

Relatório de Análise de Dados - Grupo 1 (segunda-feira / noturno)

Integrantes do grupo: Ana Julia Paiva, Sergio Gabriel, Lucas Melo e Lucas Albertini.

1. Introdução

Este relatório apresenta uma análise exploratória dos dados de Síndrome Respiratória Aguda Grave (SRAG) no Brasil, com foco nos casos relacionados à COVID-19. Os dados foram obtidos do DATASUS e abrangem o período de 2021. O objetivo é utilizar técnicas de estatística descritiva, probabilidade e inferência para responder a perguntas relevantes sobre a distribuição de casos, comorbidades, mortalidade e diferenças entre grupos.

2. Descrição dos Dados

Os dados analisados incluem informações sobre casos de SRAG no Brasil, com variáveis como idade, sexo, comorbidades, tipo de atendimento, evolução do caso (cura ou óbito), entre outras. Foram filtrados apenas os casos confirmados de COVID-19, e valores nulos foram removidos para garantir a qualidade da análise.

3. Metodologia

Foram utilizadas técnicas de:

- **Estatística Descritiva:** Média, mediana, desvio padrão, distribuição de frequências e gráficos.
- **Probabilidade:** Cálculo de probabilidades condicionais.
- **Inferência Estatística:** Testes de hipóteses (qui-quadrado, teste t, ANOVA) e intervalos de confiança.

As análises foram realizadas com Python, utilizando bibliotecas como Pandas, NumPy, Matplotlib, Seaborn e SciPy.

4. Análises e Resultados

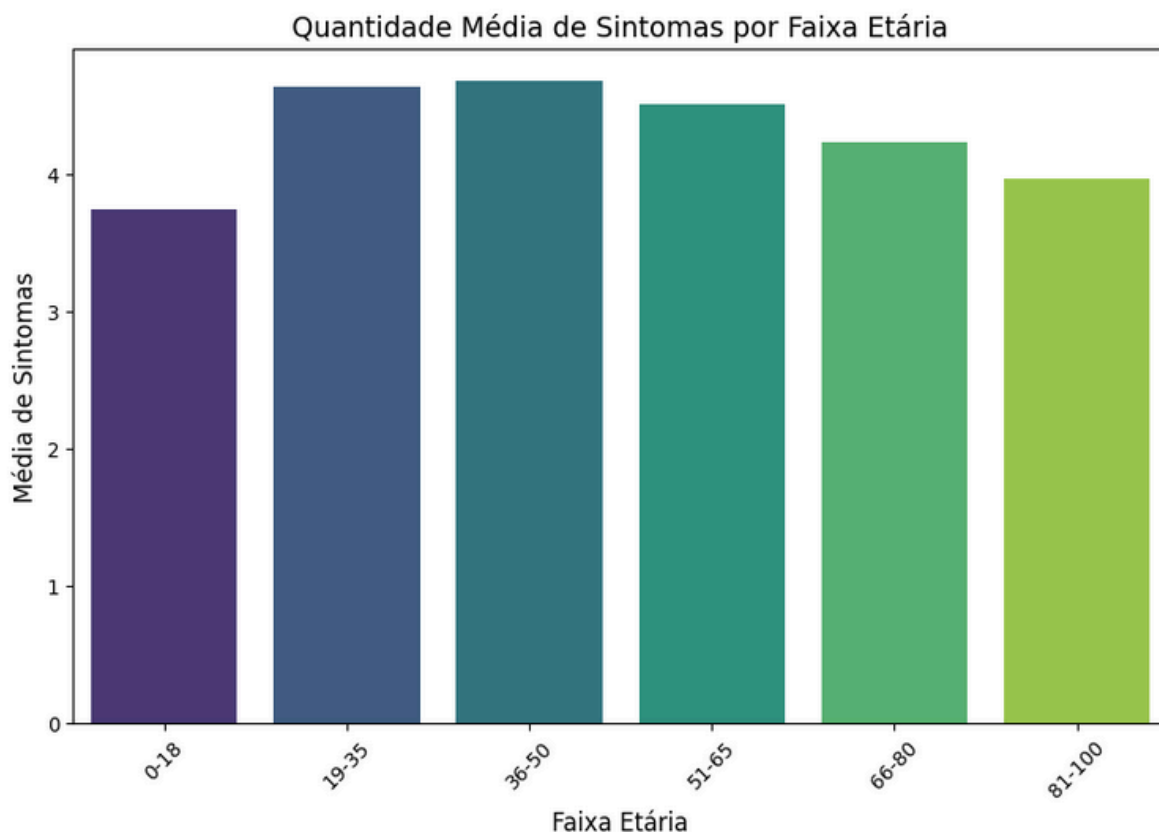
4.1 Estatística Descritiva

4.1.1 Média de sintomas por faixa de idade

- **Pergunta:** Qual é a média de sintomas reportados em diferentes faixas etárias?
- **Método:** Cálculo da **média, mediana e desvio padrão**.
- **Explicação:** A média fornece o valor central dos sintomas reportados, a mediana indica o valor que divide a distribuição ao meio, e o desvio padrão mede a dispersão dos dados. Essas métricas permitem comparar a quantidade de sintomas entre faixas etárias.

