

**INSTITUTO FEDERAL DE EDUCAÇÃO,  
CIÊNCIA E TECNOLOGIA DE SÃO PAULO  
ANÁLISE E DESENVOLVIMENTO DE SISTEMAS**

Cawan Carvalho e Silva | SP3068854

Davi Cavalcanti Queiróz Rêgo | SP3069851

Wilson da Conceição Brandão Costa | SP3076377

**PROJETO ESP1A5 - ANÁLISE EXPLORATÓRIA DE DADOS:  
VACINAÇÃO CONTRA A COVID-19**

São Paulo

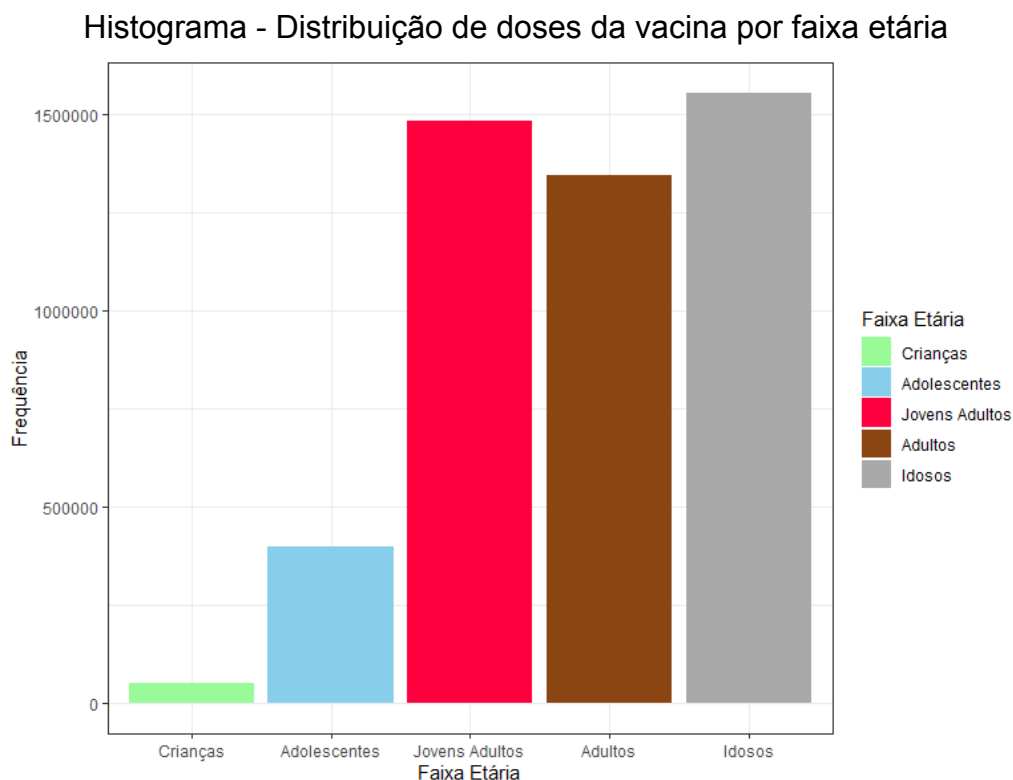
2023

## RELATÓRIO

No ano de 2021, foram iniciadas as vacinações contra o vírus SARS-CoV-2, que provocou a crise sanitária global de Covid-19 em 2020. Neste estudo serão analisados os dados disponibilizados na plataforma DATASUS sobre a vacinação na unidade federativa Distrito Federal do Brasil em 2021, a fim de entender a aderência da população a imunização contra o vírus com base em características demográficas como faixa etária, raça/cor e sexo.

A vacinação contra Covid-19 no Distrito Federal teve início em 19 de janeiro de 2021 e, até o final daquele ano, quase 5 milhões de doses foram aplicadas. Neste relatório serão abordados os resultados da análise descritiva e o teste da hipótese *“Há uma relação entre a faixa etária do vacinado e a fabricante da vacina contra a COVID-19 aplicada”* feitos em cima dos dados dessas doses.

