

# A statistical study of vaccination data in the Covid-19 pandemic

*Stephanie Lima*<sup>1</sup>

<sup>1</sup> Universidade Federal da Bahia, Salvador, Bahia, Brazil

**Abbreviated abstract:** The Severe Acute Respiratory Syndrome - SARS-CoV-2, or simply the new coronavirus, identified at the end of 2019 in China, has spread rapidly around the world, generating a global pandemic. This work aims to correlate the number of deaths and the number of people vaccinated, and to group countries that have higher numbers of vaccinated people, as well as the name of these vaccines, in relation to countries with low rates. In order to achieve these objectives, we make use of the correlation coefficients and clustering algorithms, for exploration and analysis of the data.

## Related publications:

- NICEWANDER, J. L. R. W. A. Thirteen Ways to Look at the Correlation Coefficient. 1988.
- SOUZA, L. E. P. F. de; BUSS, P. M. Desafios globais para o acesso equitativo à vacinação contra a covid-19. Cad. Saúde Pública, v. 37, n. 9, p. e00056521, 2021.



# Problem and Data

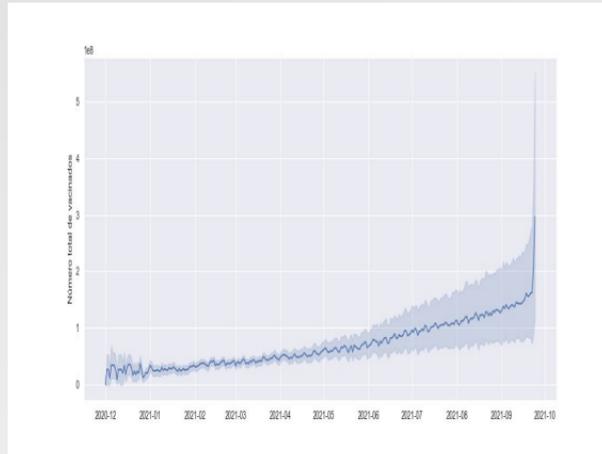
With the increasing spread of the new coronavirus, the ongoing quest to create a vaccine has become the effective defensive strategy against the coronavirus. However, the production and supply of vaccines across countries and population groups has been a major challenge due to the limited supply of vaccine doses, making distribution and equitable access between them difficult. The datasets analyzed in this work are updated daily by Our World In Data. However, the datasets that were last updated on 09/25/2021 were chosen for the development of this work

| File                | Description   |
|---------------------|---|
| vaccinations.csv    | It has country data on COVID-19 global vaccinations.  |
| owid-covid-data.csv | This file includes all historical information about the pandemic as of the last publication date. |



# Methods

- **Pearson Correlation Coefficient**



To identify the Pearson Correlation Coefficient between the variables Delayed numbers X Number of vaccinated, the `corr()` function from the Pandas library was used.

Therefore, after executing the code, the result obtained from the correlation between these two variables was equal to 0.0115, or 0.1, approximately. In view of this, it can be stated that there is a negligible correlation between these two variables.



# Results and Conclusions

- For the use of the method that calculates the Pearson Correlation Coefficient, we obtained a negligible value and for the grouping of the data, an inconclusive graph was reached because they do not present well-defined clusters. understand all the correlations between the variables and interpret them clearly, as well as a more precise definition of the attributes to be grouped for the execution of K-Means.

