

UNIVERSIDADE FEDERAL DE MINAS GERAIS
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

VICTOR FELIPE DA CRUZ OLIVEIRA

TRABALHO PRÁTICO 2 – INTRODUÇÃO A BANCO DE DADOS
ACESSO, COLETA, GERENCIAMENTO E ANÁLISE DE CONJUNTOS DE DADOS
PÚBLICOS

BELO HORIZONTE

2024

SUMÁRIO

1 APRESENTAÇÃO DA HIPÓTESE.....	3
2 DADOS UTILIZADOS	3
3 ANÁLISE DA FONTE DE DADOS	5
4 ANÁLISE EXPLORATÓRIA	7
5 CONCLUSÃO	9

1. APRESENTAÇÃO DA HIPÓTESE

Hipótese: Países com alto índice de cobertura vacinal em crianças tem menor índice de mortalidade infantil.

Objetivo: Analisar se existe correlação entre o índice de mortalidade infantil e cobertura vacinal de doenças em crianças e se há outros fatores que influenciam na diminuição da taxa de mortalidade.

2. DADOS UTILIZADOS

Cobertura de vacinação (vacinados entre 1 e 2 anos):
<https://ourworldindata.org/grapher/global-vaccination-coverage>

- Fonte dos Dados: OUR WORLD IN DATA
- Data de Obtenção: 10/06/2024
- Órgão Produtor: (WORLD HEALTH ORGANIZATION, UNICEF)
- Data de Referência (Atualização): 17/07/2022
- Limitações Registradas: Dados ausentes para certas regiões/datas
- Cobertura (Região Geográfica de Referência): Global

Taxa de mortalidade(idade menor que 5 anos):
<https://data.worldbank.org/indicator/SH.DYN.MORT?end=2022&skipRedirection=true&start=1960&type=shaded&view=map&year=2022>

- Fonte dos Dados: DATA WORLD BANK
- Data de Obtenção: 11/06/2024
- Órgão Produtor: (UNICEF, WORLD HEALTH ORGANIZATION, WORLD BANK, UNDESA POPULATION DIVISION)
- Data de Referência (Atualização): 30/05/2024
- Limitações Registradas: Dados ausentes para certas regiões/datas
- Cobertura (Região Geográfica de Referência): Global

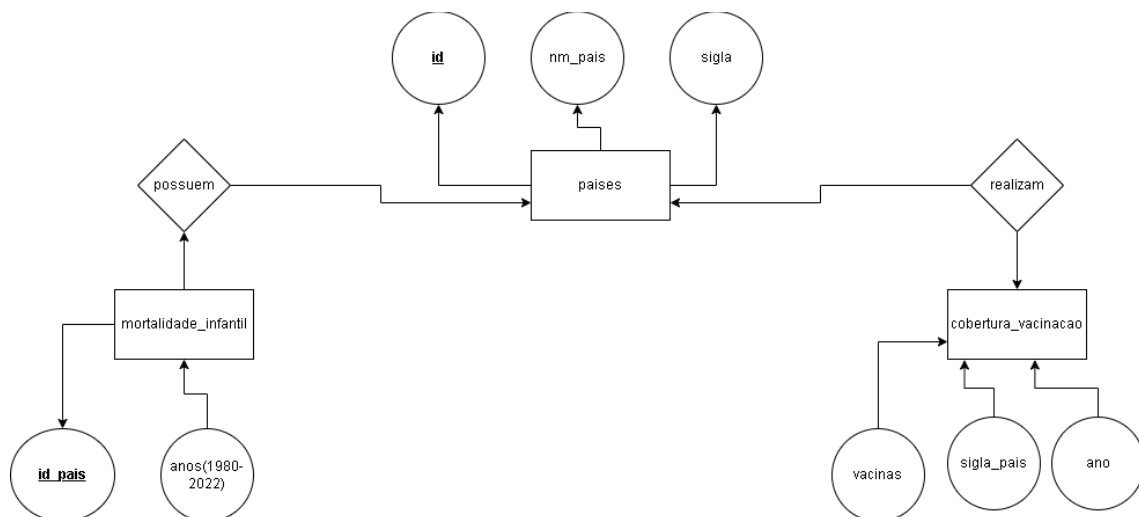
Os dados foram obtidos no formato de CSV e formatados para que pudessem ser carregados em um gerenciador, que neste caso o escolhido foi o PostgreSQL. Tanto

os originais quanto algumas tabelas geradas no fazer do TP foram publicadas no GitHub para serem visualizados, juntamente com um dump do banco que foi utilizado.