### Distribuição etária



Fonte: Os Autores

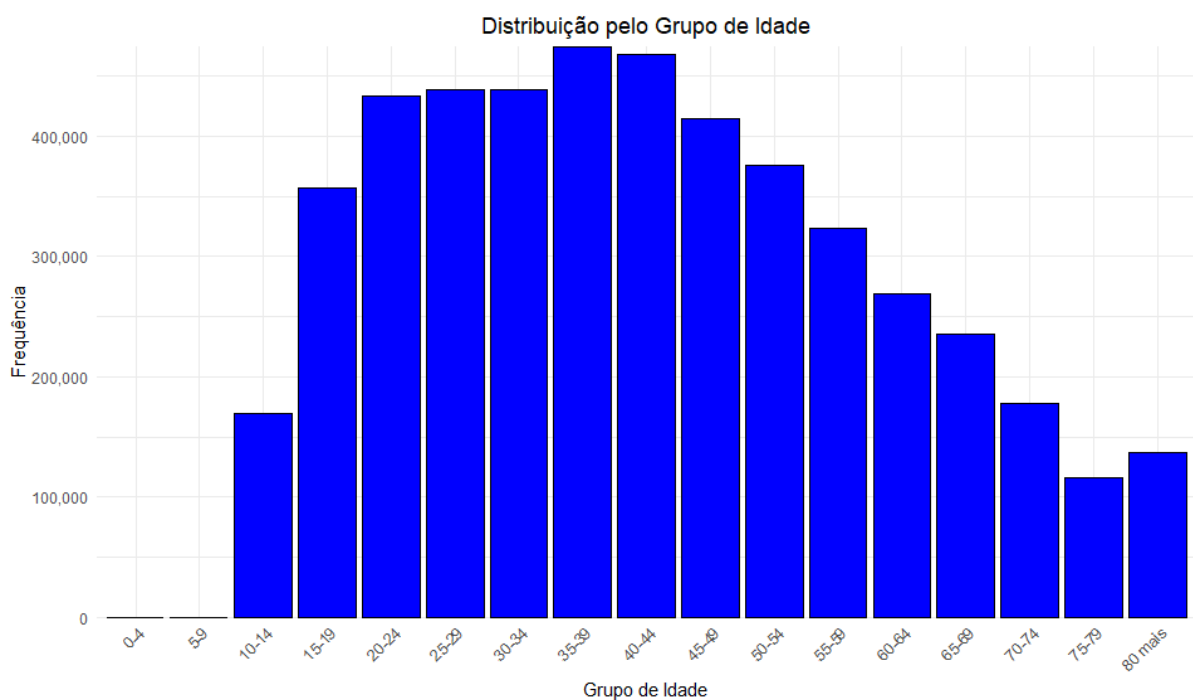
No gráfico de distribuição de doses da vacina por faixa etária apresentado acima, nós temos um filtro nos dados agrupando os vacinados em 5 diferentes fases da vida:

- Crianças (abaixo dos 12 anos de idade);
- Adolescentes (entre os 12 e 17 anos de idade);
- Jovens Adultos (entre os 18 e os 34 anos de idade);
- Adultos (entre os 35 e os 49 anos de idade);
- Idosos (acima dos 50 anos de idade).

O primeiro ponto a se destacar sobre o gráfico, é que os dados contemplados nessa análise compreendem apenas as doses distribuídas no ano de 2021, e a vacinação de crianças só foi iniciada em fevereiro do ano seguinte, o que justifica a baixa adesão desse público.

Também podemos destacar a diferença entre o número de doses distribuídas para adultos e jovens adultos, onde, apesar da liberação tardia da vacina, os jovens adultos possuem números mais elevados.

