Instituto Federal de Educação Ciência e Tecnologia de São Paulo

Câmpus São Paulo

Letícia Gonçalves Baião SP3098818

Victor Lucas Santos da Silva SP3096564

Lucas de Macedo Polezel SP305411X

Análise de casos de Febre Amarela em humanos e primatas não-humanos 1994 a 2023

São Paulo

2024

Letícia Gonçalves Baião
Victor Lucas Santos da Silva
Lucas de Macedo Polezel

Casos Febre Amarela em humanos e primatas não-humanos - 1994 a 2023

Análise e Desenvolvimento de Sistemas 5º Semestre

Trabalho da disciplina de Estatística e Probabilidade apresentado ao curso de Análise e Desenvolvimento de Sistemas do Instituto Federal de São Paulo para a análise de Casos de Febre Amarela em humanos e primatas não-humanos - 1994 a 2023

Orientador: Josceli Maria Tenorio

São Paulo
2024

Sumário

Introdução

Compreender a dinâmica da transmissão da febre amarela e avaliar as medidas de controle são passos cruciais para a formulação de políticas de saúde eficazes. Neste cenário, a utilização de métodos estatísticos robustos é essencial para explorar e interpretar os dados epidemiológicos de forma a obter uma visão clara e precisa da situação.

Estes métodos facilitam a visualização de dados complexos e a extração de informações significativas, que são fundamentais para orientar intervenções de saúde pública. Além disso, a análise estatística contribui para a avaliação da eficácia das medidas de controle implementadas, fornecendo evidências para ajustes e melhorias contínuas. Dessa forma, a estatística não apenas enriquece a compreensão científica da febre amarela, mas também apoia a tomada de decisões estratégicas que visam reduzir o impacto da doença na população.

Descrição sobre a base de dados utilizada

Dicionário de Variáveis

Para a presente análise, foi utilizada a base de dados “Febre Amarela em humanos e primatas não-humanos - 1994 a 2023” disponibilizada pelo DATASUS (Departamento de Informática do Sistema Único de Saúde). A base consiste em dois arquivos csv com informações de casos em humanos, e informações de casos em primatas respectivamente.

- Casos em Humanos

Variável	**Descrição**
ID	Identificador sequencial único
MACRORREG_LPI	Sigla da macrorregião do local provável de infecção
COD-UF_LPI	Código IBGE da Unidade Federada do local provável de infecção
UF_LPI	Sigla da Unidade Federada do local provável de infecção
COD_MUN_LPI	Código IBGE do município do local provável de infecção
MUN_LPI	Nome do município do local provável de infecção
SEXO	Sexo do indivíduo
IDADE	Idade do indivíduo
DT_IS	Data de início dos sintomas do indivíduo (dd/mm/aaaa)
SE_IS	Semana epidemiológica de início dos sintomas do indivíduo
MES_IS	Mês de início dos sintomas do indivíduo
ANO_IS	Ano de início dos sintomas do indivíduo
MONITORAMENTO_IS	Período de monitoramento de início dos sintomas do indivíduo*
OBITO	Evolução para o óbito
DT_OBITO	Data do óbito (dd/mm/aaaa)

- Casos em primatas não-humanos

Variável	Descrição
ID	Identificador sequencial único
MACRORREG_OCOR	Sigla da macrorregião do local de ocorrência
COD_UF_OCOR	Código IBGE da Unidade Federada do local de ocorrência
UF_OCOR	Sigla da Unidade Federada do local de ocorrência
COD_MUN_OCOR	Código IBGE do município do local de ocorrência
MUN_OCOR	Nome do município do local de ocorrência
DATA_OCOR	Data de ocorrência do evento (dd/mm/aaaa)
SE_OCOR	Semana epidemiológica de ocorrência do evento
MES_OCOR	Mês de ocorrência do evento
ANO_OCOR	Ano de ocorrência do evento
MONITORAMENTO_OCOR	Período de monitoramento de ocorrência do evento*

O período de monitoramento corresponde à estratificação temporal dos dados em períodos anuais com início em julho e término em junho. Cada período de monitoramento corresponde a um intervalo de 12 meses, que inclui o segundo semestre de um ano e o primeiro semestre do ano seguinte. Essa representação do componente temporal decorre do reconhecimento de um período sazonal de transmissão, entre dezembro e maio, que concentra a maior parte dos eventos registrados no país, e tem como intuito evitar a análise fragmentada dos processos de transmissão, cujo pico de ocorrência se dá geralmente na transição entre os anos. O Ministério da Saúde adota essa estratificação para fins de políticas de vigilância em saúde e análise epidemiológica.

Fonte de Dados

O Ministério da Saúde, 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), monitora, no âmbito do Sistema Nacional de Vigilância Epidemiológica da Febre Amarela (FA), as notificações de casos suspeitos da doença em humanos em primatas não-humanos (PNH; macacos). As notificações são captadas a partir do Sistema de Informação de Agravos de Notificação (Sinan Net), de instrumentos alternativos de monitoramento dos períodos sazonais (planilhas, formulários eletrônicos), além de comunicações diretas à CGARB e ao Centro de Informações Estratégicas em Vigilância em Saúde (CIEVS). Na vigilância animal, além das fontes referidas acima, as notificações também são captadas pelo Sistema de Informação em Saúde Silvestre (SISS-Geo; www.biodiversidade.ciss.fiocruz.br). A consolidação dos dados se dá sempre à época da vigência do monitoramento anual (julho a junho), uma vez que servem de subsídio à intensificação das ações durante o período sazonal de transmissão da FA (dezembro a maio), com subsequente harmonização dos registros com as Secretarias de Estado da Saúde (SES). A despeito dos esforços para consolidar bases de dados unificadas, erros de registro, dados incompletos ou ausentes e inconsistências requerem revisão e qualificação periódicas dos conjuntos de dados.

