

Raphael da Silva Ferreira Capistrano

Matheus Claudino Ribeiro

Matheus Victor Feliciano Jupi

Vitória de Lima Santos

ANÁLISE ESTATÍSTICA DA SÍNDROME GRIPAL EM SÃO PAULO (2022 E 2024)

São Paulo

Novembro, 2025

Raphael da Silva Ferreira Capistrano

Matheus Claudino Ribeiro

Matheus Victor Feliciano Jupi

Vitória de Lima Santos

ANÁLISE ESTATÍSTICA DA SÍNDROME GRIPAL EM SÃO PAULO (2022 E 2024)

Documento técnico apresentado à disciplina
de Estatística do Instituto Federal de São
Paulo, referente ao desenvolvimento e docu-
mentação da Análise proposta

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – IFSP
Campus São Paulo

São Paulo

Novembro, 2025

RESUMO

Este trabalho apresenta uma análise estatística de dados de Síndrome Gripal notificados no estado de São Paulo, comparando os anos de 2022 e 2024. Foram utilizadas técnicas de estatística descritiva, probabilidade e inferência estatística, com foco na variável idade, nos sintomas mais recorrentes e na evolução dos casos (cura, cancelado e ignorado).

Os dados foram tratados e analisados em linguagem Python, utilizando bibliotecas especializadas para manipulação, visualização e testes de hipóteses. A partir dos resultados, foi possível identificar padrões relevantes na distribuição etária, nas frequências de sintomas e nas diferenças entre as evoluções dos casos ao longo dos anos estudados.

Palavras-chave: análise de dados; estatística; síndrome gripal; Python; São Paulo.

SUMÁRIO

1	INTRODUÇÃO	4
2	DESCRIÇÃO DOS DADOS	5
2.1	Origem dos Dados	5
2.2	Estrutura das Variáveis	5
2.3	Tratamento Inicial	5
3	METODOLOGIA	7
3.1	Ferramentas Utilizadas	7
3.2	Questões de Pesquisa	7
3.3	Procedimentos de Análise	7
4	ANÁLISE EXPLORATÓRIA DOS DADOS	9
4.1	Estatística Descritiva da Idade	9
4.2	Sintomas Mais Frequentes	9
4.3	Evolução dos Casos	9
4.4	Visualizações Gráficas	9
5	RESULTADOS E DISCUSSÃO	13
5.1	Distribuição etária dos casos	13
5.2	Sintomas mais frequentes	13
5.3	Evolução dos casos	14
5.4	Avaliação da normalidade e implicações	15
5.5	Comparação entre as idades médias de 2022 e 2024	15
5.6	Síntese geral dos achados	16
6	CONCLUSÃO	17
	Referências	19

1 INTRODUÇÃO

O aumento no volume de dados em saúde pública tornou essencial a utilização de técnicas de análise estatística e ciência de dados para apoiar decisões e políticas públicas. Dentre os agravos monitorados, a Síndrome Gripal possui relevância epidemiológica por estar associada a surtos sazonais, epidemias e a doenças respiratórias de impacto coletivo.

Neste trabalho, é realizada uma análise estatística de dados de Síndrome Gripal no estado de São Paulo, considerando notificações dos anos de 2022 e 2024. O objetivo principal é comparar o comportamento da variável idade, os sintomas mais frequentes e a evolução dos casos entre os dois anos analisados.

Especificamente, busca-se:

- descrever a distribuição de idade dos casos notificados em 2022 e 2024;
- identificar os sintomas mais registrados em cada ano;
- estimar probabilidades associadas à evolução dos casos (cura, cancelado, ignorado);
- verificar, por meio de testes estatísticos, se há diferença significativa na idade entre os anos de 2022 e 2024.

A análise é conduzida utilizando linguagem Python, com apoio das bibliotecas *Pandas*, *Matplotlib*, *Seaborn*, *SciPy* e *Statsmodels*, integrando tratamento de dados, visualização gráfica e inferência estatística.

2 DESCRIÇÃO DOS DADOS

2.1 ORIGEM DOS DADOS

Conforme indicado pelo Ministério da Saúde (Saúde, 2022, 2024), os dados utilizados neste estudo foram obtidos no portal oficial de informações em saúde pública. O conjunto analisado é composto por notificações de Síndrome Gripal registradas no estado de São Paulo, disponibilizadas em arquivos no formato CSV. Para efeito de comparação temporal, selecionaram-se especificamente os registros referentes aos anos de 2022 e 2024.