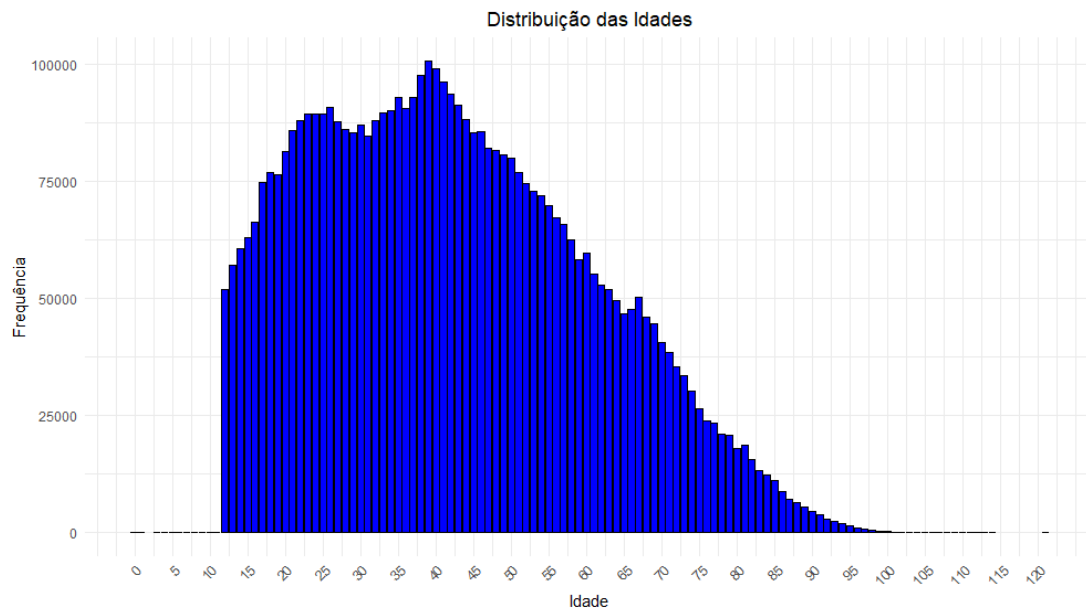
Histograma: Distribuição de Idade por Grupo



Fonte: Os Autores

O gráfico mostra uma visão mais abrangente da base de dados, agrupando os vacinados em setores de 5 anos.

## Histograma: Distribuição por idade

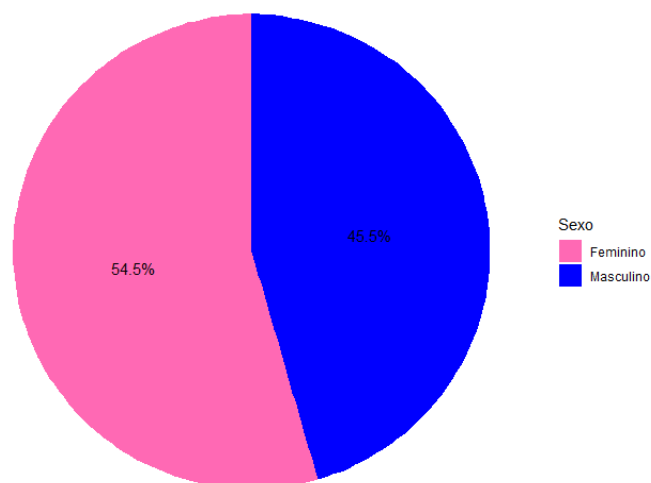


Fonte: Os Autores

O último gráfico apresenta a quantidade de doses distribuídas por idade individualmente.

