BIOLOGY HIGHER LEVEL INTERNAL ASSESMENT

An investigation of the relationship between full covid vaccination rates, total covid death rates and total covid cases per country.

**Research Question**

Is there a relationship between full covid vaccination numbers and the total cases and deaths per country?

**Personal Significance**

In December 2021, I was home for Christmas. My mother, little brother and I were all staying at my grandparent’s house since they live in a different country with the rest of my extended family. The day after Christmas my mother and brother began exhibiting symptoms of COVID-19 so they asked my grandfather to go get some antigen tests from the nearby pharmacy. That night, both of them tested positive for COVID-19 but my grandparents and I tested negative which was a big surprise seeing as we had been in very close proximity with each other. Two days later my grandparents tested positive and the day after that, I tested positive as well. Our house had become a Covid house.

Although my grandmother and grandfather are both considered to be part of a high-risk population, they ended up being practically asymptomatic whereas my mother and brother had horrible symptoms. The reason why this is related to this IA is the fact that my grandparents and I are fully vaccinated whereas my brother and mother aren’t. During the few days when my grandparents and I tested negative, we believed we were protected from the virus due to our vaccination status which then ended up not being much protection at all. This belief being proven wrong ended up becoming my motivation to conduct the research and investigations shown in this Internal Assessment.

**Introduction**

COVID-19 (severe acute respiratory syndrome coronavirus 2), also known as SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) comes from a family of viruses which are responsible for illnesses such as the common cold, severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) ("Coronavirus disease"). Currently, 57% of the world is fully vaccinated, the country with the highest vaccination rates being the United Arab Emirates or the UAE ("Tracking Coronavirus"). Coronavirus affects not only your lungs and respiratory system but it can also affect other parts of your body as well. Although most people experience fairly mild symptoms, some others can become severely ill. Coronavirus can be transmitted through droplets projected from your mouth and nose when talking, breathing, coughing, or sneezing.

The virus is called coronavirus because of its structure. *Corona* meaning crown refers to the appearance of the virus which have spike proteins sticking out of them. These spikes are what allow the virus to attach itself to human cells to infect it. Some antibodies can help you fight the virus by targeting these spikes. Through genetic surveillance, scientists who sequence the virus can identify when the virus starts to genetically change thus creating lineages which scientists call variants (‘Basics of COVID-19’).

**Background Information**

There are a total of three vaccines available in the US market.

1. BNT162b2 (Pfizer-BioNTech COVID-19 vaccine) is indicated for individuals aged 5 years or older.
2. mRNA-1273 (Moderna COVID-19 vaccine) is indicated for individuals aged 18 years or older.
3. Ad26.COV2. S (Janssen COVID-19 vaccine) is indicated for individuals aged 18 years or older.

It is recommended to get at least one of these three vaccines. Pfizer has 2 doses plus boosters, Moderna has 2 doses plus boosters and Janssen has a total of 1 dose plus boosters. These doses are what is necessary to be considered as a fully vaccinated individual. It is possible to get different vaccines for the first and second dose as long as these vaccines don’t contradict each other. These contradictions may include severe allergic reactions, a history of thrombosis etc. ("COVID-19: Vaccines"). This is why it is so important to have a post-vaccine observation period of 15-30 mins depending on each individual’s needs.

The covid vaccine can also have side effects such as pain, redness and swelling on the arm where the shot was administered as well as tiredness, headaches, muscle pain, chills, a fever and nausea. This is why after the shot it is recommended to conduct light exercise with your arms and to drink lots of fluids ("Possible Side").

**Hypothesis**

**Null Hypothesis 1:** There is no relationship between full vaccination and deaths per country.

**Null Hypothesis 2:** There is no relationship between full vaccination and covid cases per country.

**Alternate Hypothesis 1:** There is a relationship between full vaccination and deaths per country.

**Alternate Hypothesis 1:** There is a relationship between full vaccination and covid cases per country.

**Variables**

**Independent Variable:** People who are fully vaccinated.

**Dependent Variable:** Total covid cases and covid deaths.

**Acquisition of Data**

For my data collection I had to conduct extensive research in order to acquire both my raw data and background information etc. For instance, I had to begin by carrying out enough research to be able to select or rather come up with an adequate research question, variables and hypotheses. Once I had these pillars to begin my investigation I then proceeded to compile a table of raw data. I found an immense database with all sorts of data on covid and factors affecting cases, deaths and following illnesses as well as duration of hospitalization per country in *Our World in Data* which is a website with substantial amounts of data on many different things regarding the world we live in ("owid / covid-19-data.").