Os arquivos foram organizados na estrutura:

- dados/Notificações de Síndrome Gripal - 2022_MAIOR.csv;
- dados/Notificações de Síndrome Gripal - 2024_MAIOR.csv.

Para fins de teste e desempenho, foram lidas, inicialmente, amostras com até 5000 linhas de cada ano.

2.2 ESTRUTURA DAS VARIÁVEIS

Após a leitura dos arquivos, foi realizada uma padronização dos nomes das colunas, convertendo-os para letras minúsculas e removendo acentos, de forma a facilitar o tratamento no código. Entre as variáveis disponíveis, destacam-se:

- **idade**: idade do paciente no momento da notificação;
- **sintomas**: registro dos sintomas apresentados;
- **evolucaocaso**: classificação final do caso (cura, cancelado, ignorado);
- **ano**: ano ao qual pertence o registro (2022 ou 2024), criado durante o processamento.

Outras variáveis estão presentes na base, porém, para este estudo, o foco recai nas variáveis acima, por estarem diretamente relacionadas às questões de pesquisa definidas.

2.3 TRATAMENTO INICIAL

As principais etapas de tratamento dos dados foram:

- normalização dos nomes das colunas (remoção de espaços, acentos e padronização em minúsculo);
- conversão da variável `idade` para tipo numérico (`Int64`), tratando valores inválidos como ausentes;
- criação da coluna `ano` para distinguir registros de 2022 e 2024;
- junção das duas bases em um único *DataFrame*, possibilitando análises comparativas.

O tratamento foi realizado no script `analise_sindrome_gripal.py`.

3 METODOLOGIA

3.1 FERRAMENTAS UTILIZADAS

A análise foi realizada em linguagem Python, utilizando-se as seguintes bibliotecas:

- **Pandas**: para leitura, manipulação e tratamento dos dados;
- **Matplotlib e Seaborn**: para geração de gráficos e visualizações;
- **SciPy**: para testes estatísticos, como o teste t de Student e o teste de Shapiro–Wilk;
- **Statsmodels**: para geração de gráficos QQ-plot;
- **IPython.display**: para exibição de tabelas em ambiente interativo.

3.2 QUESTÕES DE PESQUISA

A metodologia foi construída com base nas seguintes questões:

- Qual é a distribuição de idade dos casos notificados em 2022 e 2024?
- Quais são os sintomas mais frequentes em cada ano?
- Como se distribuem as categorias de evolução do caso (cura, cancelado, ignorado) por ano?
- Existe diferença estatisticamente significativa entre as idades dos pacientes notificados em 2022 e 2024?

3.3 PROCEDIMENTOS DE ANÁLISE

A partir do *DataFrame* unificado, foram conduzidas as seguintes etapas:

1. seleção das variáveis de interesse: `idade`, `sintomas`, `evolucaocaso` e `ano`;
2. cálculo de estatísticas descritivas para idade por ano;
3. identificação dos sintomas mais comuns em cada ano;
4. contagem e proporção das categorias de evolução do caso;

5. cálculo de probabilidades associadas à evolução (cura, cancelado, ignorado) em cada ano;
6. verificação da normalidade da idade por ano, utilizando o teste de Shapiro–Wilk e gráficos QQ-plot;
7. aplicação do teste t de Student para amostras independentes, comparando as idades de 2022 e 2024.

4 ANÁLISE EXPLORATÓRIA DOS DADOS

4.1 ESTATÍSTICA DESCRIPTIVA DA IDADE

Para cada ano, foram calculadas medidas descritivas da variável idade, como média, mediana, desvio-padrão, quartis e valores mínimo e máximo. Essas estatísticas permitem uma visão geral do perfil etário dos pacientes.

4.2 SINTOMAS MAIS FREQUENTES

A variável `sintomas` foi agrupada por ano e foram identificados, para cada ano, os quatro sintomas mais frequentemente registrados. A apresentação foi feita em tabelas de contagem, permitindo comparar se houve alteração no padrão de sintomas entre 2022 e 2024.

4.3 EVOLUÇÃO DOS CASOS

A variável `evolucaocaso` foi analisada de forma categórica, contabilizando-se o número de registros em cada categoria (cura, cancelado, ignorado), por ano. Além das contagens absolutas, foram calculadas proporções em relação ao total de casos daquele ano.

