# TB deaths BRICSvsPA ? Gonzalo GomezMillan

November 24, 2017

# 1 Project 1: Deaths by tuberculosis

by Gonzalo Gomez Millan, 22 November 2017, based on Michel Wermelinger version of 14 July 2015

This is the personal project notebook for Week 1 of The Open University's *Learn to code for Data Analysis* course.

In 2000, the United Nations set eight Millenium Development Goals (MDGs) to reduce poverty and diseases, improve gender equality and environmental sustainability, etc. Each goal is quantified and time-bound, to be achieved by the end of 2015. Goal 6 is to have halted and started reversing the spread of HIV, malaria and tuberculosis (TB). TB doesn't make headlines like Ebola, SARS (severe acute respiratory syndrome) and other epidemics, but is far deadlier. For more information, see the World Health Organisation (WHO) page `http://www.who.int/gho/tb/en/`.

Given the population and number of deaths due to TB in some countries during one year, the following questions will be answered:

- What is the total, maximum, minimum and average number of deaths in that year?
- Which countries have the most and the least deaths?
- What is the death rate (deaths per 100,000 inhabitants) for each country?
- Which countries have the lowest and highest death rate?
- How different or similar are the BRICS and PA countries?

The death rate allows for a better comparison of countries with widely different population sizes.

## 1.1 The data

The data consists of total population and total number of deaths due to TB (excluding HIV) in 2013 in each of the BRICS (Brazil, Russia, India, China and South Africa) and PA (Pacific Alliance: Mexico, Colombia, Peru and Chile) countries of the world.

The data was taken in July 2015 from `http://apps.who.int/gho/data/node.main.POP107?lang=en` (population) and `http://apps.who.int/gho/data/node.main.593?lang=en` (deaths). The uncertainty bounds of the number of deaths were ignored.

The data was collected into two Excel files (One for BRICS and one for PA) which should be in the same folder as this notebook.

## 1.2   The BRICS countries

Let's see the data of the BRICS countries.

```
In [11]: import warnings
         warnings.simplefilter('ignore', FutureWarning)

         from pandas import *
         data = read_excel('WHO POP TB brics.xls')
         data
```

## 1.3   The range of the problem

The column of interest is the last one.

```
In [12]: tbColumn = data['TB deaths']
```

The total number of deaths in 2013 for BRICS countries is:

```
In [13]: tbColumn.sum()
```

```
Out[13]: 327400
```

The largest and smallest number of deaths in a single country are:

```
In [14]: tbColumn.max()
```

```
Out[14]: 240000
```

```
In [15]: tbColumn.min()
```

```
Out[15]: 4400
```

From 4400 to almost a quarter of a million deaths is a huge range. The average number of deaths, over all countries in the data, can give a better idea of the seriousness of the problem in each country. The average can be computed as the mean or the median. Given the wide range of deaths, the median is probably a more sensible average measure.

```
In [16]: tbColumn.mean()
```

```
Out[16]: 65480.0
```

```
In [17]: tbColumn.median()
```

```
Out[17]: 25000.0
```

The median is far lower than the mean. This indicates that some of the BRICS countries had a very high number of TB deaths in 2013, pushing the value of the mean up.

## 1.4 The most affected

To see the most affected countries, the table is sorted in ascending order by the last column, which puts those countries in the last rows.

```
In [18]: data.sort_values('TB deaths')
```

The table raises the possibility that a large number of deaths may be partly due to a large population. To compare the countries on an equal footing, the death rate per 100,000 inhabitants is computed.

```
In [19]: populationColumn = data['Population (1000s)']
         data['TB deaths (per 100,000)'] = tbColumn * 100 / populationColumn
         data
```

To see the least and most affected countries, the table is sorted in ascending order by the death rate.

```
In [20]: data.sort_values('TB deaths (per 100,000)')
```

## 1.5 The PA countries

Let's see the data of the PA countries.

```
In [21]: import warnings
         warnings.simplefilter('ignore', FutureWarning)

         from pandas import *
         data = read_excel('WHO POP TB pa.xls')
         data
```

## 1.6 The range of the problem

The column of interest is the last one.

```
In [22]: tbColumn = data['TB deaths']
```

The total number of deaths in 2013 for PA countries is:

```
In [23]: tbColumn.sum()
```

```
Out[23]: 5490
```

The largest and smallest number of deaths in a single country are:

```
In [24]: tbColumn.max()
```

```
Out[24]: 2300
```

```
In [25]: tbColumn.min()
```

```
Out[25]: 220
```

From 220 to 2300 deaths is a large range. The average number of deaths, over all countries in the data, can give a better idea of the seriousness of the problem in each country. The average can be computed as the mean or the median. Given the wide range of deaths, the median is probably a more sensible average measure.

```
In [26]: tbColumn.mean()
```

```
Out[26]: 1372.5
```

```
In [27]: tbColumn.median()
```

```
Out[27]: 1485.0
```

The median is a little higher than the average. This indicates that some of the PA countries had a very low number of TB deaths in 2013, pushing the value of the mean down.

## 1.7   The most affected

To see the most affected countries, the table is sorted in ascending order by the last column, which puts those countries in the last rows.

```
In [28]: data.sort_values('TB deaths')
```

The table raises the possibility that a large number of deaths may be partly due to a large population. To compare the countries on an equal footing, the death rate per 100,000 inhabitants is computed.

```
In [29]: populationColumn = data['Population (1000s)']
         data['TB deaths (per 100,000)'] = tbColumn * 100 / populationColumn
         data
```

To see the least and most affected countries, the table is sorted in ascending order by the death rate.

```
In [30]: data.sort_values('TB deaths (per 100,000)')
```

## 1.8   Conclusions

The BRICS countries had almost than 328 thousand deaths due to TB in 2013. The median shows that half of these coutries had fewer than 25,000 deaths. The much higher mean (over 65,480) indicates that some countries had a very high number. The least affected country was Brazil, with 4400 deaths, and the most affected was India with 240 thousand deaths in a single year. However, taking the population size into account, the most affected was South Africa with over 47 deaths per 100,000 inhabitants.

On the other hand, the PA countries had almost than 5,490 deaths due to TB in 2013. The median shows that half of these coutries had fewer than 1,485 deaths. The lower mean (over 1,372) indicates that some countries had a low number. The least affected country was Chile, with 220 deaths, and the most affected was Peu with 2300 deaths in a single year. Now, taking the

population size into account, the most affected was Peru over 7 deaths per 100,000 inhabitants having a particular dramatically situation against the other AP countries.

Looking at the numbers we can see a wide gap between average and median for the BRICS countries, unlike the AP countries, so we conclude that the health problems of the PA countries are more homogeneous than for the BRICS countries. That could be because they are geographically close to each other and share the same language and similar culture.

One should not forget that most values are estimates. Nevertheless, they convey the message that TB is still a major cause of fatalities, than there is not a correlation between deaths and size of population, and that there is a huge disparity between countries, with several ones being highly affected.