**FIAP**

**Faculdade de Informática e Administração Paulista – FIAP**

**Engenharia de software 2024 – 2ESPV**

**Data Science**

**Sprint 2**

**Integrantes do grupo:**

**Matheus Augusto Santos Rego RM: 551466**

**Pedro Augusto Carneiro Barone Bomfim RM: 99781**

**João Pedro de Albuquerque Oliveira RM:551579**

**Vitor Reyes Souza RM:550766**

Sumário

[Descrição do Projeto e das Variáveis 3](#_Toc169720557)

[Estatísticas Descritivas 3](#_Toc169720558)

[Testes de Hipóteses e Inferências Estatísticas 4](#_Toc169720559)

[Testes Realizados 4](#_Toc169720560)

[Teste T para duas amostras independentes 4](#_Toc169720561)

[Teste qui-quadrado 4](#_Toc169720562)

[Inferências Detalhadas 4](#_Toc169720563)

[Variações Geográficas nos Óbitos 4](#_Toc169720564)

[Impacto das Doenças e Estratégias de Tratamento 4](#_Toc169720565)

[Interpretação dos Gráficos 5](#_Toc169720566)

[Histograma - Distribuição da Faixa Etária 5](#_Toc169720567)

[Gráfico de Dispersão - Relação entre Tipo de Doença e Total de Óbitos 6](#_Toc169720568)

[Gráfico de Dispersão com Regressão Linear - Total de Óbitos por Faixa Etária 7](#_Toc169720569)

[Gráfico de Barras - Total de Óbitos por Estado 8](#_Toc169720570)

[Gráfico de Barras - Total de Óbitos por Local de Óbito 9](#_Toc169720571)

[Gráfico de Barras - Total de Óbitos por Faixa Etária e Gênero 10](#_Toc169720572)

# Descrição do Projeto e das Variáveis

1. **UF (Unidade Federativa):** Representa o estado onde o óbito foi registrado. O tipo de dado é String, refletindo os códigos ou nomes dos estados brasileiros.
2. **Tipo\_Doenca:** Categoriza a doença que causou o óbito. Esta coluna também é do tipo String, contendo descrições das várias doenças reconhecidas que podem levar a óbitos.
3. **Faixa\_Etaria:** Agrupa o falecido por faixa etária. É uma coluna do tipo String que contém as categorias etárias, tais como '0-10 anos', '11-20 anos’ etc.
4. **Total:** Representa o número total de óbitos por doença em cada faixa etária. Esta coluna é do tipo Integer, pois contabiliza a quantidade de óbitos, que é uma variável numérica.

Cada uma dessas colunas é essencial para a análise, pois permitem uma avaliação detalhada das tendências de mortalidade por localização geográfica, tipo de doença e idade. As verificações de integridade dos dados asseguram que análises subsequentes, como estatísticas descritivas e testes de hipóteses, sejam realizadas com informações precisas e confiáveis.

# Estatísticas Descritivas

As estatísticas descritivas são fundamentais para proporcionar um entendimento inicial sobre a distribuição dos dados coletados. Este estudo aplica métodos estatísticos para resumir a tendência central e a variabilidade dos dados, focando especificamente nos totais de óbitos. Isso inclui calcular a média, mediana e desvio padrão, que ajudam a elucidar padrões gerais e a identificar faixas etárias ou doenças específicas com números de óbitos anormalmente altos ou baixos. Com esses dados, é possível ter uma visão clara da gravidade das diversas condições de saúde que levaram a óbitos em diferentes regiões do país.