Conteúdo da base de dados

As bases de dados disponibilizadas nesta Plataforma referem-se aos dados individualizados e anonimizados (Lei 13.709/2018) de casos humanos e de epizootias em PNH confirmados para FA, com data de início dos sintomas ou de ocorrência entre 1994 e 2023 e com local provável de infecção (LPI) no território nacional. A estratégia de vigilância de PNH foi iniciada em meados de 1999, e um sistema de informação para registro e monitoramento dos eventos (Sinan) só foi implantado em meados de 2007, de modo que parte do período referido não contém dados sobre animais. O conjunto de variáveis disponibilizado corresponde àquele utilizado pela CGARB/DEDT/SVSA/MS para a construção de indicadores epidemiológicos e análises de situação divulgadas em boletins, informes, notas informativas e outras publicações oficiais, bem como para subsidiar a tomada de decisão na gestão pública, a avaliação e estratificação de risco, e a definição de áreas prioritárias para as ações de vigilância e imunização. As variáveis constantes na base de casos humanos são identificação única, macrorregião, código da unidade federada (UF) do local provável de infecção (LPI), UF do LPI, código do município do LPI, município do LPI, sexo, idade, data de início dos sintomas, semana epidemiológica de início dos sintomas, mês de início dos sintomas, ano de início dos sintomas, monitoramento de início dos sintomas, óbito, data do óbito. Os valores totais podem variar entre as variáveis, em função

de valores faltantes e registros incompletos ou ausentes. As variáveis constantes na base de epizootias em PNH são identificação única, macrorregião, código da UF de ocorrência, UF de ocorrência, código do município de ocorrência, município de ocorrência, data de ocorrência, semana epidemiológica de ocorrência, mês de ocorrência, ano de ocorrência, monitoramento de ocorrência. Os valores totais podem variar entre as variáveis, em função de valores faltantes e registros incompletos ou ausentes. Esses dados estão em constante processo de análise e revisão e, portanto, sujeitos a alterações.

Limitações

As bases de dados disponibilizadas não incluem as notificações de suspeitas não confirmadas. O número de casos humanos confirmados pode não coincidir com aquele registrado no Sinan, visto que procedimentos de inclusão, correção de erros e inconsistências e adequação dos registros entre as esferas de gestão são laboriosos e requerem permanente interação em rede. As bases de dados não incluem, por ora, dados sobre município e UF de residência e de notificação, mas incluem dados sobre o LPI, que remetem aos territórios-alvos das ações de vigilância e resposta.

Descrição prévia das análises

- Para as análises que serão apresentadas posteriormente neste documento, estamos lidando com dados do período de 1994 a 2023;
- As análises serão feitas a partir de “Perguntas” feitas à base de dados “Febre Amarela em humanos e primatas não-humanos - 1994 a 2023” disponibilizada pelo DATASUS (Departamento de Informática do Sistema Único de Saúde);
- As análises feitas para humanos e para primatas não-humanos, estão com descrição do público selecionado da análise no enunciado da questão, que se seguirá os resultados;
- As perguntas presentes neste documento podem abranger tanto o conteúdo de estatística descritiva, probabilidade ou inferência;
- O teste de Hipótese está presente na penúltima análise presente neste documento.

Análise dos dados

Bibliotecas utilizadas:

```
install.packages("readr")
install.packages("readxl")
install.packages("ggplot2")
install.packages("dplyr")
install.packages("vcd")

library(readr)
library(readxl)
library(ggplot2)
library(dplyr)
library(readr)
library(vcd)
```

Importando e tratando os dados:

```
dados <- na.omit(dados)

dados$IDADE <- as.numeric(dados$IDADE)

dados$DT_OBITO <- dmy(dados$DT_OBITO)

dados$SEXO <- factor(dados$SEXO, levels = c("M", "F"))
dados$MACRORREG_LPI <- factor(dados$MACRORREG_LPI)
dados$UF_LPI <- factor(dados$UF_LPI, levels = c("AC", "AL", "AP", "AM", "BA",
"CE", "DF", "ES", "GO", "MA", "MT", "MS", "MG", "PA", "PB", "PR", "PE", "PI",
"RJ", "RN", "RS", "RO", "RR", "SC", "SP", "SE", "TO"))
dados$OBITO <- factor(dados$OBITO, levels = c("SIM"))
```

