

Análises Estatísticas acerca da Covid-19 no Afeganistão entre os anos de 2020 a 2023

1rd Luis Carlos Rodrigues dos Anjos
Departamento de Engenharia de Computação
Universidade Federal do Ceará
Fortaleza, Ceará
lc968077@gmail.com

Resumo—Cálculos estatísticos são muito importantes para o entendimento de vários fenômenos de todos os tipos e nesse estudo aplicamos a média, o desvio padrão, correlação e covariância entre os dados sobre o número de novos casos confirmados, de novas mortes causadas pela covid-19 e de novas doses de vacinas contra o vírus entre 2020 a 2023 no Afeganistão. Além disso, foram feitas análises usando testes de hipóteses sobre alguns aspectos dos dados, bem como gráficos para visualização de dados.

Index Terms—Média, Desvio Padrão, Correlação, Covariância, Teste de Hipótese.

I. INTRODUÇÃO

O Afeganistão é um país da Ásia Central, localizado na conflituosa região do Oriente Médio, sendo uma nação predominantemente islâmica com cerca de 32 milhões de habitantes [1].

Durante o período de 2020 até 2023, ano desse estudo, o Afeganistão passou por grandes mudanças políticas e sociais, incluindo a retirada das tropas estrangeiras e a ascensão do Talibã ao controle do país em 2021, o que gerou instabilidade e incerteza em muitas áreas. O país enfrentou desafios significativos em sua infraestrutura de saúde, que já era frágil antes da pandemia. A falta de recursos, instabilidade política e conflitos internos impactaram a capacidade do país de lidar com a crise sanitária.

No início da pandemia em 2020, o Afeganistão enfrentou a propagação do vírus, com recursos limitados para testes, equipamentos de proteção e infraestrutura de saúde. As medidas de restrição foram implementadas, mas a adesão e eficácia foram desafiadas por questões socioeconômicas e políticas. Além disso, com a chegada da segunda onda e principalmente de terceira onda que teve em 2021 houve um aumento significativo no número de casos e mortes por COVID-19, e esta terceira onda juntamente com a primeira tiveram um foco maior, pois foram as duas maiores ondas de contágio segundo nosso dataset. Em suma, A instabilidade política e a transição de governo impactaram a resposta à pandemia, dificultando a implementação de estratégias eficazes de saúde pública.

Sendo assim, a proposta desse artigo é ter uma visão matemática sobre os dados da covid-19 nesse país referente aos números de contágio e morte pelo vírus cotidianamente,

bem como a quantidade de vacinas ofertadas por dia no tempo em estudo.

II. MÉTODOS

A. Pré-processamento

Os dataset utilizado tem os dados de muitos outros países e com diversas informações, porém foi devido a escolha para o estudo deste artigo apenas o país já mencionado, o Afeganistão, e os dados referentes aos números de novos casos, novas mortes e novas doses de vacinas, ambas por dia, durante Janeiro de 2020 até Junho de 2023. No entanto, outros períodos em destaques tiveram mais focos, como o período de Maio á Setembro, ambas de 2020 e 2021, tempo estes que o país teve suas duas maiores ondas de casos de coronavírus.

B. Métodos Estatísticos

Média

A média [1] de um conjunto de n observações é dada por:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

onde \bar{x} é a média e x_i são as observações.

Variança

A variação [1] é uma medida que mostra a dispersão dos dados de uma distribuição calculando a distancia desses dados à média.

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n} \quad (1)$$

Desvio Padrão

O desvio padrão [1] é uma medida de dispersão e é calculado como a raiz quadrada da variância. A fórmula para o desvio padrão é:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

onde σ é o desvio padrão e \bar{x} é a média.

Covariância

A covariância [1] entre duas variáveis X e Y é dada por:

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

onde $\text{Cov}(X, Y)$ é a covariância, x_i e y_i são as observações de X e Y , respectivamente, e \bar{x} e \bar{y} são as médias de X e Y , respectivamente.

Correlação

A correlação [1] entre duas variáveis X e Y é calculada pela fórmula:

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

onde $\rho_{X,Y}$ é a correlação, $\text{Cov}(X, Y)$ é a covariância entre X e Y , σ_X é o desvio padrão de X e σ_Y é o desvio padrão de Y .

O índice de correlação é o resultado desta conta e varia entre 1 e -1, sendo que quando cada vez mais próximo de -1 mais negativa é tal correlação e quando cada vez mais próximo de 1 mais positiva é a correlação entre as variáveis, e se for igual a 0 significa que elas não têm correlação.

