Technical exercise for Data Scientist position

Yinghui Zheng

- **Part II**

This claim in the email is trying to compare the expected life span of two different population groups with the intention to infer a conclusion that one of the groups is expected to have a longer life span. However, the claim is not comparing apple to apple. The two statistics in the comparison: **WBG Retiree Life Span Estimate** (the average life span of World Bank retirees who already passed away in the past 6 years) and **US Life Span Expectation** (the expected life span of US people who are still alive and reaching 65 years today), are not strictly comparable in the following ways. Therefore, the inferred conclusion should be carefully re-considered.

First of all, the universe based on which the two statistics were calculated are different. US Life Span Expectation is based on the entire population who reached 65 years old in the US, while WBG Retiree Life Span Estimate is only based on WBG retirees who passed away in the past 6 years without considering those that are still living. In order to make the two more comparable, it would be better to estimate the expected life span of WBG retirees also based on both living and passed away retirees.

Secondly, both of the two statistics are calculated based on conditional probability, but the conditions they used are not the same. WBG Retiree Life Span Estimate is conditional on the fact that a staff retired from WBG, while US Life Span Expectation is conditional on a person reaching the age of 65. It is obvious from the attached graph that not all WBG staff reached 65 before they retired, as the age range in the graph goes to as low as 54. In WBG Retiree Life Span Estimate, there is a mix of population who reached the age of 65 and who didn't. To improve the comparison, WBG Retiree Life Span Estimate should also only consider those who reached the age of 65.

Last but not least, the timeline based on which data was calculated do not overlap. WBG Retiree Life Span Estimate is backward looking to the past 6 years, while US Life Span Expectation is forward looking from today to the indefinite future. Besides technology innovations overtime which could help extend people's life expectation, there are a lot more uncertainties in the future compared with past 6 years which are already set in stone. To resolve this issue, one could switch to using data in the past 6 years to estimate US life span, or estimating future WBG retiree life span based on retirees who reached 65 and are still living today.