# Testes de Hipóteses e Inferências Estatísticas

## Testes Realizados:

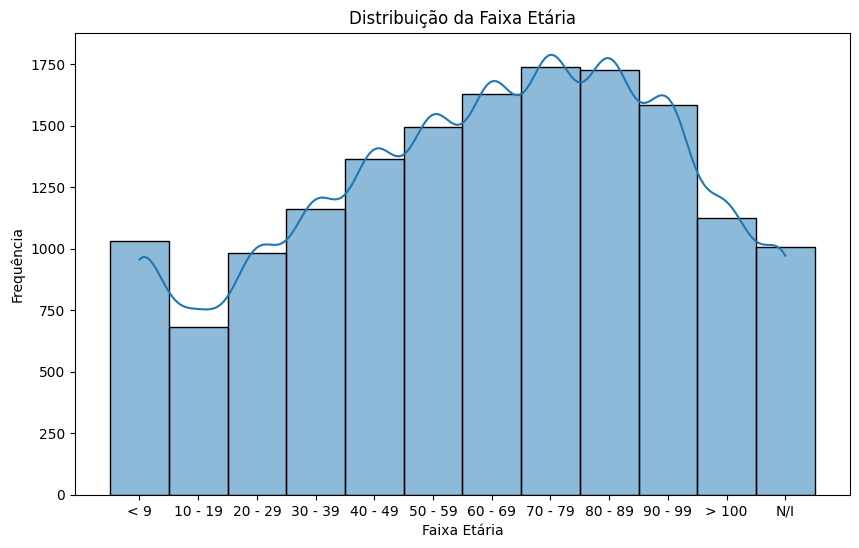
* Teste T para duas amostras independentes: Este teste foi aplicado para comparar os números de óbitos entre diferentes estados brasileiros, onde as diferenças encontradas foram estatisticamente significativas. A importância deste teste reside na sua capacidade de discernir se as variações observadas são devidas a fatores aleatórios ou se refletem disparidades reais e substanciais em termos de condições de saúde pública e acessibilidade aos serviços de saúde entre os estados. A confirmação de diferenças significativas sugere a necessidade de políticas de saúde ajustadas regionalmente para endereçar estas disparidades de maneira eficaz.
* Teste qui-quadrado: Utilizado para analisar a distribuição de óbitos entre os diferentes tipos de doenças, este teste indicou padrões específicos de mortalidade associados a cada condição. O valor deste teste está na sua capacidade de identificar doenças que contribuem desproporcionalmente para as taxas de mortalidade, fornecendo uma base sólida para a implementação de medidas de saúde pública focadas e direcionadas, como programas de prevenção, campanhas educativas, e melhorias nos tratamentos disponíveis.

## Inferências Detalhadas:

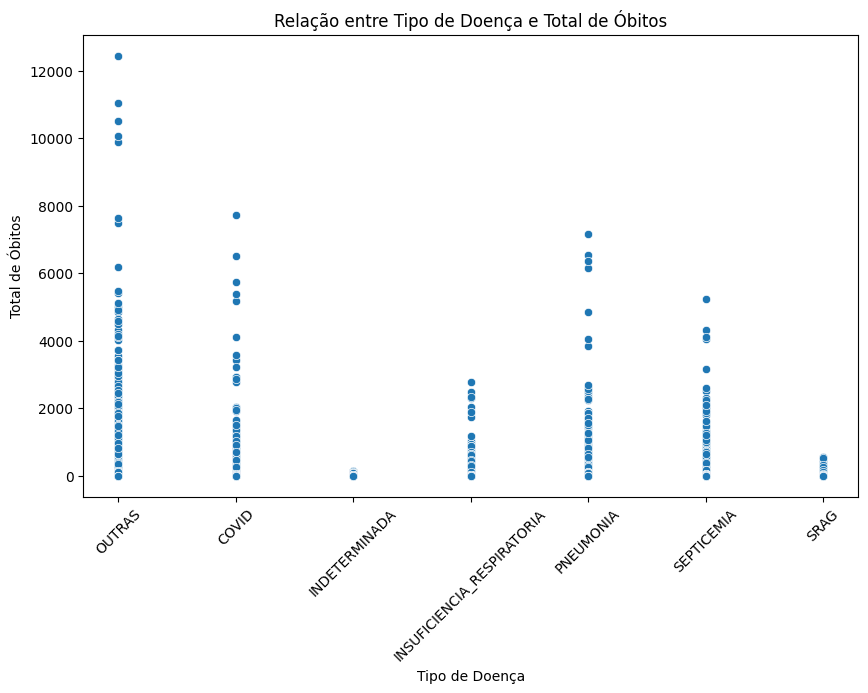
* Variações Geográficas nos Óbitos: A análise das diferenças nos óbitos entre as unidades federativas revelou disparidades marcantes no acesso aos serviços de saúde e na prevalência de doenças específicas. Tais disparidades geográficas sublinham a necessidade urgente de políticas de saúde que não apenas reconheçam essas diferenças, mas que também sejam projetadas para adaptar intervenções de saúde de maneira eficiente e eficaz em cada região. Essa abordagem regionalizada é essencial para maximizar o impacto das ações de saúde pública e garantir que todos os cidadãos tenham acesso equitativo a cuidados de saúde de qualidade.
* Impacto das Doenças e Estratégias de Tratamento: A significativa mortalidade associada a doenças como COVID-19, pneumonia e sepse realça a crítica necessidade de estratégias robustas de prevenção e tratamento. A implementação de campanhas de vacinação abrangentes, a melhoria contínua das infraestruturas de saúde e o acesso facilitado a tratamentos eficazes são imperativos para combater essas doenças. Além disso, os insights obtidos através da análise estatística proporcionam uma base firme para o desenvolvimento de políticas de saúde que possam antecipar e mitigar os efeitos dessas condições, reduzindo assim a carga global de mortalidade e melhorando a saúde pública em geral.