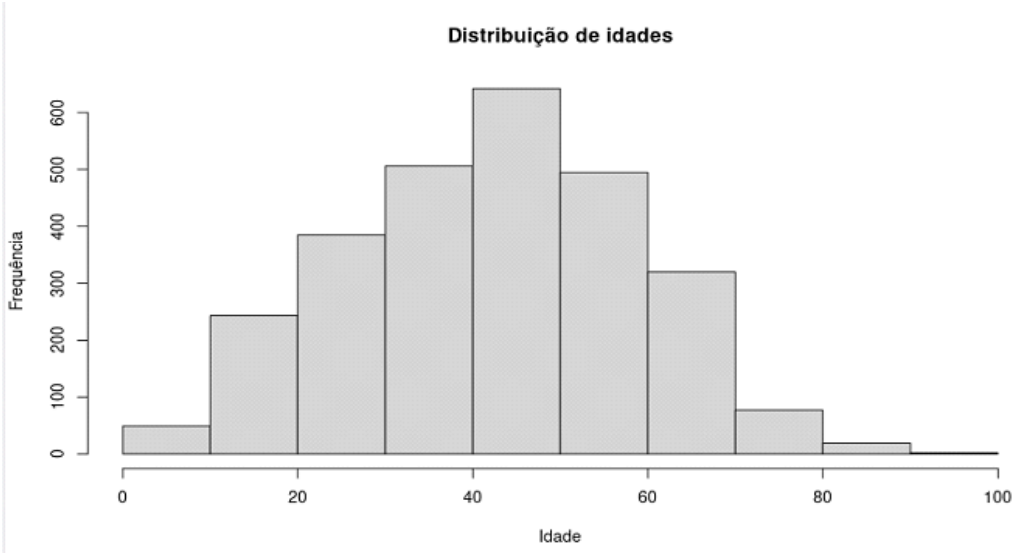
Análise Estatística

- Análise geral das principais colunas do dataset usando summary

```
      ID          MACRORREG_LPI   COD_UF_LPI       UF_LPI      COD_MUN_LPI       MUN_LPI          SEXO
Min.   :   2.0   CO: 146       Min.   :11.00   MG     :1087   Min.   :110020   Length:2738        M:2271
1st Qu.: 702.2   N : 187       1st Qu.:31.00   SP     : 687   1st Qu.:313860   Class :character   F: 467
Median :1396.5   NE:   6       Median :32.00   RJ     : 307   Median :320190   Mode  :character
Mean   :1391.7   S :  62       Mean   :32.55   ES     : 256   Mean   :327894
3rd Qu.:2080.8   SE:2337       3rd Qu.:35.00   GO     :  98   3rd Qu.:352020
Max.   :2768.0                 Max.   :53.00   PA     :  93   Max.   :530010
                                               (Other): 210
     IDADE          DT_IS              SE_IS            MES_IS           ANO_IS     MONITORAMENTO_IS
Min.   : 0.00   Length:2738        Min.   : 1.000   Min.   : 1.000   Min.   :1995   Length:2738
1st Qu.:31.00   Class :character   1st Qu.: 3.000   1st Qu.: 1.000   1st Qu.:2017   Class :character
Median :43.00   Mode  :character   Median : 6.000   Median : 2.000   Median :2018   Mode  :character
Mean   :42.73                      Mean   : 9.356   Mean   : 2.599   Mean   :2015
3rd Qu.:55.00                      3rd Qu.:10.000   3rd Qu.: 3.000   3rd Qu.:2018
Max.   :93.00                      Max.   :53.000   Max.   :12.000   Max.   :2023

  OBITO         DT_OBITO              data
SIM : 996   Min.   :1995-04-09   Min.   :NA
NA's:1742   1st Qu.:2017-01-05   1st Qu.:NA
            Median :2017-12-29   Median :NA
            Mean   :2014-07-16   Mean   :NaN
            3rd Qu.:2018-02-11   3rd Qu.:NA
            Max.   :2023-08-31   Max.   :NA
            NA's   :1772         NA's   :2738
> |
```

- Qual a faixa etária com o maior número de casos de humanos afetados pela febre amarela entre 1994 e 2023, independentemente do resultado do óbito?

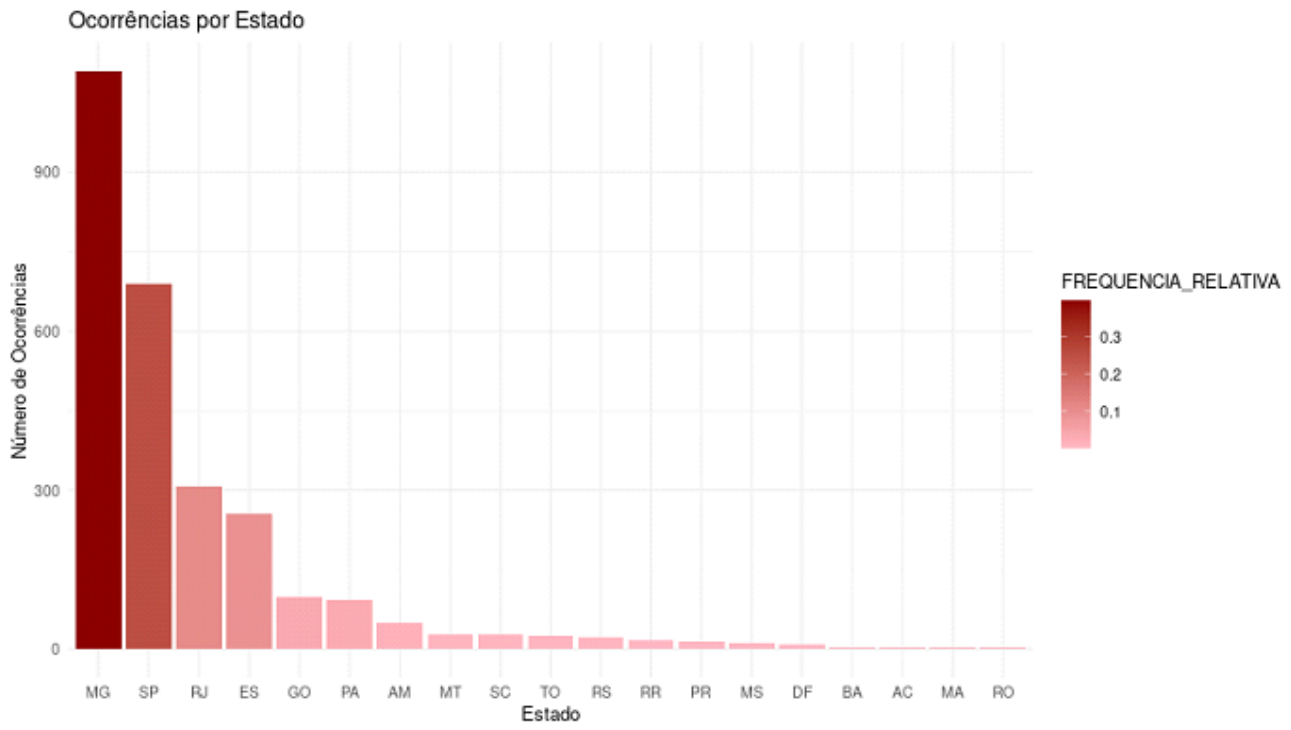
```
hist(dados$IDADE, main = 'Distribuição de idades', xlab = 'Idade', ylab =
'Frequência')
```



A partir dos resultados obtidos no gráfico gerado, é possível observar que a população de humanos mais afetada pela febre amarela se encontra na faixa dos 40 anos de idade.

- Quais estados foram mais impactados pela febre amarela no período de 1994 a 2023?

```
# Separando casos por estado
casos_x_estado <- count(dados, dados$UF_LPI, sort = TRUE)

# Adicionando frequência relativa
casos_x_estado <- casos_x_estado %>%
  mutate(frequencia_relativa = n/sum(n))

names(casos_x_estado) <- c('UF', 'NUM_OCORRENCIAS', 'FREQUENCIA_RELATIVA')

ggplot(casos_x_estado, aes(x = reorder(UF, -NUM_OCORRENCIAS), y =
NUM_OCORRENCIAS, fill = FREQUENCIA_RELATIVA)) +
  geom_bar(stat = 'identity') +
  scale_fill_gradient(low = "lightpink", high = "darkred") +
  labs(title = "Ocorrências por Estado",
       x = "Estado",
       y = "Número de Ocorrências") +
  theme_minimal()
```