- **Resultados:**

Faixa Etária	Média	Mediana	Desvio Padrão
0-18	3,74	4,0	1,91
19-35	4,64	4,0	2,23
36-50	4,68	5,0	2,16
51-65	4,51	4,0	2,09
66-80	4,23	4,0	2,01
81-100	3,98	4,0	1,93



- **Interpretação:** Para a amostra, a média de sintomas é mais alta na faixa etária de 36-50 anos (4,68) e mais baixa na faixa de 0-18 anos (3,74).

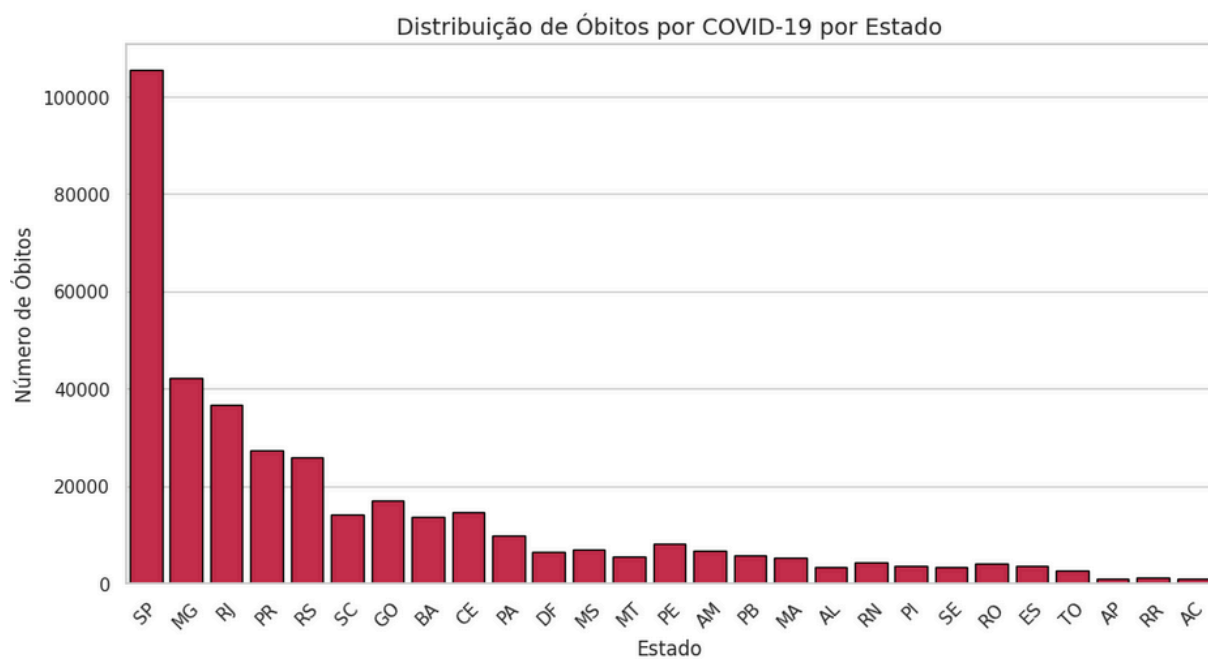
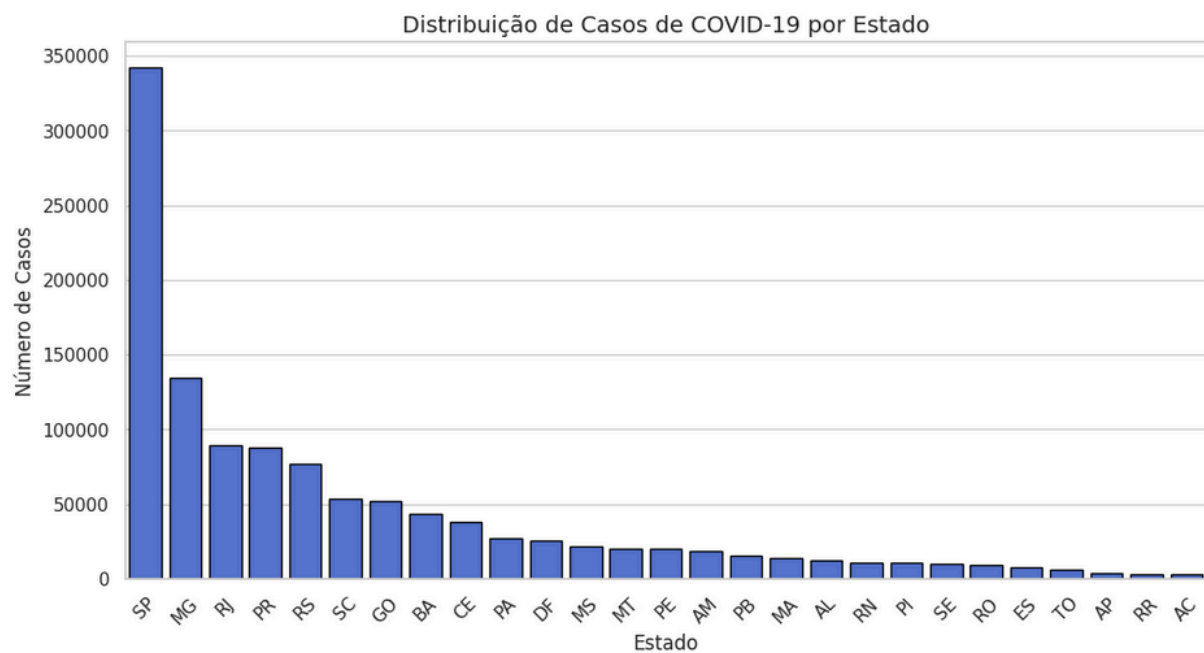
4.1.2 Distribuição geográfica de casos e óbitos