Teste de Hipótese

O teste de hipótese é um procedimento estatístico para testar afirmações sobre uma população com base em uma amostra. A formulação de hipóteses, o cálculo de estatísticas de teste (como o valor- p) e a decisão sobre a rejeição ou não-rejeição da hipótese nula são partes essenciais desse processo. Os passos comuns para um teste de hipóteses incluem:

- 1) Formulação das hipóteses nula (H_0) e alternativa (H_1).
- 2) Escolhe-se de um nível de significância (α).
- 3) Cálculo da estatística de teste (por exemplo, t , Z , F ou χ^2). A fórmula para o cálculo do valor t em relação a um valor específico X_0 é dada por:

$$t = \frac{\bar{X} - X_0}{\frac{s}{\sqrt{n}}}$$

Onde: t é o valor t ,

\bar{X} é a média da amostra,

X_0 é o valor de referência,

s é o desvio padrão amostral,

n é o tamanho da amostra.

- 4) Cálculo do valor- p .

- 5) Tomada de decisão com base no valor- p em relação ao nível de significância.

Distribuição Normal

A distribuição normal, também conhecida como distribuição de Gauss, é uma distribuição contínua simétrica que é frequentemente usada na estatística devido a sua ampla aplicabilidade. Sua função de densidade de probabilidade é definida por:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

onde μ é a média, σ^2 é a variância e x é a variável aleatória.

A distribuição normal padrão tem média (μ) igual a zero e variância (σ^2) igual a um. É comum usar a tabela ou calculadoras estatísticas para encontrar probabilidades associadas à distribuição normal padrão.

Histograma

Histograma é um gráfico de barras utilizado para demonstrar a distribuição de dados discretos ou contínuos em uma

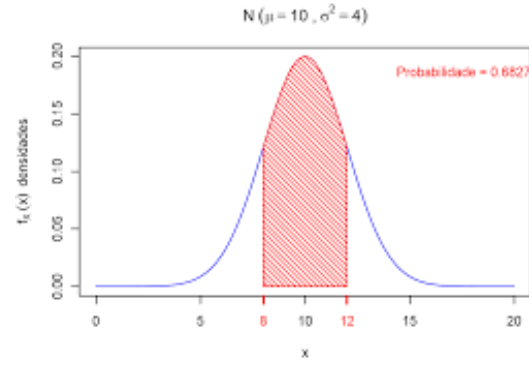


Figura 1. Exemplo de Distribuição Normal

determinada frequência, indicando a dispersão da classe dos dados.

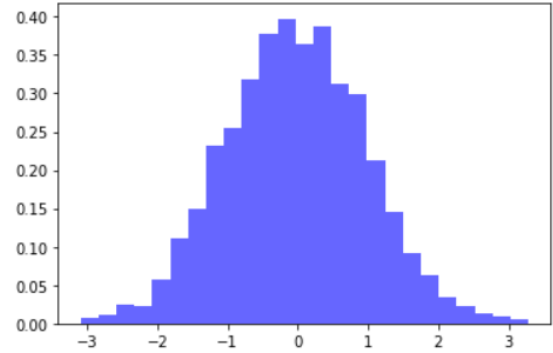


Figura 2. Exemplo de histograma gerado com 2250 dados aleatórios de média = 0, desvio padrão = 1.

III. RESULTADOS

De início, podemos ver nos gráficos das figuras 3, 4 e 5 o porquê da escolha da duas ondas mencionadas:

É fácil notar uma maior frequência de casos nos meses de maio até setembro tanto no ano de 2020 quanto no de 2021.

Além disso, podemos analisar e confirmar o que parece óbvio ao senso comum, mas que porém é preciso de uma análise para se ter uma afirmação mais concreta a cerca de tal fato, que de que o número de casos afeta o número de mortes na medida em que o primeiro cresce o segundo também aumenta. Logo, podemos observar os histogramas dos dois anos em que houve mais casos nas figuras 6 e 7, e com as 5 última figuras mencionadas temos uma grande indicação, apenas visualmente, de que o ano de 2021 teve mais casos de que 2020 e, conseqüentemente, mais mortes por dia por covis-19.

