



Estatística e Probabilidade - ESP1A5
Professora: Josceli

NATHAN NEVES PRATES - SP3094839

Análise Síndrome Respiratória Aguda
Grave
SRAG 2021

São Paulo
Junho/2024

1. SUMÁRIO

1. SUMÁRIO.....	1
2. INTRODUÇÃO.....	3
2.1. OBJETIVO.....	3
2.2.1. Origem e Desenvolvimento.....	3
2.2.2. Período Abrangido.....	3
2.2.3. Conteúdo da Base de Dados.....	3
2.2.4. Tratamento e Qualidade dos Dados.....	4
2.2.5. Finalidade.....	4
3. ANÁLISE INICIAL.....	5
3.1. CAMPOS DE INTERESSE.....	5
3.2. SANITIZAÇÃO DOS CAMPOS.....	5
3.3. SUMMARY.....	6
4. HIPÓTESES.....	8
4.1. HIPÓTESE 1.....	9
4.1.1. Definição das Variáveis de Interesse.....	9
4.1.2. Preparação e Limpeza dos Dados.....	9
4.1.3. Exploração Inicial dos Dados.....	10
4.1.4. Teste de Associação.....	11
4.1.5. Interpretação e conclusão.....	11
4.2. HIPÓTESE 2:.....	12
4.2.1. Definição das Variáveis de Interesse.....	12
4.2.2. Preparação e Limpeza dos Dados.....	12
4.2.3. Exploração Inicial dos Dados.....	13
4.2.4. Teste de Associação.....	14
4.2.5. Interpretação e conclusão.....	14
4.3. HIPÓTESE 3.....	15

4.3.1. Definição das Variáveis de Interesse.....	15
4.3.2. Preparação e Limpeza dos Dados.....	15
4.3.3. Exploração Inicial dos Dados.....	16
4.3.4. Teste de Associação.....	17
4.3.5. Interpretação e conclusão.....	17
4.4. HIPÓTESE 4.....	18
4.4.1. Definição das Variáveis de Interesse.....	18
4.4.2. Preparação e Limpeza dos Dados.....	18
4.4.3. Exploração Inicial dos Dados.....	19
4.4.4. Teste de Associação.....	20
4.4.5. Interpretação e conclusão.....	20

2. INTRODUÇÃO

2.1. OBJETIVO

Este relatório tem como objetivo realizar uma análise detalhada dos dados de Síndrome Respiratória Aguda Grave (SRAG) no Brasil para o ano de 2021, com especial enfoque na incidência e impacto da COVID-19, uma vez que esta doença foi incorporada à vigilância de SRAG desde o início da pandemia. A análise busca identificar padrões epidemiológicos, avaliar a eficácia das medidas de vigilância e monitoramento implementadas, e fornecer insights que possam contribuir para o aprimoramento das estratégias de saúde pública no enfrentamento de pandemias e outras emergências sanitárias.

2.2. SOBRE A BASE DE DADOS

A base de dados utilizada nesta análise é proveniente do Ministério da Saúde (MS) do Brasil, especificamente desenvolvida pela Secretaria de Vigilância em Saúde (SVS). Esta base integra a vigilância da Síndrome Respiratória Aguda Grave (SRAG), uma iniciativa que começou durante a pandemia de Influenza A(H1N1)pdm09 e foi posteriormente ampliada para incluir outros vírus respiratórios, incluindo a COVID-19.

2.2.1. Origem e Desenvolvimento

A vigilância de SRAG foi inicialmente implementada na rede de vigilância de Influenza e outros vírus respiratórios. Em 2020, com o surgimento da COVID-19, a vigilância desta nova infecção foi incorporada ao sistema existente.

O Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP-Gripe) é o sistema oficial para o registro de casos e óbitos por SRAG.

2.2.2. Período Abrangido

A base de dados disponibiliza informações desde 2009 até os dias atuais (2024), com dados específicos para o ano de 2021 sendo o foco desta análise.

2.2.3. Conteúdo da Base de Dados

Inclui registros detalhados de casos e óbitos por SRAG, abrangendo diversas informações epidemiológicas.