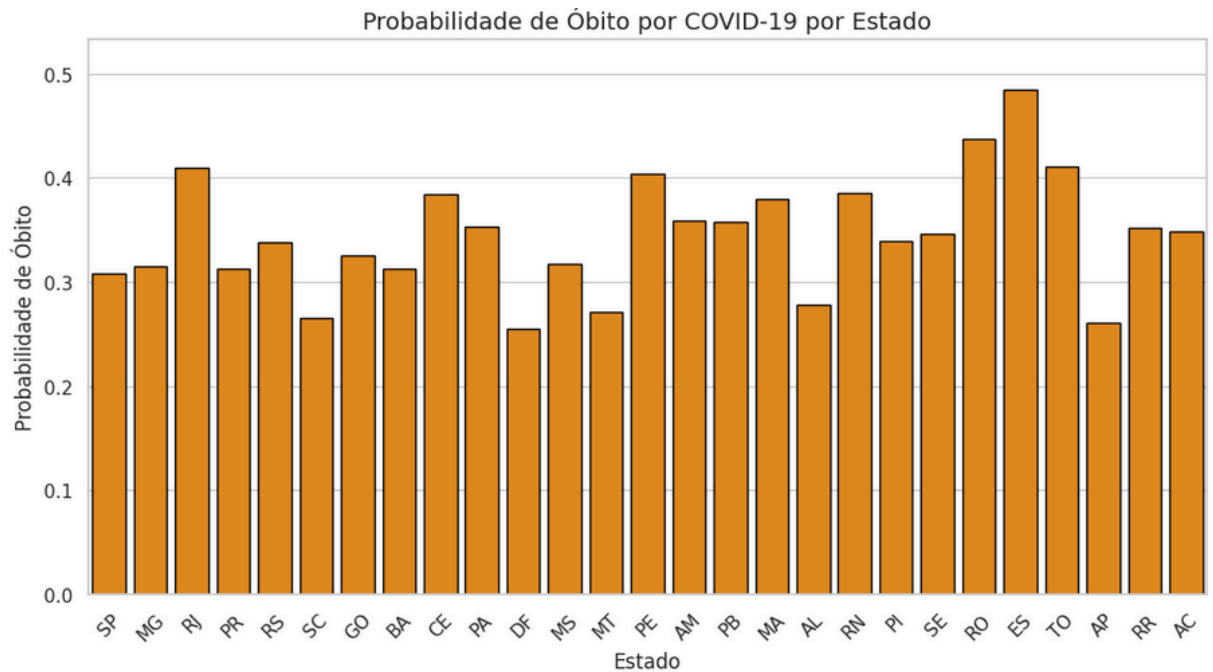
- **Pergunta:** Como os casos e óbitos de COVID-19 estão distribuídos geograficamente?
- **Método:** Tabelas de frequência e gráficos de barras.
- **Explicação:** As tabelas de frequência resumem o número de casos e óbitos por estado, enquanto os gráficos de barras permitem visualizar a distribuição geográfica de forma clara.
- **Resultados:**

Estado	Casos	Óbitos	Probabilidade de Óbito
SP	342.210	105.435	30,81%

MG	134.344	42.269	31,46%
RJ	89.664	36.754	40,99%
PR	87.778	27.408	31,22%
RS	76.715	25.957	33,84%
SC	53.435	14.174	26,52%
GO	52.466	17.065	32,52%
BA	43.285	13.545	31,29%
CE	38.177	14.659	38,37%
PA	27.453	9.701	35,34%
DF	25.599	6.543	25,56%
MS	21.597	6.857	31,75%
MT	20.409	5.533	27,11%
PE	20.271	8.184	40,37%
AM	18.469	6.632	35,91%

PB	15.746	5.631	35,76%
MA	14.131	5.359	37,92%
AL	12.096	3.365	27,82%
RN	10.869	4.192	38,57%
PI	10.780	3.654	33,90%
SE	9.712	3.366	34,66%
RO	9.141	3.998	43,74%
ES	7.393	3.584	48,48%
TO	6.114	2.512	41,08%
AP	3.640	949	26,07%
RR	3.051	1.073	35,17%
AC	2.702	941	34,83%





- Interpretação:** Na amostra, São Paulo tem o maior número de casos e óbitos, enquanto o Espírito Santo tem a maior probabilidade de óbito (48,48%) apesar da pouca quantidade de casos. Esse aumento na probabilidade de óbito no ES pode ter ocorrido pela omissão de notificação de casos de covid-19.

