Report on the variation of Covid-19 Case types on population of different Country Regions.

Shellah Nsubuga, Joyce Adee, Atim Sarah, Annet Balinabyo.

**Introduction**

This is a report on *the variation of Covid-19 Case types on population of different Country Regions.*

COVID-19, the respiratory disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first reported in Wuhan, China in late December 2019

[[1]](https://globalizationandhealth.biomedcentral.com/articles/10.1186/s12992-020-00643-7#ref-CR1). On January 30, 2020, SARS-CoV-2 was declared a public health emergency of international concern, having spread to multiple countries outside of China [[2]](https://globalizationandhealth.biomedcentral.com/articles/10.1186/s12992-020-00643-7#ref-CR2). By March 11, 2020, it was declared a global pandemic, with approximately 118,000 confirmed cases and nearly 4300 deaths on all continents except Antarctica [[3]](https://globalizationandhealth.biomedcentral.com/articles/10.1186/s12992-020-00643-7#ref-CR3).

The COVID-19 pandemic is a catastrophe taking an enormous toll on humanity disrupting lives and livelihoods. The scale and severity of COVID-19 is unprecedented. Different regions have been affected unequally by COVID-19.

In a fast-evolving pandemic it is not a simple matter to identify the countries that are most successful in making progress against Covid-19. For a comprehensive assessment, *w*e based our attention on specific country regions driven by the following questions;

* How many new cases are being confirmed each day?
* How many cases have been confirmed since the pandemic started?
* How is the number of cases changing?
* How many deaths from COVID-19 have been reported?
* Is the number of deaths rising or falling?
* How does the death rate compare to other countries?

We are more interested in knowing how different covid-19 case types varied in different country regions. Basing on the Covid-19 Case.csv dataset, we analyse the cases that were involved in different regions across the globe.

**Data**

Our main interest is in knowing how different covid-19 case types varied on population in different country regions. Basing on the Covid-19 Case.csv dataset, we analyse the cases that were involved in different regions across the globe.

We used numerical data variable of Confirmed cases, Deaths and Population 2019 based on our dataset.

We investigated the data on confirmed cases, deaths and the population 2019 of different regions across the globe.

The following table shows the descriptive statistics for our covid\_19 data set.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variables | mean | median | Standard deviation | Minimum  (Afghanistan)  (High-income) | Maximum (Zimbabwe)    (Upper\_middle\_income) |
| Population density 2018 | 2.096 | 8.395 | 6.423 | 2.041 | 7952.998 |
| Mortality Rate | 2.552 | 2.004 | 2.807 | 0.000 | 28.916 |
| GDP\_2019 | 5.037 | 4.732 | 2.061 | 429016605.2 | 21400000000000.0 |
| GDP\_per\_ca p\_2019 | 2.218 | 1.475 | 2.203 | 782.817 | 121292.739 |
| Avg temp | 1.851 | 2.185 | 8.180 | -5.1 | 28.290 |
| Normalised cases | 5.142 | 2.313 | 7.289 | 0.000 | 0.044 |
| Normalised death | 1.256 | 3.815 | 2.078 | 0.000 | 0.001 |
| Population\_2 019 | 4.177 | 9.758 | 1.495 | 33860 | 1397715000 |
| Deaths | 5.453 | 2.280 | 2.108 | 0 | 204486 |
| Confirmed  cases | 1.799 | 1.338 | 7.728 | 19 | 7078039 |

This means that death has the highest mean with 5.453 and the lowest being normalised deaths with 1.256

The variable with the highest standard deviation is Avg temp with 8.180 (average temperature) and the variable with the lowest standard deviation is population with 1.495.

The country with the highest covid\_19 Confirmed cases is Zimbabwe which is 7078039 a middle-income country

The country with the lowest covid\_19 Confirmed cases is Afghanistan which is 19.

**The following table shows the descriptive statistics for population 2019, death and confirmed cases as our chosen variables**.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Standard deviation | Minimum | Maximum |
| Population\_2019 | 41767618.132 | 9758033.0 | 149477394.325 | 33860 | 1397715000 |
| Death | 5435.187 | 228.0 | 21079.534 | 0 | 204486 |
| Confirmed cases | 179868.121 | 13379.5 | 772821.589 | 19 | 7078039 |

Visualizing our chosen variables

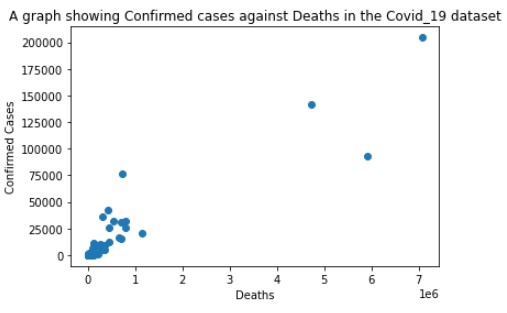
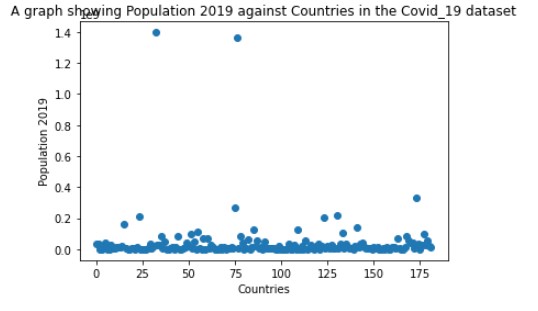
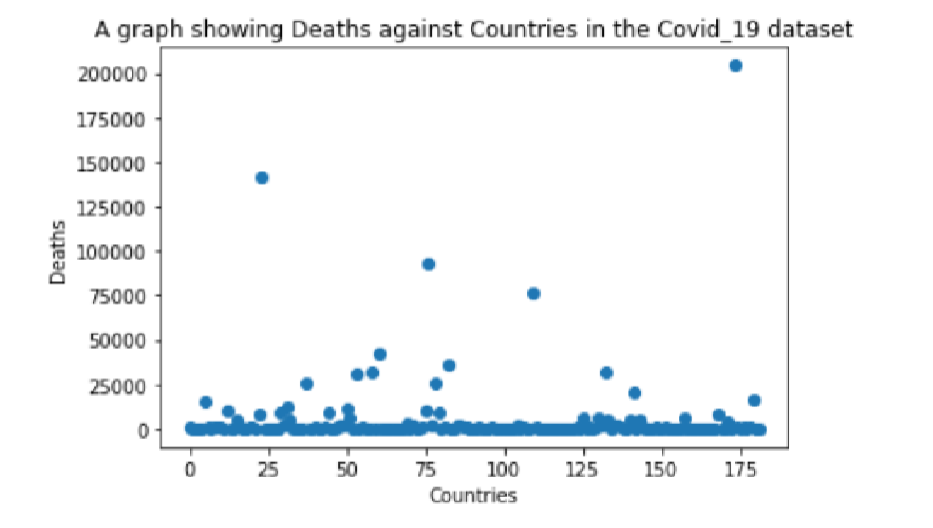
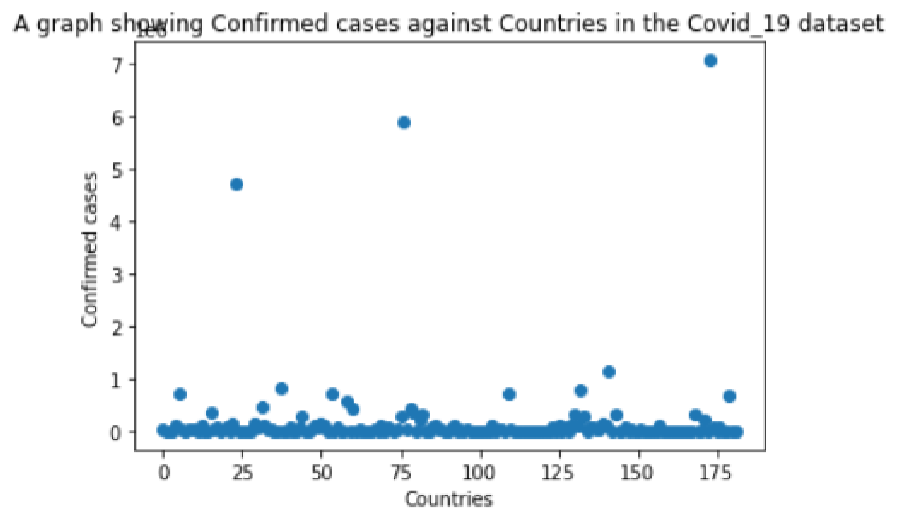
a)

b)