## Distribuição por sexo

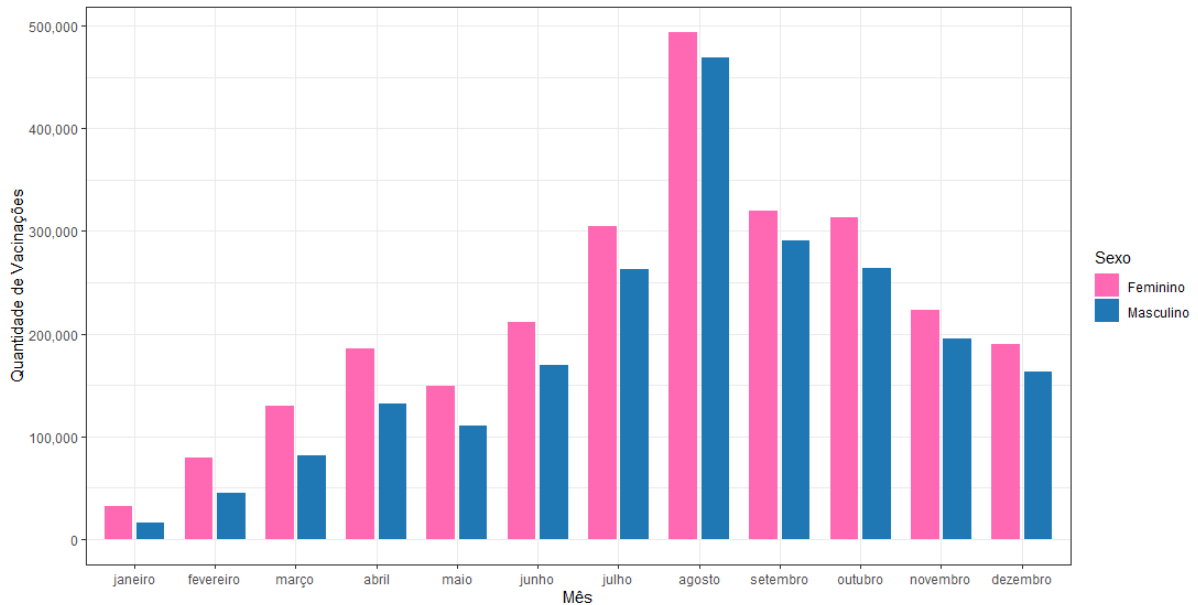
### Histograma: Distribuição por Sexo



Fonte: Os Autores

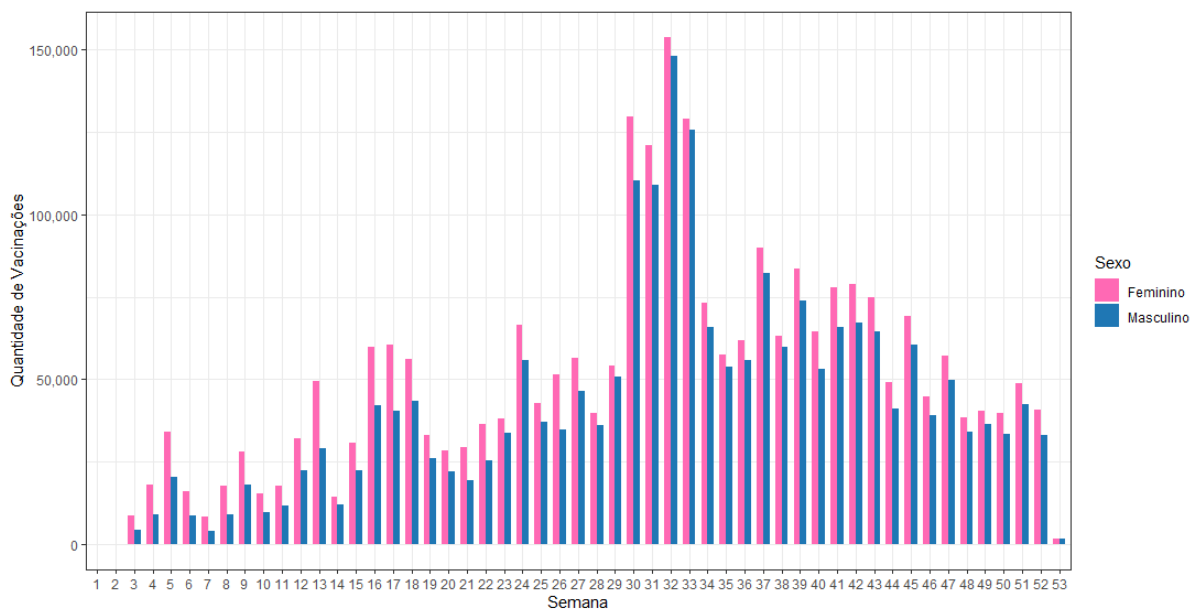
O gráfico acima apresenta a proporção entre doses distribuídas para homens e mulheres, onde notamos uma participação mais expressiva das mulheres na imunização contra o vírus.