4.1.3 Taxa de mortalidade por faixa etária e sexo

- Pergunta:** Qual é a taxa de mortalidade por faixa etária e sexo?
- Método:** Cálculo da **taxa de mortalidade** (proporção de óbitos em relação ao total de casos) e **gráficos comparativos**.
- Explicação:** A taxa de mortalidade mede o risco de óbito para cada grupo, e os gráficos comparativos permitem visualizar diferenças entre sexos e faixas etárias.
- Resultados:**

Faixa Etária	Sexo	Taxa de Mortalidade
--------------	------	---------------------

0-18	F	6,75%
------	---	-------

0-18	M	6,39%
------	---	-------

19-35	F	13,21%
-------	---	--------

19-35	M	12,93%
-------	---	--------

36-50	F	20,14%
-------	---	--------

36-50	M	18,93%
-------	---	--------

51-59	F	27,59%
-------	---	--------

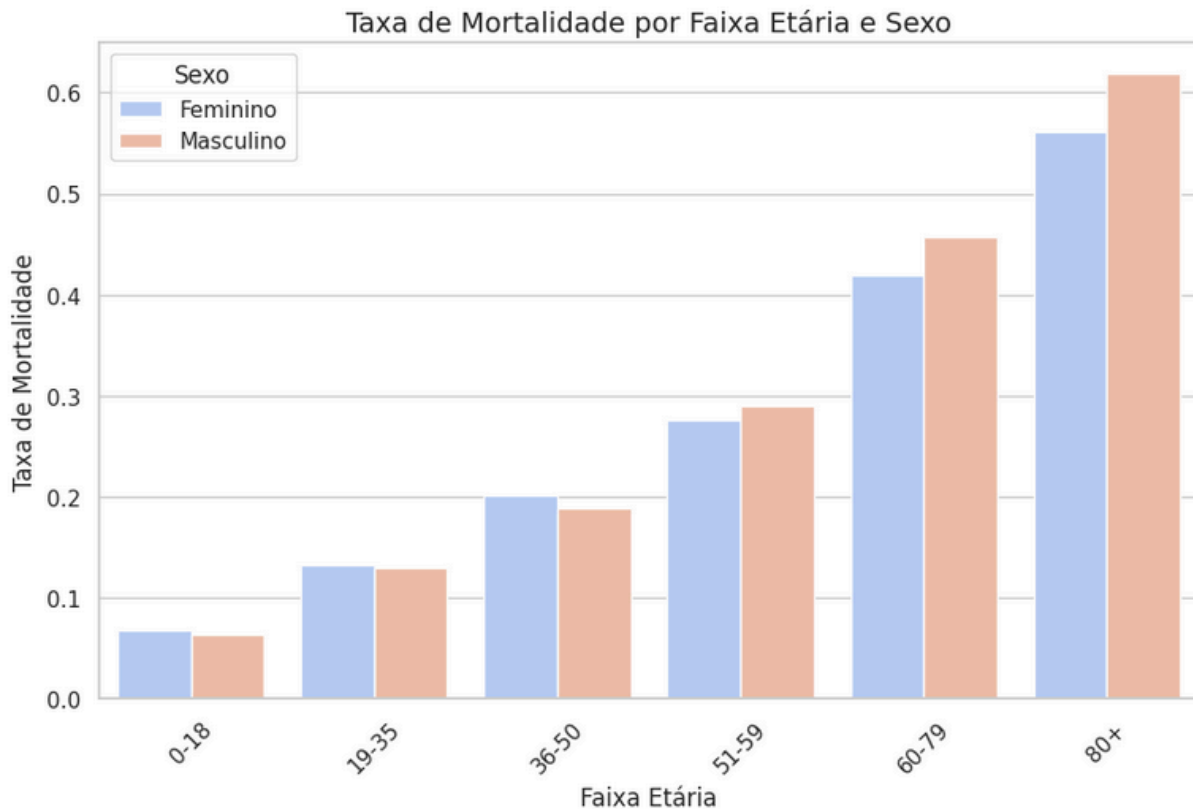
51-59	M	29,04%
-------	---	--------

60-79	F	41,97%
-------	---	--------

60-79	M	45,75%
-------	---	--------

80+	F	56,17%
-----	---	--------

80+	M	61,83%
-----	---	--------



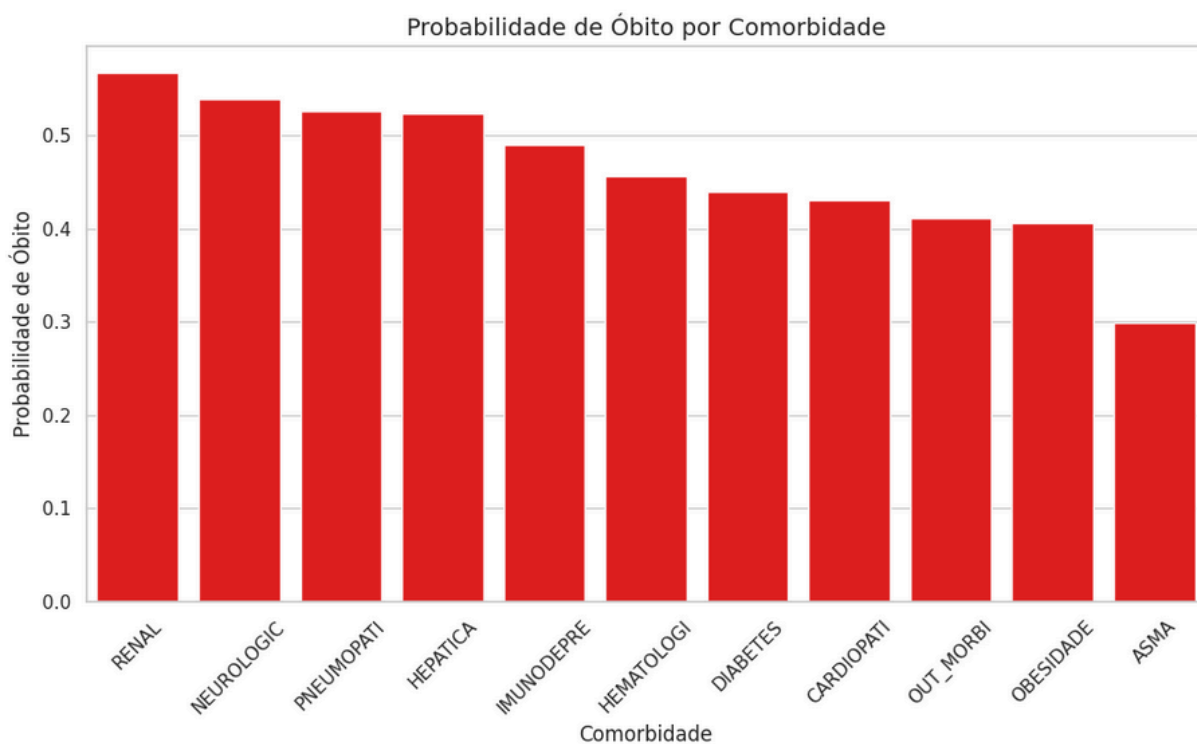
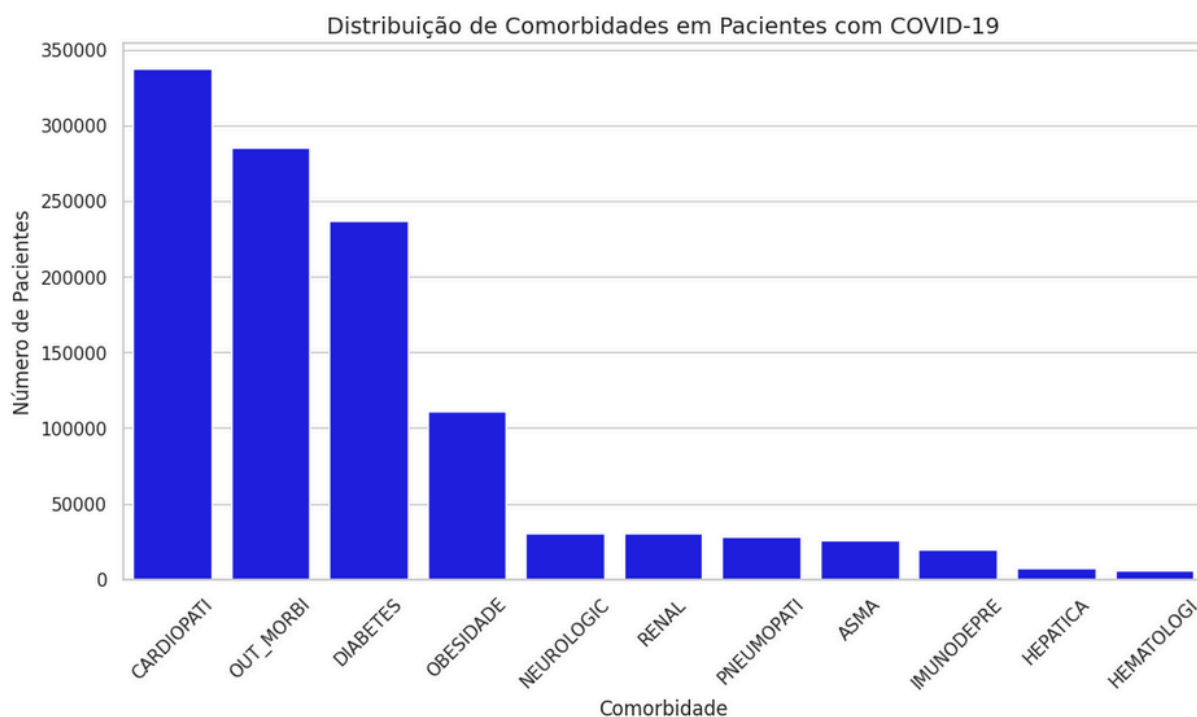
-

Interpretação: De acordo com a amostra, os dados indicam que a mortalidade aumenta significativamente com a idade, sendo mais alta para homens acima de 80 anos (61,83%). É interessante notar que na amostra de 2021, a taxa de mortalidade feminina se mantém acima da masculina até a faixa de idade 36-50, após isso, a taxa masculina se sobressai.