Mas, podemos ter mais certeza disso utilizando os conhecimentos de covariância e correlação [1], pois o índice de correlação que usa da covariância, citado nos métodos, obtido usando os dados de novos casos e mortes equivale a 0,842, o que significa que eles têm uma alta correlação positiva pois

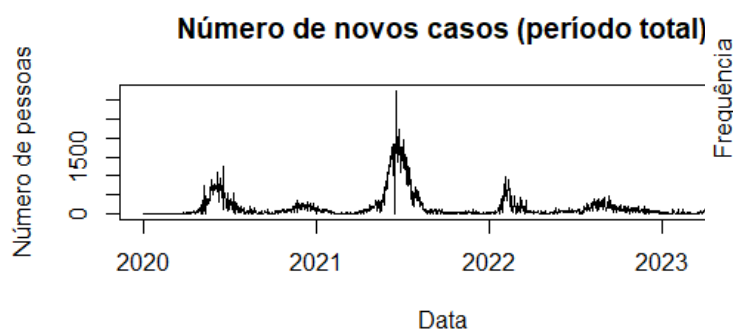


Figura 3. Gráfico do número de novos casos de covid-19 do período

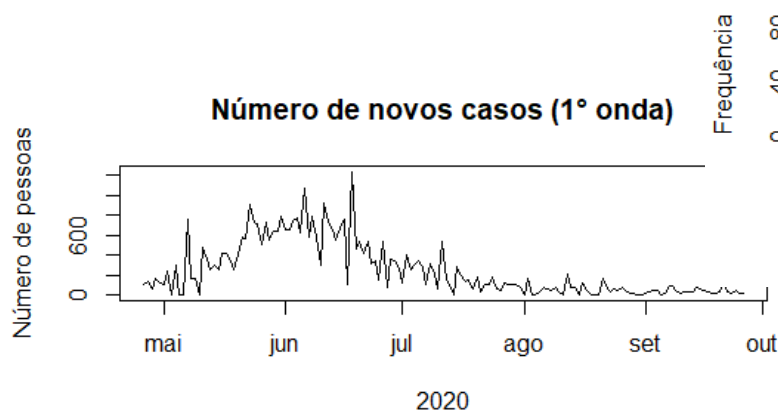


Figura 4. Gráfico do número de novos casos de covid-19 na Primeira onda (2020)

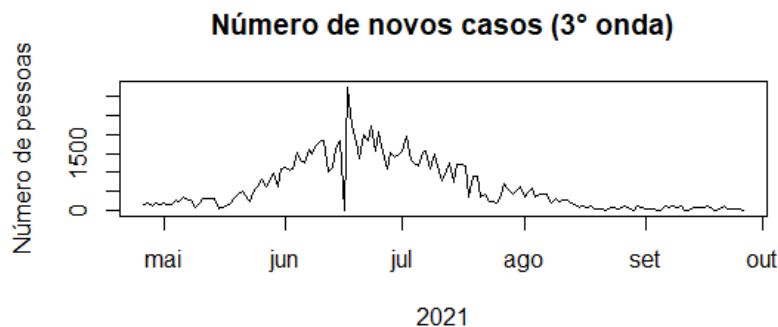


Figura 5. Gráfico do número de novos casos de covid-19 na Terceira onda (2021)

Histograma dos dados de novos casos primeira onda

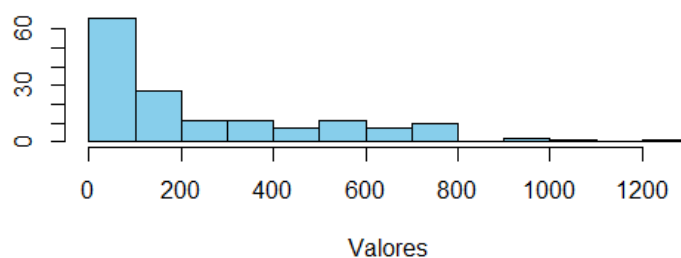


Figura 6. Histograma do número de novas mortes por dia de 2020

Histograma dos dados de novos casos terceira onda

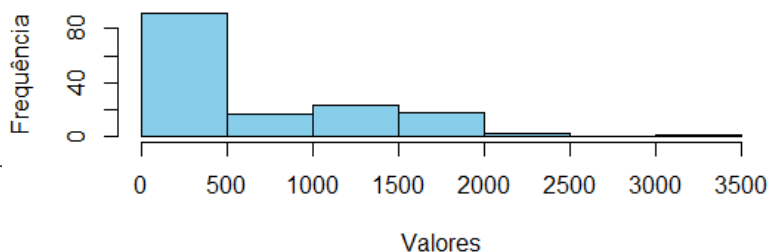


Figura 7. Histograma do número de novas mortes por dia de 2021

está bem próximo de 1, isto é, se um cresce o outro também tende crescer.