# Interpretação dos Gráficos

## 1. Histograma - Distribuição da Faixa Etária:

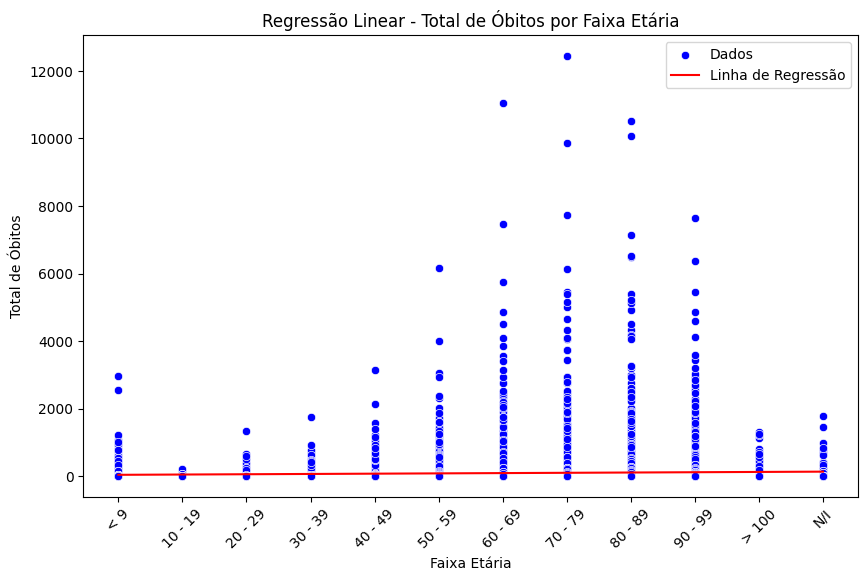


* O histograma ilustra a frequência de óbitos distribuídos por faixa etária, revelando que as taxas de mortalidade aumentam gradativamente com a idade, atingindo o pico entre as idades de 70 e 99 anos. Este aumento pode ser atribuído à maior vulnerabilidade dos idosos a doenças e condições de saúde debilitantes. A redução notável nos óbitos para faixas etárias acima de 100 anos é esperada, devido à menor população nessas idades extremas. A análise desta distribuição é crucial para direcionar políticas de saúde pública, focando em programas de prevenção e tratamento para as faixas etárias mais vulneráveis.