Os dados incluem informações sobre definições de casos, critérios de confirmação e encerramento dos casos, conforme o Guia de Vigilância Epidemiológica Emergência de Saúde Pública de Importância Nacional pela Doença pelo Coronavírus 2019.

2.2.4. Tratamento e Qualidade dos Dados

Os dados são sujeitos a alterações decorrentes de investigações adicionais ou correções de erros de digitação pelas equipes de vigilância epidemiológica.

As bases de dados passam por um processo de tratamento que inclui a anonimização, em conformidade com a Lei 13.709/2018, assegurando a privacidade dos indivíduos.

2.2.5. Finalidade

A base de dados visa fornecer um legado epidemiológico abrangente e atualizado da vigilância de SRAG no Brasil, auxiliando na análise histórica e na formulação de estratégias de saúde pública.

Esta base de dados é essencial para entender a evolução e o impacto das síndromes respiratórias graves no Brasil, fornecendo uma plataforma para análises detalhadas que podem informar políticas e práticas de saúde pública.

3. ANÁLISE INICIAL

3.1. CAMPOS DE INTERESSE

Abaixo está uma lista dos campos, referentes ao paciente, que são de interesse para a análise

- CS_SEXO (Sexo);
- CS_GESTANT (Gestante);
- AVE_SUINO (Contato com aves/suínos/outros);
- FEBRE (Febre);
- TOSSE (Tosse);
- DISPNEIA (Dispneia - dificuldade respiratória);
- DESC_RESP (Desconforto respiratório);
- SATURACAO (Saturação oxigênio menor que 95%);
- PERD_OLFT (Perda de olfato);
- FATOR_RISC (Fatores de risco);
- IMUNODEPRE (Imunodeficiência ou imunodepressão);
- OBESIDADE (Paciente obeso);
- HOSPITAL (Paciente internado);
- UTI (Internado em UTI);
- EVOLUCAO (Evolução do caso);

3.2. SANITIZAÇÃO DOS CAMPOS

Todos os campos que foram analisados passaram por um processo de sanitização, limpando seus dados e transformando seus tipos. Exemplo:

```
dados_considerados <- dados_considerados %>%  
  mutate(EVOLUCAO = recode(EVOLUCAO,  
    `1` = 'cura',  
    `2` = 'óbito',  
    `3` = 'óbito-outras-causas',  
    `9` = 'ignorado'))
```

```
dados_considerados <- dados_considerados %>%  
  mutate(EVOLUCAO = as.factor(EVOLUCAO))
```

3.3. SUMMARY

```
> summary(dados_considerados$CS_SEXO)
      F      I      M
788311  283 942696
> summary(dados_considerados$CS_GESTANT)
1-trimestre      1708
2-trimestre      4453
3-trimestre      9979
Idade-gestacional-ignorada      857
ignorado          74736
não              541951
nao-aplica       1097399
NA's             207
> summary(dados_considerados$AVE_SUINO)
ignorado      não      sim      NA's
 256437  1134436   15466   324951
> summary(dados_considerados$FEBRE)
ignorado      não      sim      NA's
 21021   495240   904156   310873
> summary(dados_considerados$TOSSE)
ignorado      não      sim      NA's
 16830   316673  1163153   234634
> summary(dados_considerados$DISPNEIA)
ignorado      não      sim      NA's
 14838   293422  1197747   225283
> summary(dados_considerados$DESC_RESP)
ignorado      não      sim      NA's
 19631   418932   943729   348998
> summary(dados_considerados$SATURACAO)
ignorado      não      sim      NA's
 19952   335922  1086479   288937
> summary(dados_considerados$PERD_OLFT)
ignorado      não      sim      NA's
 45214   959929   138694   587453
> summary(dados_considerados$FATOR_RISC)
      não      sim
764985  966305
> summary(dados_considerados$IMUNODEPRE)
ignorado      não      sim      NA's
 13488   534876   35664   1147262
```

```

> summary(dados_considerados$OBESIDADE)
ignorado      não      sim      NA's
  15334    468165    135627    1112164
> summary(dados_considerados$HOSPITAL)
ignorado      não      sim      NA's
   3042    31554  1650346    46348
> summary(dados_considerados$UTI)
ignorado      não      sim      NA's
  34317   955216   507700   234057
> summary(dados_considerados$EVOLUCAO)
Cura                1071423
Óbito                438872
Óbito-outras-causas      22768
Ignorado              43065
NA's                  155162