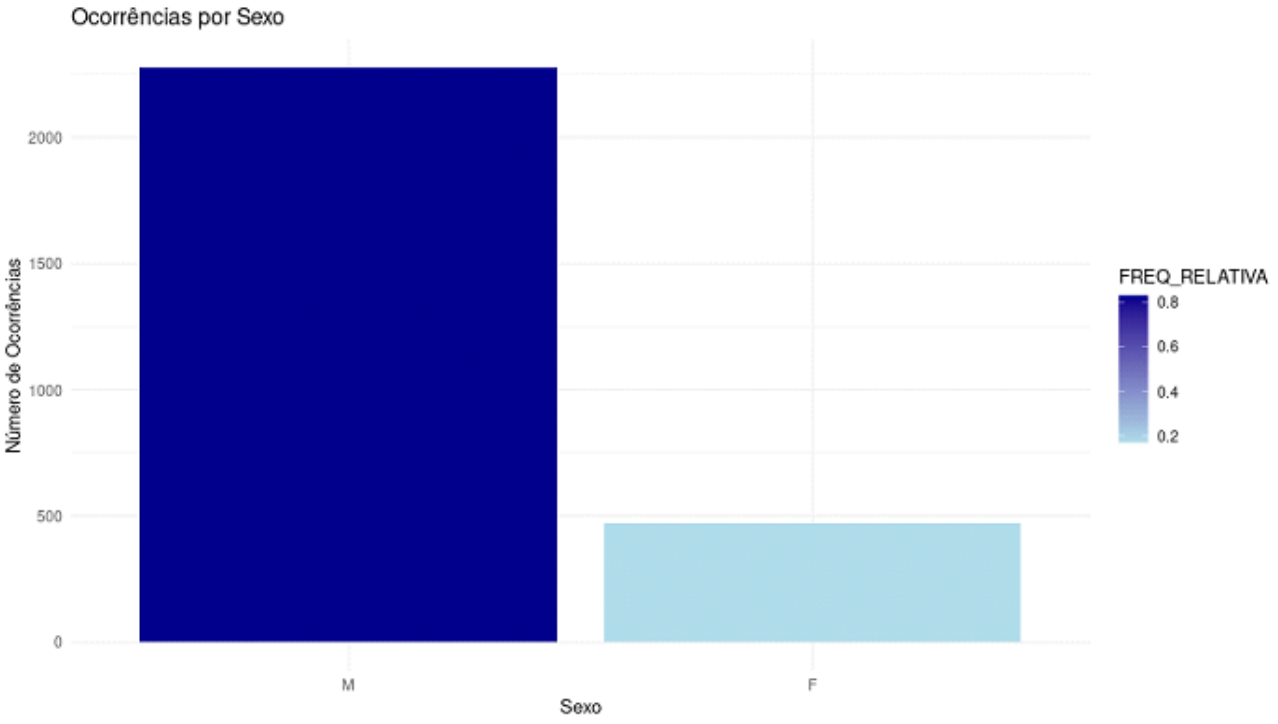


Verificando os resultados obtidos a partir do gráfico, é possível concluir que o estado com maior incidência de humanos afetados pela febre amarela foi Minas Gerais, com mais de 30% dos óbitos, seguido de São Paulo e Rio de Janeiro.

- Qual a proporção de homens e mulheres atingidos pela febre amarela?

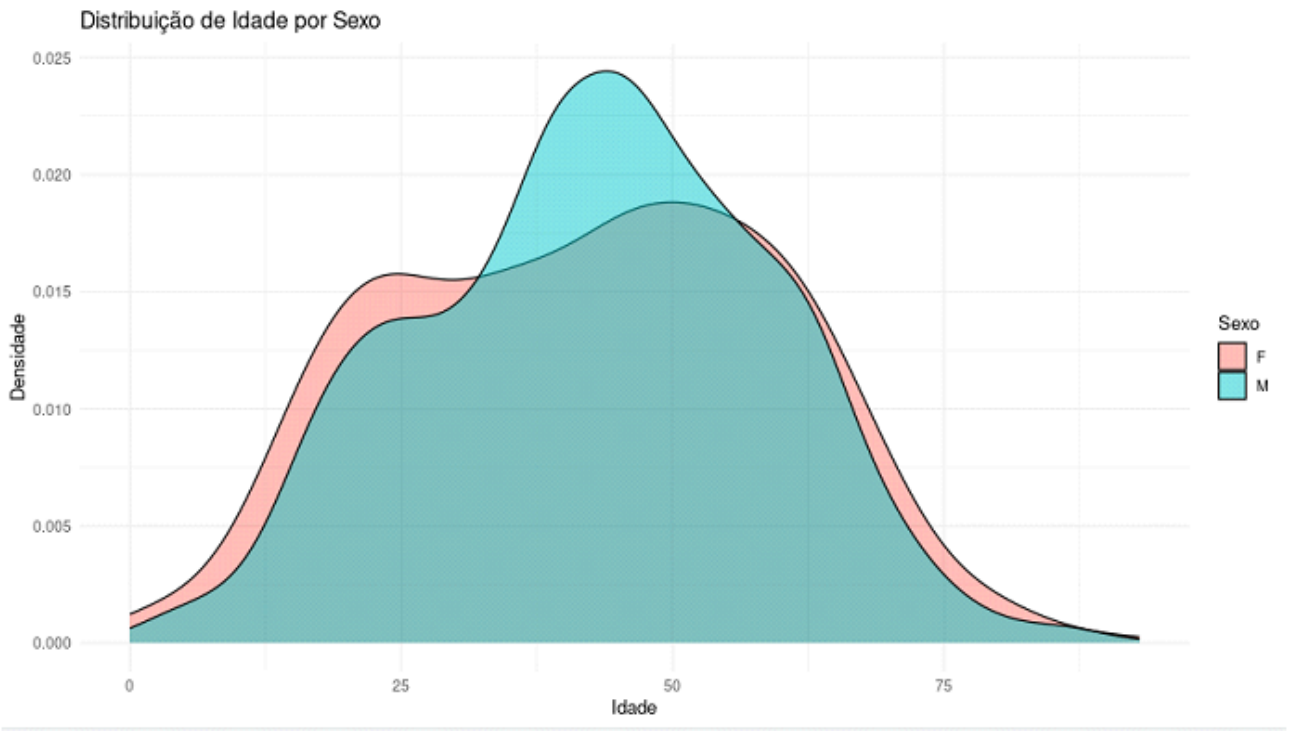
Para essa análise foram feitos dois gráficos, sendo eles:

```
casos_x_sexo <- count(dados, dados$SEXO)

casos_x_sexo <- casos_x_sexo %>%
  mutate(frequencia_relativa = n/sum(n))

names(casos_x_sexo) <- c('SEXO', 'OCORRENCIAS',
'FREQ_RELATIVA')

ggplot(casos_x_sexo, aes(x = reorder(SEXO, -OCORRENCIAS), y =
OCORRENCIAS, fill = FREQ_RELATIVA)) +
  geom_bar(stat = 'identity') +
  scale_fill_gradient(low = "lightblue", high = "darkblue") +
  labs(title = "Ocorrências por Sexo",
       x = "Sexo",
       y = "Número de Ocorrências") +
  theme_minimal()
```



Analisando a frequência de casos a partir do gráfico gerado, é possível verificar que em geral os homens correspondem a aproximadamente 80% dos infectados pela febre amarela.

