O cenário da Tuberculose no Brasil: uma análise da doença que ainda assola o país

Henrique Diniz¹ - 89373, Nádia Oliveira¹ - 5162, Natália Camillo¹ - 5201

¹Universidade Federal de Viçosa

1. Introdução

A tuberculose é uma doença que tem diminuído significativamente, porém é ainda considerada uma das doenças transmissíveis mais perigosas do mundo (CAZABON et al., 2017), sendo que o Brasil é um dos países onde a carga de tuberculose ainda é bastante alta.

A tuberculose (TB) é uma doença infecciosa potencialmente grave que afeta principalmente os pulmões. A bactéria que causa a tuberculose é transmitida de pessoa para pessoa por meio de pequenas gotículas liberadas no ar por meio de tosses e espirros. A TB se torna ainda mais grave devido a possibilidade de acometer outros órgãos e sistemas (CLINIC, 2021).

A forma extrapulmonar - que atinge outras áreas - porém, é mais comum em pacientes com sistema imunológico comprometido, como no caso de pessoas que vivem com HIV, já que enfraquece o sistema imunológico de uma pessoa, esta não pode combater os germes da tuberculose (CLINIC, 2021).

Embora seu corpo possa abrigar a bactéria que causa a tuberculose, seu sistema imunológico geralmente pode evitar que você adoeça. Por esse motivo, os médicos fazem uma distinção entre: TB latente no qual existe uma infecção de TB, mas as bactérias no corpo são inativas e não causam sintomas, não é contagiosa porém pode virar TB ativa; TB ativa deixa a pessoa doente e é contagiosa.

Por conta de fatores como os aspectos socioculturais, estrutura política, serviços de saúde que podem variar consideravelmente entre os estados brasileiros, a diminuição na incidência de tuberculose tem sido lenta (CAZABON et al., 2017; CORTEZ et al., 2021). O presente trabalho tem como objetivo poder contribuir para o conjunto de evidências disponíveis na literatura sobre a tuberculose e seus indicadores durantes os anos.

2. Motivação

A tuberculose ainda é a doença infecciosa que mais mata no mundo todo (USP, 2020). Com o avanço da pandemia, os compromissos que foram assumidos para eliminar a tuberculose como problema de saúde estão sendo ameaçados. Segundo Cruz (2021), a pandemia eliminou doze anos de progressos na luta global contra a tuberculose.

Em relação ao Brasil, o país continua entre os 30 países de alta carga para tuberculose e para coinfecção TB-HIV, sendo, portanto, considerado prioritário para o controle da doença no mundo pela Organização Mundial de Saúde (OMS) (SAÚDE, 2021).

Apesar da COVID-19 estar sendo a principal preocupação, outras doenças transmissíveis como a tuberculose, não devem ser negligenciadas. O intuito do estudo é analisar a doença no Brasil, quais fatores pudessem ter aumentado ou diminuído os casos e

quais indicadores sofreram com quedas, aumento ou ajustes durante a pandemia e suas respectivas razões.

3. Escolha dos dados e planejamento

3.1. Conjunto de Dados

Os dados utilizados neste projeto foram obtidos do TabNet - DATASUS ¹, um banco de dados disponibilizado pelo departamento de informática do Sistema Único de Saúde do Brasil. Entre os dados disponibilizados, encontramos informações de saúde (indicadores de saúde, assistência à saúde, informações epidemiológicas e de morbidade, informações sobre a rede de assistência à saúde, estatísticas vitais, informações demográficas e socioeconômicas) e informações financeiras (referentes aos recursos do Fundo Nacional de Saúde transferidos aos municípios, aos créditos aos prestadores de serviços de saúde, aos orçamentos públicos de saúde declarados pelos Estados, pelo Distrito Federal e pelos Municípios).