4.1.4 Distribuição de comorbidades e probabilidade de óbito

- **Pergunta:** Qual é a frequência de comorbidades e a probabilidade de óbito associada?
- **Método:** Tabelas de frequência e cálculo de probabilidades condicionais.
- **Explicação:** A frequência de comorbidades mostra quais condições são mais comuns, e a probabilidade de óbito condicional mede o risco associado a cada comorbidade.
- **Resultados:**

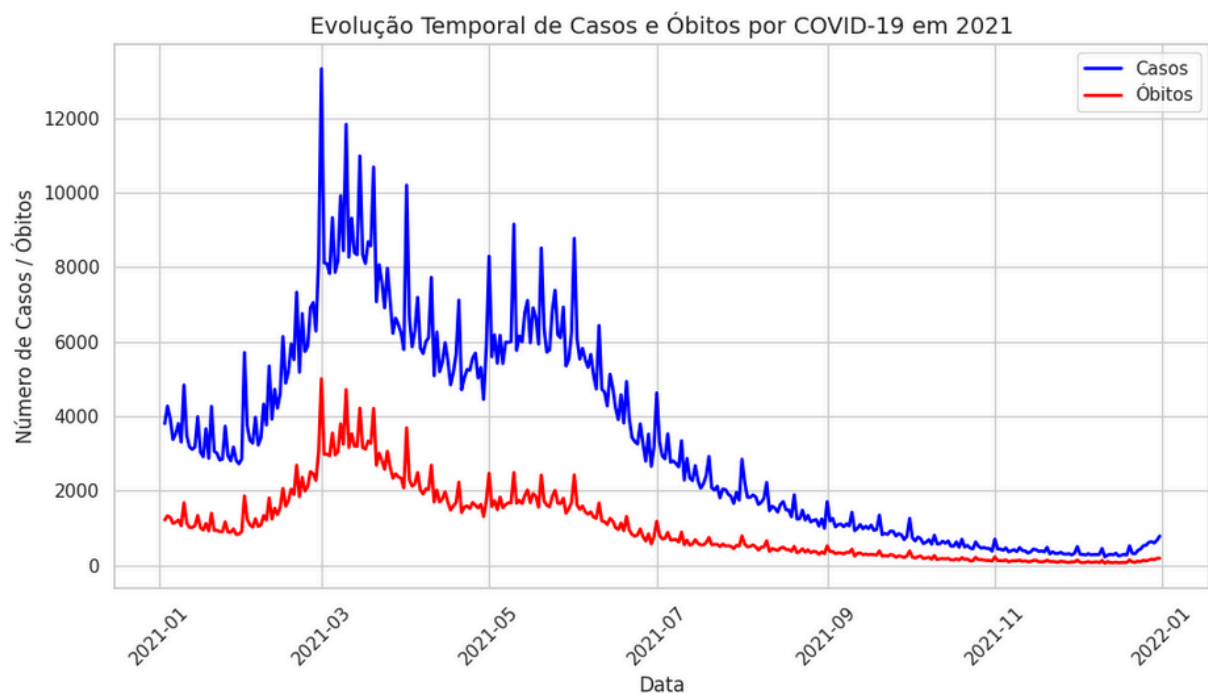
Comorbidade	Frequência	Probabilidade de Óbito
CARDIOPATI	337.585	43,07%
DIABETES	236.686	43,94%
OBESIDADE	110.822	40,56%
NEUROLOGIC	30.603	53,87%
RENAL	30.497	56,79%
PNEUMOPATI	28.228	52,65%
ASMA	26.014	29,91%
IMUNODEPRE	19.959	49,00%
HEPATICA	7.196	52,35%
HEMATOLOGI	5.837	45,61%
OUT_MORBI	284.824	41,17%



- Interpretação:** Doenças renais (56,79%) e neurológicas (53,87%) apresentam as maiores taxas de mortalidade entre os pacientes com COVID-19 na amostra de 2021.

4.1.5 Análise temporal de casos e óbitos

- **Pergunta:** Como evoluíram os casos e óbitos ao longo de 2021?
- **Método:** Gráficos de linha.
- **Explicação:** Os gráficos de linha mostram a evolução de casos e óbitos ao longo do tempo, permitindo identificar padrões sazonais ou picos de infecção.
- **Resultados:**
 - Gráfico de linha mostrando picos de casos e óbitos em determinados períodos.