My next step was selecting what parts of the database I would be using for my IA seeing as not all rows or columns of data applied to my research question. In order to do this more efficiently I used some of my previous knowledge on python, the programming language to select and compile my final table of raw data. I used a simple code that I made which basically told me which columns and rows were null and which ones had the most data for statistical analysis. For example, I initially wanted to include people with booster shots in the analysis but the column had far too many null values to be able to use it. The word null is used to describe a section of the database which has insufficient data (or no data) to conduct any sort of analysis (“Handling Null”). This code linked below in the appendix. Hence, I made a spreadsheet where I placed my raw data which will also be linked in the appendix (*Table 1*).

Graphical user interface, application, table, Excel

Description automatically generated Even after this data selection, there were still over a hundred thousand country samples from multiple dates which were more than valid for statistical analysis, thus I conducted further data selection. The way I did this was by applying a filter on excel (“Filter data”). I decided to only select countries with a life expectancy over 79.10, thus narrowing down my data to a total of 52 countries. Lastly, I made a pivot table with excel in order to show my raw data in an organized way which was more visually appealing (“Create a PivotTable”).

*Table 1: Excel spreadsheet – unprocessed raw data*

**Ethical Concerns**

The main ethical concern was accidentally collecting biased or inaccurate data which would ultimately alter and invalidate my results. I needed to ensure that all of my data was from reliable and fair sources and also I had to make sure that none of my analysis or collection of the data was somehow oblique towards a specific point of view or final result. In order to achieve this, I used python to ensure that I was not biased throughout my data selection process since this program is unbiased and preprogrammed to be as efficient and fair as possible for a computational language.

**Data Presentation and Analysis**

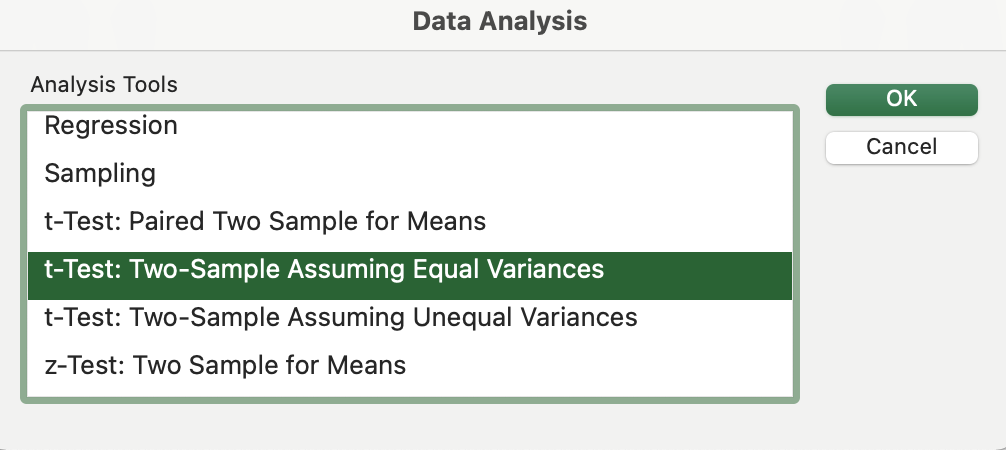
As already mentioned in the acquisition of data section, I filtered my raw data to only select countries which had a life expectancy of over 79.10 which left me with 52 countries to make a pivot table (*Table 2*) (“Create a PivotTable”).