O único *dataset* o qual não foi possível encontrar no TabNet, foi o da População Privada de Liberdade, que foi retirado de Infopen (2017). Os restantes dos conjuntos de dados, foram retirados do TabNet e estes foram:

- Casos de TB.
- Cobertura Vacinal de TB.
- Imunização de TB.
- Casos Novos de TB.
- Casos Novos confirmados por laboratório.
- Casos Novos de TB da PPL (População Privada de Liberdade).
- Casos Novos de TB da PR (População de Rua).
- Casos Novos de TB de Profissionais de Saúde.
- Casos Novos de TB de Imigrantes.
- Casos Novos Raça Branca, Parda, Preta, Amarela e Indígena, respectivamente.
- Casos Novos de TB Drogarresistente.
- Casos de Óbitos por TB.
- Casos Novos de TB encerrados em CURA.
- Casos Novos de TB encerrados em ABANDONO.
- Casos Novos de TB em 2020 (Meses)
- Casos Novos de TB em 2019 (Meses)
- Casos de AIDS
- Pib per capita
- Estimativa Populacional

Com o tema bem estabelecido, o cenário da Tuberculose no Brasil, foi possível fazer questionamentos e criar perguntas para com que com os conhecimentos obtidos na matéria de CCF425, iria ser possível responder e trazer conclusões. As seguintes perguntas foram levantadas:

¹Disponível em: DATASUS - TabNet

	Pergunta				
1	Qual período teve a maior taxa de doses aplicadas de Tuberculose?				
2	Quais os estados que mais vacinaram durante os últimos anos?				
3	Qual período teve a menor taxa de doses aplicadas de TB?				
4	Quais os estados que menos vacinaram durante os últimos anos?				
5	Qual a taxa de incidência de TB por ano?				
6	Qual estado teve o maior percentual de vacinação per capta? (Vacina BCG)				
7	Com base nas análises anteriores, o coeficiente de incidência de TB no país aumentou ou diminuiu?				
8	Durante o começo da pandemia em 2020, o número de casos diminuiu				
	significativamente?				
9	Qual a letalidade média por ano da doença?				
10	Qual estado brasileiro que mais apresenta casos e óbitos da doença durante				
10	os últimos anos?				
11	Existe uma relação entre as doses aplicadas das vacinas e as taxas de				
	incidência e óbito?				
12	Existe uma relação entre a taxa de pobreza e a incidência de TB?				
13	Há alguma ligação entre o número de população privada de liberdade e				
	o número de casos novos de TB?				
14	Existe uma relação entre a taxa de incidência de AIDS com a população privada				
	de liberdade (PPL)?				
15	A taxa de TB drogarresistente cresceu nos últimos anos?				
16	Houve muitos casos de cura da doença?				
17	Como é a relação das populações vulneráveis com a doença?				
18	Há alguma relação que raças diferentes com o número de casos novos				
	de tuberculose?				
19	A confirmação por laboratório da doença diminuiu ou aumentou?				
20	O que se pode concluir com as análises feitas?				

3.2. Preparação dos dados

É necessário preparar o ambiente para que a análise dos mesmos seja realizada. Sendo assim é necessário entender os atributos e objetos dos dados, os tipos de cada atributo, verificar se existe informações ausentes e buscar técnicas para resolver esse problema, entre outras transformações.

Primeiramente, houve um tratamento no qual todos os arquivos CSV (valores separados por vírgulas) da fonte de dados continham texto com algumas explicações. A exclusão desses textos foram feitas de maneira manual.

Como no presente trabalho havia uma discussão sobre como os estados brasileiros lidavam com a doença, muito dos conjuntos de dados traziam o código da UF junto ao nome do Estado. Por esta razão, foi feito uma função para separar esse código da UF, e foi feita uma função para adicionar a sigla do Estado em vez do nome inteiro, como pode ser observado na Figura 5. Dessa forma, na visualização de gráficos e para analisar certas instâncias, ficaria fácil apenas com a sigla do Estado.

```
def adiciona_coluna_estado_codigo_sigla(dataframe, colname):
  dataframe.drop(25, inplace=True)
  nomes_estados = []
  codigos_estados = []
  siglas_estados = []
  sigla = {
   '11':'RO','12':'AC','13':'AM','14':'RR','15':'PA','16':'AP','17':'
   '21':'MA','22':'PI','23':'CE','24':'RN','25':'PB','26':'PE','27':'
   AL',
   '28':'SE','29':'BA','31':'MG','32':'ES','33':'RJ','35':'SP','41':'
   '42':'SC','43':'RS','50':'MS','51':'MT','52':'GO','53':'DF','Total'
   . . .
  }
  for uf in dataframe[colname]:
   if len(uf.split(" ")) == 2:
     nomes_estados.append(uf.split(" ")[1])
     nomes_estados.append(" ".join(uf.split(" ")[1:]))
    codigos_estados.append(uf.split(" ")[0])
    siglas_estados.append(sigla[uf.split(" ")[0]])
  dataframe.drop([colname], axis=1, inplace=True)
  dataframe.insert(loc=0, column='Estado', value=nomes_estados)
  dataframe.insert(loc=1, column='Código do Estado', value=
   codigos estados)
  dataframe.insert(loc=1, column='Sigla', value=siglas_estados)
```

Figura 1. Função de adicionar sigla do estado

Foi feito uma exclusão da coluna 2021 de alguns *datasets* pois a maioria só vai até o ano de 2020. Uma das transformações mais importantes foi a de fazer uma função para converter os dados das colunas de anos, que estavam originalmente como strings, para o tipo int. A Figura 2 mostra a função de conversão.

```
def converte_coluna_de_anos_em_inteiro(dataframe):
   inicio = 2001
   final = 2020
   for ano in range(inicio, final + 1):
      if {str(ano)}.issubset(dataframe.columns):
        dataframe[str(ano)] = dataframe[str(ano)].astype(int)
```

Figura 2. Código para transformar string em int

Havia uma linha e coluna para o Total dos dados em alguns dos *datasets* porém houve a decisão de excluir essa linha e coluna, e calcular o total de dada informação de

acordo com os anos. Na Figura 3 é possível ver um exemplo desse cálculo para saber o número de população privada de liberdade em cada ano.

```
ano_inicial = 2016
ano_final = 2020

nome_anos = []
soma_anos = []

for i in range(5):
   nome_anos.append(str(ano_inicial + i))
   soma_anos.append(casos_novos_ppl[str(ano_inicial + i)].sum())

casos_novos_ppl_periodo = pd.DataFrame(list(zip(nome_anos, soma_anos)),
        columns=['Anos', 'PopulaçÃfo Privada de Liberdade'])
```

Figura 3. Código para somar os dados das colunas de anos

Houve a união de vários conjuntos em um só. Um dos conjuntos que foram feitos a partir de unificação, foi o da população vulnerável da Figura 4, o qual juntou-se os *datasets* da PPL, população de rua, imigrantes e profissionais da saúde. Outro conjunto oriundo de mesclagem foi o *dataset* das raças, que continham conjunto de dados de pessoas brancas com TB, pessoas pardas, pretas, amarelas e indígenas.

```
casos_novos_polvul = pd.concat([casos_novos_ppl_periodo.set_index('Anos
'), casos_novos_prof_saude_periodo.set_index('Anos'),
    casos_novos_pr_periodo.set_index('Anos'), casos_novos_imi_periodo.
    set_index('Anos')], axis=1)
```

Figura 4. Código para junção dos conjuntos de dados

4. Análise exploratória e extração de conhecimento

A fase de exploração dos dados é uma das mais importantes quando se trabalha com ciência de dados. É através dela que se pode analisar tendências e entender o cenário no qual se está trabalhando.

