

**Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Campus São
Paulo**

Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas

Trabalho de Estatística e Probabilidade
Febre Amarela em humanos e primatas não-humanos - 1994 a 2023

Wesley Vieira dos Santos SP3083896

Orientadora:
Josceli Maria Tenorio

São Paulo
2024

Sumário

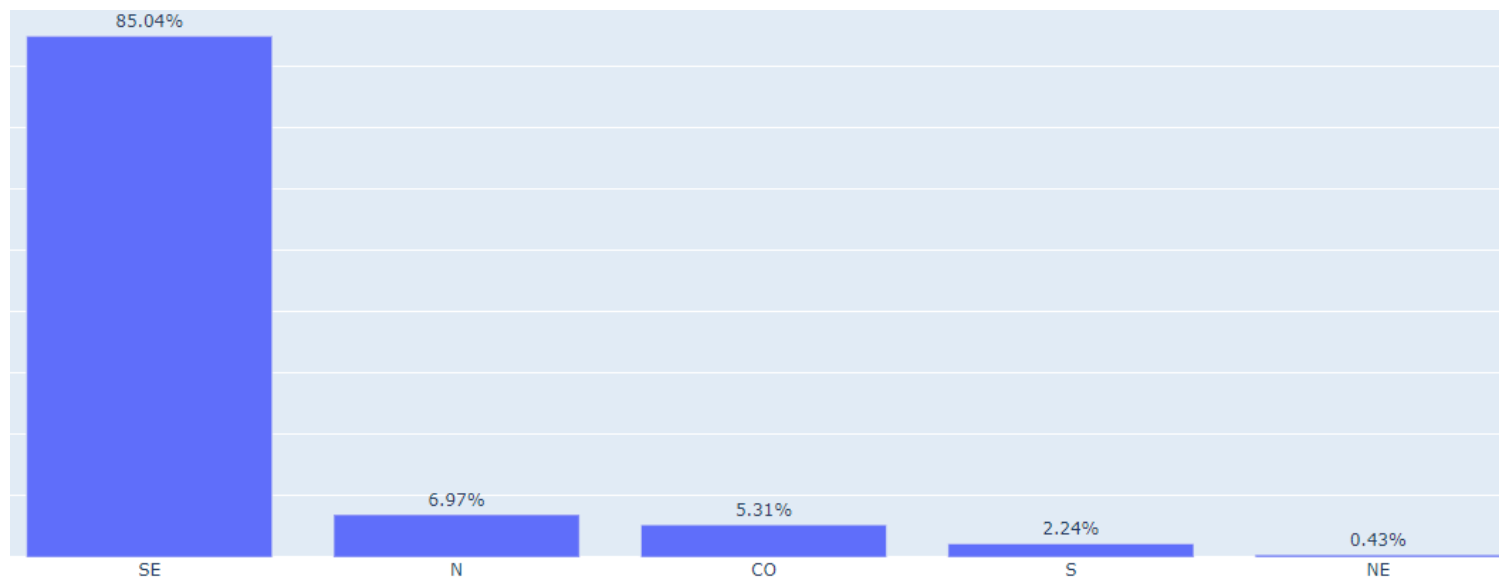
1. Introdução.....	3
2. Regiões e Estados.....	4
3. Gênero.....	6
4. Idades.....	7
4.1 Box Plot das Idades.....	9
5. Análise das Probabilidades.....	12
6. Conclusão.....	13
6.1 Recomendações.....	13