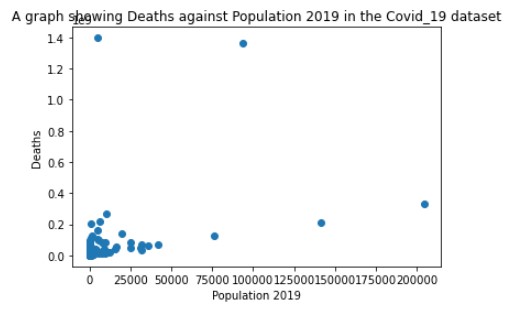
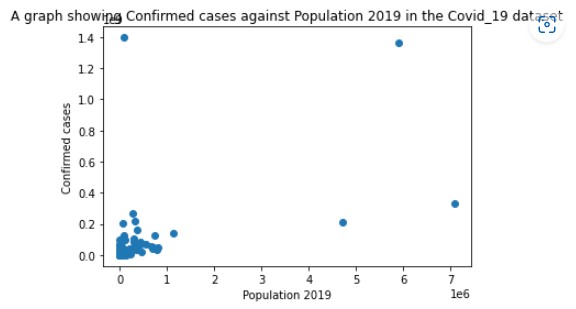
c)

d

)



**e) f)**



**Description of the data according to the used variables;**

* Population 2019
* Confirmed cases
* Deaths

1. The number of confirmed cases is very high in countries between 150 and 175. Basing on the scatter graph of confirmed cases against the population 2019, there is a positive relationship because as the population increases, the confirmed cases are also increasing due to related factors of heavy population in different regions across the globe.
2. Deaths are very high in low income countries. With the variation of deaths and population 2019, the number of deaths increases as the population increases especially in high income countries. In low income countries across the globe, there were relatively low deaths rates though they had relatively high population.
3. Basing on our analysis population is relatively many countries across the globe. Population does not rely on the income group of the countries.
4. Variation of confirmed cases and deaths shows that as the confirmed cases increase, the deaths also relatively increase in some regions across the globe. The deaths rates relatively decrease in some regions though there are high confirmed cases.

