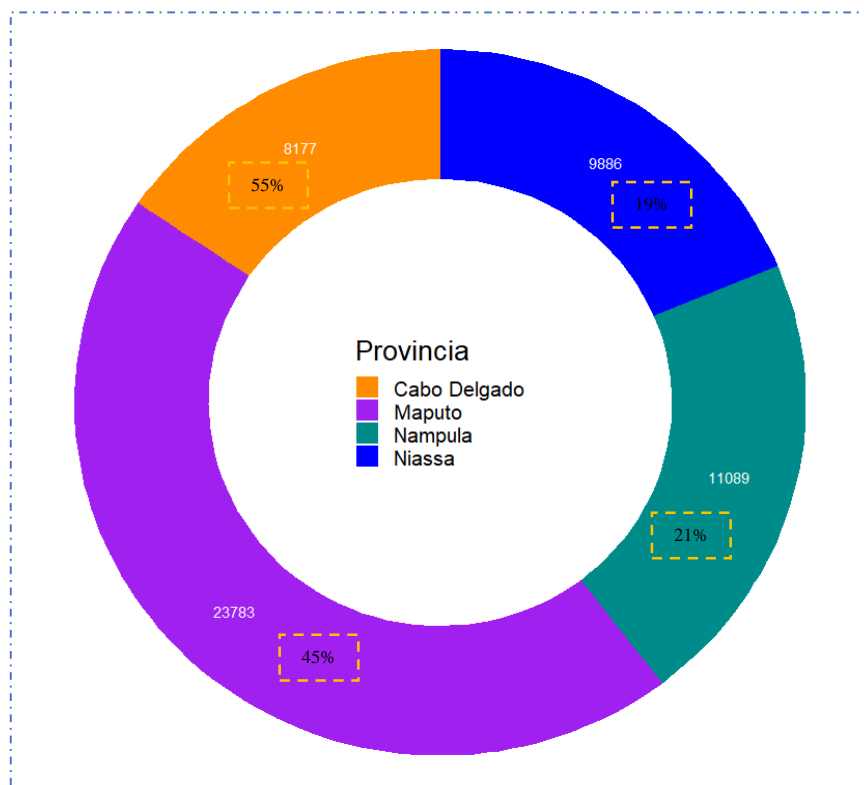
○ *Gráfico Circular ou de Pizza*

Os *gráficos de pizza* mostram a proporção ou percentagem total das categorias. Este tipo de gráfico procura traduzir a composição de dados categóricos com cada segmento proporcional à quantidade que representa.

O mesmo conjunto de dados pode ser visualizado de diferentes formas. Não diferente, a comparação da proporção de ocorrências de diferentes categorias pode ser feita de várias maneiras em um gráfico circular simples bem como em um gráfico de setores *donut chat*.

O gráfico abaixo, ilustra a comparação da proporção de indivíduos de cada uma das 4 províncias em referência na base de dados utilizada para análise do perfil epidemiológico do COVID-19 em Moçambique.

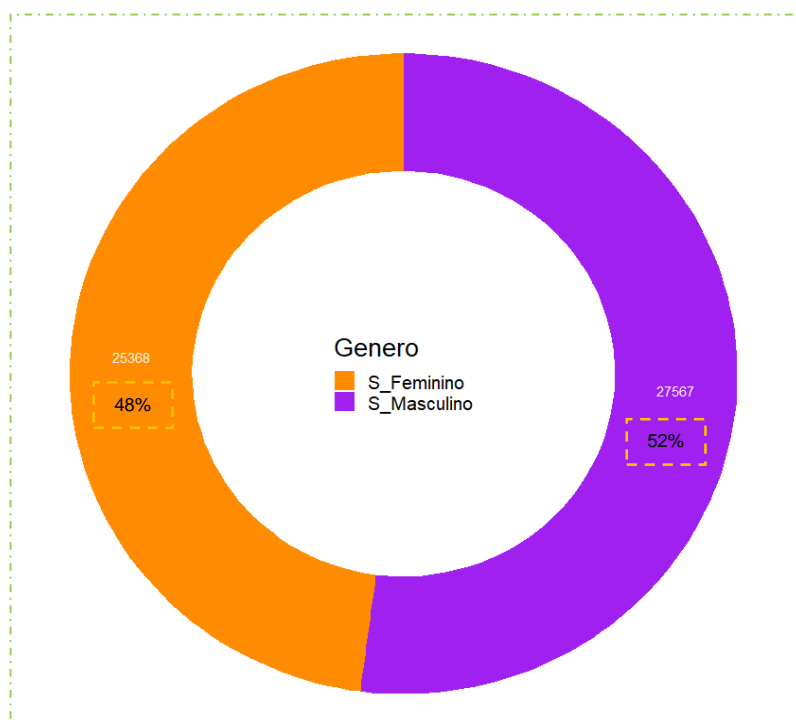
Gráfico 15: Distribuição de Casos Positivos de COVID-19 por Área Geográfica



Fonte: MISAU - Base de dados de COVID-19

O gráfico acima ilustra a distribuição e concentração geográfica de número de pessoas diagnosticadas positiva para a COVID-19 no país, que de forma particular no contexto da análise das províncias amostrais de zona norte e sul de moçambique, com destaque para as províncias de Maputo (Cidade de Matola) e Nampula (Cidade de Nampula) com elevado número de casos notificados e reportados.

Gráfico 16: Distribuição de Casos Positivos Confirmados de COVID-19 por Género



Fonte: MISAU - Base de dados de COVID-19

Os dados ora apresentados no gráfico 16, mostram que até o período da corte, os indivíduos de género feminino foram as menos afetadas pelo COVID-19 em detrimento de pessoas de género masculino.

De acordo com os resultados apresentados no gráfico 7, é notória que a população de sexo masculino foi a mais afetada pelo vírus corona nas províncias em análise, com 52% do total do número de casos de COVID-19.

A luta pela sobrevivência pode explicar de forma substancial a variabilidade dos casos notificados positivos em indivíduos de sexo masculino em todas as províncias do país e em particular nas províncias destacadas.

[d] Mapas

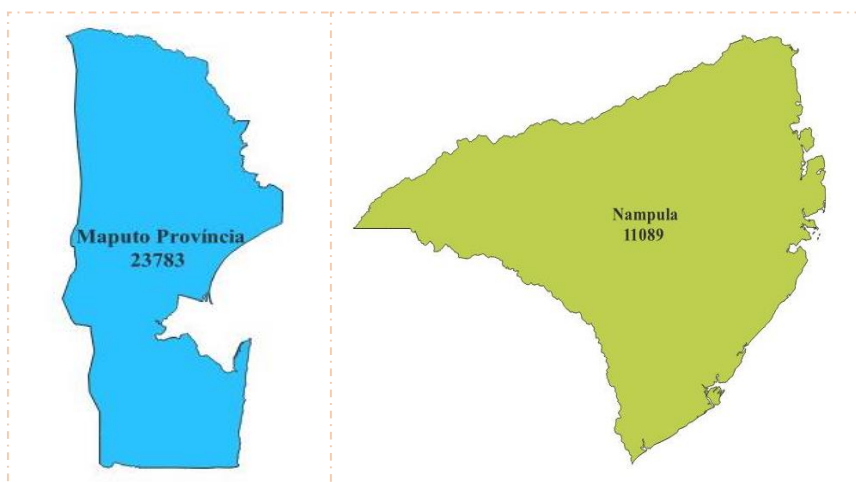
Na atualidade e com os avanços tecnológicos, o mundo digital veio transformar a realidade num ambiente digital e, a complexidade dos sistemas tornaram-se mais simplificados da carga de tempo a depender para visualização de massas de dados. Os mapas, têm uma história antiga e continuam sendo uma ferramenta essencial nos dias de

hoje, quer que seja para servir de guia durante viagens ou para analisarmos a distância entre lugares.

[b].1. Visualização de dados em Mapas com recurso a Software QGIS

Nos momentos atuais, os mapas tornaram-se referência na visualização de dados, possibilitando o homem a ver dados de localização representados em um mapa e incluídos em visualizações que permitam o entendimento de fenómenos e casos de estudo.