Ademais, pode-se também analisar as médias e o desvio padrão no número de casos e mortes de 2020 e 2021, a saber: 2020 - média de 143,76 casos com desvio padrão de 205,66 e média de 7,06 mortes com desvio padrão de 9,24; 2021 - média de 289,66 casos com desvio padrão de 506,66 e média de 14,15 mortes com desvio padrão de 24,34. Logo, a média de casos e de mortes aumentaram em 2021 em relação a 2020. Além disso, tivesse-se: 2022 - média de 135,60 casos com desvio padrão de 143,07 e média de 1,35 mortes com desvio padrão de 2,40; 2023 - média de 90,058 casos com desvio padrão de 97,64 e média de 0,42 mortes com desvio padrão de 0,81.

Com isso, podemos ver que em 2020 e em 2022 tiveram-se média e desvio padrão bem próximos, no entanto, em 2022 teve uma média bem menor de mortes e isso se deve porque o ano de 2022 foi o que teve mais vacinas, 68 085 com média de 186,53 doses por dia enquanto que em 2020 ainda não tinha vacinação contra a covid-19. Essas doses que começaram em 2021, com média de apenas 18,83, geraram um wfwito positivo em 2023, pois este ano teve uma média de mortes muito menor que dos outros anos desde a chegada do vírus ao Afeganistão, 0,42, e embora a média de doses diárias de 2023 ainda fossem baixas ,em comparação a 2022, de 19,39, temo-

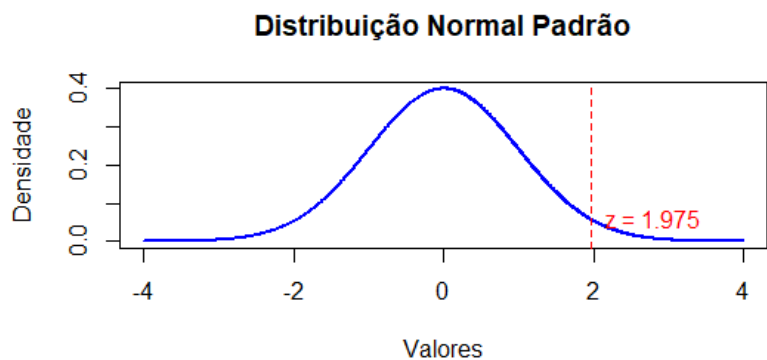


Figura 8. Distribuição normal referente ao teste de hipótese

se assim um argumento a favor do uso da vacina, apesar do número menor médio de novos casos também têm influenciado na média de óbitos por este vírus.

Além disso, aqui foram feitos testes de hipóteses: hipótese nula é de que em 2020 a média é de 7,06 e a hipótese alternativa é de que a média de mortos por dia aumenta em grandes ondas de casos. O nível de significância usado foi de 5%, com 153 amostras (número de dias para os 5 meses de onda de covid) e desvio padrão amostral obtido de 11,12 e uma média amostral de 9,16. Assim, utilizando-se os métodos estatísticos sobre teste de hipóteses, o valor de t obtido foi de 2,335 que é maior que o valor de $t = 1,975$ (obtido pela distribuição t de Student [3]) que é para 5% nível de significância. Dessa forma, a hipótese nula foi rejeitada e o mesmo ocorreu para a terceira grande onda de casos que ocorreu em 2021, pois o valor de t obtido foi de 6,331, que é maior que 1,975. Então, sobre isto podemos observar a seguinte distribuição normal para melhorar nossas análises:

Conclusão

Em suma, neste trabalho foi possível dar mais solidez por meios matemáticos estatísticos às confirmações sobre o aumento de mortes decorrente do número de casos, principalmente em relação ao que ocorre em grandes ondas de contágio, em que as médias de casos e mortes aumentam. Além disso, foi possível ver o impacto positivo das vacinas em relação ao número de óbitos a longo prazo e isso tudo tendo um país como foco que enfrentou e enfrenta muitos problemas [2], ou seja, o Afeganistão mostra que apesar das dificuldades de um país, a vacina é sim eficaz. Todos estes dados foram obtidos pelo algoritmo em R feitos pelo autor.

REFERÊNCIAS

- [1] Montgomery, Douglas C., George C. Runger e Norma F. Hubele. Probabilidade e Estatística para Engenharia. 6ª ed., LTC Editora, 2018.
- [2] RFI. (2023, 15 de agosto). Afeganistão: dois anos após volta de talibãs ao poder, miséria e cerco às mulheres reinam no país. G1 Recuperado de [https://g1.globo.com/mundo/noticia/2023/08/15/afeganistao-dois-anos-apos-volta-de-talibas-ao-poder-miseria-e-cerco-as-mulheres-reinam-no-pais.ghtml].
- [3] Student. (1915). The Probable Error of a Mean. Biometrika, 8(1), 1-25.