Table

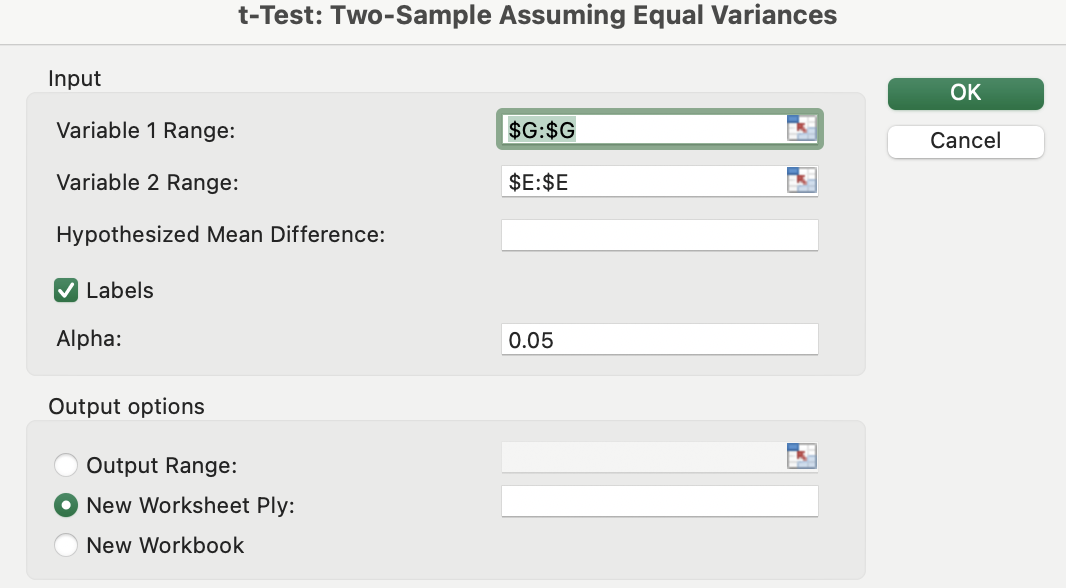
Description automatically generated*Table 2: Pivot Table – Processed Data*

At first glance, the number of vaccinations does not seem to affect the total covid cases per country but on the other hand, the higher the number of fully vaccinated individuals per country, the lower the death rates seem to be (*Table 2*).

To conduct statistical analysis I used an excel formula. I basically separated the analysis into different parts. First of all, I used the T-test for my inferential statistics which I learned how to carry out on excel (*Figure 1, Figure 2*) through further research ("Running a t-test in Excel").



*Figure 1: Excel step 1 for t-test*



*Figure 2: Excel step 2 for t-test*

Since the t-test is only used when you have two sets of data and I have three I made the decision of first conducting the t-test on sum of people fully vaccinated and sum of total cases and once that was done, repeating the process but this time with sum of people fully vaccinated and sum of total deaths (*Table 2*).

Table

Description automatically generated

*Table 3: T-test 1 people fully vaccinated vs total cases*

The first t-test shows a P value of 0 which means that there is no correlation between the two sets of data as I observed from the processed data table (*Table 3*). Table

Description automatically generatedNext I conducted the test on the sum of people fully vaccinated and sum of total deaths (*Table 2*).

*Table 4: T-test 2 people fully vaccinated vs total deaths*

Just as predicted from my observations, there is a positive correlation between the amount of fully vaccinated people and the total deaths per country (*Table 4*). The reason why this statistical test was chosen was because it calculates a correlation between two datasets which is what I needed to answer the research question.

There is some slight uncertainty in my raw data. Since the data is categorized by country, this means that the data is incorrect if only the registered residents were considered for the data set. If there happened to be some unregistered immigrant or unregistered child being born etc. that would make the data inaccurate. Furthermore, there’s a possibility that some people that may have had covid did not go get tested and hence, these cases wouldn’t be recorded. Lastly, often we see cases of people who die and their deaths are not recorded. All of these cases could render the data uncertain or inaccurate.

**Evaluation**

To conclude, I will firstly reinstate my hypothesis.

**Null Hypothesis 1:** There is no relationship between full vaccination and deaths per country.

**Null Hypothesis 2:** There is no relationship between full vaccination and covid cases per country.

**Alternate Hypothesis 1:** There is a relationship between full vaccination and deaths per country.

**Alternate Hypothesis 2:** There is a relationship between full vaccination and covid cases per country.

The null hypothesis 2 is accepted but on the other hand the alternate hypothesis 1 was accepted. So, to answer the research question; Is there a relationship between full covid vaccination numbers and the total cases and deaths per country? There is a relationship between the full covid vaccinations and deaths but there is no correlation between full vaccinations and total cases. This is visible from the P value in the statistical analysis (*Table 3, Table 4*). The reason for this is that vaccines help fight disease not prevent it (“Coronavirus disease”).

The data used was appropriately chosen due to the fact that COVID-19 is the cause of a global pandemic and hence, it is necessary to address all the countries with available data. Again, as already mentioned, there are many factors that could lead to uncertainty of the data since I don’t know the collection criteria for the creator of the data base.