Figura 03: Áreas de Alto Risco de Transmissão do COVID-19



Fonte: MISAU – Resultados da pesquisa

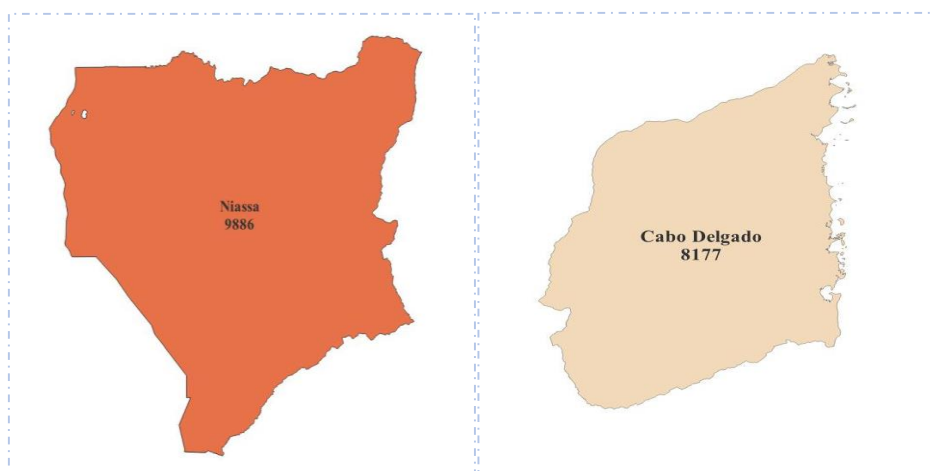
De acordo com a figura 03, os mapas visualizam o número total de casos notificado e reportado nos dois distritos de alta endemicidade e com características quase similares com histórico de transmissão comunitária do vírus corona nas primeiras 5 semanas após a notificação do 1º caso da doença. Com a transmissão comunitária, as províncias de Maputo e Nampula passaram a ser consideradas de áreas de alto risco de transmissão da COVID-19.

Os polígonos da figura 03 (Províncias de Maputo e Nampula) por apresentarem uma concentração de massa populacional regular nas zonas urbanas, com maior destaque para os locais públicos representados pelos mercados existentes nas cidades, proporcionaram vários aglomerados populacionais suscetíveis de contrair a doença, como

maior ênfase na província de Nampula com a receção de deslocados de Cabo Delgado como resultado dos conflitos armados (Insurgentes). Também, algumas áreas de baixa densidade populacional (zonas rurais) apresentaram uma elevada concentração descontrolada de pessoas nos mercados, igrejas e campos de futebol proporcionando uma proliferação de pessoas a procura de autossustento e lazer.

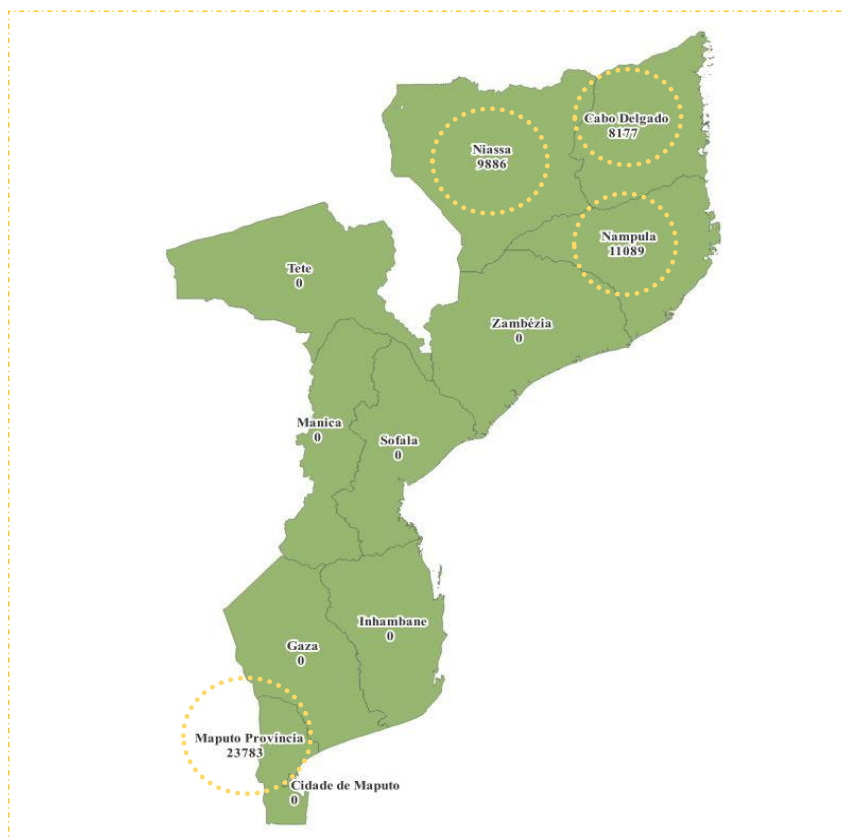
O controlo da transmissibilidade por COVID-19 nos postos de vigilância ativa criados pelas entidades de saúde no país e em particular nas províncias para dar resposta a identificação de casos suspeitos de COVID-19 bem como as pessoas de contacto com casos confirmados positivos para o rastreio e seguimento clínico, proporcionou grandes avanços na efetivação do plano de resposta do COVID-19 em áreas de alta transmissibilidade.

Figura 04: Áreas de baixo e/ou medio Risco de Transmissão da COVID-19



A figura 04 ilustra as regiões do país que registaram baixo e médio risco de transmissibilidade devido a dimensão regional e por outro lado os movimentos migratórios devido a insurgência registada na província de Cabo Delgado, reduzindo o risco de transmissibilidade do COVID-19 como consequência da baixa mobilidade interna e externa.

Figura 05: Distribuição Geográfica de Casos e Níveis de Risco de transmissão



Fonte: Resultados da pesquisa

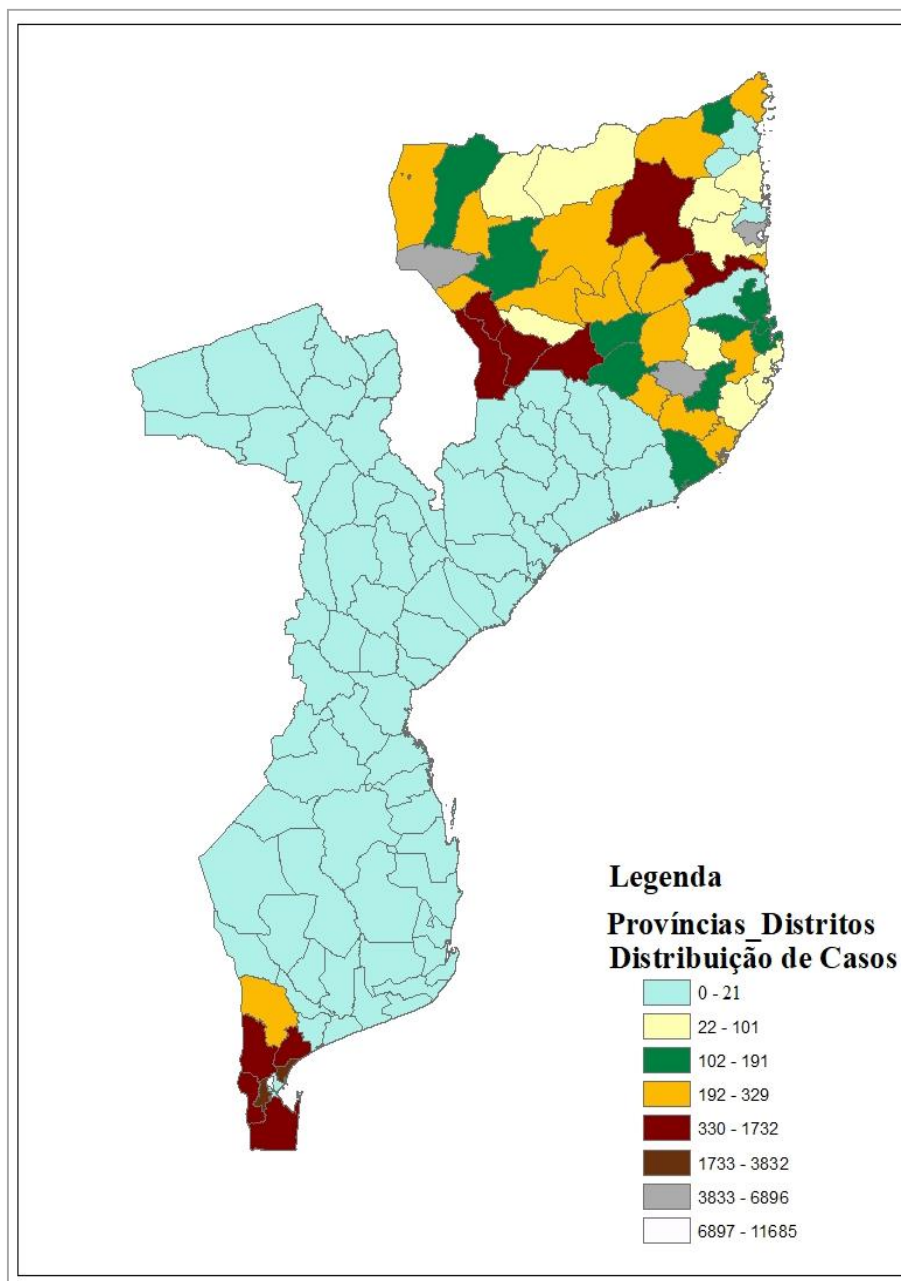
O mapa apresentado acima ilustra a distribuição dos níveis de risco, casos confirmados positivos para o COVID-19. Apesar de a cidade capital do país – Maputo seja a primeira área geográfica do país a notificar o caso da COVID-19 em Moçambique, as duas províncias do país (Nampula e Província de Maputo) apresentaram alta endemicidade com casos reportados que em algumas áreas geográficas chegaram a reportar cerca de 6896 e 11685 casos confirmados positivos, respetivamente e alto risco de transmissibilidade da pandemia bem como declaradas províncias com transmissão comunitária.