1. Introdução

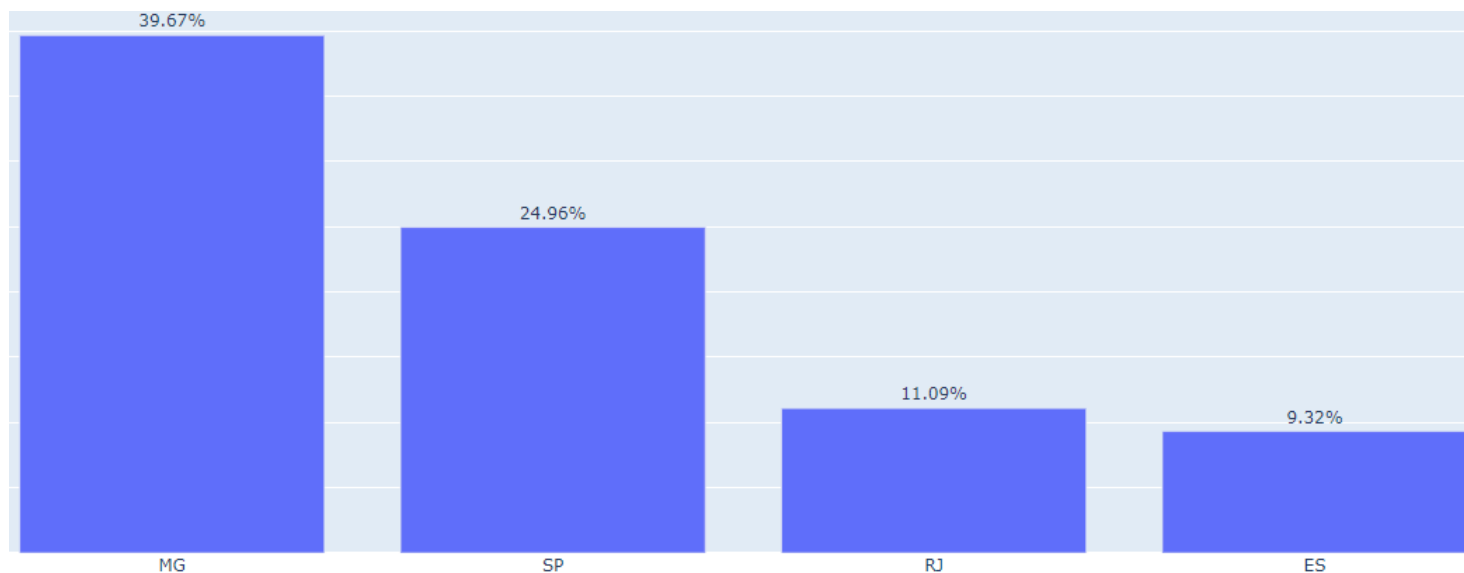
Por meio da Coordenação-Geral de Vigilância de Arboviroses (CGARB) do Departamento de Doenças Transmissíveis (DEDT) da Secretaria de Vigilância em Saúde e Ambiente (SVSA), foi obtida uma base de dados sobre a Febre Amarela em humanos e primatas não-humanos, abrangendo o período de 1994 a 2023. Esta base de dados inclui informações de 2768 indivíduos.

Com base nesses dados, serão analisadas as regiões e estados mais afetados pela doença, o gênero com maior incidência de Febre Amarela e as faixas etárias dos indivíduos contagiados, juntamente com as frequências das idades mais afetadas e a construção de um box plot para melhor visualização dos resultados.