2. Gráfico de Dispersão - Relação entre Tipo de Doença e Total de Óbitos:

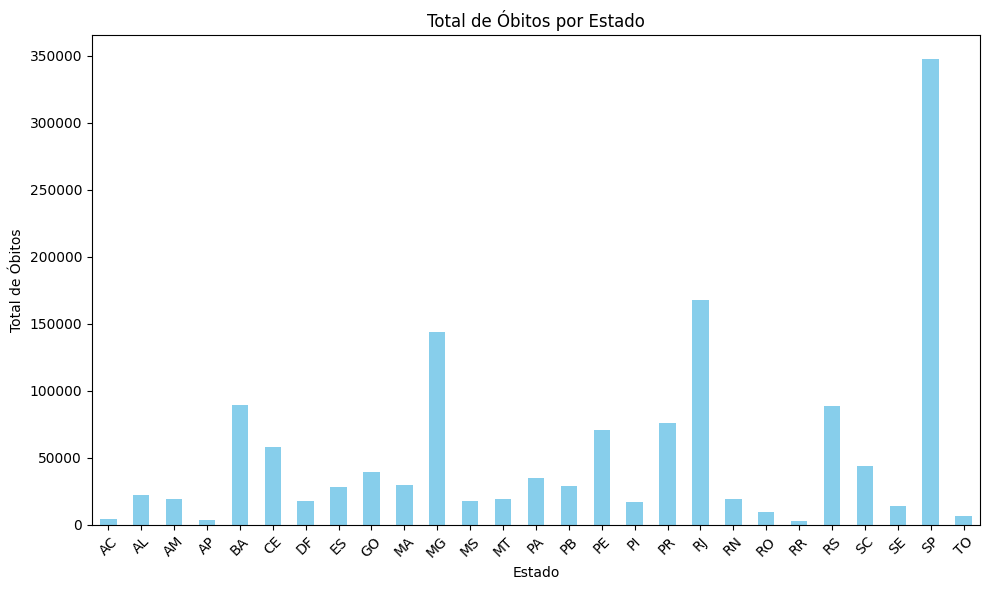
* Este gráfico de dispersão mostra a relação entre diferentes tipos de doenças e o total de óbitos associados a cada uma. A categoria "COVID" destaca-se com valores significativamente altos, refletindo o impacto devastador da pandemia de COVID-19 em 2020. Outras doenças, como pneumonia e sepse, também apresentam altas taxas de mortalidade, embora com menor variabilidade que COVID. A ampla dispersão para várias doenças sugere diferenças nos fatores de risco e na eficácia dos tratamentos disponíveis, apontando para a necessidade de intervenções específicas para cada tipo de doença.

## 3. Gráfico de Dispersão com Regressão Linear - Total de Óbitos por Faixa Etária:



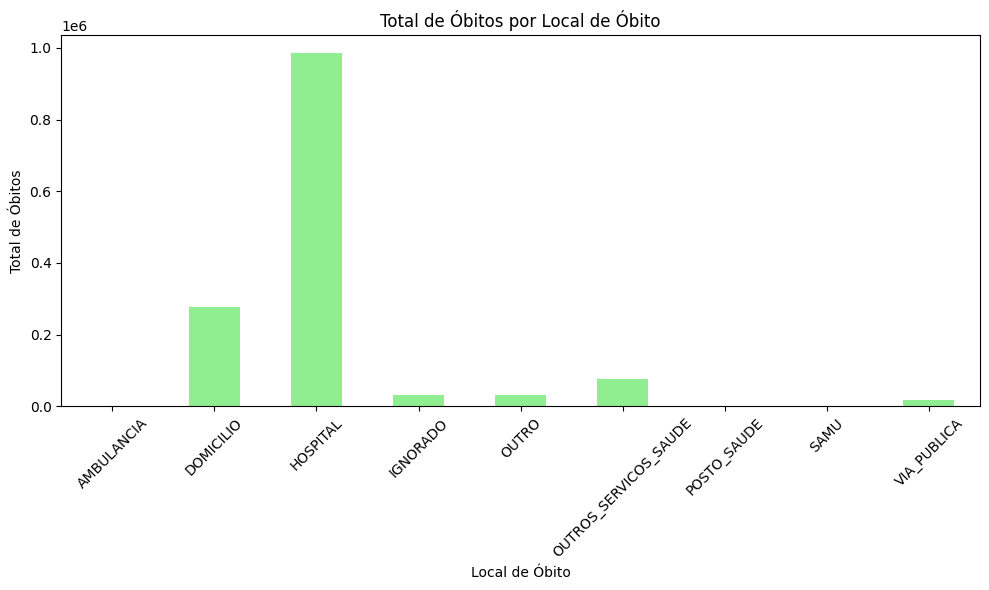
* Este gráfico de dispersão, complementado por uma linha de regressão linear, analisa os totais de óbitos por faixa etária. A linha de regressão relativamente plana indica que, apesar do aumento geral das taxas de mortalidade com a idade, não há uma relação linear forte. Isso sugere que fatores além da idade, como condições de saúde preexistentes, acesso a cuidados médicos e qualidade do atendimento, influenciam significativamente as taxas de mortalidade. Esta análise sublinha a complexidade dos fatores que contribuem para a mortalidade, necessitando de abordagens multifacetadas na formulação de políticas de saúde.