[c].1. Visualização de dados em Mapas com recurso a Software ArcGIS

Como se referiu na introdução de modelos e técnicas de visualização de dados, os dados de COVID-19 reportados durante o período de corte ora referenciado ao longo do trabalho, podem ser visualizados em mapas coropléticos (mapa temático em que cores,

sombreamentos ou padrões são usados para representar o número de casos de COVID-19) simbolizando o número de casos e sua concentração em cada região.

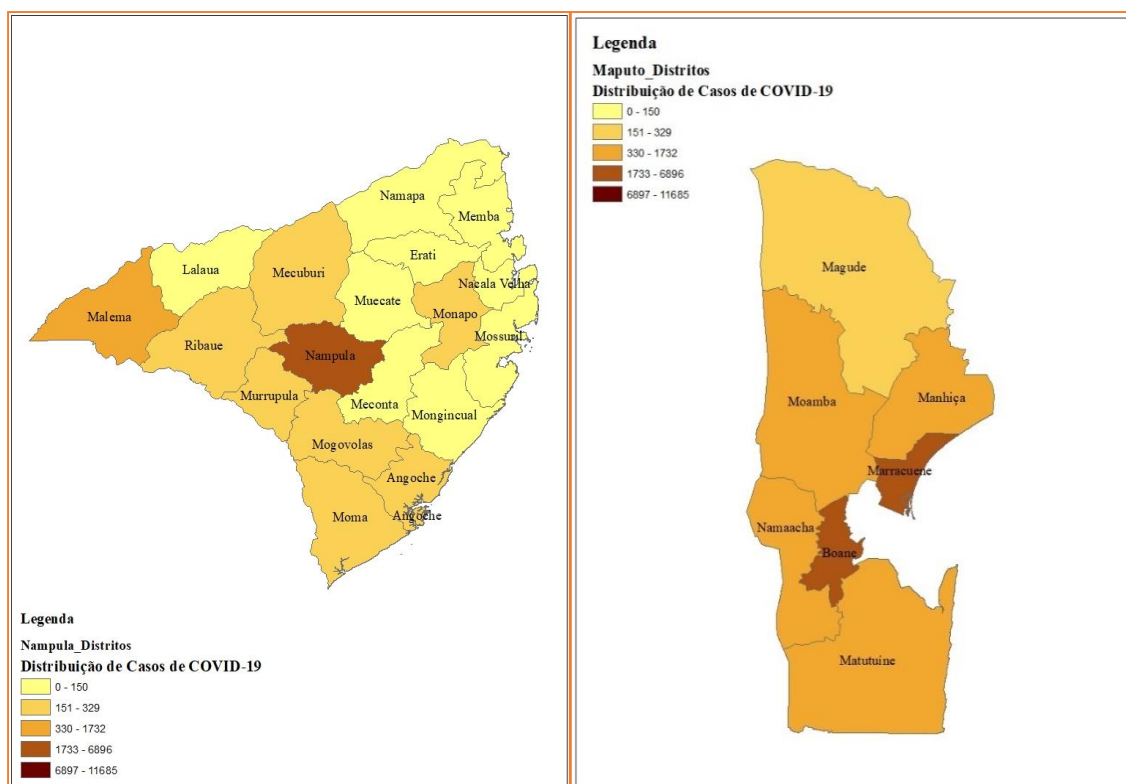
Figura 06: Distribuição Geográfica de Casos de COVID-19 no País - Moçambique



Fonte: Resultados da pesquisa

A figura acima representa o mapa de moçambique no contexto de visualização de mapas coropléticos. Moçambique tem 11 províncias e deste foram seleccionados 4 para se analisar os dados da pandemia de COVID-19. As províncias do centro e sul com exceção

Figura 08: Distribuição Geográfica de Casos de COVID-19 – Concentração de Casos em Coroplético



A análise de dados em mapas de cores ou simplesmente coropléticos no contexto da COVID-19 envolve a interpretação das informações apresentadas no mapa para extrair grandes significados sobre a transmissibilidade e a propagação do vírus corona por áreas geográficas, permitindo também visualizar a distribuição dos casos, sua tendência e seus impactos geográficos.

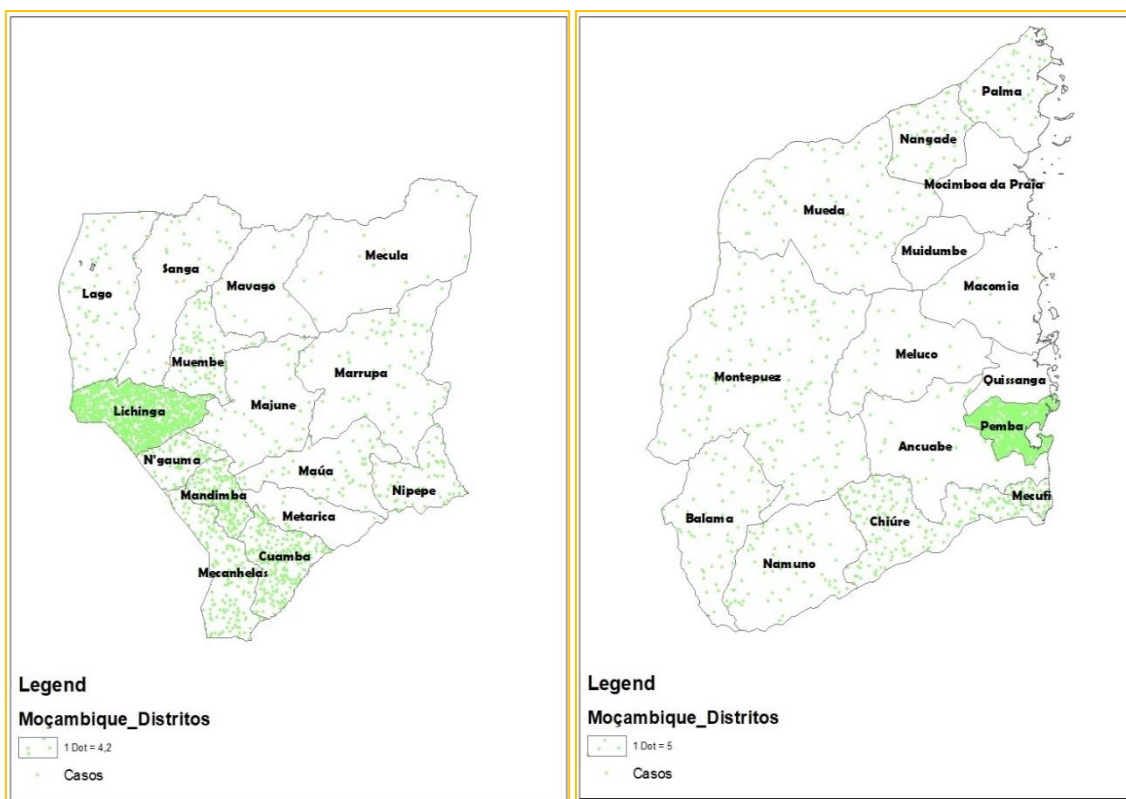
O gráfico 08 elucida a distribuição espacial de casos de COVID-19 nas duas grandes províncias classificadas de províncias com transmissão comunitária da doença. Para entender corretamente o significado dos mapas, há que realçar:

- A dispersão dos casos de COVID-19 nos distritos e/ou nas províncias em análise (conforme os mapas acima) é representada pelas cores desde a amarela até ao castanho-escuro (veja a legenda).
- As cores mais claras representam a densidade de casos da doença não consideradas críticas e as cores mais escuras representam uma concentração média e alta. Os casos estão parametrizados em escalas que permita identificar as zonas (distritos) com casos entre 0 a 150 casos reportados de COVID-19, consideradas de baixa transmissibilidade; 151 a 329 casos da doença como

distritos de média transmissibilidade e, as com 330 e mais casos são zonas consideradas de alto risco de transmissibilidade (distritos com a transmissão comunitária). Como podemos ver na figura acima, os distritos de Nampula (na província de Nampula) e Distrito da Matola (província de Maputo) apresentam a cor mais intensas – escura realçando áreas dos mapas com inúmeros casos (acima de 6900 casos) colocando estes distritos na situação de distritos com altas taxas de transmissibilidade, colocando o sector da saúde a assumir maior preocupação em termos das intervenções de saúde pública para dar resposta integrada e sustentável para o controle da doença.

- Também, pode-se perceber que a distribuição e/ou a propagação de COVID-19 no país e em particular nas províncias em análise, a transmissibilidade da doença tem uma relação direta com a mobilidade de pessoas fundamentalmente em zonas urbanas colocando uma disparidade de contaminação de pessoa com as zonas rurais.