2. Regiões e Estados



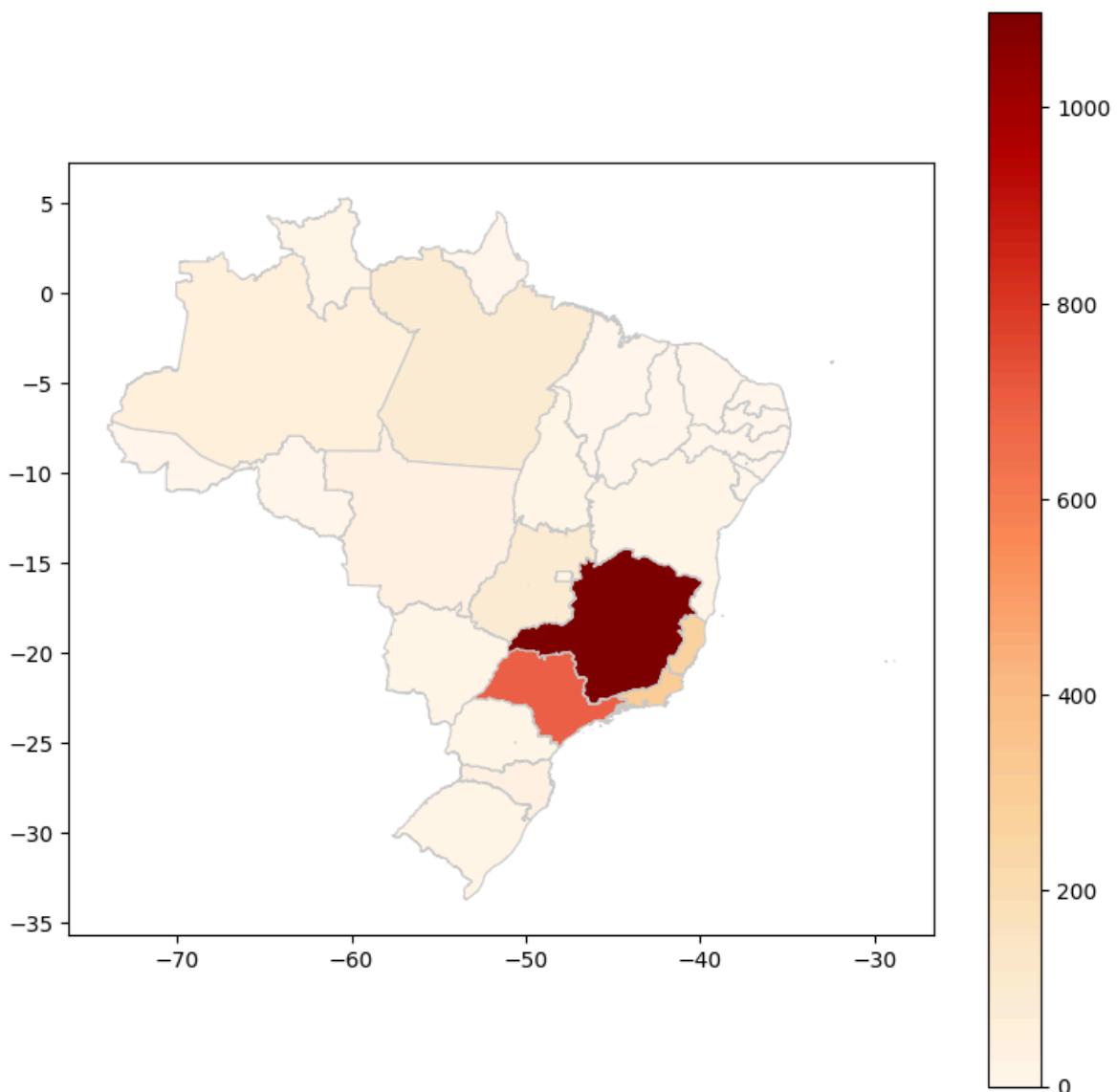
Com base no histograma, observou-se que as regiões impactadas pela incidência de casos de febre amarela foram o Sudeste (SE), Norte (N), Centro-Oeste (CO), Sul (S) e Nordeste (NE). Dentre essas, a Sudeste foi a mais significativamente afetada, com aproximadamente 85% dos casos registrados, seguida pela região Norte, que apresentou cerca de 7% de incidência da doença.