The main improvement that I would suggest is a list of the data collection criteria for the data base. Outside factors were possibly not considered in the database such as unrecorded deaths, illegal immigration and emigration, unrecorded births and so on. These limitations should be pre-stated in the data base for ethical and fair use of the data. Furthermore, the lack of graphical visuals in this IA is due to the nature of the data. It is so spread out yet so concentrated in certain areas that when attempting to graph it, a very badly formed graph would appear. Hence the lack of this element.

For future investigations I would most likely recommend not only conducting inferential statistics with fully vaccinated individuals but also include partially vaccinated people and people who received booster shots. It is important for people in the medical field to conduct investigations such as this one in order to keep people properly informed and as safe as possible.

**Reference list**

"Basics of COVID-19." *CDC*, [www.cdc.gov/coronavirus/2019-ncov/your-health/about-covid-19/](http://www.cdc.gov/coronavirus/2019-ncov/your-health/about-covid-19/)

basics-covid-19.html. Accessed 25 Feb. 2022.

"Coronavirus disease 2019 (COVID-19)." *Mayo Clinic*, [www.mayoclinic.org/diseases-condition](http://www.mayoclinic.org/diseases-condition) s/coronavirus/symptoms-causes/syc-20479963#:~:text=In%202019%2C%20a%20new COVID%2D19). Accessed 23 Feb. 2022.

"Coronavirus disease (COVID-19): Herd immunity, lockdowns and COVID-19." *Word Health*

*Organization*, www.who.int/news-room/questions-and-answers/item/ herd-immunity-lo ckdowns-and-covid-19#:~:text=Vaccines%20train%20our%20immune%20systems,wit hout%20making%20us%20sick. Accessed 1 Mar. 2022.

"COVID-19: Vaccines." *UpToDate*, www.uptodate.com/contents/covid-19-vaccines. Accessed 27 Feb. 2022.

"Create a PivotTable to analyze worksheet data." *Microsoft*, support.microsoft.com/en-us/office/ create-a-pivottable-to-analyze-worksheet-data-a9a84538-bfe9-40a9-a8e9-f9913445657 6. Accessed 28 Feb. 2022.

"Filter data in a range or table." *Microsoft*, support.microsoft.com/en-us/office/filter-data-in-a-ra nge-or-table-01832226-31b5-4568-8806-38c37dcc180e. Accessed 28 Feb. 2022.

"Handling Null Values." *Microsoft*, docs.microsoft.com/en-us/dotnet/framework/data/adonet/sql/ handling-null-values.Accessed 28 Feb. 2022.

"owid / covid-19-data." *Our World in Data*, ourworldindata.org/explorers/coronavirus-data-expl orer?facet=none&Metric=Confirmed+deaths&Interval=7-day+rolling+average&Relati ve+to+Population=true&Color+by+test+positivity=false&country=USA~ITA~CAN~ DEU~GBR~FRA~JPN~AFG~Africa~ALB~OWID\_WRL. Accessed 22 Feb. 2022.

"Possible Side Effects After Getting a COVID-19 Vaccine." *CDC*, [www.cdc.gov/coronavirus/20](http://www.cdc.gov/coronavirus/20) 19-ncov/vaccines/expect/after.html.

"Running a t-test in Excel." *rwu.edu*, [www.rwu.edu/sites/default/files/downloads/fcas/mns/runni](http://www.rwu.edu/sites/default/files/downloads/fcas/mns/runni) ng\_a\_t-test\_in\_excel.pdf. Accessed 1 Mar. 2022.

"Tracking Coronavirus Vaccinations Around the World." *The New York Times*, [www.nytimes.com/interactive/2021/world/covid-vaccinations-tracker.html. Accessed 25](http://www.nytimes.com/interactive/2021/world/covid-vaccinations-tracker.html.%20Accessed%2025) Feb. 2022.

**Appendix**Table 1……………………………………………………………………………………………4

Table 2……………………………………………………………………………………………5

Table 3……………………………………………………………………………………………7

Table 4……………………………………………………………………………………………7

Figure 1…………………………………………………………………………………………...6

Figure 2…………………………………………………………………………………………...6

**Google Drive folder**

<https://drive.google.com/drive/folders/15CGO5sb8Z3hxi4wlWUmUxxTR1qZZ1w1u?usp=sharing>

**Contents**

* Python Code
* Python Code Screenshot
* Raw Data Spreadsheet
* Pivot Table of Processed Data