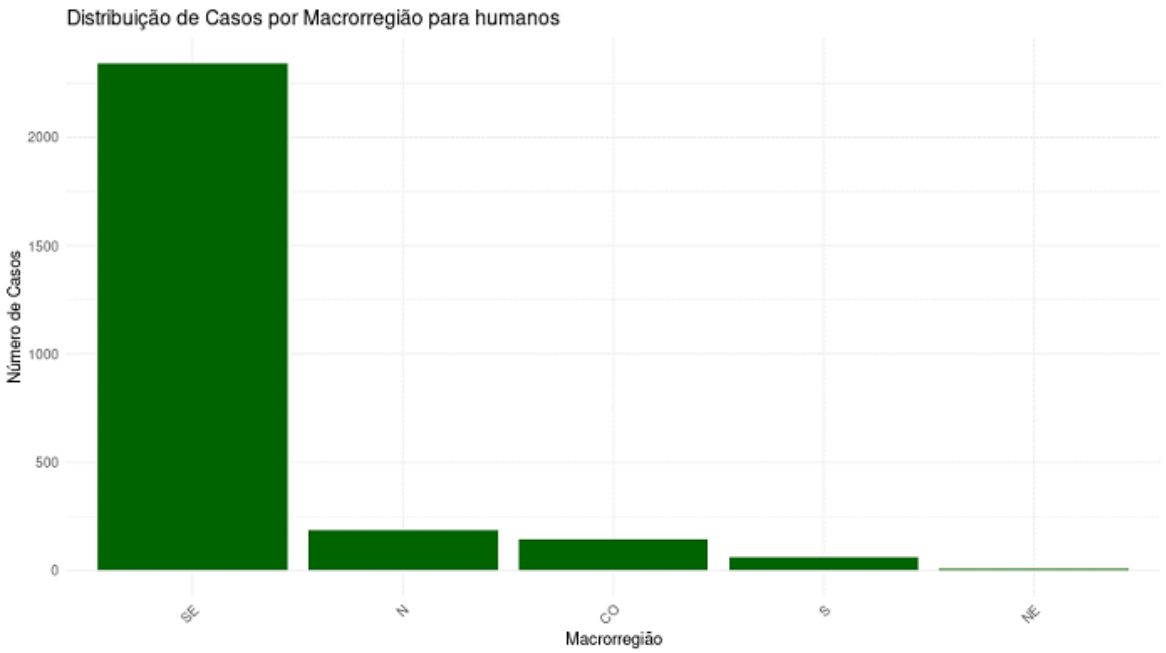
```
ggplot(dados, aes(x = dados$IDADE, fill = dados$SEXO)) +
  geom_density(alpha = 0.5) +
  labs(title = "Distribuição de Idade por Sexo",
       x = "Idade",
       y = "Densidade",
       fill = "Sexo") +
  theme_minimal()
```



Analisando a distribuição da frequência de acordo com a idade temos que apenas por volta da faixa etária dos 40 anos que os homens superam as mulheres no número de infectados. Desta maneira, por essa faixa etária ser a mais atingida conforme analisado no item 1, esta análise corrobora a ideia de que os homens são os mais atingidos pela doença.

- Qual a região do Brasil com maior incidência de humanos afetados pela febre amarela no Brasil?

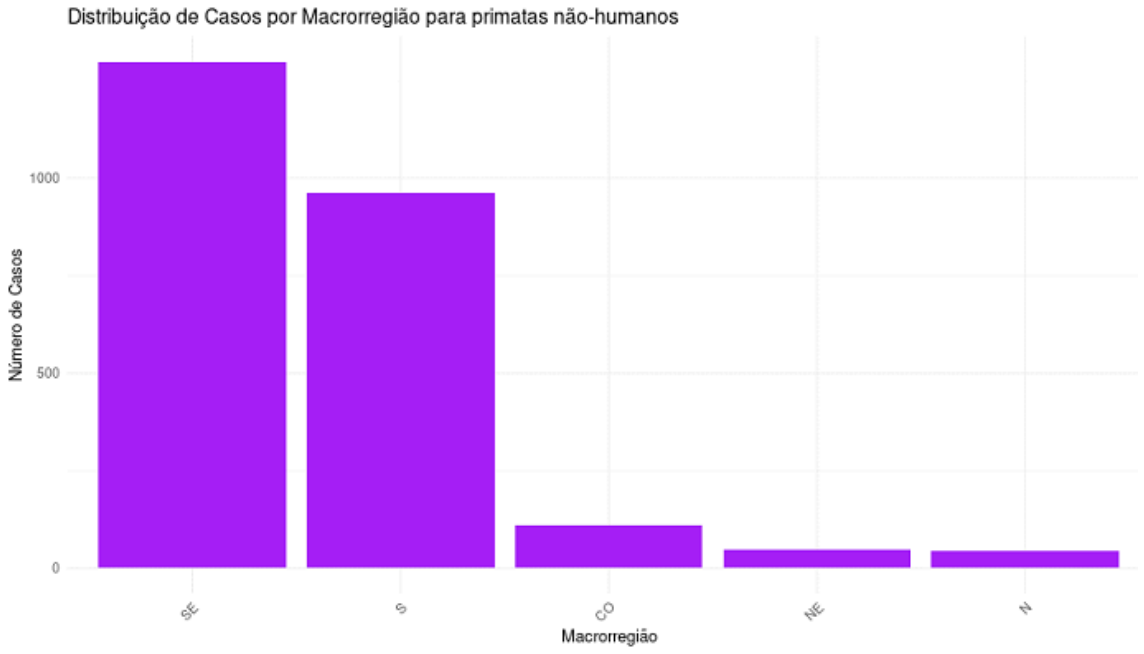
```
casos_por_macrorregiao <-
as.data.frame(table(dados$MACRORREG_LPI))
# Renomear colunas para legibilidade
colnames(casos_por_macrorregiao) <- c("Macrorregião", "Número de
Casos")

# Gráfico de barras para distribuição de casos por macrorregião
ggplot(casos_por_macrorregiao, aes(x = reorder(Macrorregião,
-`Número de Casos`), y = `Número de Casos`)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "Distribuição de Casos por Macrorregião",
       x = "Macrorregião",
       y = "Número de Casos") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



A região com maior incidência de casos de humanos afetados pela febre amarela no período analisado, foi a região Sudeste, composto pelos estados Espírito Santo (ES), Minas Gerais (MG), Rio de Janeiro (RJ) e São Paulo (SP)

- Quantos casos de primatas não humanos afetados pela febre amarela em cada região do Brasil?