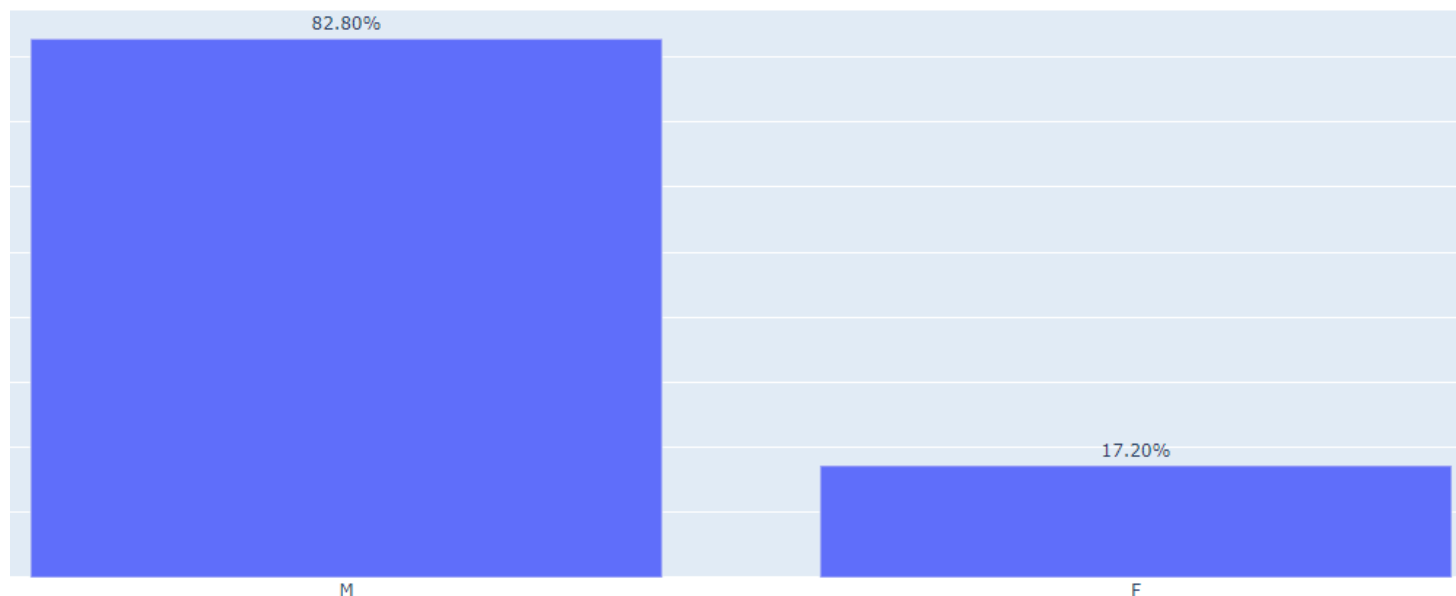
Com base na análise regional, os estados mais impactados pela febre amarela na região Sudeste foram Minas Gerais (MG), São Paulo (SP), Rio de Janeiro (RJ) e Espírito Santo (ES). Entre esses, Minas Gerais se destacou com cerca de 40% dos casos registrados, seguido por São Paulo, que apresentou aproximadamente 25% da incidência da doença.

Uma análise estatística adicional, utilizando o teste de qui-quadrado, indicou que há uma associação significativa entre a localização geográfica (estado) e a incidência de febre amarela. Os resultados do teste (Estatística qui-quadrado: 2558, Valor-p: 0) sugerem que a distribuição dos casos de febre amarela não é uniforme entre os estados.

Isso significa que alguns estados têm significativamente mais ou menos casos do que seria esperado, reforçando a ideia de que a febre amarela afetou de maneira desigual os diferentes estados do Brasil. O mapa de calor logo abaixo mostra a distribuição geográfica dos casos de febre amarela no Brasil.



