

Project 1: Deaths by tuberculosis

by Michel Wermelinger, 14 July 2015

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/> (<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.

The data

The data consists of total population and total number of deaths due to TB (excluding HIV) in 2013 WHO TB in Europe Nordic Countries.

The data was taken from <http://apps.who.int/gho/data/node.main.POP107?lang=en> (<http://apps.who.int/gho/data/node.main.POP107?lang=en>) (population) and <http://apps.who.int/gho/data/node.main.593?lang=en> (<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 [36]: from pandas import *
data = read_excel('WHO_nordic.xls')
data.sort('TB deaths')
```

```
Out[36]:
```

	Country	Population (1000s)	TB deaths
2	Iceland	330	0.93
3	Norway	5043	4.40
4	Sweden	9571	13.00
1	Finland	5426	17.00
0	Denmark	5619	24.00

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 [37]: deathsColumn = data['TB deaths']
         populationColumn = data['Population (1000s)']
         rateColumn = deathsColumn * 100 / populationColumn
```

The least and most affected countries.

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 [38]: data['TB deaths (per 100000)'] = rateColumn
         data.sort('TB deaths (per 100000)')
```

```
Out[38]:
```

	Country	Population (1000s)	TB deaths	TB deaths (per 100000)
3	Norway	5043	4.40	0.087250
4	Sweden	9571	13.00	0.135827
2	Iceland	330	0.93	0.281818
1	Finland	5426	17.00	0.313306
0	Denmark	5619	24.00	0.427122

The sum of Population in the Nordic countries

```
In [39]: popColumn = data['Population (1000s)']
         popColumn
         popColumn.sum()*1000
```

```
Out[39]: 25989000
```

The sum of death in the Nordic countries

```
In [40]: tbColumn = data['TB deaths']
         tbColumn.sum()
```

```
Out[40]: 59.33
```

The mean of a collection of numbers is the sum of those numbers divided by how many there are.

```
In [41]: tbColumn.mean()
```

```
Out[41]: 11.866
```

The median of a collection of numbers is the number in the middle, i.e. half of the numbers are below the median and half are above.

```
In [42]: tbColumn.median()
```

```
Out[42]: 13.0
```

Conclusions

The Europe Nordic Countries countries had a total population of 25 million inhabitants and about 60 deaths due to TB in 2013. The median shows that half of these countries had fewer than 13 deaths. The lower mean (11,8) indicates that some countries had a lower number. The least affected were Norway and Iceland, with 1 and 4 deaths respectively, and the most affected were Finland and Denmark with 17 and 24 deaths in a single year. However, taking the population size into account, the least affected were Norway and Sweden with less than .2 deaths per 100 thousand inhabitants, and the most affected were Finland and Denmark with over 0.30 deaths per 100,000 inhabitants.

One should not forget that most values are estimates, and that the chosen countries are a small sample of all the world's countries. Nevertheless, they convey the message that TB is still a major cause of fatalities, and that there is a huge disparity between countries, with several ones being highly affected.