Figura 09: Distribuição Geográfica de Casos e Níveis de Risco de transmissão



A análise das regiões das províncias do Niassa e de Cabo Delgado no contexto epidemiológico são caracterizadas por um perfil sanitário similar a de Nampula e Maputo, olhando nas zonas de concentração de casos de COVID-19. A disseminação de informação sobre a prevenção e cumprimento de medidas de prevenção para satisfazer a capacidade do sistema de saúde e a eficácia das medidas de controle da pandemia.

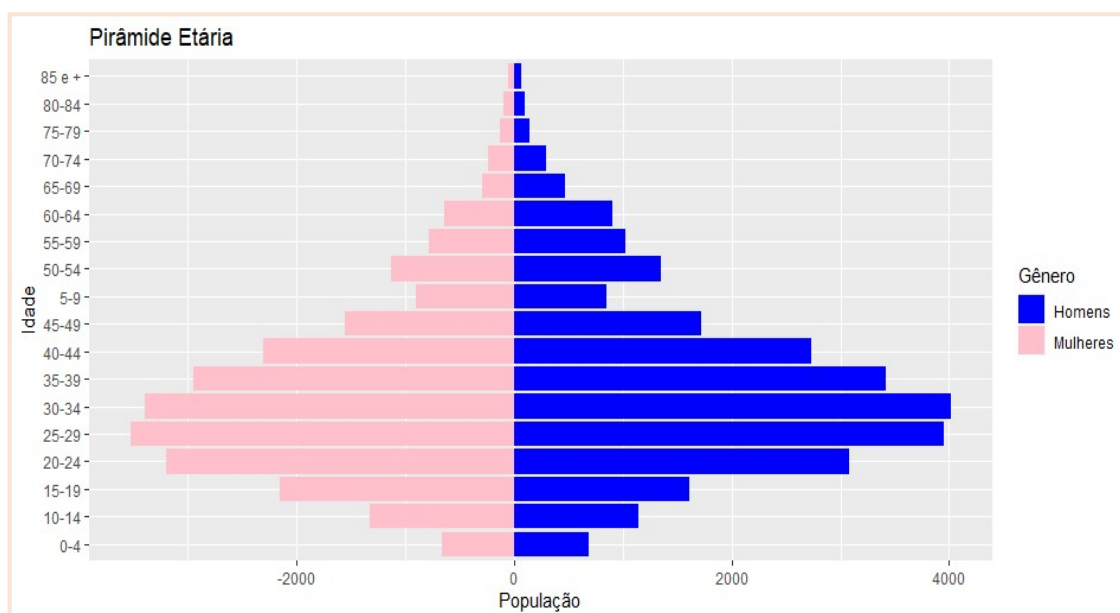
Conforme a figura 09, é notória a concentração de casos da doença nas zonas urbanas como consequência da densidade populacional elevada e dinâmica populacional evolutiva marcada por movimentos forçados da população vítima de terrorismo nos distritos periféricos da província de Cabo Delgado.

[e] Pirâmide Etária

A pirâmide etária no contexto de visualização de dados de COVID-19 permitirá a representação gráfica dos dados da pandemia mostrando a distribuição de casos de COVID-19 em diferentes faixas etárias da população infetada. É útil para entender como a pandemia afeta diferentes grupos de pessoas por idade e sexo que, por conseguinte, irá auxiliar na formulação de políticas de saúde que visem o controlo da propagação da pandemia.

○ Pirâmide Etária em Rstudio

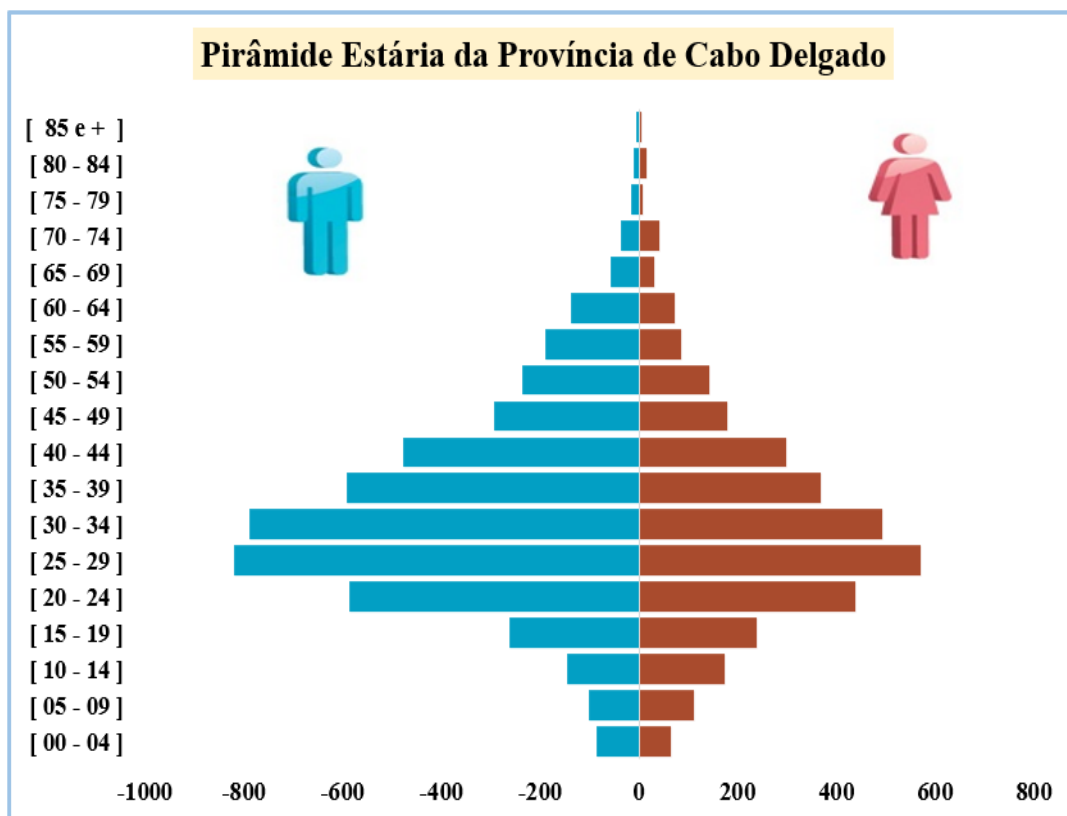
Figura 10: Ilustração gráfica da Pirâmide Etária dos dados cumulativos das 04 províncias em análise



Duma forma geral, os dados mostram que a camada juvenil e adulta até 39 anos de idade é a faixa etária mais infetada pelo vírus corona nos dois gêneros (masculino e feminino). Pela característica da população do país (menor população idosa), a população de idade compreendida entre 25 a 39 anos de idade e de sexo masculino é a que esteve mais exposta a fator de risco de contrair o COVID-19 nas 4 províncias em análise, o que pode ser justificado pela mobilidade desta camada a procura de condições de vida sem medir esforços sobre o lema “*fica em casa*” como um dos métodos usado para o controle da propagação da doença.

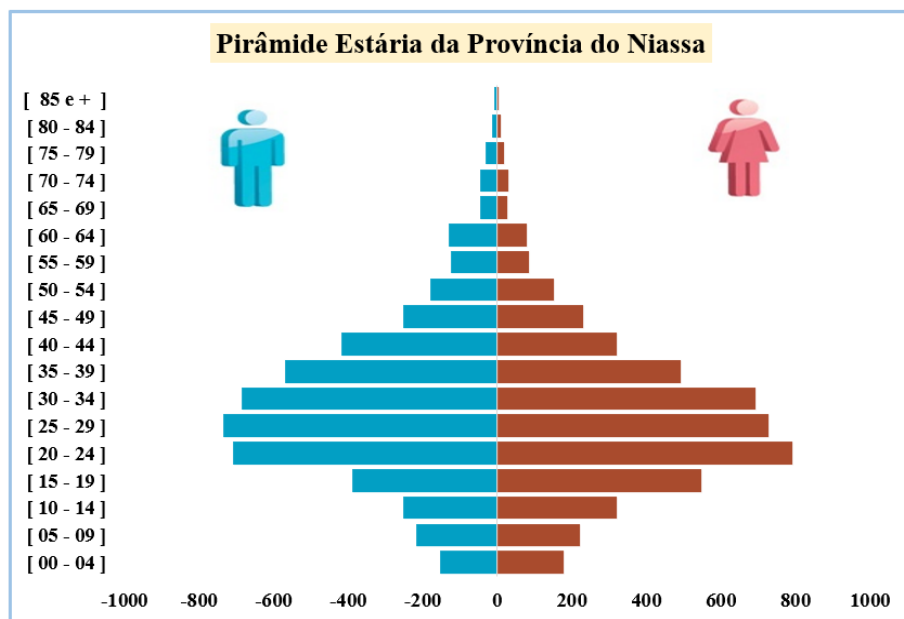
- **Pirâmide Etária em Folha de Cálculo (Excel)**

Figura 11: Ilustração gráfica da Pirâmide Etária de casos de COVID-19 na província de Cabo Delgado



