Does Schooling impact Life Expectancy?

Executive Summary

By: Chaitanya

"They suggest that improving equity in access to and quality of education is something tangible that can help reverse this troubling trend in reduction of life expectancy among middle-aged adults." The researchers at a multi-institution study led by the Yale School of Medicine and University of Alabama-Birmingham report that 'Level of education, and not race, is the best predictor of who will live the longest'.[1]

Undoubtedly, the improvements in mortality stem from postponing it and thus prolonging both lifespan and healthspan. People are living longer due to being healthier and thus aging diseases are pushed back, developing later in life. Hence in this project, we seek to observe how education or schooling impacts life expectancy through regression analysis. For our *null hypothesis*, we assume that schooling and life expectancy have no relationship between each other. To test this hypothesis, we believe that a data-driven approach will be useful to obtain insights as currently , we do not have any insight into the relationship between education and life expectancy, or how they may impact each other.

Why Data?
We need real-world evidence or data to gain an understanding of these topics and make conclusions about their relationship. With our data fetched from Kaggle[2] – Global Health Repository under World Health Organization (WHO), we aim to test our hypothesis that education impacts human's life expectancy and validate these results with statistically significant tests.

There are 3 possibilities :

    a.   **There is no relationship between education and life expectancy.** Education is an independent factor that does not impact human life expectancy or mortality in any way. In older times, irrespective of the educational status, people have lived for a 100 years and in contrast have also died at a younger age. Hence, it is a possibility that there is no relationship between education and life-expectancy.

    b.   **Education and life expectancy are positively correlated.** An educated person has a better understanding of healthy well-being. We are also assuming that an educated person will have a better job than someone without education, and hence better access to healthcare facilities. Additionally, a well-educated society leads to innovation, thereby making the healthcare industry more advanced and accessible. Hence, we can expect life expectancy to go up with the increase in education.

    c.   **Education and life expectancy are negatively correlated.** A well-educated person is expected to have a better job than someone with lesser education. The better the job is, the more complicated one's work life becomes which consecutively increases

their work pressure and the issues with respect to work-life balance. Thereby, adversely affecting their mental and physical health. People become more focused on making their careers, and money, barely taking any effort for their well-being. Hence, although education brings monetary satisfaction, life expectancy may gradually decrease.
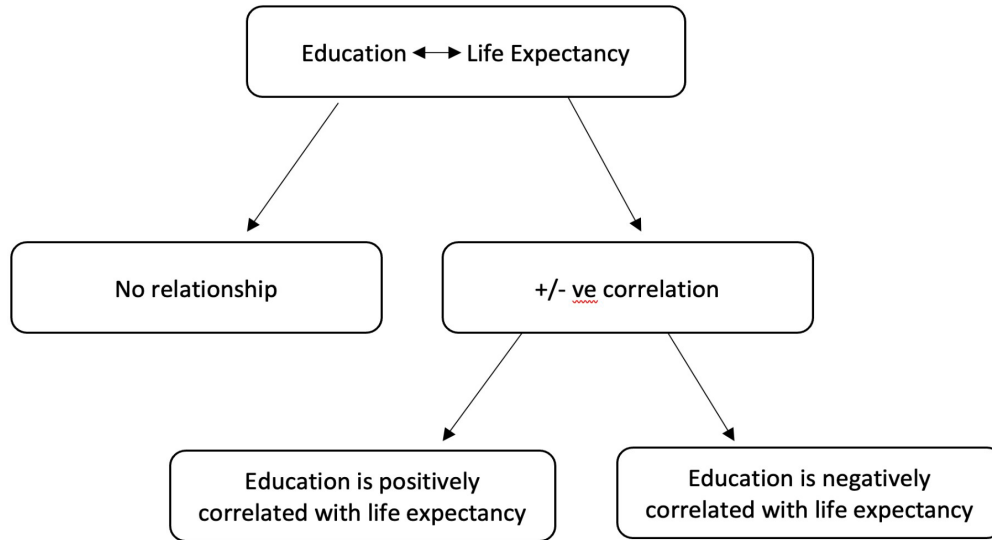
To test these possibilities, an ideal experiment should have a large sample size in order to make the results reliable. They should be random and should get rid of all factors leading to a bias. In this case, the following considerations would make the experiment random:

1. *Impact of Demography* - People within the same region should be taken into consideration because different regions can impact health in different ways as there could be region specific illnesses and also the standard of living would vary.
2. *Genetics bias* - This can play a role in life expectancy regardless of schooling as some people may be prone to terminal genetic diseases which could impact LE negatively.

Cleaning
For our analysis, we are considering a dataset that consists of 22 Columns and 2938 rows which represents 21 predicting variables and each row represents the data recorded during a specific year within a geographic region. We are aggregating based on a country's development *status* and our main independent variable is *schooling* whose impact we will test on our dependent variable - *life expectancy* (LE).
To prevent any potential skewing we used the *Interquartile rule* for outlier detection and handled them by filtering the data which were outside the 1.5IQR range. Missing data can have the potential to reduce the statistical power of the analysis, which can distort the validity of our results. Hence we used *median imputation technique* in which the missing values are replaced with the median value of the entire feature column thus preventing the skewing of our data.

Methodology and Results

Our correlation matrix indicates a positive 0.73 correlation between schooling and life expectancy. Since our hypothesis mainly focuses on schoolings' impact on life expectancy, we use *schooling* as our main independent variable. The variables *income composition of resources* and *alcohol* have the next highest correlations of 0.68 and 0.39 respectively. Hence we keep them as our control variables in our experiment to observe how they impact the relation between *schooling* and *life expectancy*.