Procurou-se nessa fase responder as perguntas propostas no capítulo 3, onde gerou-se diversos gráficos e tabelas para entender o cenário que buscou-se compreender. Além disso, estudou-se a natureza dos dados, ou seja, o tipo de cada coluna, operações possíveis de se fazer, dentre outros.

Em relação a código, algumas funções foram utilizadas para reutilização no desenho dos gráficos, como por exemplo a responsável por elaborar um gráfico de barras que automaticamente pinta de cores diferentes o maior e o menor valor encontrado, como pode ser visto no trecho de código abaixo:

```
def plota_grafico_destacando_maior_e_menor_valor(dataframe, titulo,
   eixo_x, eixo_y, label_x, label_y, largura, altura, orientacao,
   minimo_parametro, escala, sobra):
 rcParams['figure.figsize'] = largura, altura
 cores = []
 maximo = max(dataframe[eixo_y])
 minimo = min(dataframe[eixo_y])
 for anos in dataframe[eixo_y]:
   if (anos == maximo):
     cores.append('#C70039')
   elif (anos == minimo):
     cores.append('#FFBD39')
   else:
     cores.append('#11144C')
 cores_da_legenda = {
   'Maior Taxa': '#C70039',
   'Menor Taxa': '#FFBD39',
 labels=["Maior Taxa", "Menor Taxa"]
 if orientacao == 'v':
   ax = sns.barplot(x=eixo_x, y=eixo_y, data=dataframe, palette=cores,
    orient=orientacao)
 if orientacao == 'h':
   ax = sns.barplot(y=eixo_x, x=eixo_y, data=dataframe, palette=cores,
    orient=orientacao)
 ax.set_title(titulo, fontsize=15)
 patches = [matplotlib.patches.Patch(color=v, label=k) for k,v in
   cores_da_legenda.items()]
 plt.legend(handles=patches, fontsize=12)
 plt.xticks(fontsize=12)
 plt.yticks(fontsize=12)
 ax.set_xlabel(label_x, size = 12 )
 ax.set_ylabel(label_y , size = 12)
 ax.set(yticks=[i for i in range(minimo + escala, max(dataframe[eixo_y
  ]) + escala, escala)])
 ax.set(ylim=(minimo_parametro , maximo + sobra))
 labels = ax.bar_label(ax.containers[0], padding=4, fontsize=12)
 plt.show()
```

Figura 5. Função que desenha o gráfico de barras destacando seu maior e menor valor

Partindo disso e respondendo as perguntas da maior e menor taxa de cobertura vacinal da TB, vê-se que os anos de 2001 e 2020 destacaram-se respectivamente, como é

mostrado na imagem abaixo:

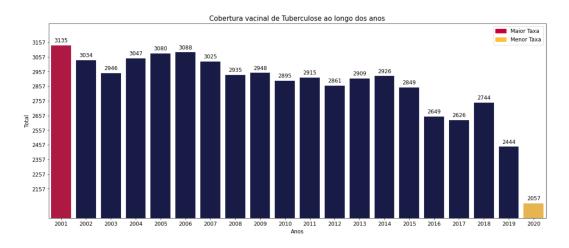


Figura 6. Maior e menor taxa de cobertura vacinal de TB

Isso pode ser devido a vários aspectos, como por exemplo: tamanho da população, campanhas políticas da época, dentre outras possibilidades para o aumento ou não aumento da taxe de doses aplicadas. Já em relação aos estados que mais e menos vacinaram nos últimos anos, optou-se pelo auxílio de manipulação de dataframes para listar os 5 estados que mais e menos vacinaram nos últimos anos, gerando dessa forma uma tabela com os dados, sendo o Maranhão o estado que mais vacinou e o Sergipe o estado que menos vacinou ao longo dos anos de 2001 a 2020.