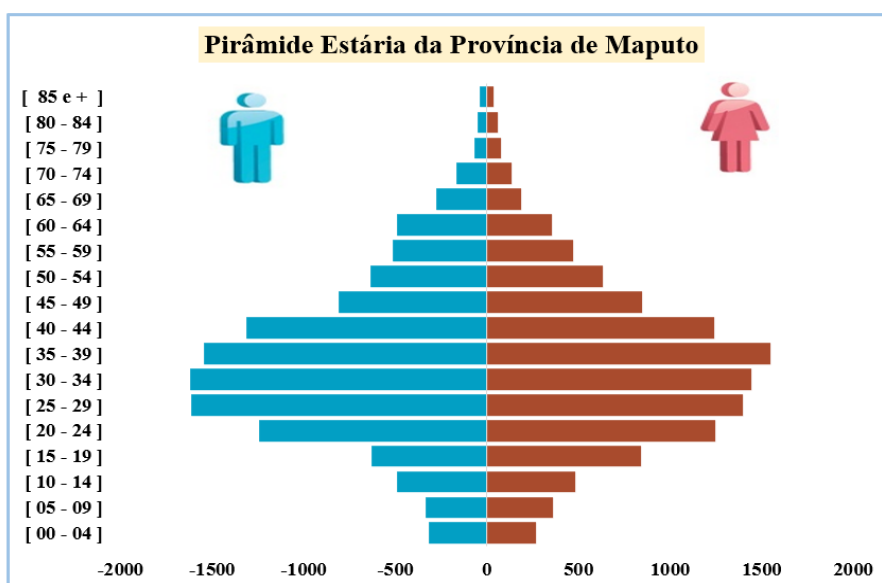
A figura acima ilustra a distribuição de casos de COVID-19 na província de Cabo Delgado. É notória que a população da faixa etária compreendida entre 25 a 34 anos de idades é a mais infetada pelo COVID-19 com uma percentagem de cerca de 33% do total de casos repostados pela província, com destaque para o sexo masculino com uma percentagem de cerca de 20%.

Figura 12: Ilustração gráfica da Pirâmide Etária de casos de COVID-19 na província do Niassa



A pirâmide etária acima (figura 12) ilustra a distribuição por idade e sexo de casos de COVID-19 reportados na província do Niassa. Ao contrário da província de Cabo Delgado, na província do Niassa a população considerada vulnerável à COVID-19 foi maioritariamente de sexo feminino das faixas etárias de 20 a 34 anos de idade. Dos 44% de população jovem que ficou infetada pelo vírus corona, cerca de 22% deste grupo foram mulheres.

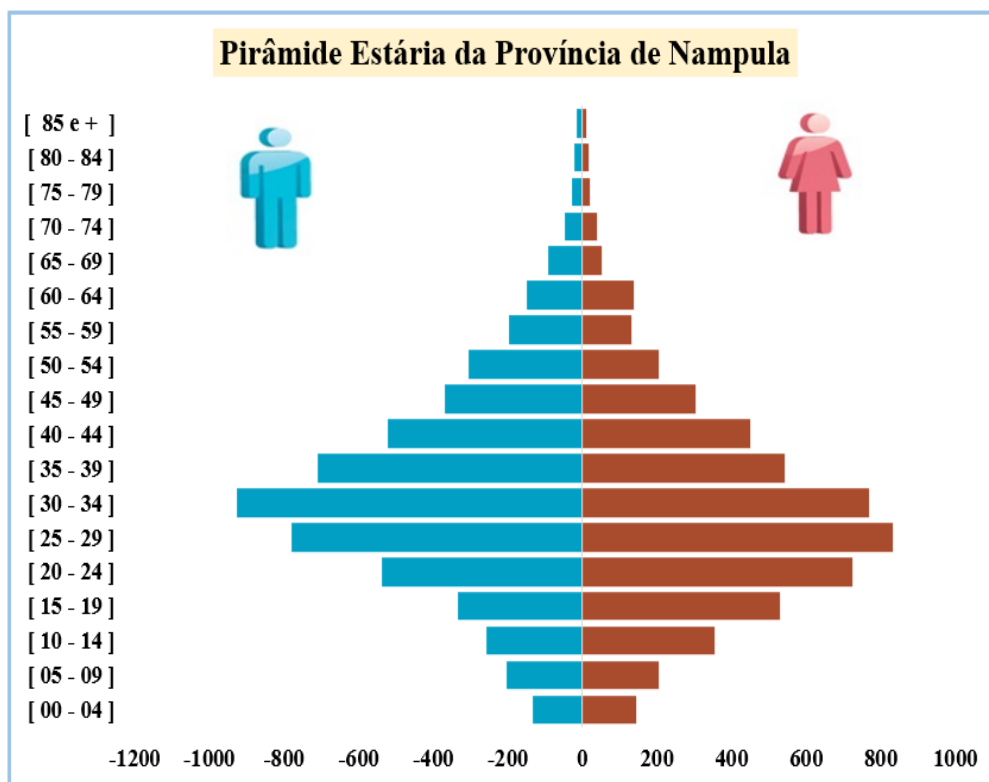
Figura 13: Ilustração gráfica da Pirâmide Etária de casos de COVID-19 na província de Maputo



A característica da pirâmide etária da província de Maputo ilustrada acima apresenta uma concentração populacional de alto risco para as faixas etárias de 20 a 44 anos de idade ambos sexos.

A distribuição dos casos de COVID-19 nesta província não apresenta maior variabilidade entre o sexo feminino a masculino apesar o domínio de casos na faixa etária de 40 a 44 anos seja de pessoas de sexo feminino. É notória que a população da faixa etária compreendida entre 25 a 34 anos de idades e de sexo masculino é a que suaviza a curva de casos da doença na população jovem mais infetada pelo COVID-19 com uma percentagem de 14%.

Figura 14: Ilustração gráfica da Pirâmide Etária de casos de COVID-19 na província de Nampula



A figura 14 ilustra a distribuição da transmissão de COVID-19 na província de Nampula. Os dados mostram um quadro de contaminação similar à das outras áreas geográficas (Cabo Delgado, Niassa e Maputo) em que a população mais afetada pela

doença situa-se na faixa etária entre os 20 a 44 anos de idade e maior peso para os indivíduos de sexo masculino.

É de notar que a população da faixa etária compreendida entre 25 a 34 anos de idades é a mais vulnerável e infetada pelo COVID-19 com uma percentagem de cerca de 30% do total de casos repostados na província, com destaque para o sexo masculino com uma percentagem de cerca de 15%.

4.3. Conclusões

O presente trabalho teve como principal objetivo de identificar e selecionar os modelos de análise e mecanismo de visualização de dados de COVID-19 em Moçambique – caso particular para as províncias de Maputo, Nampula, Cabo Delgado e Niassa.

- Os modelos de Regressão Linear Múltipla e Epidemiológicos – Modelo SIR se adequam e caracterizam-se como modelos flexíveis e simples para análise do comportamento da pandemia de COVID-19;
- Os modelos de análise e de visualização de dados permitem a transformação de uma quantidade de dados complexos em simples e mais amigável.
- Uma forma acessível de visualizar e entender exceções, tendências e padrões nos dados é utilizando modelos matemáticos e elementos visuais, em modelos tais como: dashboards, tabelas, diagramas, gráficos, figuras e mapas.
- Não é suficiente analisar a problemática da pandemia de COVID-19 olhando apenas as atitudes e comportamento da população suscetível e exposta à doença, mas também é fundamental analisar o problema em estudo na ótica ambiental, isto é, os modelos de análise devem incorporar como variáveis a região (espaço físico) e a temperatura, entre outras.

Contudo, aceita-se afirmar que os modelos de visualização de dados permitem a tomada de decisões com base em evidências e transforma uma quantidade de dados complexos em simples e mais amigável, proporcionando uma compreensão rápida e direcionada.

4.4. Perspetivas

Neste subcapítulo, debruçar-se-á sobre o pensamento do “amanhã”.

- O futuro da visualização de dados, bem como o pensamento de haver novas técnicas, ferramentas ou abordagens que podem transformar como lidamos com dados e informações visuais no mundo digital (Futuro da Visualização de Dados);
- Os desafios que a visualização de dados pode enfrentar no futuro, com maior foco para questões éticas, privacidade e segurança de dados, sem se esquecer dos desafios técnicos no concernente à como lidar com conjuntos de dados cada vez maiores e mais complexos;
- Como a visualização de dados pode ser usada para tornar as informações mais acessíveis ao público, bem como o interesse de melhorar a alfabetização pessoas em dados e capacitar as pessoas a tomar decisões informadas com base em informações visuais (Acesso à Informação);
- Qual será o impacto da visualização de dados na tomada de decisões, tanto individualmente quanto organizacional? Uma reflexão de como as visualizações de dados podem influenciar como entendemos problemas e formulamos estratégias para a solução dum determinado problema de saúde pública (Impacto na tomada de decisões).