4.4 VISUALIZAÇÕES GRÁFICAS

Para apoiar a interpretação dos resultados, foram gerados os seguintes gráficos:

- **Boxplot da idade por ano**, evidenciando a distribuição e possíveis outliers;
- **Histogramas da idade** para 2022 e 2024, mostrando a forma da distribuição;
- **QQ-plots** da idade em 2022 e 2024, utilizados para avaliar a normalidade da distribuição.

Os gráficos foram salvos automaticamente na pasta `output/` pelo script em Python e são apresentados a seguir.

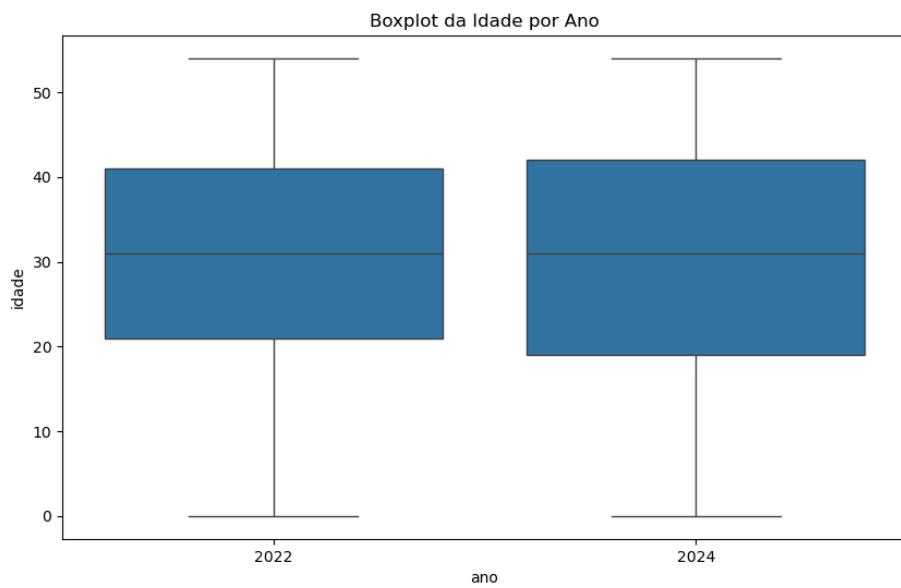


Figura 1 – Boxplot da idade por ano (2022 e 2024)