- **Interpretação:** Houve picos de casos e óbitos durante ondas da pandemia, como no primeiro semestre de 2021.

4.1.6 Proporção de óbitos por idade e comorbidades

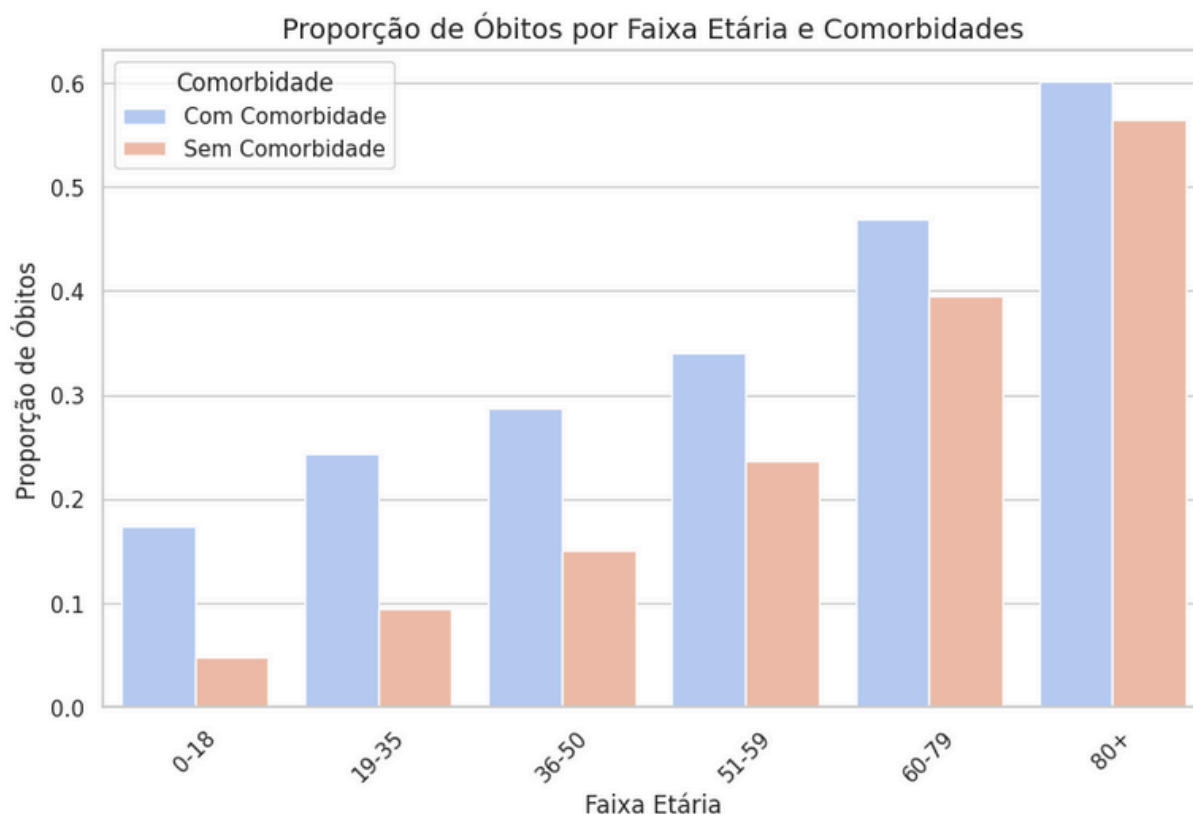
- **Pergunta:** Qual é a proporção de óbitos por faixa etária e presença de comorbidades?
- **Método:** Tabelas de contingência e cálculo de proporções.
- **Explicação:** As tabelas de contingência comparam a proporção de óbitos entre pacientes com e sem comorbidades, permitindo avaliar o impacto das comorbidades na mortalidade.
- **Resultados:**

Faixa Etária

Comorbidade

Proporção de Óbitos

0-18	Com Comorbidade	17,39%
0-18	Sem Comorbidade	4,79%
19-35	Com Comorbidade	24,37%
19-35	Sem Comorbidade	9,53%
36-50	Com Comorbidade	28,78%
36-50	Sem Comorbidade	15,13%
51-59	Com Comorbidade	34,12%
51-59	Sem Comorbidade	23,73%
60-79	Com Comorbidade	46,88%
60-79	Sem Comorbidade	39,50%
80+	Com Comorbidade	60,10%
80+	Sem Comorbidade	56,51%



- Interpretação:** Para a amostra de pacientes de 2021, a presença de comorbidades aumenta significativamente a proporção de óbitos em todas as faixas etárias. As maiores proporções de óbito se mantêm na população idosa, com ou sem comorbidades. A diferença da proporção de óbitos de pacientes com comorbidades e sem comorbidades diminui conforme a faixa de idade aumenta.