```
casos_por_macrorregiao_primatas <-
as.data.frame(table(dadosPrimatas$MACRORREG_OCOR))

#Renomear as colunas para legibilidade
colnames(casos_por_macrorregiao_primatas) <- c("Macrorregião",
"Número de Casos")

#Gráfico de barras para distribuição de casos por macrorregião
ggplot(casos_por_macrorregiao_primatas, aes(x =
reorder(Macrorregião, -`Número de Casos`), y = `Número de Casos`))
+
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "Distribuição de Casos por Macrorregião",
       x = "Macrorregião",
       y = "Número de Casos") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

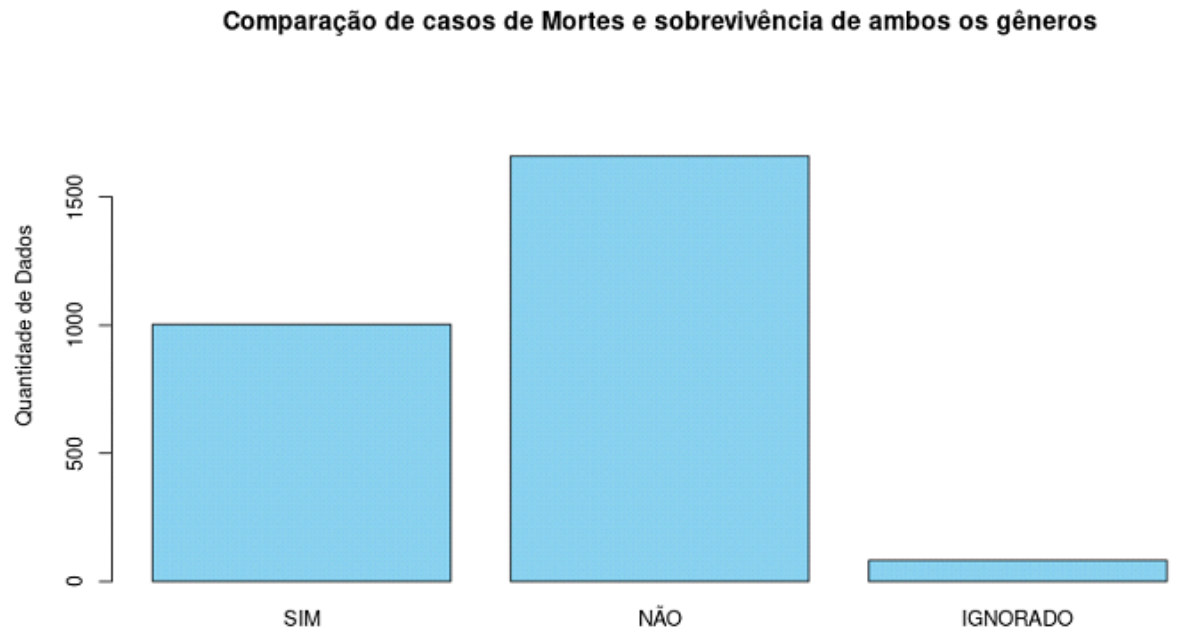


A região com maior incidência de casos de primatas-não humanos afetados pela febre amarela no período analisado, foi a região Sudeste, resultado semelhante à mesma análise

feita com seres humanos. Dessa maneira é possível identificar uma maior incidência de casos nessa região do Brasil para os dois grupos inseridos na análise.

- Qual o número de casos fatais de humanos afetados pela febre amarela, independentemente do gênero?

```
contagem_obito <- table(dados$OBITO)
# Acessar a quantidade de dados na categoria "SIM"
quantidade_sim <- contagem_obito["SIM"]
quantidade_nao <- contagem_obito["N\xc3O"]
quantidade_ignorado <- contagem_obito["IGN"]
# Vetor com os dados
quantidades <- c(quantidade_sim, quantidade_nao,
quantidade_ignorado)
nomes_variaveis <- c("SIM", "NÃO", "IGNORADO")

# Criar o gráfico de barras
barplot(quantidades, names.arg = nomes_variaveis, col = "skyblue",
        main = "Comparação de casos de Mortes e sobrevivência",
        ylab = "Quantidade de Dados", ylim =c(0, max(quantidades)
* 1.2))
```



Verificando o gráfico gerado, é possível concluir que a incidencia de humanos (independente do sexo) com febre amarela, e que vieram a óbito, tem como resultado uma quantidade menor de casos, em comparação com a incidencia de humanos que sobreviveram.

Análise Probabilística

- Qual a probabilidade de uma mulher, afetada pela febre amarela, vir a óbito?

```
#P(A|B)=P(A∩B)/ P(B)
#P(A) = Pessoa vir a óbito
#P(B) = Pessoa com febre amarela ser do sexo feminino
#P(A|B) = pessoa vir a obito e ser do sexo feminino

table(dados$SEXO)
prob_M <- mean(dados$SEXO == "M")
prob_F <- mean(dados$SEXO == "F")
mean(dados$SEXO == "F" & dados$OBITO == "SIM")
```

A partir dos resultados obtidos com a análise feita a partir do grupo de humanos, foi verificado que a probabilidade de uma mulher com febre amarela vir a óbito é aproximadamente 24.09%.