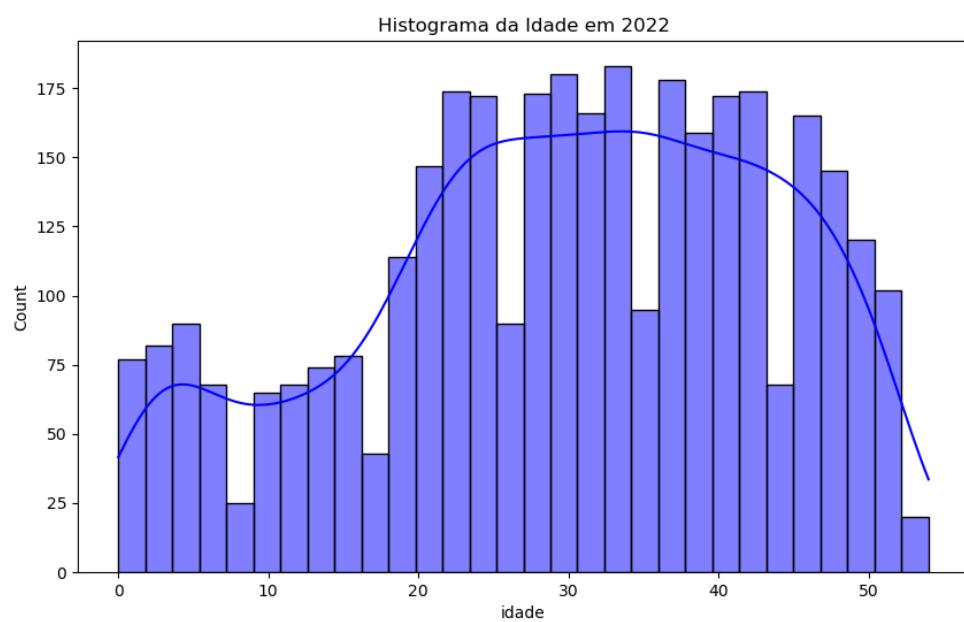


Figura 2 – Histograma da idade em 2022

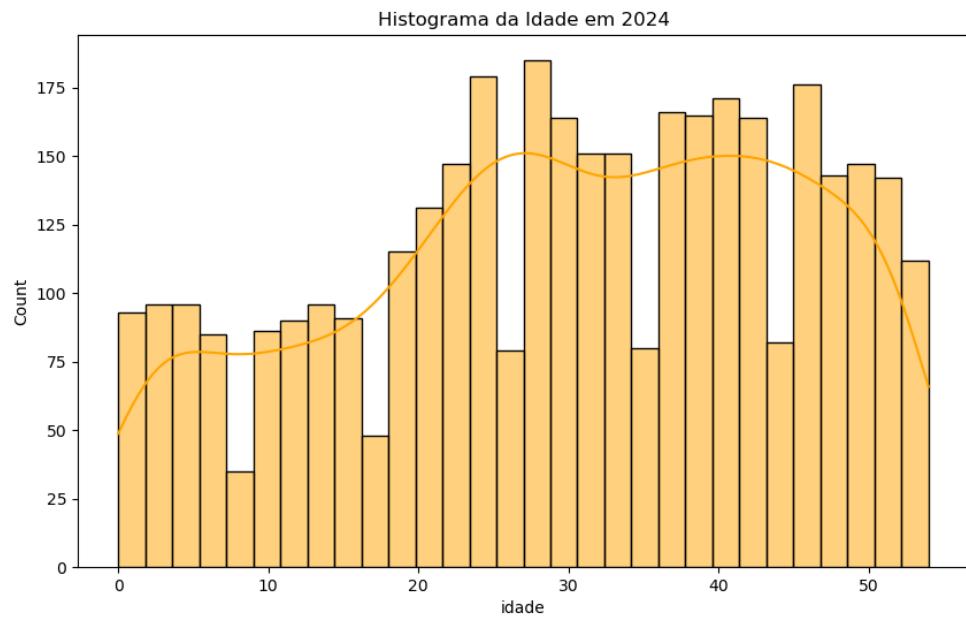


Figura 3 – Histograma da idade em 2024

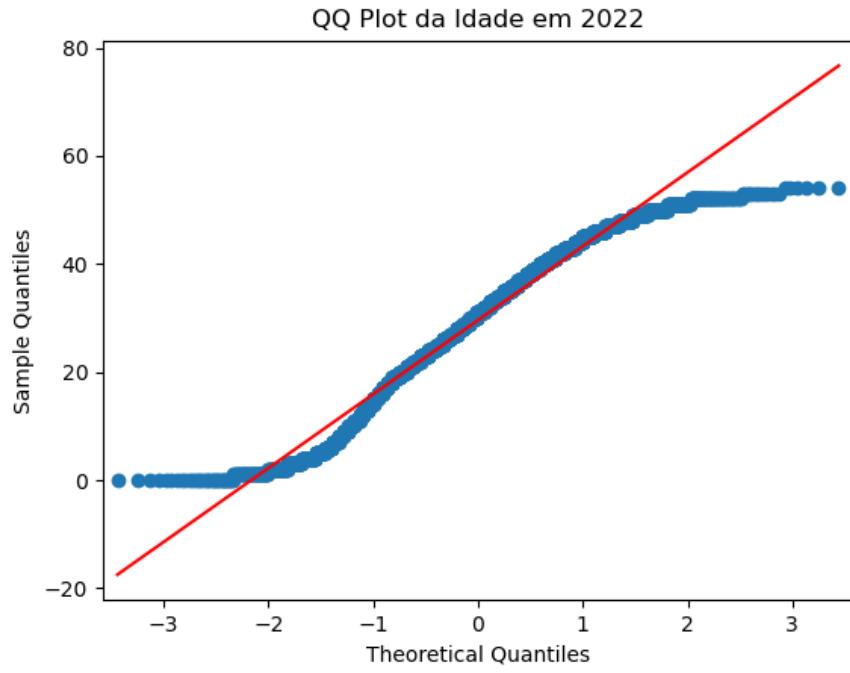


Figura 4 – QQ-Plot da idade em 2022

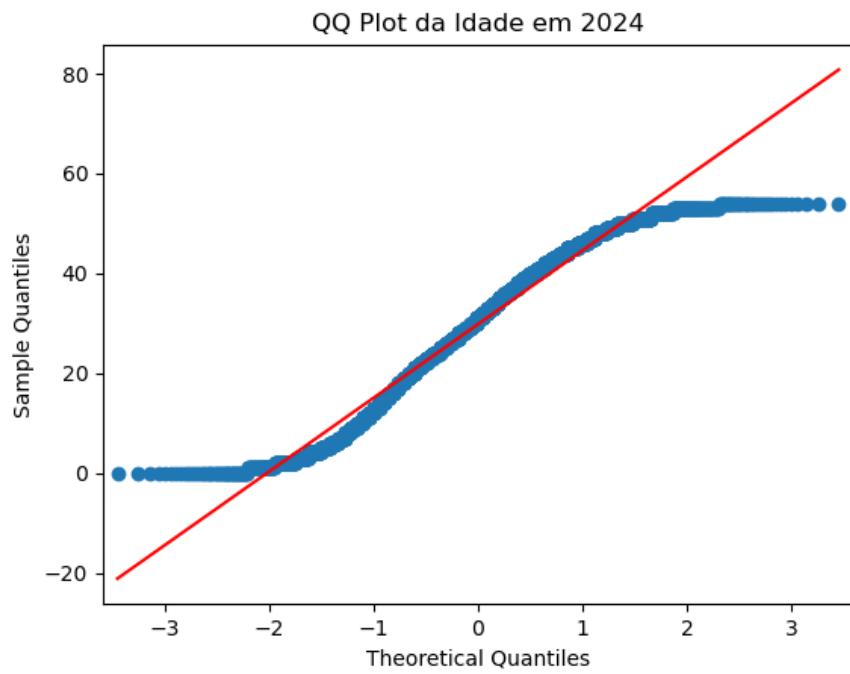


Figura 5 – QQ-Plot da idade em 2024

5 RESULTADOS E DISCUSSÃO

5.1 DISTRIBUIÇÃO ETÁRIA DOS CASOS

A Tabela de estatísticas descritivas mostrou que a idade dos pacientes notificados em 2022 e 2024 apresenta perfis muito semelhantes. Em 2022, a média foi de aproximadamente 29,6 anos, enquanto em 2024 a média foi de 29,8 anos. Em ambos os anos, a mediana situou-se em 31 anos, com primeiro e terceiro quartis em torno de, respectivamente, 21 e 41 anos em 2022 e 19 e 42 anos em 2024. Os valores mínimo e máximo observados variaram de 0 a 54 anos em ambos os períodos.

O *boxplot* da idade por ano (Figura 1) reforça essa semelhança: as distribuições apresentam amplitudes e intervalos interquartis próximos, com alguns valores extremos em idades muito baixas, possivelmente relacionados a lactentes e crianças pequenas. Observa-se também discreto aumento da dispersão em 2024, evidenciado pelo desvio-padrão ligeiramente maior nesse ano.

Os histogramas da idade (Figuras 2 e 3) indicam uma distribuição assimétrica, com maior concentração de casos em faixas etárias jovens e adultas, e cauda que se estende para idades mais elevadas. Visualmente, o padrão geral se mantém entre 2022 e 2024, sem mudanças estruturais relevantes.

5.2 SINTOMAS MAIS FREQUENTES

A análise da variável **sintomas** mostrou que, em ambos os anos, o padrão sintomático é bastante semelhante. Em 2022, os sintomas mais frequentes foram:

- **Assintomático** (721 registros);
- **Coriza e Tosse** (178 registros);
- **Tosse isolada** (148 registros);
- **Outros** (128 registros).

Em 2024, a ordem das maiores frequências se repete, com pequenas variações de contagem:

- **Assintomático** (611 registros);
- **Coriza e Tosse** (183 registros);

- **Tosse** isolada (175 registros);
- **Outros** (108 registros).

Esses resultados sugerem que o perfil clínico da Síndrome Gripal se manteve estável no período analisado, com importante proporção de indivíduos assintomáticos e predominância de sintomas respiratórios clássicos (coriza e tosse).