**Model**

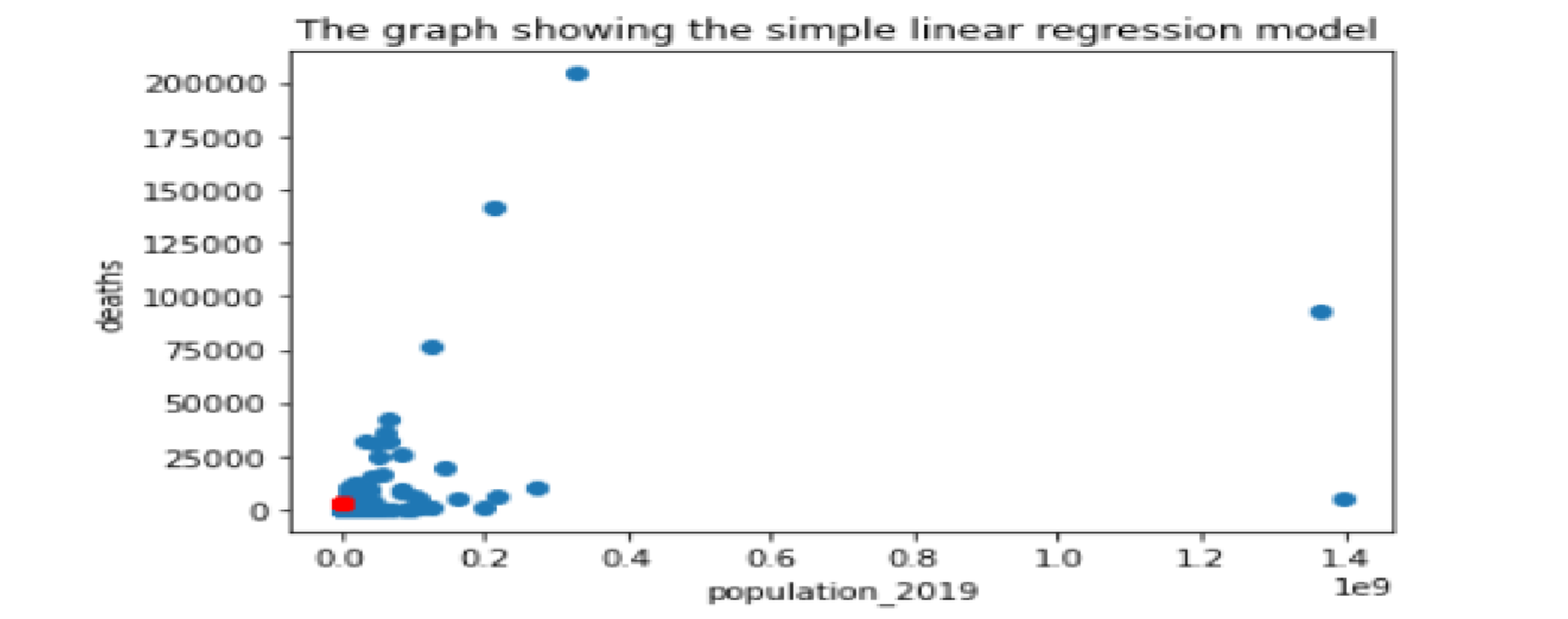
We want to regress the population 2019 by the deaths of Covid-19 because with our analysis, we realised that most countries across the globe suffered relatively high death rates. There was a relatively positive relationship between the Population 2019 and the deaths as our variables basing on the equation: **y=mx + c** using the scatter plot.

Basing on the results, Deaths are very high in low income countries. With the variation of deaths and population 2019, the number of deaths increases as the population increases especially in high income countries. In low income countries across the globe, there were relatively low deaths rates though they had relatively high population.

We perform a hypothesis test on the chosen variables:

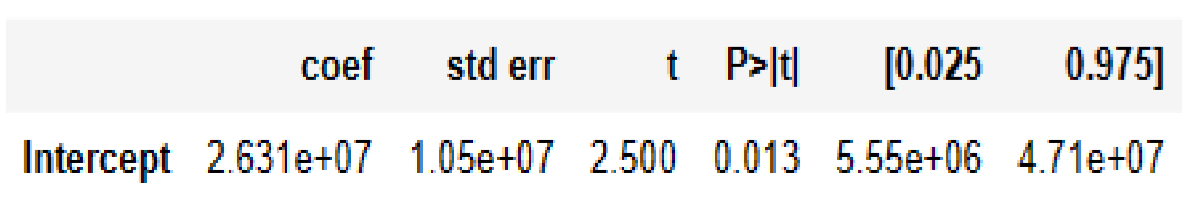
* Population 2019
* Deaths

We perform a simple linear regression model after combining the actual values and the predicted values as shown using the scatter plot below.



With the predicted values in red color and actual values in blue color, the plot shows a relatively perfect regression coz our model is not over predicted with no outliers and predicted values in line with the actual values at the starting point.

The table below shows the results of the summary of the used variables of population 2019 and deaths.



Basing on the results, the summary includes the difference in sample means of (2.631e+07) the p-values have been appropriately rounded and the 5% confidence intervals around the difference in means. Here we 95% certain that the true difference between population 2019 and deaths is somewhere between 4.71e+07 and 5.55e+06.

1. Standard Errors assume that the covariance matrix of the errors is correctly specified.
2. The condition number is large, 2.24e+04. This might indicate that there are strong multicollinearity or other numerical problems.

**Conclusion**

We have analysed the relationship between the population of different countries across the globe and the different Covid -19 case types. Using a linear model, we arrived to the conclusion, that relatively high population in different countries across the globe where highly affected by the different case types; deaths and confirmed cases. Changing the model specification does not seems to have a significant effect on the parameter of interest.

The results of the study basing on the dataset will be beneficial to medical personals who would want to know how different regions were affected by covid-19.