- Qual a probabilidade de um homem, afetado pela febre amarela, vir a óbito?

```
table(dados$SEXO)
prob_M <- mean(dados$SEXO == "M")
mean(dados$SEXO == "M" & dados$OBITO == "SIM")
```

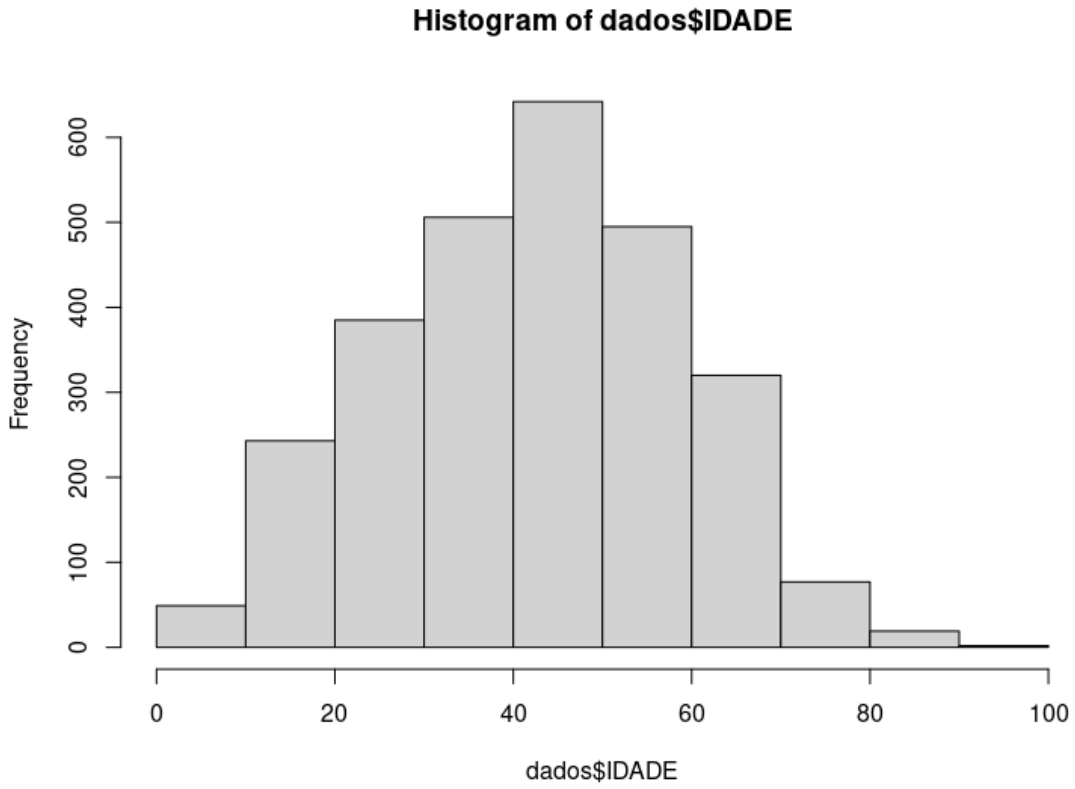
A probabilidade de um homem com febre amarela vir a óbito é aproximadamente 39.12%. Comparando os resultados de ambos os gêneros (Feminino e Masculino), é possível perceber que a probabilidade de óbito é significativamente maior entre os homens (39.12%) do que entre as mulheres (24.09%). Isso indica que homens têm um risco maior de morrer de febre amarela comparado às mulheres.
Em resumo, os homens têm uma probabilidade mais alta de óbito por febre amarela comparado às mulheres. Este resultado pode refletir fatores como maior exposição ao mosquito transmissor devido a atividades ao ar livre, diferenças biológicas, acesso aos cuidados de saúde, ou comportamento de busca de tratamento.
As campanhas de prevenção e controle da febre amarela devem focar mais nos homens, que apresentam um risco maior de mortalidade, já os programas de conscientização devem abordar os fatores que podem estar contribuindo para a maior taxa de mortalidade entre homens, incentivando práticas preventivas e busca precoce por tratamento.