5.3 EVOLUÇÃO DOS CASOS

A variável **evolucaocaso** foi utilizada para caracterizar o desfecho dos atendimentos. Em 2022, destacam-se:

- **Cura:** 2719 casos;
- **Ignorado:** 309 casos;
- **Cancelado** (encerramento não favorável, associado a óbito ou perda de seguimento): 82 casos;
- **Em tratamento domiciliar:** 57 casos;
- **Internado:** 4 casos;
- **Internado em UTI:** 1 caso.

Em 2024, o padrão permanece semelhante, com leve aumento no número absoluto de casos encerrados como cura:

- **Cura:** 3178 casos;
- **Ignorado:** 382 casos;
- **Em tratamento domiciliar:** 83 casos;
- **Cancelado:** 71 casos;
- **Internado:** 4 casos.

Ao transformar essas contagens em probabilidades em relação ao total de notificações de cada ano, obtém-se, aproximadamente, para 2022:

$$P(\text{Cura}) \approx 0,54, \quad P(\text{Cancelado}) \approx 0,02, \quad P(\text{Ignorado}) \approx 0,06,$$

e, para 2024:

$$P(\text{Cura}) \approx 0,64, \quad P(\text{Cancelado}) \approx 0,01, \quad P(\text{Ignorado}) \approx 0,08.$$

Em ambos os anos, observa-se que a maior parte dos casos é encerrada como cura, com proporção ligeiramente maior em 2024. As categorias *Cancelado* (associada a desfechos desfavoráveis) e *Internado/UTI* apresentam probabilidades baixas, sugerindo menor gravidade entre os casos notificados na amostra analisada. Por outro lado, a presença de registros *Ignorado* em torno de 6 a 8% indica uma fração não desprezível de desfechos sem encerramento adequado, o que pode impactar a completude das informações.

5.4 AVALIAÇÃO DA NORMALIDADE E IMPLICAÇÕES

O teste de Shapiro-Wilk aplicado à idade nos dois anos resultou em p-valores inferiores a 0,05, levando à rejeição da hipótese de normalidade para ambas as distribuições. Esse comportamento é consistente com os QQ-plots apresentados nas Figuras 4 e 5, nos quais se observa desvio em relação à linha de referência, especialmente nas caudas da distribuição.

Apesar disso, o grande tamanho amostral (mais de três mil observações de idade em cada ano) reduz o impacto da não normalidade sobre a estimativa das médias, em virtude do Teorema Central do Limite. Ainda assim, essa característica deve ser considerada como limitação na interpretação de testes paramétricos.

5.5 COMPARAÇÃO ENTRE AS IDADES MÉDIAS DE 2022 E 2024

Para avaliar se houve mudança significativa na idade média dos pacientes entre 2022 e 2024, foi aplicado o teste t de Student para amostras independentes. As hipóteses consideradas foram:

$$H_0 : \mu_{2022} = \mu_{2024} \quad \text{vs.} \quad H_1 : \mu_{2022} \neq \mu_{2024}.$$

O teste resultou em estatística $t \approx -0,63$ e p-valor de aproximadamente 0,53, valor superior ao nível de significância adotado ($\alpha = 0,05$). Assim, **não há evidência estatística para rejeitar a hipótese nula**, indicando que as idades médias dos pacientes notificados em 2022 e 2024 podem ser consideradas estatisticamente semelhantes.

Esse resultado está em consonância com as medidas descritivas e com o *boxplot* da idade por ano, que já sugeriam grande proximidade entre as distribuições etárias. Do ponto de vista epidemiológico, isso indica que o perfil etário dos indivíduos acometidos por Síndrome Gripal se manteve estável no período analisado.

5.6 SÍNTESE GERAL DOS ACHADOS

De forma geral, os resultados obtidos permitem destacar que:

- o perfil etário dos casos de Síndrome Gripal é semelhante em 2022 e 2024, com idade média em torno de 30 anos e grande participação de adultos jovens;
- o padrão de sintomas permanece estável, com predominância de casos assintomáticos e sintomas respiratórios típicos (coriza e tosse);
- a maioria dos casos é encerrada como cura, com probabilidades de cancelamento/óbito e internação relativamente baixas nos dois anos;
- embora a distribuição de idade não seja estritamente normal, o tamanho amostral elevado permite comparar as médias com razoável robustez, não sendo observada diferença estatisticamente significativa entre 2022 e 2024.