GitHub: <https://github.com/victorxkd/TP-2-IBD>

Matrícula: 2022035687

Abaixo é possível verificar um simples modelo, no formato que os dados foram obtidos. Da forma que os dados foram obtidos, foi possível observar bastante redundância de dados, foram agrupados todos em uma única tabela, que de certa forma pode facilitar alguns aspectos e dificultar outros na implementação.





OBS: Algumas informações foram emitidas/simplificadas para facilitar a leitura.

Resumidamente foi dessa forma que recebi os dados, porém trabalhar com eles dessa forma pode ser bastante difícil, talvez seja desejável normalizar esses modelos para facilitar as consultas e evitar redundâncias.

3. ANÁLISE DA FONTE DE DADOS

Considerando o fato de que meu objeto inicial era realizar uma análise comparativa ao longo dos anos, de modo que permitisse criar uma relação inversamente proporcional entre a taxa de mortalidade em crianças menores de 5 anos e a taxa de cobertura de vacinação em crianças nesta faixa, foi possível observar

alguns padrões que de certo modo tornam a conclusão final menos confiável se não observado juntamente com outros dados. Por exemplo para alguns países temos uma quantidade de valores nulos exorbitante para certos tipos de vacinas em determinados países ou anos, de tal forma que se faz necessário levar em conta outros aspectos não levantados inicialmente tal como contexto histórico/geográfico e socioeconômico. Em respeito a formatação dos dados, devido a grande quantidade de países, anos e por vacinas diferentes, a visualização desses dados deve ser tratada de forma laboriosa para que permita que os usuários sejam capazes de tomar alguma conclusão sobre os dados analisados, como por exemplo utilizar médias, averiguar o espaçamento de 10 em 10 anos, verificar para somente x vacina ou x país, as consultas aos dados devem ser arquitetadas para permitir uma análise tanto de forma geral como centralizada em uma data ou local específico. Vale ressaltar que os dados são por países, se fosse necessário fazer verificações por regiões e/ou continentes, seria necessário estar modificando as tabelas originais para que contivessem esses dados ou incluir todos os países específicos na hora de consultar, o que pode trazer dificuldades nas consultas, por conta das redundâncias que iriam eventualmente ocorrer por conta da forma que os dados foram coletados.

Em minha experiência pessoal, a parte mais difícil foi a de cruzamento dos dados, pois a forma que eu obtive os dados iria me dificultar na análise final, na forma de gerar tabelas ou gráficos, pois eu teria muitas colunas, eu tentei carregar uma planilha específica para este documento Word e não consegui, pois ela era muito grande, então eu pude averiguar que a visualização dos resultados foi uma tarefa bem difícil para mim, tanto que eu me resumi de propósito em algumas análises para que fosse possível manter dentro deste documento.

Inicialmente após obter os dados, pude reparar que em ambos os conjuntos, eles possuíam a coluna da sigla em comum, para poder tirar proveito deste fato, eu organizei as tabelas em ordem alfabética pela sigla, e então retirei de uma das tabelas todas as siglas que não se encontravam na outra tabela, desta forma pude obter tabelas que estavam organizadas alfabeticamente pelas siglas dos países e que possuíam todos os países em comum. Acredito que esta etapa tenha sido necessária pois alguns países/regiões não continham dados relevantes para as análises dos problemas que iam ser observados, também entrariam nos cálculos das médias sem agregar nenhum valor.

4. ANÁLISE EXPLORATÓRIA

Fazer análises mais sofisticadas de todos os dados consumiria mais tempo para que fosse possível obter um resultado mais simples de ser representado, para fins didáticos, optei por me limitar a alguns anos e países, por exemplo, vamos fazer uma análise das taxas de mortalidade e cobertura de vacinação em crianças menores de 5 anos nos anos de 1980, 2000 e 2020 para o Brasil;

sigla_pais	mortalidade_1980	mortalidade_2000	mortalidade_2020
BRA	95.9	34.6	14.5

Podemos verificar que a média da taxa de mortalidade (lembrando que esse valor é para cada 1000 crianças), vem diminuindo ao longo do tempo, onde em 1980 esse valor era de 95.9 em 1000, e em 2020 era de 14.5 em 1000.