4.5. Considerações para o futuro

Um aspecto de suma importância a ser contemplado é a análise de resíduos, elemento fundamental em estudos epidemiológicos, que proporciona uma avaliação rigorosa da adequação dos modelos estatísticos empregados. Para além da aplicabilidade à regressão linear em geral, também nas situações que envolvem análise conjunta de regressão o papel da análise dos resíduos pode tornar-se fundamental, ver, por exemplo, Oliveira et al. (2008), Morgado et al. (2010).

Em investigações futuras, pretende-se também prosseguir com uma análise mais detalhada de métodos estatísticos alternativos, procurando avaliar a respetiva adequação nas esferas de interesse delineadas na dissertação. Uma área importante diz respeito à investigação do produto de distribuições normais, conforme abordado, por exemplo, por Seijas-Macias et al. (2020), Oliveira et al. (2013) e Oliveira et al. (2016). Esta área pode elucidar a complexidade e a variabilidade também inerentes aos dados do tipo epidemiológico, uma vez que a consideração da estrutura de dependência entre variáveis pode tornar-se particularmente pertinente, possibilitando perceções sobre a interação entre fatores de risco e os desfechos na saúde.

Também a consideração de modelos de Planeamento de Experiências, como os planos em blocos, conforme delineado por Oliveira (2010), e dos planos hierárquicos, apresentados em Valente e Oliveira (2009), pode enriquecer substancialmente a análise ao permitir um controle mais rigoroso das variáveis. Tais abordagens propiciam a maximização da eficiência estatística, utilizando técnicas de aleatorização e estratificação cruciais para a validação dos estudos.

Outra vertente de estudos promissora é a adaptação dos modelos de Aprendizagem Baseada em Problemas (PBL), conforme apresentado em Nunes et al. (2017), que oferece uma perspetiva inovadora e interdisciplinar na abordagem de questões mais complexas em epidemiologia. Essa metodologia pode ser instrumental na formação de hipóteses e na conceção de estudos que se alinhem com as exigências contemporâneas de evidência científica.

Estas direções futuras visam não apenas aprimorar os métodos estatísticos, mas também promover uma compreensão mais robusta e integrada dos dados, fundamental para a interpretação e aplicação dos resultados em contextos epidemiológicos. A ênfase

na análise de resíduos permitirá uma avaliação crítica da adequação dos modelos, contribuindo para a qualidade e validade das inferências estatísticas, ambicionando conduzir a prática epidemiológica a novos patamares de rigor e sofisticação analítica.

[5]. **Referências Bibliográficas**

1. Agresti, A. (2013). *Categorical data analysis*. John Wiley & Sons.
2. Alberto Da Silva, C., Castelo, F., De, R., Carvalheira, M., De, P., & Capucho Paulino, J. (n.d.), (2021), *COVID 19 e os constrangimentos do mundo do trabalho dos profissionais de saúde: Contributos para a revisão da literatura [*]*.
3. Alberto, C., Carlos, S., Sanquetta, R., Rex, F. E., Fernanda, I., Pfutz, P., De, R., & Macedo, C. (2020). *Explorando o QGIS 3.X Dalla Corte et al. Ana Paula Dalla Corte*.
4. Anderson, R. M., & May, R. M. (1992). *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press.
5. Aven, T., & Vinnem, J. E. (2007). *Risk Management with Applications from the Offshore Petroleum Industry*. Springer.
6. Brauer, F., & Castillo-Chávez, C. (2012). *Mathematical Models in Population Biology and Epidemiology*. Springer.
7. Campos, M. C., Ferreira, S., Doutora, P., Costa, T. P., & Oliveira, A. (2013). *Modelos de Regressão: uma aplicação em Medicina Dentária*. Universidade Aberta
8. Campos, M. R., De Andrade Schramm, J. M., Emmerick, I. C. M., Rodrigues, J. M., De Avelar, F. G., & Pimentel, T. G. (2020). *Burden of disease from COVID-19 and its acute and chronic complications: Reflections on measurement (DALYs) and prospects for the Brazilian Unified National Health System. Cadernos de Saúde Pública, 36(11). <https://doi.org/10.1590/0102-311X00148920>*
9. Cavalcanti, I. (2020). *Conhecendo a Covid-19*. In *Conhecendo a Covid-19*. Rfb Editora. <https://doi.org/10.46898/rfb.9786599097867>
10. Chein, F. (2019). *Introdução aos modelos de regressão linear: um passo inicial para compreensão da econometria como uma ferramenta de avaliação de políticas públicas*. In *Coleção Metodologias de Pesquisa*. [https://repositorio.enap.gov.br/bitstream/1/4788/1/Livro_Regressão Linear.pdf](https://repositorio.enap.gov.br/bitstream/1/4788/1/Livro_Regressão%20Linear.pdf)
11. Cordeiro, G. M. (2010). *Modelos Lineares Generalizados e Extensões*.

12. Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons.
13. Cristina, S., & Rodrigues, A. (2012). *Modelo de Regressão Linear e suas Aplicações*. Universidade da Beira Interior Ciências.
14. De Fora, J., & Gerais, M. (2013). *A Visualização de dados na Teoria da Comunicação*.
15. De, A., & Vieira, O. (2016). *Estudo sobre Modelos Matemáticos Aplicados à Epidemiologia: Modelo SIR, SIR com vacinação e SIRS*.
16. De, R., Ministério, M., & Saúde, D. A. (2019). *COVID-19 em Moçambique Relatório do 1º Ano*.
17. Diekmann, O., Heesterbeek, J. A. P., & Metz, J. A. J. (1990). *On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations*. *Journal of Mathematical Biology*, 28(4), 365–382.
18. Diggle, P. J., & Ribeiro Jr, P. J. (2007). *Model-based Geostatistics*. Springer.
19. Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford University Press.
20. Federighi, E., & Chagas, B. (2017). *Módulo 5 - Análise Multivariada no SPSS Análise Multivariada no SPSS*. <https://doi.org/10.13140/RG.2.2.10857.90727>
21. Felipe De Oliveira, P., Guerra, S., & Mcdonnell, R. (2018). *Ciência de Dados com R*. Introdução autores. www.ibpad.com.br
22. Fernandez, P. J., & Yohai, V. (2014). *Introdução à Análise Exploratória de Dados Multivariados*.
23. Ferraz, D. R. (2018). *Princípios de Visualização de dados aplicados no software de gestão financeira binfolio*.
24. Friendly, M., & SAS Institute. (2005). *Visualizing categorical data*. Books24x7.com.
25. Gonçalves, A. C., & Sousa, A. M. O. (2017). *Tecnologias de Informação Geográfica ARCGIS 10*.

26. Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge University Press.
27. Hoffmann, R. (2017). *Análise de regressão: uma introdução à econometria*. Universidade de São Paulo. Escola Superior de Agricultura Luiz de Queiroz.
28. Hosmer Jr., D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.
29. Imperial College COVID-19 Response Team. (26 de March de 2020). <https://bit.ly/Covid-Imperial-12>. Obtido de <https://bit.ly/Covid-Imperial-12>: <https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-Global-Impact-26-03-2020.pdf>
30. Infarmed. (27 de 05 de 2020). *SNS - Sistema Nacional de Saúde*. Obtido de SNS - Sistema Nacional de Saúde: https://www.infarmed.pt/web/infarmed/infarmed/-/journal_content/56/15786/3700712
31. INS: Análise da situação epidemiológica SARS-CoV-2/COVID-19, acessido a 17 de Maio em <https://covid19.ins.gov.mz/estatisticas/>
32. Ivo, P., Machado, P., Santos, R., & Rio De Janeiro, C. (2013). *Modelos Lineares Generalizados Aplicados À Precificação Em Seguro Saúde*.
33. Kabacoff, R. (2011). *R in action: data analysis and graphics with R*.
34. Kenny, D. A. (2017). *ebook the guilford press Regression Analysis and Linear Models Methodology in the Social Sciences*. www.guilford.com/MSS
35. Kermack, W. O., & McKendrick, A. G. (1927). *A Contribution to the Mathematical Theory of Epidemics*. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 115(772), 700–721.
36. Kerns, G. J. (2010). *Introduction to probability and statistics using R*. [s.n.].
37. Kleinbaum, D. G. (2010). *Logistic regression: A self-learning text*. Springer.
38. Lanum, C. L. (2016). *MEAP Edition Manning Early Access Program Visualizing Graph Data*. www.manning.com<https://forums.manning.com/forums/visualizing-graph-data>