4.1.7 Proporção de mulheres idosas infectadas

- Pergunta:** Qual é a proporção de mulheres idosas entre os infectados?
- Método:** Cálculo de proporções.
- Explicação:** A proporção de mulheres idosas entre os infectados é calculada dividindo o número de mulheres idosas infectadas pelo total de infectados.
- Resultados:**

 - Número total de mulheres idosas infectadas: 254.342.
 - Proporção de mulheres idosas entre os infectados: 21,02%.
- Interpretação:** Mulheres idosas representam uma parcela significativa dos infectados na amostra de pacientes infectados por covid-19 em 2021.

4.2 Probabilidade

4.2.1 Probabilidade de um paciente ser homem

- **Pergunta:** Qual é a probabilidade de um paciente infectado ser homem?
 - **Método: Probabilidade condicional.**
 - **Explicação:** A probabilidade condicional é calculada dividindo o número de homens infectados pelo total de infectados.
 - **Resultados:**
 - Total de infectados: 1.211.609.
 - Total de homens infectados: 674.803.
 - Probabilidade: 55,69%.
 - **Interpretação:** A maioria dos pacientes infectados é do sexo masculino nessa amostra.
-

4.3 Inferência Estatística

4.3.1 Análise de casos por tipo de atendimento

- **Pergunta:** Há diferença na distribuição dos tipos de atendimento entre homens e mulheres?
 - **Método: Teste qui-quadrado.**
 - **Explicação:** O teste qui-quadrado compara as frequências observadas com as frequências esperadas sob a hipótese nula de independência entre as variáveis (sexo e tipo de atendimento).
 - **Resultados:**
 - Estatística Qui-Quadrado: 262,1259.
 - p-valor: 0,0000.
 - Conclusão: Há diferença significativa na distribuição dos tipos de atendimento entre homens e mulheres.
-

4.3.2 Testes de hipóteses sobre mortalidade

- **Pergunta:** Há diferença na mortalidade entre homens e mulheres e entre faixas etárias?
 - **Método: Teste t de Student e ANOVA.**
 - **Explicação:**
 - O **teste t de Student** compara as médias de mortalidade entre dois grupos (homens e mulheres).
 - A **ANOVA** compara as médias de mortalidade entre múltiplos grupos (faixas etárias).
 - **Resultados:**
 - Teste t: p-valor < 0,05 (diferença significativa entre homens e mulheres).
 - ANOVA: p-valor < 0,05 (diferença significativa entre faixas etárias).
 - **Interpretação:** A mortalidade varia significativamente entre sexos e faixas etárias.
-

4.3.3 Análise de diferenças entre regiões

- **Pergunta:** Há diferença na proporção de óbitos entre regiões?
 - **Método:** Teste qui-quadrado e ANOVA.
 - **Explicação:**
 - O teste qui-quadrado compara a proporção de óbitos entre regiões.
 - A ANOVA compara as médias de proporção de óbitos entre regiões.
 - **Resultados:**
 - Estatística Qui-Quadrado: 8.893,9396.
 - p-valor: 0,0000.
 - Conclusão: Há diferença significativa na proporção de óbitos entre regiões.
-

4.3.4 Intervalo de confiança para taxa de mortalidade

- **Pergunta:** Qual é o intervalo de confiança para a taxa de mortalidade?
 - **Método:** Intervalo de confiança (IC) para proporções.
 - **Explicação:** O IC 95% estima a faixa de valores dentro da qual a taxa de mortalidade populacional provavelmente está, com 95% de confiança.
 - **Resultados:**
 - Taxa de mortalidade estimada: 32,78%.
 - IC 95%: [32,6940%, 32,8650%].
 - **Interpretação:** A taxa de mortalidade está entre 32,69% e 32,87% com 95% de confiança.
-

5. Discussão

A escolha dos métodos foi guiada pelas perguntas de pesquisa e pela natureza dos dados. Por exemplo:

- Métodos descritivos (média, mediana, gráficos) foram usados para resumir e visualizar os dados.
- Testes de hipóteses (qui-quadrado, teste t, ANOVA) permitiram comparar grupos e identificar diferenças significativas.
- O intervalo de confiança forneceu uma estimativa da taxa de mortalidade com uma medida de precisão.

Estes métodos permitiram responder às perguntas de forma confiável, fornecendo insights sobre a distribuição de casos, comorbidades, mortalidade e diferenças entre grupos.

6. Conclusão

Este trabalho demonstra a importância da análise de dados para entender a dinâmica da COVID-19 no Brasil. As descobertas podem auxiliar na tomada de decisões em saúde pública, como a alocação de recursos e a priorização de grupos de risco.

7. Referências

- DATASUS: <https://opendatasus.saude.gov.br/dataset>
- BIBLIOTECA PANDAS: <https://pandas.pydata.org/docs/>
- BIBLIOTECA NUMPY: <https://numpy.org/doc/stable/>
- BIBLIOTECA SCIPY: <https://docs.scipy.org/doc/scipy/reference/stats.html>
- BIBLIOTECA MATPLOTLIB:
https://matplotlib.org/3.5.3/api/_as_gen/matplotlib.pyplot.html
- BIBLIOTECA SEABORN: <https://seaborn.pydata.org/>