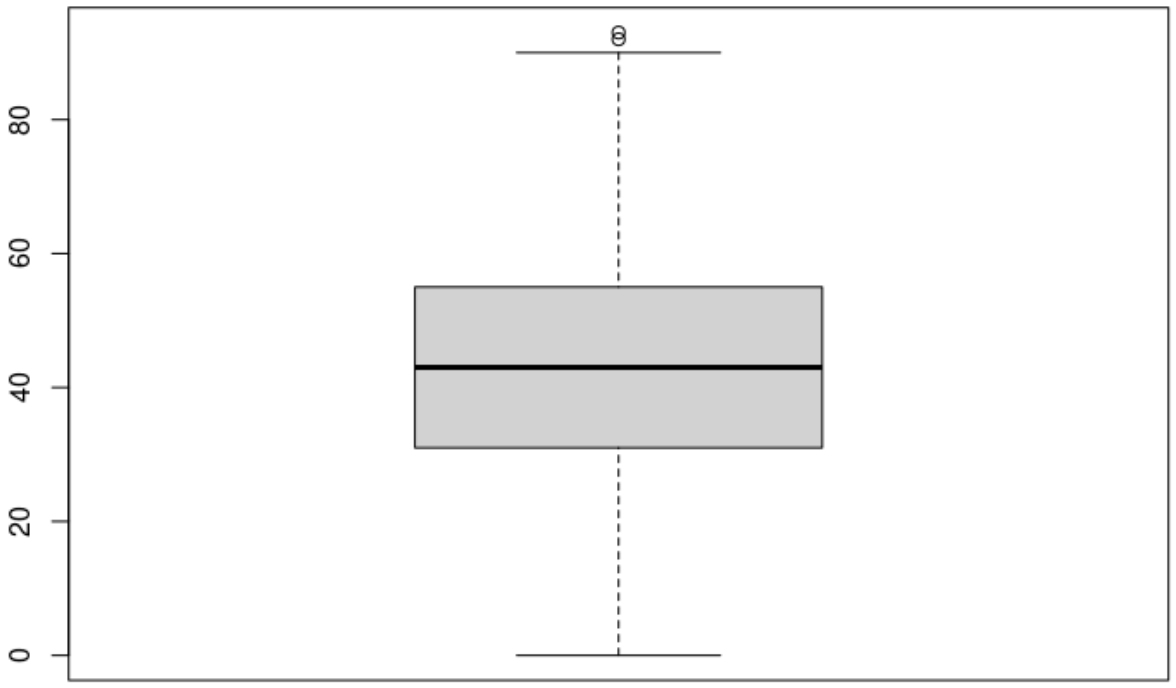
Análise Inferencial

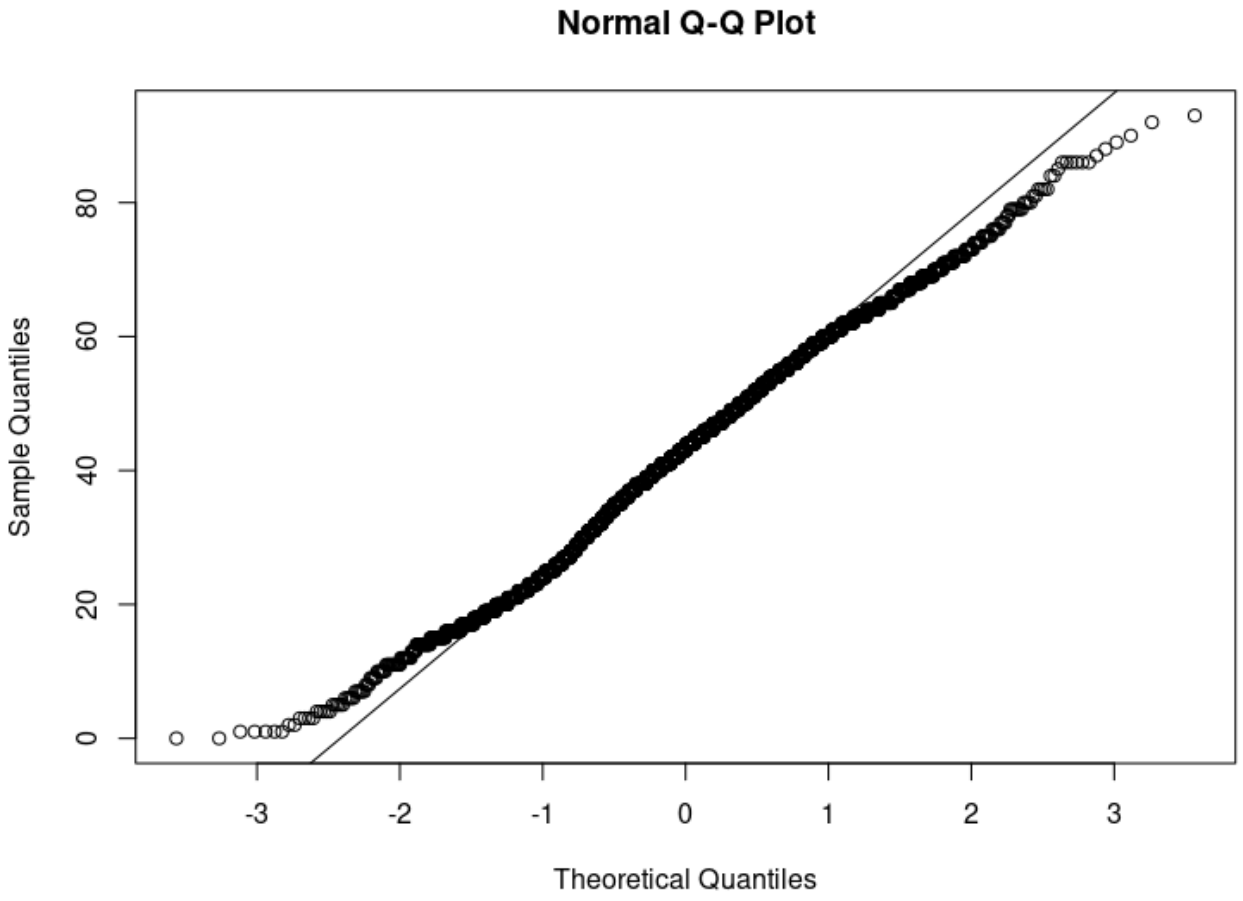
- Há diferença na média de idades entre os humanos afetados pela febre amarela, para ambos os gêneros?

```
##Determinando se a distribuição de idade é normal:
hist(dados$IDADE)
boxplot(dados$IDADE)
qqnorm(dados$IDADE)
qqline(dados$IDADE)
shapiro.test(dados$IDADE)
```

Com base no histograma e no QQplot podemos observar que a distribuição de idades se comporta como semelhantemente a uma distribuição normal e confirmamos isso com o valor de p resultante do teste de Shapiro Wilk, que é p = 0,518. ou seja, p > 0,05. Além disso, com o boxplot podemos ver que não há presença de outliers significativos.




0
20
40
60
80


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
0
20
40
60
80
-3
-2
-1
0
1
2
3

```
t.test(IDADE ~ SEXO, data = dados)
```

***teste de hipótese**

As médias das idades dos grupos femininos e masculinos são muito próximas (42.06 anos para mulheres e 42.86 anos para homens). Além disso, o valor p (0.3718) resultante do teste T é maior do que alpha, dessa maneira, rejeitamos a hipótese e concluímos que a diferença média de idade entre os dois grupos é pequena e não significativa do ponto de vista estatístico.

- Há diferença na proporção de casos e óbitos por gênero para humanos?

```
#Tabela de contingência para a variável OBITO em relação à SEXO
table(dados$OBITO, dados$SEXO)
assoc_measure <- assocstats(table(dados$OBITO, dados$SEXO))
assoc_measure$chisq

#ESTADO
table(dados$OBITO, dados$UF_LPI)

assoc_measure <- assocstats(table(dados$OBITO, dados$UF_LPI))
assoc_measure$chisq

#Faixa_etaria
table(dados$OBITO, dados$IDADE)
assoc_measure <- assocstats(table(dados$OBITO, dados$IDADE))
assoc_measure$chisq

df <- as.data.frame(assoc_measure)
df$IDADE <- rownames(df)
```

A tabela de contingência apresenta a distribuição dos desfechos de febre amarela por gênero (F: Feminino, M: Masculino):

OBITO \ SEXO	F	M
IGNORADO	20	62
NÃO	336	1324
SIM	113	890

```
chisq.test(data$SEXO, data$OBITO, data = data)
```

O teste Pearson Chi-Square resultou em um valor p extremamente baixo (próximo de zero), indicando que a relação entre o desfecho de óbito e o gênero é altamente significativa. Isso significa que a distribuição dos óbitos em relação ao gênero não é aleatória.

Referências Bibliográficas

Febre Amarela em humanos e primatas não-humanos - 1994 a 2023 - OPENDATASUS. Disponível em: <https://opendatasus.saude.gov.br/dataset/febre-amarela-em-humanos-e-primatas-nao-humanos>. Acesso em: 15 jun. 2024.