39. Luís, J., & Manjate, S. (2020). *Conhecimentos, Atitudes e Práticas dos Funcionários Públicos de Moçambique em relação à Prevenção da COVID-19* Knowledge, Attitudes and Practices of Mozambican Public Employees in relation to the Prevention of COVID-19. <https://orcid.org/0000-0002-6209-2807>
40. Malthus, T. R. (1798). *An Essay on the Principle of Population*. J. Johnson.
41. Martins, H. F. B., & Hansine, R. (2020). *Análise epidemiológica e demográfica da COVID-19 em África* COVID-19's epidemiological and demographic analysis in Africa *Artigo Original*. <https://doi.org/10.25761/anaisihmt.353>
42. Martynova, E., West, S. G., & Liu, Y. (2018). Review of Principles and Practice of Structural Equation Modeling. In *Structural Equation Modeling* (Vol. 25, Issue 2). <https://doi.org/10.1080/10705511.2017.1401932>
43. MF Teodoro, T Oliveira, E. Taero (2023). *COVID-19 infection and risk analysis: a short introduction*. *Biometrics & Biostatistics International Journal*, Ed. MedCrave. <http://hdl.handle.net/10400.2/15624>
44. MF Teodoro, TA Oliveira, F Arune (2024). *Biometrics & Biostatistics International Journal* 13 (1), 7-14. Ed. MedCrave. <https://medcraveonline.com/BBIJ/BBIJ-13-00408.pdf>
45. Miguel Elias Temer Lulia, M., Esteves Pedro Colnago Júnior, D., Roberto Luís Olinto Ramos Diretor-Executivo Fernando José de Araújo Abrantes, P., Sant, J., Bevilaqua, A., & Wu Tai, D. (2018). *Presidente da República Instituto Brasileiro De Geografia E Estatística-IBGE*.
46. Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2008). *Introduction to time series analysis and forecasting*. John Wiley & Sons.
47. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. Wiley.
48. Morgado, M.L., Teodoro, M.F., Perdicoulis, T.A. (2010). *Métodos Estatísticos em Ciências Biomédicas, Série didáctica Ciências Aplicadas, 400*. Vila Real, Portugal: Universidade de Trás os Montes e Alto Douro (UTAD). Morgado et al.
49. Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press.

50. Oliveira, A., Oliveira, T.A., Seijas-Macias, A. (2013). *The influence of ratios and combined ratios on the distribution of the product of two independent gaussian random variables*. Proceedings of the 59th World Statistics Congress of the International Statistical Institute, p.25-30, Hong Kong, China.
51. Oliveira, A., Oliveira, T.A., Seijas-Macias, A. (2016). *Skewness into the product of two normally distributed variables and the risk consequences*. Revstat-Statistical Journal Volume 14, Number 2, April 2016, 119–138.
52. Oliveira, T.A.(2010). *BIB Designs with Repeated Blocks: Review and perspectives*, Proceedings of ICCS-X Conference, Tenth Islamic Countries Conference on Statistical Sciences– Statistics for Development and Good Governance. Editors Zeinab Amin and Ali, S., Hadi. The American University in Cairo,1, 82–96.
53. OMR. (2022). *Caracterização das condições socioeconómicas dos deslocados internos no norte de moçambique ao longo do ano de 2021*. 28-07-2022. Disponível em <https://omrmz.org/wp-content/uploads/2022/08/OR-127>.
54. Pacifico Filho, M., Iwamoto, H. M., Borges, T. P., & Cançado, A. C. (2022). *Disseminação da COVID-19: contágio tardio em centros locais no sudeste do Tocantins*. Revista Brasileira de Estudos Urbanos e Regionais, 24(1). <https://doi.org/10.22296/2317-1529.rbeur.202203pt>
55. Patrícia, S., Santana, B., Prof, O., Co-Orientação Prof, V. G., Cristina, M., & Raposo, F. (2004). “*Modelos estatísticos para análise de dados longitudinais categorizados ordinais*.”
56. Petersen, M. S., Kristiansen, M. F., Hanusson, K. D., Danielsen, M. E., Á Steig, B., Gaini, S., Strøm, M., & Weihe, P. (2021). *Long COVID in the Faroe Islands: A Longitudinal Study among Nonhospitalized Patients*. Clinical Infectious Diseases, 73(11), E4058–E4063. <https://doi.org/10.1093/cid/ciaa1792>
57. Pimpler, E. (2022). *Data Visualization and Exploration with R: A practical guide to using R, RStudio, and Tidyverse for data visualization, exploration, and data science applications*.

58. Pires Brito, S. B., Braga, I. O., Cunha, C. C., Palácio, M. A. V., & Takenami, I. (2020). *Pandemia da COVID-19: o maior desafio do século XXI. Vigilância Sanitária Em Debate*, 8(2), 54–63. <https://doi.org/10.22239/2317-269x.01531>
59. Ribeiro, D. M. (2009). *Visualização de dados: mapas e cartografias do ciberespaço*. Rio de Janeiro: ISBN.
60. Ricardo Malheiro, D., António Vieira dos Santos, F., Rodrigues de Oliveira Araruna, A., & Ribeiro Francellino Sampaio, J. (2021). *Perspetivas Socioambientais Sobre A COVID-19 Olhares interdisciplinares em ambiente e saúde*.
61. Ricardo, A., Grégio, A., Pereira De Carvalho Filho, B., Montes, A., & Santos, R. (2009). *Capítulo 5 Técnicas de Visualização de Dados aplicadas à Segurança da Informação*.
62. Rocha, C. R., Oliveira, I., Marina, M., Rocha, Z., Luíza, A., Lucas, A., & Da Rosa, P. (2020). *Modelagem Dinâmica Aplicada à COVID-19 Sebastião Cícero Pinheiro Gomes*.
63. Rosa, A. (2014). *Dados espaciais disponibilizados pelo sistema Geobases-Es e II*. <http://www.mundogeomatica.com.br>
64. Rosa, A., Santos, D., António, C., Soares, A., Gilberto, R., Sediyaama, C., Batista, J., Peluzio, E., Macedo, J. E., & Rosembergue Bragança, P. (2015). *Autores Principais*. <http://www.mundogeomatica.com.br>
65. Seijas-Macías, A., Oliveira, A., Oliveira, T. A., & Leiva, V. (2020). *Approximating the distribution of the product of two normally distributed random variables*. *Symmetry*, 12(1201), 1–13.
66. Simão Carvalho Adérito Fernandes Marcos, E., & Técnico, R. (2009). *Visualização da Informação*.
67. Taero, E. (2025). *Análise e Avaliação do Risco de contágio da COVID-19 em Moçambique*. Tese de Doutoramento em Modelação Matemática e Aplicações. Universidade Aberta.

68. Tavares Alvarenga, A. M. (2015). *Modelos lineares generalizados: aplicação a dados de acidentes rodoviários* Dissertação Mestrado em Gestão de Informação Especialização em Gestão e Análise de Dados.
69. Tomaz, D. (2015). *Criando Dashboards no Excel*. www.infopratica.net
70. Torres, A. J., Denise, S., & Pimenta, N. (2020). *Coleção História Do Tempo Presente: Volume III A Pandemia De Covid-19: Interseções E Desafios Para A História Da Saúde E Do Tempo Presente* Rita de Cassia Marques.
71. Valente, V, Oliveira, T. (2009). *Hierarchical linear models in education sciences: an application*. *Biometrical Letters*; 46:71–86.
72. Vanessa Rodrigues da Silva Santana, V., Cristina Pereira do Nascimento Oliveira, K., da Silva Mendonça, K., Tavares da Silva, D., & Laisa de Lima Dantas, H. (2020). *Artigo 6 Revisão Integrativa De Literatura Fatores De Risco Para O Agravamento Da Covid-19 Em Indivíduos Jovens Artigo 6-Original Revisão Integrativa De Literatura Fatores De Risco Para O Agravamento Da Covid-19 Em Indivíduos Jovens*. In *Enferm. Foco* (Vol. 11, Issue 2). <https://orcid.org/0000-0003-3740-9670>
73. Watson, G. S. (1986). *Generalized Linear Models (P. McCullagh and J. A. Nelder)*. In *SIAM Review* (Vol. 28, Issue 1). <https://doi.org/10.1137/1028043>
74. WHO. (2020). *Coronavirus disease 2019 (COVID-19) Situation Report – 57*. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200317-sitrep57-covid-19.pdf?sfvrsn=a26922f2_4 (accessed March 18, 2020).
75. WHO. (2020). *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)*. [https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-\(covid-19\)](https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19)) (accessed March 22, 2020).
76. Wickham, H. (2009). *ggplot2*. In *ggplot2*. Springer New York. <https://doi.org/10.1007/978-0-387-98141-3>
77. Yan, X., & Gang Su, X. (2009). *Linear Regression Analysis Theory and Computing*.