Em relação a taxa de incidência de Tuberculose por ano, foi feita uma análise gráfica onde é possível perceber que de 2007 a 2016 houve uma baixa significativa dos números de casos. Porém, ao longo dos últimos 4 anos essa taxa teve aumento significativo, sendo 2019 o ano de maior número de casos da doença. Os casos novos da doença, tirando então pessoas que tiveram reingresso após abandono, transferência e pós óbito, foram bem baixo nos últimos anos com exceção de 2019:

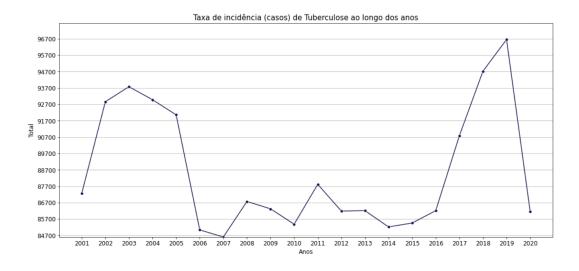


Figura 7. Taxa de incidência de TB ao longo dos anos

Também foi possível analisar qual estado teve o maior percentual de vacinação per capta, sendo eles São Paulo, seguido de Minas Gerais e Rio de Janeiro. Nessa abordagem o tipo de gráfico se diversificou, sendo escolhido um de pizza para mostrar a maior parte percentual para São Paulo - o estado que mais vacinou nos últimos anos:



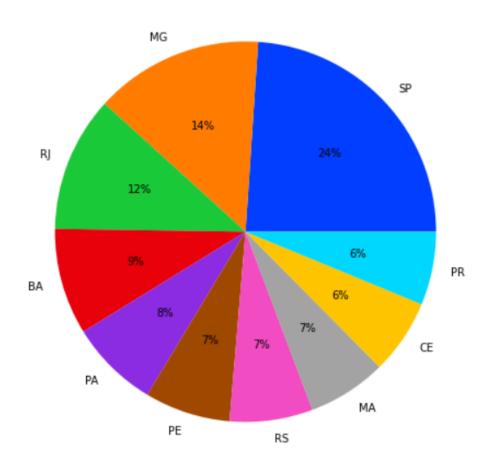


Figura 8. 10 estados com maiores percentuais de vacinação de Tuberculose

Buscando responder sobre o cenário atual da Pandemia de Covid-19, observouse que durante o período da pandemia de 2020 o número de casos de TB teve picos no começo do ano, de janeiro a março. Após isso, a partir de outubro teve quedas. Encontra partida, no ano de 2019 o maior pico não chegou perto do picos grandes de 2020:

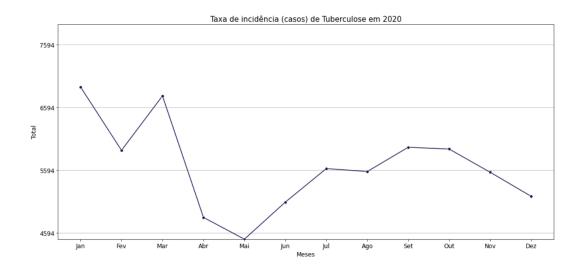


Figura 9. Taxa de incidência de TB no ano de 2020

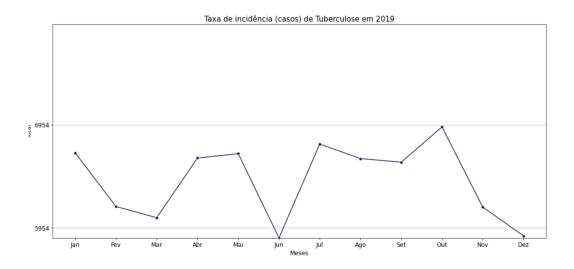


Figura 10. Taxa de incidência de TB no ano de 2019

Base de dados de óbitos também foram obtidas e através delas foi possível retirar a informação de que dos anos de 2007 a 2020 a letalidade média varia entre 4 a 49 mortes em média por ano. Nesse dataset houve uma discrepância em relação aos outros anos, que comumente estava se analisando de 2001 a 2020, mas, por falta de dados na base do governo, apenas os anos de 2007 a 2020 foram disponibilizados. E ainda sobre óbitos, vale ressaltar que o estado de São Paulo é o que mais possui casos e óbitos. Vale a pena ressaltar que São Paulo é um dos maiores estados do país.