Histograma: Quantidade de Vacinas X Sexo (Mensal)



Fonte: Os Autores

Histograma: Quantidade de Vacinas X Sexo (Semanal)

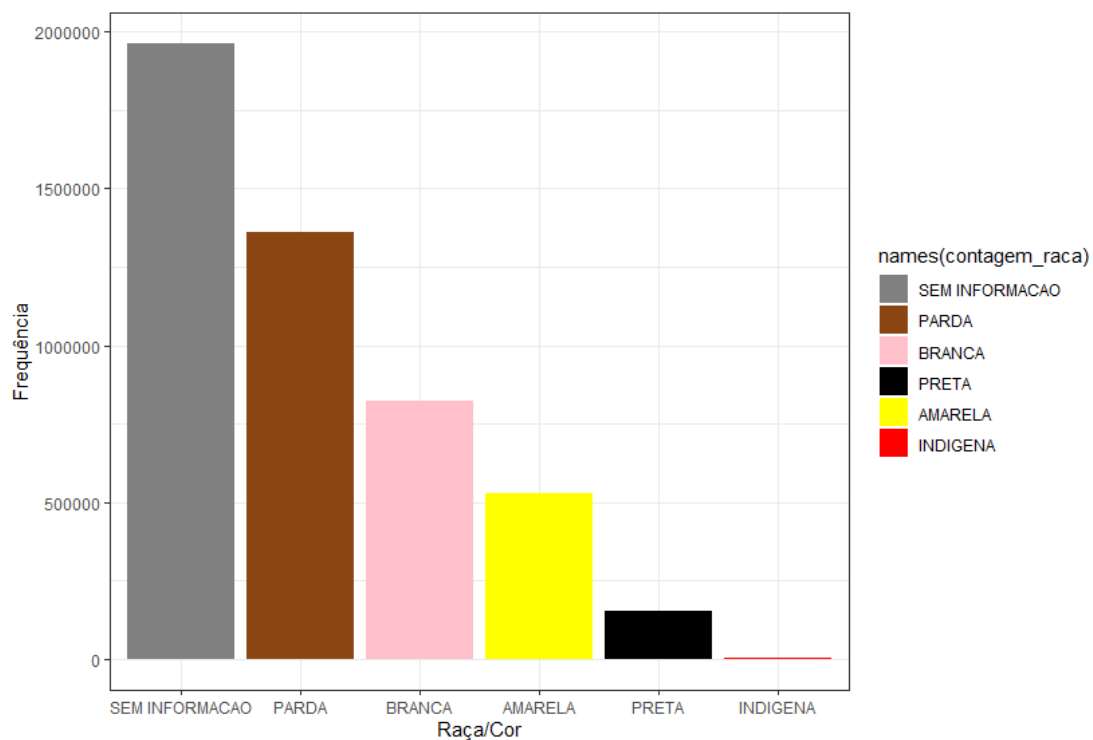


Fonte: Os Autores

Os dois gráficos acima exibem a distribuição de doses mensalmente e semanalmente para homens e mulheres, onde percebemos um fenômeno interessante, onde o sexo feminino predomina em todos os períodos determinados, principalmente nos primeiros meses, podendo apontar uma adesão mais rápida das mulheres aos imunizantes.