3. Gênero

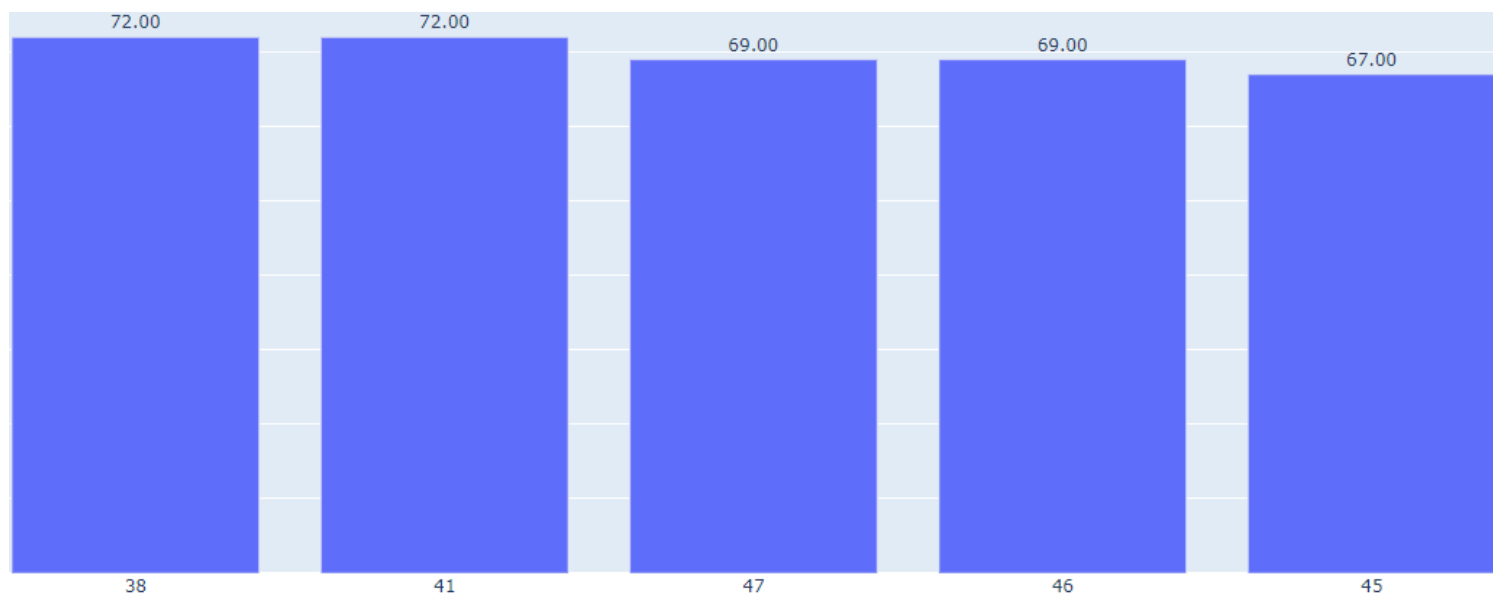


Entre 1994 e 2023, a febre amarela afetou predominantemente o sexo masculino, representando cerca de 82% dos casos registrados, enquanto o sexo feminino correspondia a aproximadamente 17%. Na prática, os homens têm cerca de cinco vezes mais chances de contrair a doença em comparação com as mulheres.

Análises estatísticas adicionais revelaram uma diferença significativa na incidência de febre amarela entre os sexos (Estatística do teste: 48,79; Valor-p: 0,0). Esses dados sugerem fortemente uma disparidade substancial na frequência da doença, com uma predominância significativa de casos entre os homens.

Nos estados de Minas Gerais (MG) e São Paulo (SP), onde a incidência da febre amarela foi mais alta, um homem tem 84% de chance de contrair a doença em MG e 82% em SP. Em contraste, uma mulher tem 15% de chance de contrair a doença em MG e 17% em SP.

4. Idades



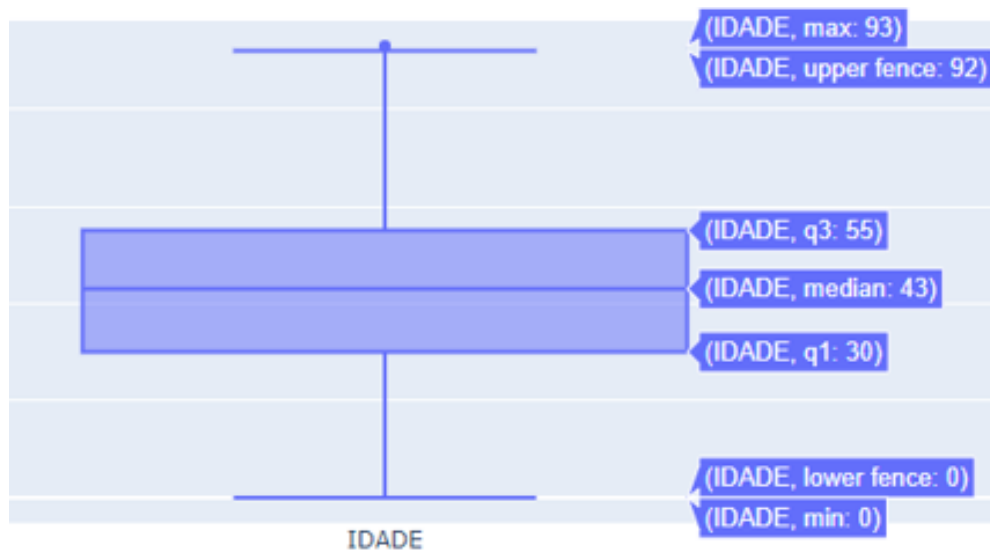
Com base no histograma, observa-se que a faixa etária mais impactada pela febre amarela é aquela próxima ou acima dos 40 anos. Em particular, os grupos de 38 e 41 anos se destacam, com cerca de 72 pessoas afetadas pela doença em cada grupo em relação aos 2768 indivíduos. Além disso, de acordo com os resultados do teste de média, não há diferença estatisticamente significativa na idade média entre homens e mulheres afetados pela febre amarela, com um valor-p de 0.31. Isso sugere que a idade média não é um fator determinante na suscetibilidade à doença entre homens e mulheres.