Outro tipo de gráfico utilizado foi o de comparação entre duas variáveis, através dele foi relacionado a taxa de doses aplicadas e óbitos, levando a concluir que existe sim uma relação entre doses aplicadas, ou seja a cobertura vacinal, e o número de óbitos por tuberculose. Quando o número da cobertura começou a cair em 2015, conseguimos ver o aumento dos óbitos. A caída significativa de 2019 e 2020, mostraram um cenário o qual a vacinação de tuberculose ficou de lado:

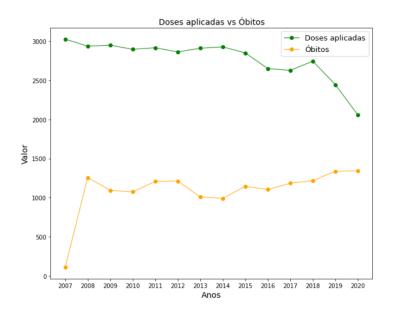


Figura 11. Doses aplicadas vs Óbitos

Outro cenário interessante é analisar a população privada de liberdade (no popular diz-se presos) e analisou-se que a PPL é uma das populações vulneráveis que mais tem casos de TB. Uma avaliação das populações vulneráveis é feita mais adiante. Não podemos afirmar que o aumento em si de casos novos são por conta, na maioria, da PPL. Para isso temos que analisar os casos novos vs os casos novos da PPL:

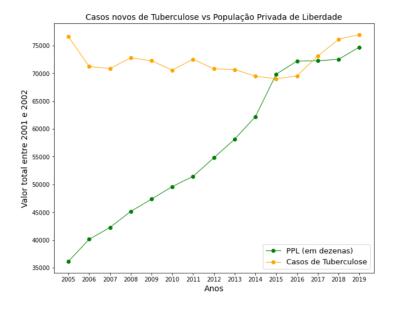


Figura 12. Casos de Tuberculose", "Casos novos de Tuberculose vs População Privada de Liberdade

Os casos de cura também foram levantados, sendo em 2005 a maior taxa de cura e em 2020 a menor. Fato no qual pode estar ligado com o movimento anti-vacina devido a pandemia da Covid-19 que instaurou um certo receio em vacinar.

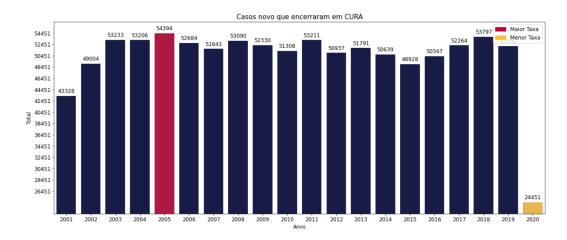


Figura 13. Casos novo que encerraram em Cura

A relação das populações vulneráveis da doença como PPL, profissionais da saúde, população de rua e imigrantes teve um aumento no período mostrado, dos anos de 2016 a 2020. A maior variação de casos nesse período foi da PPL (população privada de liberdade). O número de casos novos pela PPL diminuiu do ano 2019 para o ano de 2020 por conta da pandemia e a parada de muitos serviços no país.

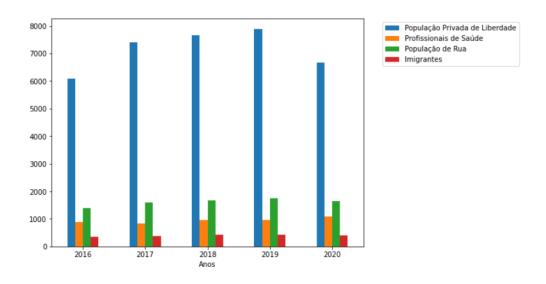


Figura 14. TB na população vulnerável

Por fim, uma estatística observada importante é a confirmação por laboratório, que aumentou durante os anos. Casos novos confirmados por critério laboratorial significa que há menos um resultado positivo nos exames: de baciloscopia de escarro, teste rápido molecular para tuberculose (TRM-TB) ou cultura.