```


4. HIPÓTESES

Baseando-se nos dados apresentados, uma série de hipóteses foi formulada. Para comprovar ditas hipóteses, um procedimento e 5 passos foi adotado:

- 1° **Definição das Variáveis de Interesse:** Definimos quais das variáveis serão utilizadas para comprovar a hipótese e descrevemos o que elas representam;
- 2° **Preparação e Limpeza dos Dados:** Transformamos os dados contidos nas tabelas em dados legíveis e sensíveis, caso haja a necessidade;
- 3° **Exploração Inicial dos Dados:** Analisamos individualmente as variáveis a fim de compreender melhor sua importância e contexto;
- 4° **Teste de Associação:** Realizamos a associação das variáveis de interesse e anotamos os resultados, realizando testes que indicam correlação significativa entre elas;
- 5° **Interpretação e Conclusão:** Interpretamos os resultados baseando-se nos testes estatísticos e análises realizadas, interpretando se há evidências suficientes para apoiar ou rejeitar sua hipótese inicial. Após isto, geramos uma conclusão referente a hipótese, confirmatória ou não, com base nos resultados da análise estatística.

O 4° (quarto) passo, envolve testar se há uma associação significativa entre as variáveis. Para garantir que haja esta correlação, utilizamos o teste estatístico qui-quadrado. Como referência de interpretação dos dados que serão obtidos deste teste, segue uma descrição das variáveis e seu significado:

- **p-value:** O valor-p é a probabilidade de obter um valor pelo menos tão extremo quanto o observado, assumindo que a hipótese nula (de que não há associação entre as variáveis) é verdadeira. Um valor-p muito pequeno (menor que qualquer nível de significância comum, como 0.05) indica que é extremamente improvável que a associação observada seja devida ao acaso.
- **X-squared (X^2):** Esse é o valor do teste qui-quadrado. Ele indica a magnitude da diferença observada entre as frequências esperadas e observadas na tabela de

contingência. Valores maiores sugerem uma maior discrepância entre as frequências observadas e esperadas.

- **df (Degrees of Freedom):** Esse é o número de graus de liberdade do teste. É calculado com base no número de categorias das variáveis. Para uma tabela de contingência $r \times c$, os graus de liberdade são $(r - 1) \times (c - 1)$.

4.1. HIPÓTESE 1

Pacientes com saturação de oxigênio menor que 95% têm maior probabilidade de serem internados em UTI.

4.1.1. Definição das Variáveis de Interesse

- **SATURACAO:** Nível de saturação de oxigênio menor que 95% (sim / não / ignorado) .
- **UTI:** Indicador de internação em UTI (sim / não / ignorado) .

4.1.2. Preparação e Limpeza dos Dados

```
dados_considerados <- dados_considerados %>%  
  mutate(CS_GESTANTE = recode(STATURACAO,  
    `1` = 'sim',  
    `2` = 'não',  
    `9` = 'ignorado'))
```

```
dados_considerados <- dados_considerados %>%  
  mutate(STATURACAO = as.factor(STATURACAO))
```

```
dados_considerados <- dados_considerados %>%  
  mutate(UTI = recode(UTI,  
    `1` = 'sim',  
    `2` = 'não',  
    `9` = 'ignorado'))
```

```
dados_considerados <- dados_considerados %>%  
  mutate(UTI = as.factor(UTI))
```

4.1.3. Exploração Inicial dos Dados

Exploramos a distribuição dos dados.

```
> summary(dados_considerados$UTI)
ignorado      não      sim      NA's
   34317   955216   507700   234057
> summary(dados_considerados$SATURACAO)
ignorado      não      sim      NA's
   19952   335922  1086479   288937
```

Comparamos a incidência de internação em UTI com a saturação do paciente

Comando:

```
table(
dados_considerados$UTI,
dados_considerados$SATURACAO
)
```

Saida:

	ignorado	não	sim
ignorado	2452	7208	19142
não	9122	225492	594369
sim	4909	71815	365920

4.1.4. Teste de Associação

Para testar se há uma associação significativa entre ter saturação baixa e a internação em UTI, utilizamos o teste estatístico qui-quadrado.