We ran a linear regression model against *schooling* and *life expectancy* to reveal that the OLS model had a variability of 0.533 and an *almost zero p-value* leading to reject the null hypothesis and show that schooling impacts life expectancy. With every one year increase in schooling, life expectancy would increase by 2.3 years.
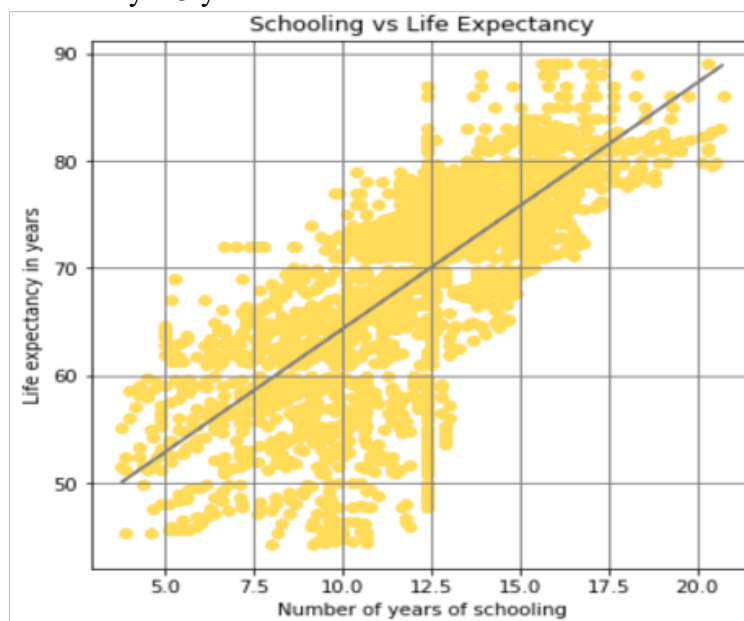
After learning that schooling is a significant factor, we wanted to observe its impact country status wise and hence we introduced the *status* factor as a control variable. The *status* variable is categorical with *developing* and *developed* as its levels. This model had an *r-square* value of 0.564 which was more than the initial model indicating that this model had more variability. And we also observed that the interaction term of *schooling* and *status* had an *almost zero p-value* showing that the *status* term was highly significant in our experiment. The results indicated that, for a developed country, with 1 year increase in schooling the life expectancy increased by 0.7 years and for a developing country, with 1 year increase in schooling the life expectancy increased by 2.2 years.

For our final model we further wanted to observe how schooling would affect LE when controlled by the next highest correlated factors such as *income composition of resources* and *alcohol*. The results had variability of 0.606 and an adjusted r square of 0.605 which tells us that the model was highly reliable and precise.

The interaction between *alcohol* and *status* variable had a *p-value* of 0.334 telling us that distribution between the *status* category with respect to *alcohol* is insignificant. The *schooling* had a positive coefficient of 1.67 and *p-value of almost zero* indicating that when controlled for country *status*, *alcohol* and *income composition* of a country, the positive effect of schooling on life expectancy was still significant. Also, when controlled for *income composition*, *country status,* and *alcohol consumption*, for every unit increase in alcohol consumption, life expectancy decreased.

Conclusion

Schooling greatly influences the life expectancy of a person. More so, in a developing country when compared to a developed country. Reason being, chances of accessing most of the new age amenities increases with education in developing countries, whereas, people already have access to better healthcare, better pharma, better standard of living by default in developed countries.

A developing country is constantly educating themselves, making changes towards advancement and striving to make their amenities better, in turn offering a better chance at healthcare, hence we can expect a better effect of schooling on LE in a developing country over an already developed country.
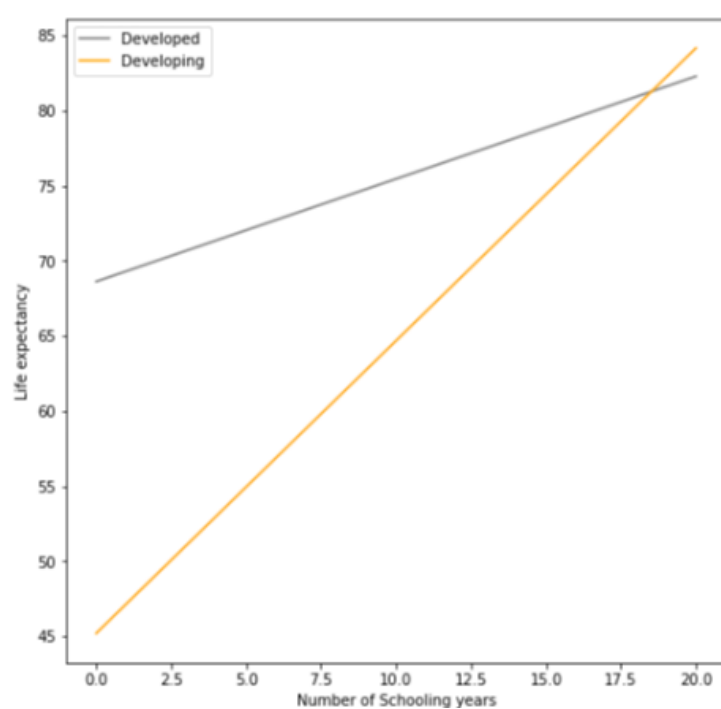
People in developed countries do not need to do schooling mandatorily to bring their life expectancy longer, there may be other factors that are driving them to live longer as we can see that people with least schooling live up to > 60 years, but for a person in developing country, they need to do 10 years of schooling to bring their lifeline above 60 years. A person who does 8 years of schooling in developed countries lives 14 years longer than a person who does the same years of schooling in undeveloped countries.