pol3_1980	pol3_2000	pol3_2020	dtp3_1980	dtp3_2000	dtp3_2020
0.69	0.99	0.74	0.37	0.98	0.77

Valores em porcentagem para crianças entre 12 e 24 meses

Em um exemplo, utilizando somente duas vacinas, da Poliomelite – POL3 e Tríplice Bacteriana Acelular - DTP3, podemos verificar que houve uma diminuição da mortalidade com o aumento da cobertura da vacina (para o ano de 2020 é importante lembrar que foi o início da pandemia da Covid-19, que causou turbulências no setor de saúde). Claro que uma análise desse escopo é muito vaga em um âmbito geral, levando em consideração apenas duas vacinas em um único país, é somente um exemplo de como seria feita essa análise. Podemos utilizar também dados gerais, por exemplo verificar a média da taxa de mortalidade e cobertura vacinal de vários países de uma única vez, porém isso não faria muito sentido levando o contexto, o panorama social dos países é diferente demais para essas junções.

Leve este outro exemplo em consideração, para o país do Afeganistão, temos a seguinte relação de taxa de mortalidade infantil e cobertura vacinal para a POL3 e DTP3:

sigla_pais	mortalidade_1980	mortalidade_2000	mortalidade_2020
AFG	246	131.6	61.6

pol3_1980	pol3_2000	pol3_2020	dtp3_1980	dtp3_2000	dtp3_2020
NULL	0.24	0.75	0.04	0.24	0.7

Pode-se perceber um declínio considerável na mortalidade infantil desse país, ao mesmo tempo um aumento quantioso na quantidade de crianças vacinadas para respectivas doenças.

Após um estudo mais aprofundado dos dados, torna-se possível concluir que quanto mais crianças vacinadas contra doenças, menor será o índice de mortalidade infantil. Minha intenção com a análise não é comparar os países ou chamar a atenção para os valores absolutos em si, e sim demonstrar a relação entre diminuição da taxa de mortalidade infantil e aumento de cobertura vacinal em crianças, lembrando que estou levando somente esses dois em consideração, mas facilmente poderia ser incluído uma avaliação de PIB anual dos países também.

Dentre os países com as menores taxas de mortalidade infantil, podemos destacar o Japão, Suécia e Finlândia, que constantemente se mantêm nos índices das menores taxas de mortalidade e maiores taxas de cobertura vacinal, é possível verificar estes fatos facilmente comparando tabelas geradas, irei enviar tais tabelas junto a este documento, a quantidade de linhas das mesmas a tornam ruins de visualizar por aqui. Por outro lado, na tentativa de gerar tabelas mais simples de visualizar, como a correlação de Pearson, muitos dados ficaram nulos, devido ter sido considerado muitos países que inicialmente já não possuíam esses dados.

Vacina	Correlacao_pearson
BCG	0.38
HepB3	NULL
Hib3	NULL
IPV1	NULL
MCV1	0.32
PCV3	NULL
Pol3	0.42
RCV1	0.83
RotaC	NULL
YFV	NULL
DTP3	0.46

****Levando em consideração a diferença de mortalidade de 1980-2020 e cobertura vacinal 1980-2020, os resultados iriam mudar levando outro espaçamento de tempo****

5. CONCLUSÃO

Após a análise realizada foi possível observar uma relação entre a expansão da cobertura vacinal em crianças e a significativa redução da taxa de mortalidade infantil, que pode ser observada como decrescente em todos os países, ainda que outros aspectos influenciam nesse dado. Dessa forma pode-se concluir que os programas de vacinação têm desempenhado um papel crucial na melhora da saúde infantil, prevenindo doenças desde simples até fatais. As implicações dessas conclusões são claras, é essencial continuar e aumentar os programas de vacinação, a fim de garantir que sejam acessíveis a toda população alvo, independente de localização geográfica ou condição econômica. Em análises futuras, seria interessante analisar outros aspectos juntamente com os já obtidos, com o objetivo de verificar quais outros fatores influenciam positivamente/negativamente na redução da mortalidade, para que seja possível gerar insights valiosos para a formulação de políticas públicas que visem atender aquela necessidade específica.