Comando:

```
contingency_table <- table(dados_considerados$UTI,  
dados_considerados$SATURACAO)  
chi_sq <- chisq.test(contingency_table)  
chi_sq
```

Saída:

```
Pearson's Chi-squared test  
  
data:  contingency_table  
X-squared = 31997, df = 4, p-value < 2.2e-16
```

Interpretando os valores obtidos, “p-value” do teste qui-quadrado indica se há uma associação significativa entre possuir saturação baixa e ser internado em uma UTI. Um valor-p (2.2×10^{-16}) menor que 0.05 sugere que há evidências significativas para rejeitar a hipótese nula de que não há associação.

4.1.5. Interpretação e conclusão

Rejeição da Hipótese Nula: Existe uma associação estatisticamente significativa entre UTI (indicador de internação em UTI) e SATURACAO (indicador de saturação de oxigênio no sangue do paciente). Os resultados evidenciam que a probabilidade de ser internado em uma UTI está associada à baixa saturação.

Hipótese Alternativa: Como a hipótese nula foi rejeitada, podemos aceitar a hipótese alternativa de que pessoas com saturação baixa possuem uma chance maior de desenvolver sintomas respiratórios graves em comparação às pessoas com saturação normal.

4.2. HIPÓTESE 2:

Gestantes têm maior probabilidade de desenvolver sintomas respiratórios graves em comparação com mulheres não grávidas.

4.2.1. Definição das Variáveis de Interesse

- CS_GESTANTE: Indicador de se a paciente está gestante (sim/não/ignorado).
- DESC_RESP: Variável que pode indicar a presença de sintomas respiratórios graves, como desconforto respiratório (sim/não/ignorado).

4.2.2. Preparação e Limpeza dos Dados

```
dados_considerados <- dados_considerados %>%
  mutate(CS_GESTANTE = recode(CS_GESTANTE,
    `1` = '1-trimestre',
    `2` = '2-trimestre',
    `3` = '3-trimestre',
    `4` = 'Idade-gestacional-ignorada',
    `5` = 'não',
    `6` = 'não aplica',
    `9` = 'ignorado'))

dados_considerados <- dados_considerados %>%
  mutate(CS_GESTANTE = as.factor(CS_GESTANTE))

dados_considerados <- dados_considerados %>%
  mutate(DESC_RESP = recode(DESC_RESP,
    `1` = 'sim',
    `2` = 'não',
    `9` = 'ignorado'))

dados_considerados <- dados_considerados %>%
  mutate(DESC_RESP = as.factor(DESC_RESP))
```

4.2.3. Exploração Inicial dos Dados

Exploramos a distribuição dos dados.

```
> summary(dados_considerados$DESC_RESP)
ignorado      não      sim      NA's
   19631   418932   943729   348998
> summary(dados_considerados$CS_GESTANT)
1-trimestre      1708
2-trimestre      4453
3-trimestre      9979
Idade-gestacional-ignorada  857
ignorado          74736
não              541951
nao-aplica       1097399
NA's             207
```

Comparamos a incidência de sintomas respiratórios graves entre gestantes e não gestantes.

Comando:

```
table(
dados_considerados$CS_GESTANT,
dados_considerados$DESC_RESP
)
```

Saida:

	ignorado	não	sim
1-trimestre	10	708	651
2-trimestre	43	1503	2127
3-trimestre	98	4445	3572
idade-gestacional-ignorada	10	249	311
ignorado	2213	15503	33610
não	5041	131367	301913
nao-aplica	12215	265106	601454

4.2.4. Teste de Associação

Para testar se há uma associação significativa entre ser gestante e a presença de sintomas respiratórios graves, utilizamos o teste estatístico qui-quadrado.