Em 2020, o número teve uma queda significante porém lembrando do cenário de 2020 podemos fazer especulações, as quais sendo que a infraestrutura do setor de saúde estava esgotada.

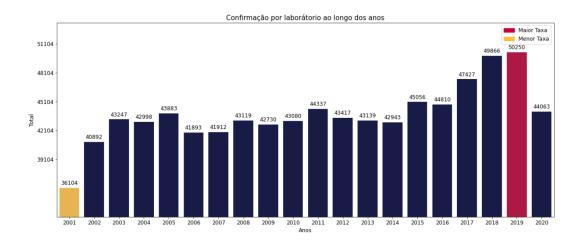


Figura 15. Confirmação por laboratório ao longo dos anos

No notebook disponibilizado para consulta, mais gráficos poderão ser analisados, nesse relatório focou-se no conhecimento extraído e nos gráficos que tiveram maior impacto para as considerações conclusivas sobre a Tuberculose no país.

5. Análise preditiva

No presente foi usado a regressão, que é o estudo de uma ou mais variáveis que dependem uma da outra com o objetivo de prever, baseados nas variáveis preditoras, um dado no futuro. No caso deste trabalho, foi usado Regressão Linear Múltipla. O objetivo da regressão era de tentar prever o número de óbitos que talvez iriam acontecer no futuro em relação aos números das outras variáveis.

Em primeira vista, foi feito uma união das principais variáveis observadas que poderiam influenciar na previsão de óbitos pela tuberculose. Logo depois da união, foi feito uma medida para verificar a relação entre as variáveis, o coeficiente de Pearson, como mostra a Figura 16

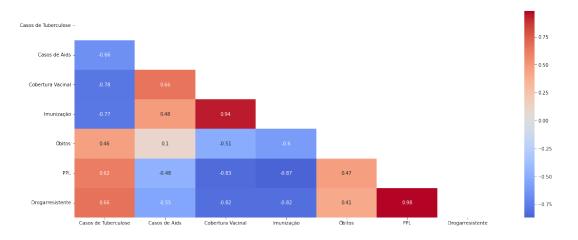


Figura 16. Coeficiente de Pearson com as variavéis escolhidas

É possível ver a relação próxima entre as variáveis de casos de tuberculose,

população privada de liberdade e casos de tuberculose drogarresistente como sendo as mais fortes em relação ao óbito.

A biblioteca statsmodels.api consegue trazer mais detalhes estatísticos para melhor o modelo.Na Figura 17 é possível ver que os números de p não são grandes, o que é bom já que queremos que as variáveis preditoras expliquem a variável alvo.

	coef	std err	t	P> t	[0.025	0.975]		
const	-1087.0958	481.083	-2.260	0.065	-2264.264	90.072		
Casos de Tuberculose	0.0750	0.029	2.545	0.044	0.003	0.147		
Casos de Aids	0.1881	0.054	3.510	0.013	0.057	0.319		
Cobertura Vacinal	-4.1321	1.758	-2.350	0.057	-8.434	0.170		
Imunização	0.0047	0.002	1.974	0.096	-0.001	0.010		
PPL	0.0006	0.000	1.985	0.094	-0.000	0.001		
Drogarresistente	-5.6841	2.890	-1.967	0.097	-12.755	1.387		

Figura 17. Resultados da Regressão OSL

Foi possível obter um coeficiente R² de 0.8082917037060047, e nisso o coeficiente diz o quanto o modelo explica os resultados. Os coeficientes de cada variável podem ser vistos na Figura 18