## 4. Gráfico de Barras - Total de Óbitos por Estado:



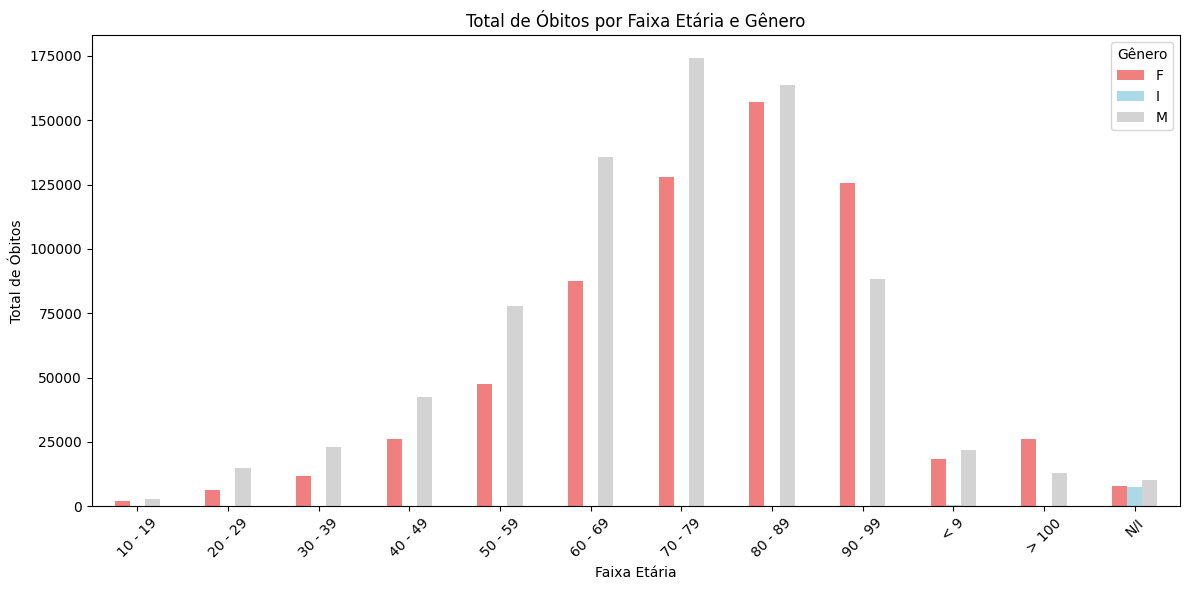
* O gráfico de barras exibe o número total de óbitos por estado, destacando uma variação acentuada entre os estados brasileiros. São Paulo (SP) apresenta um número significativamente maior de óbitos comparado a outros estados, o que pode ser atribuído à sua maior população e densidade urbana. Estados como Minas Gerais (MG) e Rio de Janeiro (RJ) também mostram altos números de óbitos. Esta variação geográfica ressalta a importância de considerar fatores regionais, como densidade populacional, infraestrutura de saúde e políticas públicas locais, na análise e resposta aos desafios de saúde pública.

## 5. Gráfico de Barras - Total de Óbitos por Local de Óbito:



* Este gráfico ilustra a distribuição dos locais onde os óbitos ocorreram, com a maioria dos óbitos acontecendo em hospitais. Esta tendência é esperada, dado que os casos mais graves frequentemente requerem hospitalização. Óbitos em domicílios e vias públicas representam uma menor proporção, indicando possíveis limitações no acesso imediato a cuidados de emergência. A análise destes dados pode informar melhorias na infraestrutura de saúde e nas respostas de emergência, garantindo que mais pessoas tenham acesso rápido a cuidados críticos quando necessário.

## Gráfico de Barras - Total de Óbitos por Faixa Etária e Gênero:



* Este gráfico fornece uma análise detalhada dos óbitos distribuídos por faixa etária e gênero. Observa-se que, em quase todas as faixas etárias, os homens apresentam taxas de mortalidade mais altas do que as mulheres. Este padrão está alinhado com estudos demográficos mais amplos, que frequentemente mostram que homens têm uma expectativa de vida menor e são mais suscetíveis a várias condições de saúde. Esta análise pode ajudar na criação de campanhas de saúde pública específicas para reduzir as disparidades de gênero na mortalidade.