Comando:

```
contingency_table <- table(dados_considerados$CS_GESTANT,  
dados_considerados$DESC_RESP)  
chi_sq <- chisq.test(contingency_table)  
Chi_sq
```

Saída:

```
Pearson's Chi-squared test  
  
data:  contingency_table  
X-squared = 6208.6, df = 12, p-value < 2.2e-16
```

Interpretando os valores obtidos, “p-value” do teste qui-quadrado indica se há uma associação significativa entre ser gestante e a presença de sintomas respiratórios graves. Um valor-p ($2.2 \cdot 10^{-16}$) menor que 0.05 sugere que há evidências significativas para rejeitar a hipótese nula de que não há associação.

4.2.5. Interpretação e conclusão

Rejeição da Hipótese Nula: Existe uma associação estatisticamente significativa entre CS_GESTANTE (indicador de gestante) e DESC_RESP (indicador de sintomas respiratórios graves). Em outras palavras, os resultados fornecem forte evidência de que a probabilidade de desenvolver sintomas respiratórios graves está associada ao estado de gravidez.

Hipótese Alternativa: Como a hipótese nula foi rejeitada, podemos aceitar a hipótese alternativa de que gestantes têm uma probabilidade diferente (nesse caso, potencialmente maior) de desenvolver sintomas respiratórios graves em comparação com mulheres não grávidas.

4.3. HIPÓTESE 3

Pacientes com perda de olfato têm uma evolução clínica menos grave.

4.3.1. Definição das Variáveis de Interesse

- **PERD_OLFT:** Sintoma de perda de olfato (sim / não / ignorado) .
- **EVOLUCAO:** Indica a evolução e conclusão de um caso de SRAG (cura / óbito / óbito-outras-causas / ignorado)

4.3.2. Preparação e Limpeza dos Dados

```
dados_considerados <- dados_considerados %>%  
  mutate(EVOLUCAO = recode(EVOLUCAO,  
    `1` = 'cura',  
    `2` = 'obito',  
    `3` = 'obito-outras-causas',  
    `9` = 'ignorado'))  
  
dados_considerados <- dados_considerados %>%  
  mutate(EVOLUCAO = as.factor(EVOLUCAO))  
  
dados_considerados <- dados_considerados %>%  
  mutate(PERD_OLFT = recode(PERD_OLFT,  
    `1` = 'sim',  
    `2` = 'não',  
    `9` = 'ignorado'))  
  
dados_considerados <- dados_considerados %>%  
  mutate(PERD_OLFT = as.factor(PERD_OLFT))
```


4.3.3. Exploração Inicial dos Dados

Exploramos a distribuição dos dados.

```
> summary(dados_considerados$EVOLUCAO)
Cura                1071423
Óbito              438872
Óbito-outras-causas 22768
Ignorado           43065
NA's               155162
> summary(dados_considerados$PERD_OLFT)
ignorado      não      sim      NA's
  45214    959929  138694  587453
```

Comparamos a incidência de perda de olfato com a evolução do caso

Comando:

```
table(
dados_considerados$EVOLUCAO,
dados_considerados$PERD_OLFT
)
```

Saida:

	ignorado	não	sim
cura	23999	620723	94834
óbito	15207	241138	29914
óbito-outras-causas	882	14179	470
ignorado	1059	20187	3108

4.3.4. Teste de Associação

Para testar se há uma associação significativa entre a perda de olfato e a evolução do caso, utilizamos o teste estatístico qui-quadrado.

Comando:

```
contingency_table <- table(dados_considerados$EVOLUCAO,  
dados_considerados$PERD_OLFT)  
chi_sq <- chisq.test(contingency_table)  
Chi_sq
```

Saida:

```
Pearson's Chi-squared test  
  
data:  contingency_table  
X-squared = 4575.4, df = 6, p-value < 2.2e-16
```

Interpretando os valores obtidos, “p-value” do teste qui-quadrado indica se há uma associação significativa entre possuir perda de olfato e a evolução do caso. Um valor-p (2.2×10^{-16}) menor que 0.05 sugere que há evidências significativas para rejeitar a hipótese nula de que não há associação.