Abaixo, é possível verificar as idades, frequência e frequência relativa das 5 faixas etárias mais afetadas.

Idade	Frequência	Frequência Relativa
38	72	0.0260
41	72	0.0260
47	69	0.0249
46	69	0.0249
45	67	0.0242

A partir dos dados coletados, a média das idades afetadas é de 42,5 e a mediana é igual a 43, o que significa que os valores estão relativamente uniformemente distribuídos. Além disso, a probabilidade de uma pessoa entre 38 e 45 anos contrair a Febre Amarela é cerca de 18% em relação ao total de indivíduos.

4.1 Box Plot das Idades



Com base no box plot fornecido, podemos fazer diversas observações sobre a distribuição dos dados:

Valores Mínimo e Máximo

- X menor (mínimo): 0
- X maior (máximo): 92 (sem considerar o outlier)

Quartos

- Primeiro Quarto (Q1): 30
- Mediana (Q2): 43
- Terceiro Quarto (Q3): 55

Distribuição dos Dados

A mediana (43) está mais próxima de Q1 (30) do que de Q3 (55), sugerindo que a distribuição dos dados pode ser levemente assimétrica, com uma cauda mais longa à direita (distribuição assimétrica positiva).

- A presença de um outlier (93) reforça essa assimetria positiva, indicando que há um valor extremo no lado superior da distribuição.

Resumo da Distribuição

- A maior parte dos dados (50%) está entre 30 e 55.
- 25% dos dados estão abaixo de 30 e 25% estão acima de 55.
- A mediana de 43 indica que metade dos dados está abaixo desse valor e a outra metade está acima.

Interpretação Visual do Box Plot

O box plot ilustra a dispersão e os possíveis outliers dos dados, a linha no centro da caixa representa a mediana (43), as extremidades da caixa representam Q1 (30) e Q3 (55), os "bigodes" se estendem do 0 até 92, excluindo outliers e o ponto ou asterisco fora dos "bigodes" (93) indica um outlier.

Essas informações permitem uma análise rápida e visual da distribuição dos dados, destacando a centralidade, a dispersão e a presença de valores extremos.

Variância Amostral

- A variância amostral é 273. Ela mede a dispersão dos dados em relação à média.
- Uma variância alta indica que os dados estão mais espalhados em torno da média.

Desvio Padrão

- O desvio padrão é 17. Ele é a raiz quadrada da variância e também mede a dispersão dos dados, mas está na mesma unidade dos dados.
- Um desvio padrão de 17 indica uma dispersão significativa em relação à média, mas ele é mais interpretável diretamente no contexto dos dados.

Análise Geral

A análise combinada do box plot e das medidas de dispersão revela uma distribuição com uma forte concentração central entre 30 e 55, uma leve assimetria positiva, e uma dispersão considerável, evidenciada pelo desvio padrão de 17. A presença de um outlier em 93 destaca a existência de valores extremos que aumentam a variabilidade dos dados.

Apesar disso, a maioria dos dados permanece dentro de um intervalo bem definido, com uma mediana que indica um valor central robusto. Essas características sugerem que, enquanto os dados são majoritariamente consistentes e concentrados, há outliers que influenciam a distribuição geral, causando uma dispersão maior.

5. Análise das Probabilidades

Nesta análise, investigamos as probabilidades de contrair febre amarela em diferentes faixas etárias - crianças, jovens e adultos - e em duas regiões geográficas específicas: Sudeste (SE) e Norte (N) do Brasil, que foram as mais afetadas pela doença.