Esses achados sugerem que, na amostra analisada, não houve mudança estrutural importante no perfil dos casos de Síndrome Gripal entre os anos estudados, ainda que pequenas variações nas proporções de cura, cancelamento e registros ignorados possam refletir diferenças na dinâmica de atendimento ou na qualidade do preenchimento das notificações.

6 CONCLUSÃO

Este trabalho apresentou uma análise estatística de dados de Síndrome Gripal notificados no estado de São Paulo, com foco comparativo entre os anos de 2022 e 2024. A partir do tratamento das bases em Python e da aplicação de técnicas de estatística descritiva, probabilidade e inferência, foi possível caracterizar o perfil etário dos casos, os sintomas mais frequentes e a evolução dos atendimentos.

Os resultados indicaram que o **perfil etário** dos pacientes se manteve estável nos dois anos analisados. As idades médias ficaram em torno de 30 anos, com mediana de 31 anos e amplitude semelhante entre 2022 e 2024. O teste *t* de Student não apontou diferença estatisticamente significativa entre as idades médias, o que sugere ausência de mudança estrutural no grupo etário mais atingido pela Síndrome Gripal no período considerado.

Em relação aos **sintomas**, observou-se predominância de indivíduos assintomáticos e de manifestações respiratórias clássicas, como coriza e tosse, em ambos os anos. Esse padrão reforça a consistência clínica dos casos notificados e indica que o quadro típico de Síndrome Gripal permaneceu similar entre 2022 e 2024.

Quanto à **evolução dos casos**, a maior parte das notificações foi encerrada como cura, com aumento da proporção de casos curados em 2024. As categorias associadas a desfechos mais graves ou desfavoráveis, como cancelado, internado e internado em UTI, apresentaram probabilidades baixas nos dois anos, o que sugere, na amostra analisada, predomínio de quadros menos graves. Por outro lado, a presença de uma fração não desprezível de registros classificados como ignorado evidencia limitações na completude das informações e aponta para a necessidade de aprimoramento nos processos de registro e encerramento dos casos.

A avaliação da normalidade da idade, por meio do teste de Shapiro-Wilk e dos gráficos QQ-plot, indicou que as distribuições de idade não seguem exatamente o modelo normal. Entretanto, o grande tamanho amostral confere robustez às estimativas de média e permite a utilização de testes paramétricos com cautela, desde que as limitações sejam explicitadas.

Como **limitações** deste estudo, destacam-se: (i) o uso de amostras parciais das bases originais (subconjuntos de registros para cada ano), o que pode não capturar integralmente todas as nuances do banco de dados; e (ii) a dependência da qualidade do preenchimento das notificações, sujeitas a erros, omissões e inconsistências.

Como **trabalhos futuros**, sugerem-se:

- ampliar o período analisado, incorporando outros anos para avaliar tendências de

longo prazo;

- incluir variáveis adicionais, como sexo, raça/cor, região de residência e presença de comorbidades, permitindo análises mais detalhadas de subgrupos populacionais;
- aplicar modelos de regressão ou técnicas de aprendizado de máquina para investigar fatores associados a desfechos específicos, como necessidade de internação ou cancelamento do caso;
- explorar indicadores de qualidade do dado, visando aprimorar a completude e a confiabilidade das notificações.

Em síntese, a análise realizada evidencia a importância do uso combinado de ferramentas computacionais e métodos estatísticos na interpretação de bases de dados em saúde pública. Mesmo com as limitações apontadas, os resultados obtidos contribuem para a compreensão do comportamento da Síndrome Gripal em São Paulo e reforçam o potencial dos dados de vigilância epidemiológica como apoio à tomada de decisão em políticas de saúde.

REFERÊNCIAS

SAÚDE, Ministério da. **Base de Dados de Síndrome Gripal.** [S. l.: s. n.], 2022.

Acesso em 2025. Disponível em:

<https://opendatasus.saude.gov.br/dataset/notificacoes-de-sindrome-gripal-leve-2022/resource/9585044e-6555-477f-b103-12310b24e9b8>.

SAÚDE, Ministério da. **Base de Dados de Síndrome Gripal.** [S. l.: s. n.], 2024.

Acesso em 2025. Disponível em:

<https://opendatasus.saude.gov.br/dataset/notificacoes-de-sindrome-gripal-leve-2024/resource/db7e8e94-37e2-4b52-9aa6-db141dc71722>.