4.3.5. Interpretação e conclusão

Rejeição da Hipótese Nula: Existe uma associação estatisticamente significativa entre EVOLUCAO (indicador da conclusão do caso) e PERD_OLF (sintoma de perda de olfato). Os resultados evidenciam que a evolução do caso está associada à perda de olfato.

Hipótese Alternativa: Como a hipótese nula foi rejeitada, podemos aceitar a hipótese alternativa de que pessoas com perda de olfato possuem uma chance maior de concluírem o quadro da infecção de maneira menos grave.

4.4. HIPÓTESE 4

A obesidade está associada a um aumento na taxa de mortalidade em pacientes hospitalizados com infecções respiratórias.

4.4.1. Definição das Variáveis de Interesse

- **OBESIDADE:** Indica se o paciente é obeso - consequentemente, de um grupo de risco (sim / não / ignorado) .
- **EVOLUCAO:** Indica a evolução e conclusão de um caso de SRAG (cura / óbito / óbito-outras-causas / ignorado)

4.4.2. Preparação e Limpeza dos Dados

```
dados_considerados <- dados_considerados %>%  
  mutate(EVOLUCAO = recode(EVOLUCAO,  
    `1` = 'cura',  
    `2` = 'obito',  
    `3` = 'obito-outras-causas',  
    `9` = 'ignorado'))
```

```
dados_considerados <- dados_considerados %>%  
  mutate(EVOLUCAO = as.factor(EVOLUCAO))
```

```
dados_considerados <- dados_considerados %>%  
  mutate(OBESIDADE = recode(OBESIDADE,  
    `1` = 'sim',  
    `2` = 'não',  
    `9` = 'ignorado'))
```

```
dados_considerados <- dados_considerados %>%  
  mutate(OBESIDADE = as.factor(OBESIDADE))
```

4.4.3. Exploração Inicial dos Dados

Exploramos a distribuição dos dados.

```
> summary(dados_considerados$EVOLUCAO)
Cura                1071423
Óbito              438872
Óbito-outras-causas 22768
Ignorado           43065
NA's               155162
> summary(dados_considerados$OBESIDADE)
ignorado      não      sim      NA's
15334      468165  135627  1112164
```

Comparamos a incidência de obesidade com a evolução do caso

Comando:

```
table(
  dados_considerados$EVOLUCAO,
  dados_considerados$OBESIDADE
)
```

Saida:

	ignorado	não	sim
cura	7862	278933	76214
óbito	5285	145804	48037
óbito-outras-causas	347	9924	1070
ignorado	409	8240	2168

4.4.4. Teste de Associação

Para testar se há uma associação significativa entre a perda de olfato e a evolução do caso, utilizamos o teste estatístico qui-quadrado.

Comando:

```
contingency_table <- table(dados_considerados$EVOLUCAO,  
dados_considerados$OBESIDADE)  
chi_sq <- chisq.test(contingency_table)  
Chi_sq
```

Saida:

```
Pearson's Chi-squared test  
  
data: contingency_table  
X-squared = 2086.2, df = 6, p-value < 2.2e-16
```

Interpretando os valores obtidos, “p-value” do teste qui-quadrado indica se há uma associação significativa entre ser obeso e evoluir o caso de forma negativa. Um valor-p (2.2×10^{-16}) menor que 0.05 sugere que há evidências significativas para rejeitar a hipótese nula de que não há associação.

4.4.5. Interpretação e conclusão

Rejeição da Hipótese Nula: Existe uma associação estatisticamente significativa entre EVOLUCAO (indicador da conclusão do caso) e OBESIDADE (que indica se o paciente é de grupo de risco de obesidade). Os resultados evidenciam que a evolução do caso está associada à obesidade.

Hipótese Alternativa: Como a hipótese nula foi rejeitada, podemos aceitar a hipótese alternativa de que pessoas obesas possuem uma chance maior de concluir o quadro da infecção com óbito.