Apêndice

Código RStudio para a obtenção do Modelo de Regressão Linear Múltipla:

```
##### Código R: Modelo de Regressão Linear Múltipla #####

library(pacman)
pacman::p_load(dplyr, car, rstatix, lmtest, ggpubr,
               QuantPsyc, psych, scatterplot3d)

##### Construindo o Modelo Matemático #####
dados_mrlm <- read_excel("/Users/Arune/Documents/DADOSCOVID19_DISSERT.xls",
                        sheet=1)
names(dados_mrlm)
modelo <- lm(formula = Numero_Casos ~ Provincia + Genero + Idades, data
             = dados_mrlm)
```

Código RStudio para a obtenção do teste de análise da Multicolinearidade entre as variáveis:

```
##### Analise da Multicolinearidade #####
pairs.panels(dados_mrlm)
vif(modelo)

               vif(modelo)
Provincia      Genero      Idades
1.095266  1.056087  1.041817
```

Código RStudio para a obtenção do teste de Normalidade dos Resíduos:

```
##### Código R: Teste de normalidade dos Resíduos #####

shapiro.test(modelo$residuals)

      Shapiro-Wilk normality test

      data:  modelo$residuals

      W = 0.662, p-value = 1.127e-10
```

Código RStudio para análise de Outliers dos Resíduos:

```
##### Código R: Analise de Outliers nos Resíduos #####
summary(rstandard(modelo))

      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
-0.959701 -0.478994 -0.149038  0.005046  0.119905  5.403029
```

Código RStudio para obtenção do teste de análise da Independência dos Resíduos:

```
##### Código R: Analise da Independência dos Resíduos #####
durbinWatsonTest(modelo)

      lag   Autocorrelation D-W Statistic   p-value
1      0.1192868      1.745134           0.26

Alternative hypothesis: rho != 0
```

Código RStudio para obtenção do teste de análise da Homocedasticidade:

```
##### Código R: Analise da Homocedasticidade _ Breusch_Pagan #####
bptest(modelo)

Studentized Breusch-Pagan test

      data: modelo
BP = 8.1979, df = 3, p-value = 0.04209
```

Código RStudio para obtenção do modelo compartimental SIR:

```
##### CARREGAMENTO DE PACOTES NECESSÁRIOS #####

install.packages("deSolve")
library(deSolve)
library(reshape2)
library(ggplot2)
library(tidyverse)

##### DEFINIÇÃO DE PARÂMETROS DO MODELO #####

beta <- 1/3 # Taxa de transmissão
gamma <- 1/7 # Taxa de recuperação
```

```
##### DEFINIÇÃO DO MODELO SIR #####

sir_model <- function(time, state, parameters) {
  with(as.list(c(state, parameters)), {
    dS <- -beta * S * I / N
    dI <- beta * S * I / N - gamma * I
    dR <- gamma * I

    return(list(c(dS, dI, dR)))
  })
}

##### DEFINIÇÃO DE CONDIÇÕES INICIAIS #####

initial_state <- c(S = 52953, I = 1, R = 0)
parameters <- c(beta = beta, gamma = gamma, N = sum(initial_state))

##### DEFINIÇÃO DO INTERVALO DE TEMPO #####

time <- seq(0, 790, by = 1) # Simular por 2 anos e 2 meses

##### RESOLVENDO O SISTEMA DE EQUAÇÕES DIFERENCIAIS #####

results <- ode(y = initial_state, times = time, func = sir_model, parms = parameters)

##### CONVERSÃO DE RESULTADOS EM DATA FRAME #####
results_df <- as.data.frame(results)

##### PLOTANDO OS RESULTADOS #####

plot(results_df$time, results_df$S, type = "l", col = "blue", xlab = "Tempo", ylab =
"Número de Pessoas", ylim = c(0, 55000))
lines(results_df$time, results_df$I, col = "red")
lines(results_df$time, results_df$R, col = "green")
legend("right", legend = c("Suscetíveis", "Infectados", "Recuperados"), col =
c("blue", "red", "green"), lty = 1)
```

Código RStudio para obtenção da Tabela Resumida de casos de COVID-19:

```
##### Código R: Tabela Simples ou Resumida #####

df <- data.frame(

  Provincias = c("Cabo Delgado", "Maputo", "Nampula", "Niassa", "Total Geral"),

  Numero_Casos = accounting(c(8177, 23783, 11089, 9886), format = "d")

)

View(df)
```

Código RStudio para obtenção da Tabela de frequências da Positividade de Casos:

```
##### Código R: Tabela de Frequências Percentual #####

df <- data.frame(
  Provincias = c("Maputo", "Nampula", "Cabo Delgado", "Niassa", "Total
Geral"),
  Numero_Casos = accounting(c(8177, 23783, 11089, 9886, 52935), format =
"d"),
  Percentagem = accounting(c(45, 21, 15, 19, 100), format = "d")
)

View(df)
```

Código RStudio para obtenção da Tabela de Casos de COVID-19 por área geográfica e género:

```
##### Código R: Tabela cruzada #####

dados <- read_excel("/Users/Arune/Documents/DADOSCOVID19_DISSERT.xls", sheet
= 1)

view(dados)
glimpse(dados)
table(dados$Provincia)
table(dados$Genero)
table(dados$Numero_Casos)
table(dados$Provincia, dados$Numero_Casos)
view(table)
```

Código RStudio para obtenção da Curva Cumulativa de Casos de COVID-19:

```
##### Código R: Gráfico de Linha - Dados Cumulativos. #####

dis <- read_excel("/Users/Arune/Documents/DADOSCOVID19_DISSERT.xls", sheet = 9)
dis
head(dis)
```

```
ggplot(dis) + aes(Semana, Casos_Cumulativos) +
geom_line(col="blue") +
theme_minimal()
```

Código RStudio para obtenção da Série Temporal de Casos de COVID-19:

```
##### Código R: Gráfico de Linha - Serie Temporal. #####

dis <- read_excel("/Users/Arune/Documents/DADOSCOVID19_DISSERT.xls",
sheet = 9)
dis
ggplot(dis) + aes(Semana, Numero_Casos) +
  geom_line(col="blue") +
  theme_minimal()
```

Código RStudio para obtenção do gráfico de Barras de Casos de COVID-19:

```
##### Código R: Gráfico de Barras - Casos Positivos por Província. #####

dis <- read_excel("/Users/Arune/Documents/DADOSCOVID19_DISSERT.xls",
sheet = 11)
dis
ggplot(data = dis, aes(x = Provincia, y = CasosPositivos, fill =
Provincia)) +
  geom_bar(stat = "identity", width = .6) +
  coord_flip() +
  geom_label(aes(label = CasosPositivos), fill = "white") +
  labs(x = "CasosPositivos", y = "Provincia", fill = "CasosPositivos")
```

Código RStudio para obtenção do gráfico de Distribuição de Casos Positivos Confirmados de COVID-19 por Área Geográfica:

Código R: Gráfico Circular - Casos Positivos por Província.

```
dis <- read_excel("/Users/Arune/Documents/DADOSCOVID19 DISSERT.xls", sheet = 11)
head(dis)
ggplot(data = dis, aes(x = 2, y = CasosPositivos, fill = Provincia)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = paste0(CasosPositivos)), color = "white",
            position = position_stack(vjust = .5), size = 4) +
  scale_fill_manual(values = c("darkorange", "purple", "cyan4", "blue")) +
  xlim(0, 2.5) +
  coord_polar(theta = "y", start = 0) +
  theme_void() +
  theme(legend.position = c(.5, .5),
        legend.title = element_text(size = 20),
        legend.text = element_text(size = 15)) +
  labs(fill = "Provincia")
```

Código RStudio para obtenção do gráfico de Distribuição de Casos Positivos Confirmados de COVID-19 por Género:

Código R: Gráfico Circular - Casos Positivos por Género.

```
dis <- read_excel("/Users/Arune/Documents/DADOSCOVID19 DISSERT.xls", sheet = 12)
head(dis)
ggplot(data = dis, aes(x = 2, y = CasosPositivos, fill = Genero)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = paste0(CasosPositivos)), color = "white",
            position = position_stack(vjust = .5), size = 4) +
  scale_fill_manual(values = c("darkorange", "purple")) +
  xlim(0, 2.5) +
  coord_polar(theta = "y", start = 0) +
```

```

theme_void() +
theme(legend.position = c(.5, .5),
      legend.title = element_text(size = 20),
      legend.text = element_text(size = 15)) +
labs(fill = "Genero")

```

Código RStudio para obtenção da Pirâmide Etária de Casos de COVID-19:

```

##### Código R: Pirâmide Etária Geral #####
dados <- data.frame(
  Idade = c("0-4", "5-9", "10-14", "15-19", "20-24", "25-29", "30-34", "35-39",
"40-44", "45-49", "50-54", "55-59", "60-64", "65-69", "70-74", "75-79", "80-84", "85 e
+"),
  Homens = c(690, 854, 1143, 1613, 3078, 3949, 4016, 3416, 2730, 1722, 1354, 1025,
901, 472, 292, 144, 99, 69),
  Mulheres = c(658, 900, 1331, 2155, 3199, 3529, 3394, 2948, 2309, 1558, 1131,
779, 646, 296, 244, 127, 102, 57)
)
piramide_etaria <- ggplot(dados, aes(x = Idade)) +
  geom_bar(aes(y = Homens, fill = "Homens"), stat = "identity") +
  geom_bar(aes(y = -Mulheres, fill = "Mulheres"), stat = "identity") +
  coord_flip() +
  scale_fill_manual(values = c("Homens" = "blue", "Mulheres" = "pink")) +
  labs(title = "Pirâmide Etária",
       x = "Idade",
       y = "População",
       fill = "Gênero")
print(piramide_etaria)

```