Crianças (até 14 anos):

- **Região Sudeste (SE):** A probabilidade de uma criança contrair febre amarela na região Sudeste foi calculada em 51.85%.
- **Região Norte (N):** Na região Norte, a probabilidade foi ligeiramente menor, estimada em 34.26%.
- **Observação:** A probabilidade é maior na região SE em comparação com a região N para crianças.

Jovens (de 15 anos até 24 anos):

- **Região Sudeste (SE):** A probabilidade de um jovem contrair febre amarela na região Sudeste foi significativamente alta, em torno de 66.19%.
- **Região Norte (N):** Em contraste, a probabilidade na região Norte foi consideravelmente mais baixa, aproximadamente 22.44%.
- **Observação:** Verifica-se uma disparidade marcante entre as regiões, com uma probabilidade muito maior na região Sudeste em comparação com a região Norte para jovens.

Adultos (acima de 24 anos):

- **Região Sudeste (SE):** A probabilidade de adultos contrair febre amarela na região Sudeste foi excepcionalmente alta, em torno de 89.51%.
- **Região Norte (N):** Por outro lado, a probabilidade na região Norte foi notavelmente baixa, cerca de 3.30%.
- **Observação:** Destaca-se uma discrepância impressionante entre as regiões, com uma probabilidade extremamente alta na região Sudeste para adultos, enquanto é muito baixa na região Norte.

6. Conclusão

A análise estatística da incidência da Febre Amarela entre 1994 e 2023 revelou padrões significativos que merecem atenção. Inicialmente, a distribuição geográfica dos casos demonstrou que a região Sudeste foi a mais impactada, seguida pela região Norte, evidenciando uma distribuição não uniforme da doença pelo território brasileiro. Além disso, a análise por gênero revelou uma disparidade substancial, com uma prevalência significativamente maior de casos entre os homens. Essa discrepância foi notável nos estados mais afetados, como Minas Gerais e São Paulo.

Quanto às faixas etárias, os dados indicaram uma maior incidência da doença entre indivíduos próximos ou acima dos 40 anos. Embora a análise das probabilidades tenha mostrado uma variação nas chances de contrair a doença entre diferentes faixas etárias e regiões geográficas, a predominância de casos entre os adultos, especialmente na região Sudeste, foi notável. O box plot e as medidas de dispersão proporcionaram uma compreensão mais aprofundada da distribuição das idades afetadas, destacando uma concentração central entre 30 e 55 anos, embora com uma dispersão considerável e a presença de outliers.

Em suma, os resultados apontam para a complexidade da dinâmica da Febre Amarela, influenciada por variáveis geográficas, demográficas e de gênero. Essas descobertas são fundamentais para orientar estratégias de prevenção e controle da doença, destacando a importância da vigilância contínua e de intervenções direcionadas para grupos de maior risco.

6.1 Recomendações

Logo abaixo segue algumas recomendações para o controle e a diminuição da doença no Brasil:

- **Foco Regional:** Concentrar os esforços de prevenção e controle na região Sudeste, com atenção especial a Minas Gerais e São Paulo. Dada a significativa concentração de casos nessas áreas, estratégias direcionadas de vigilância e intervenção podem ser mais eficazes para conter a propagação da Febre Amarela.

- **Campanhas Direcionadas:** Implementar campanhas de conscientização e prevenção específicas para homens e para pessoas na faixa etária de 38 a 45 anos. Dado o padrão observado de maior incidência entre homens e indivíduos nessa faixa etária, abordagens direcionadas podem aumentar a conscientização sobre os riscos e promover comportamentos preventivos.
- **Inclusividade na Vacinação:** Garantir que os programas de vacinação sejam abrangentes, incluindo todas as faixas etárias e ambos os sexos, com prioridade para os grupos mais afetados. Uma abordagem inclusiva na vacinação é crucial para alcançar a imunização adequada da população e reduzir a transmissão da doença. Priorizar os grupos mais vulneráveis identificados nesta análise pode ajudar a maximizar os benefícios das campanhas de vacinação.