	Coefficient
Casos de Tuberculose	0.075010
Casos de Aids	0.188062
Cobertura Vacinal	-4.132066
lmunização	0.004688
PPL	0.000595
Drogarresistente	-5.684149

Figura 18. Equação linear

A equação linear para prever o número de óbitos é mostrado na Figura 19

```
Equação linear: ÓBITOS = [0.01906674] * Casos de Tuberculose + 0.21098315539003598 Equação linear: ÓBITOS = [0.01560474] * Casos de Aids + 0.009970114731039703 Equação linear: ÓBITOS = [-0.94725317] * Cobertura Vacinal + 0.25833192419470896 Equação linear: ÓBITOS = [-0.00084642] * Imunização + 0.3583221684757579 Equação linear: ÓBITOS = [4.47948295e-05] * PPL + 0.21971623639057292 Equação linear: ÓBITOS = [0.43637754] * Drogarresistente + 0.16905553300061224
```

Figura 19. Equação linear

6. Conclusão

O presente trabalho foi estruturado para analisar os períodos de Tuberculose, em particular incluímos um cenário atípico, ou seja, a pandemia que começou no ano 2020.

Vale ressaltar também a importância de dados públicos para que essa análise fosse feita. Em alguns cenários há dados faltantes, o que impossibilitou uma avaliação mais

detalhada ao longo dos anos, porém, com os dados coletados foi possível entender que os dados apresentados no último ano demonstram um comportamento diferente em relação ao que foi observado durante um bom tempo da doença, com uma queda acentuada da incidência e uma piora de indicadores como o de aumento do abandono.

Não é possível dizer se a pandemia influenciou em algum aspecto da doença, na sua gravidade. É incerto também concluir com 100% de certeza se os indicadores pioraram por conta de aspectos externos como a sobrecarga de hospitais e centros de saúde. Isso potencialmente pode trazer uma qualidade não tão satisfatória dos dados.

Na presente data (28/10/2021) a pandemia não acabou, mesmo com as vacinas ainda existe um descontrole em certos países como também a chegada de variantes. Esse trabalho está no começo da demonstração do comportamento da TB na pandemia. Podendo ser útil para análise comparativa entre os anos antes da pandemia e pós-pandemia.

Referências

CAZABON, D. et al. Quality of tuberculosis care in high burden countries: the urgent need to address gaps in the care cascade. *International Journal of Infectious Diseases*, Elsevier, v. 56, p. 111–116, 2017.

CLINIC, M. *Tuberculosis*. 2021. (https://www.mayoclinic.org/diseases-conditions/tuberculosis/symptoms-causes/syc-20351250). Acesso: 14-09-2021.

CORTEZ, A. O. et al. Tuberculose no brasil: um país, múltiplas realidades. *Jornal Brasileiro de Pneumologia*, SciELO Brasil, v. 47, 2021.

CRUZ, F. *Pandemia limita ações para a erradicação da tuberculose*. 2021. (https://portal.fiocruz.br/noticia/pandemia-limita-acoes-para-erradicacao-da-tuberculose). Acesso: 14-09-2021.

INFOPEN. *Peitenciário Nacional*. 2017. (http://antigo.depen.gov.br/DEPEN/depen/sisdepen/infopen). Acesso: 15-09-2021.

SAÚDE, S. de Vigilância em Saúde — Ministério da. *Boletim Epidemiológico Tuberculose 2021*. 2021. (https://www.gov.br/saude/pt-br/media/pdf/2021/marco/24/boletim-tuberculose-2021_24.03). Acesso: 14-09-2021.

USP, J. da. *Tuberculose é a doença infecciosa que mais mata no mundo*. 2020. (https://jornal.usp.br/atualidades/tuberculose-e-a-doenca-infecciosa-que-mais-mata-no-mundo/). Acesso: 14-09-2021.