TB deaths all world ? Gonzalo GomezMillan

November 24, 2017

1 Project 1: Deaths by tuberculosis

by Michel Wermelinger, 14 July 2015, with minor edits on 5 April 2016 and 22 Novembre 2017

This is the 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?

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 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 an Excel file which should be in the same folder as this notebook.

```
In [32]: import warnings
         warnings.simplefilter('ignore', FutureWarning)

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

1.2   The range of the problem

The column of interest is the last one.

```
In [33]: tbColumn = data['TB deaths']
```

The total number of deaths in 2013 is:

```
In [34]: tbColumn.sum()
```

```
Out[34]: 1072677.97
```

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

```
In [35]: tbColumn.max()
```

```
Out[35]: 240000.0
```

```
In [36]: tbColumn.min()
```

```
Out[36]: 0.0
```

From 0 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 [37]: tbColumn.mean()
```

```
Out[37]: 5529.267886597938
```

```
In [38]: tbColumn.median()
```

```
Out[38]: 315.0
```

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

1.3   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 [39]: 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 [40]: 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 [41]: data.sort_values('TB deaths (per 100,000)')
```

1.4 Conclusions

The whole countries of the world had more than 1 million deaths due to TB in 2013. The median shows that half of these coutries had fewer than 350 deaths. The much higher mean (over 5,500) indicates that some countries had a very high number. The least affected were San Marino and Monaco, with 0 deaths respectively, and the most affected were Nigeria and India with 160 thousand and 240 thousand deaths in a single year. However, taking the population size into account, the most affected were Nigeria and Djibouti with over 90 deaths per 100,000 inhabitants.

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.

In [0]: