**“Is Awareness an Important factor in high number of AIDS related deaths?”**

HIV/AIDS has been ranked 4th in Top causes of Death worldwide, it is large and growing causes of death and disease burden in the whole world, especially in sub-Saharan Africa. Few countries even have their 25 % of population infected with HIV and few have even less than 0.1 %. We are just trying to understand what reason of such big differences could be, what all factors are important in having such high number of HIV cases and Finally, we are going to see if Awareness is big factor in affecting the high number of AIDS related deaths. Dataset was prepared which has data for 108 different countries, columns like Number of HIV deaths, Awareness percentage, Country GDP, Life expectancy of countries and average number of schooling years people receive in that country were captured. Models were built using linear regression to predict the number of deaths due to AIDS taking expectancy, GDP, Education and Awareness as independent variables. Most of the countries in the Dataset are developing, hence, their Life expectancy, GDP and Education is going to be low comparatively but the interesting variable here is % Awareness. I would like to see how it behaves in different Linear models and whether it is significant factor or not in determining the Number of deaths per 100,000 people in different countries.

**Problem Question**: Is Awareness (Correct knowledge of HIV) a significant factor in determining the number of AIDS related deaths?

or

Can HIV awareness in the country also reduce the risks of new infection across all ages?

**Data:**

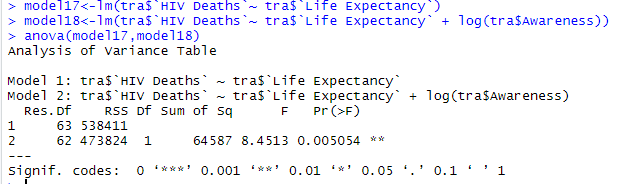
TOP 20 Countries with highest number of HIV Death rate in HIV Data-set :

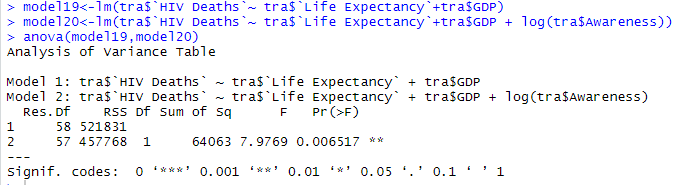
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Country | Year | HIV Deaths (Total number per 100,000 individuals) | Awareness % (correct knowledge of HIV) | Life Expectancy (in ages) | GDP (USD) | Education (years) |
| Lesotho | 2015 | 555.68859 | 31.55 | 53.7 | 173.8289 | 10.7 |
| Swaziland | 2015 | 413.4952 | 0.04 | 58.9 | 3136.925 | 11.4 |
| South Africa | 2015 | 333.7744 | 37.725 | 62.9 | 5769.773 | 13 |
| Mozambique | 2015 | 297.31177 | 32.9425 | 57.6 | 528.3126 | 9.1 |
| Botswana | 2015 | 262.44002 | 0.08 | 65.7 | 6532.651 | 12.6 |
| Namibia | 2015 | 215.48398 | 50.516667 | 65.8 | 4737.67 | 11.7 |
| Central African Republic | 2015 | 188.1316 | 21.385 | 52.5 | 348.3814 | 7.1 |
| Malawi | 2015 | 183.78211 | 43.184091 | 58.3 | 362.6575 | 10.8 |
| Zimbabwe | 2015 | 173.9874 | 53.043333 | 67 | 118.6938 | 10.3 |
| Zambia | 2015 | 168.27112 | 38.58 | 61.8 | 1313.89 | 12.5 |
| Kenya | 2015 | 147.84414 | 53.226667 | 63.4 | 1349.971 | 11.1 |
| Cameroon | 2015 | 138.34643 | 30.173333 | 57.3 | 1244.429 | 10.4 |
| Nigeria | 2015 | 126.87363 | 26.696364 | 54.5 | 2655.158 | 10 |
| Congo | 2015 | 122.12494 | 24.244 | 64.7 |  | 11.1 |
| Guinea-Bissau | 2015 | 120.50259 | 18.756667 | 58.9 | 596.8717 | 9.2 |
| Uganda | 2015 | 95.643026 | 38.933333 | 62.3 | 693.8964 | 10 |
| Togo | 2015 | 93.971306 | 34.71 | 59.9 | 551.1383 | 12 |
| Gabon | 2015 | 77.541776 | 32.85 | 66 | 7388.984 | 12.6 |
| Angola | 2015 | 70.432555 | 32.9 | 52.4 | 3695.794 | 11.4 |
| Gambia | 2015 | 63.893592 | 30.5 | 61.1 |  | 8.9 |

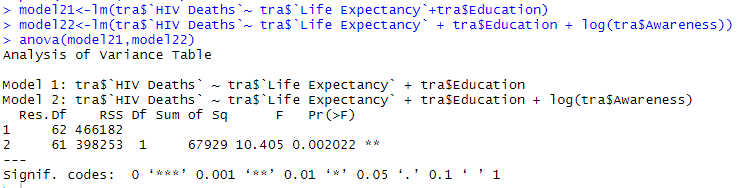
**Technique:**

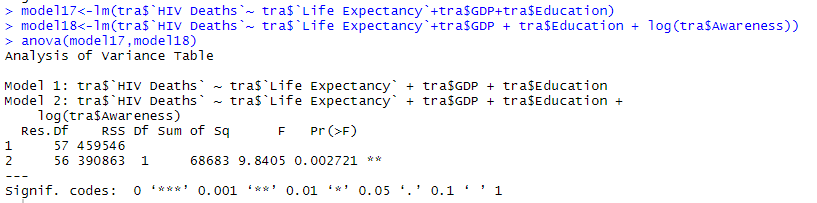
Linear Regression and ANOVA

Different scenarios:









**Limitations/Complications:**

* Data Cleaning and Data Preparation were big tasks to do here, There were lot of missing and improper data which needed to be cleaned
* I collected data from different sources such as UNICEF and Kaggle, there may be people who have HIV infections, but they are not aware of it. So, the numbers can be more from what the data suggests here