## Análise da raça/cor

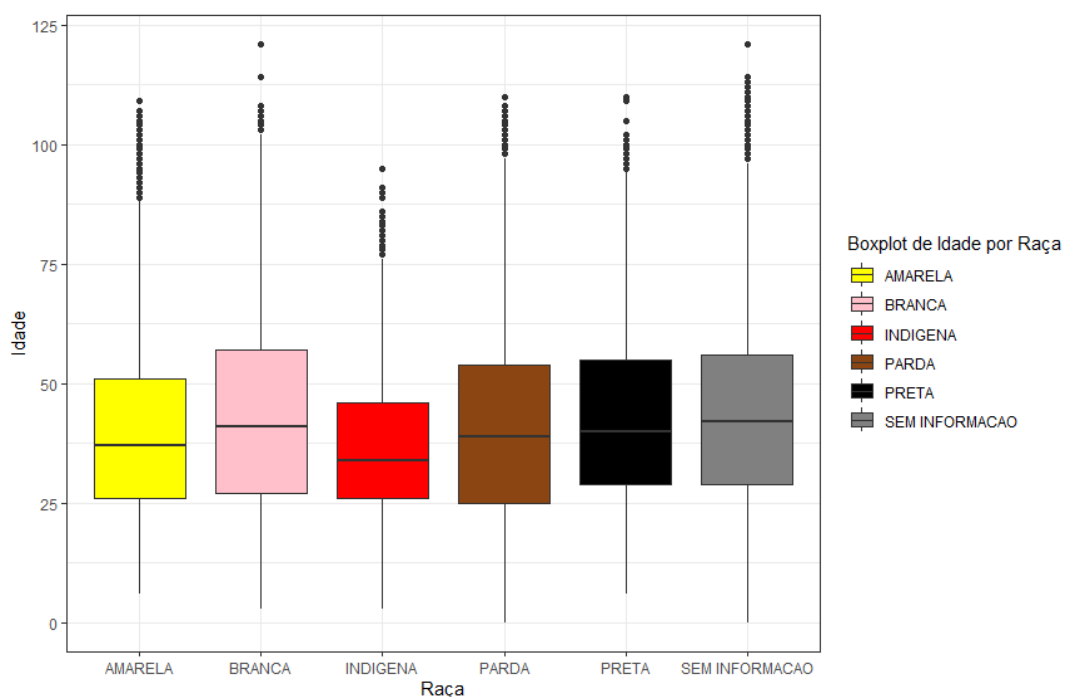
Histograma: Doses X Raça/Cor



Fonte: Os Autores

A análise anterior de caracterização dos vacinados do Distrito Federal é representada pelo gráfico acima e ela mostra a relação entre quantidades de vacinas distribuídas e a raça/cor do vacinado. Nesse histograma, o interessante a ser demonstrado é a quantidade de indivíduos que não estão classificados no espectro racial do país.

Boxplot: Distribuição de Vacinas Aplicadas por Idade X Raça



Fonte: Os Autores

Para uma visualização diferente da relação entre doses distribuídas para diferentes raças/cores, o boxplot acima relaciona, também, a idade dos vacinados, exibindo a mediana, os quartis e os outliers das idades para cada raça/cor.

Tabela: Análise Probabilística Dose X Raça/Cor

Análise Probabilística - Dose x Raça/Cor					
Dose	Preta	Amarela	Branca	Parda	Indígena
1ª	5,42%	19,04%	27,62%	47,83%	0,07%
2ª	5,36%	18,44%	28,87%	47,44%	0,07%
Reforço	4.99%	14.94%	35.95%	44.07%	0.05%

Fonte: Os Autores

Acima vemos uma tabela com a probabilidade de que uma dose em cada um dos estágios tenha sido aplicada em um indivíduo de cada uma das cinco raças/cores, o cálculo da probabilidade foi realizado dividindo-se o número de ocorrências da raça

em questão na dose, pelo número total de vacinações na mesma dose, excluindo-se as vacinações em que a raça não estava informada. Sobre essa tabela, é interessante notar como ao mesmo tempo em que, em relação a primeira e segunda dose, a raça branca subiu em percentual na dose de reforço, todas as demais tiveram seu percentual diminuído individualmente.

### **TESTES DE HIPÓTESE - Há uma relação entre a faixa etária do vacinado e a fabricante da vacina contra a COVID-19 aplicada**

Nesta seção, será abordado o processo de teste da hipótese para determinar se a idade de um indivíduo influencia na marca da vacina que ele recebeu. A partir da base disponibilizada, foram realizadas análises estatísticas e cálculos utilizando a linguagem R para investigar essa associação.

Primeiramente, após a importação da base de dados foi criado um *dataframe* contendo os dados completos. Em seguida, foi ajustada as variáveis "paciente\_grupo\_idade" e "vacina\_codigo" para o tipo categórico utilizando a função *factor()*. Essa etapa foi importante para garantir que as variáveis fossem tratadas como categóricas, facilitando a análise estatística posterior. Após o ajuste das variáveis, foram removidos os registros com valores nulos usando a função *complete.cases()*. Essa etapa garantiu que os dados utilizados para o teste de hipótese estivessem completos e não apresentassem valores faltantes.

Em seguida, foi criada uma tabela de contingência chamada "TESTE\_contingencia" para investigar a relação entre as variáveis categóricas "paciente\_grupo\_idade" e "vacina\_codigo". Essa tabela foi construída utilizando a função *table()*, que contou as ocorrências de cada combinação de valores das duas variáveis. A tabela de contingência mostrou a distribuição das marcas de vacina em cada faixa etária, sendo as marcas das vacinas representadas pelos códigos:

- 85 - Astrazeneca/Fiocruz - Covishield
- 86 - Sinovac/Butantan - Coronavac
- 87 - Pfizer - Comirnaty
- 88 - Janssen - Ad26.Cov2.S
- 89 - Astrazeneca - Chadox1-S



A tabela de contingência "TESTE\_contingencia" mostra a quantidade de indivíduos em cada faixa etária que receberam cada uma das marcas de vacina (códigos 85, 86, 87, 88 e 89) e apresentou os seguintes resultados.

Tabela de Contingência - Grupos de Idade X Marcas Vacinas

	85	86	87	88	89
10-14 anos	33	90	169532	9	3
15-19 anos	10952	44131	296710	1539	3460
20-24 anos	45461	94597	272882	7411	13076
25-29 anos	52149	89896	272385	10212	14000
30-34 anos	57391	89314	270614	11969	9494
35-39 anos	145623	124803	173323	16412	13719
40-44 anos	193051	85409	159812	15865	13677
45-49 anos	142868	79736	169442	9676	13144
50-54 anos	189430	51653	117888	3903	13003
55-59 anos	190649	30445	88325	1772	12064
60-64 anos	192754	18148	51164	217	6585
65-69 anos	58670	113584	60519	101	2029
70-74 anos	5330	120520	51923	63	300
75-79 anos	40202	34725	32831	19	7824
80 mais	57546	40147	38272	23	905

Fonte: Os Autores

Após a criação da tabela de contingência, foram removidas as colunas "98" e "99", que provavelmente representavam categorias irrelevantes para o teste de hipótese. Além disso, foram excluídas as linhas "0-4 anos" e "5-9 anos", devido à insuficiência de dados nessas faixas etárias.

Em seguida, foi realizado o teste qui-quadrado de independência utilizando a função `chisq.test()`. Esse teste avalia se a associação observada entre as variáveis é estatisticamente significativa. O resultado desse teste é apresentado a seguir:

Resultado qui-quadrado de independência

Pearson's Chi-squared test

```
data: TESTE_contingencia
X-squared = 1598776, df = 56, p-value < 2.2e-16
```

Fonte: Os Autores

Os resultados do teste qui-quadrado de independência mostram que existe uma forte associação entre a idade dos indivíduos e a marca da vacina contra COVID-19 que eles receberam durante o ano de 2021. O valor extremamente baixo do p-value ( $< 2.2e-16$ ) indica que a associação observada não ocorre por acaso. Em outras palavras, a probabilidade de que a associação entre idade e marca da vacina seja resultado do acaso é extremamente baixa.

Esses resultados sugerem que a idade teve uma influência significativa na escolha da marca da vacina contra COVID-19 durante o ano de 2021. No entanto, é importante ressaltar que o teste qui-quadrado de independência não fornece informações sobre a direção ou magnitude dessa associação. Portanto, embora possamos concluir que a idade está associada à marca da vacina, não podemos determinar se a idade influencia na escolha da marca ou se a escolha da marca é influenciada pela idade.

É necessário considerar que outros fatores, como disponibilidade das vacinas em determinadas faixas etárias e políticas de distribuição podem estar contribuindo para essa associação observada. Porém, com base nos resultados do teste qui-quadrado de independência, concluímos que a idade dos indivíduos apresentou uma associação estatisticamente significativa com a marca da vacina contra COVID-19, essa associação pode estar relacionada aos efeitos colaterais e a efetividade das vacinas, que podem influenciar na escolha dessas diferentes faixas etárias.