**Luisa M. Mimmi 22 of February, 2018**

**Part II**

**QUESTION**

A document containing a few paragraphs (Word document or similar).

|  |
| --- |
| **How long do World Bank retirees live?**  We collected information about all World Bank retirees who died within the past 6 years.  We compiled that information, and came up with the following distribution of **919** deceased retirees.  https://lh5.googleusercontent.com/pGabA2oYERnLrrdD-ewnmRub4xCDM9FyOgcrTFXky2f01Rt13N_bhZ30CDCwI8jMoNUpLmh9etnUjIk7GJEN2eJtaJciUXeP4fHHP2h_1bLyNYhaE56MtAgwmVDHeyfwWbl4K44I  It’s a wide distribution. Some were unfortunate enough to expire in their 50s. At the other extreme, one hardy soul lasted until 104! The **mean for this distribution is 81 years old**.  This is not very impressive: compare it to [the US](https://www.ssa.gov/planners/lifeexpectancy.html), where a man and woman **reaching age 65 today can expect to live**, on average, to ages 84.4 and 86.6, respectively. |
| What do you think of the claims made in that email?  Please review and critique. Your response should be concise (~300 to 500 words) |

**ANSWER**

I am afraid my colleague is making the very common mistake of “comparing apples and oranges”. The proposed comparison is not very meaningful for 2 reasons:

1. **The very parameter estimated in the two situations is not the same.** Specifically, in the WBG case, we are looking at the entire population of “World Bank retirees who died within the past 6 years” or, in other words, at the average *life expectancy at birth* of this specific population calculated between the years from 2011 – 2017. (Assuming that mortality patterns –which we know, given that we were able to count the survivors– at the time of birth remain constant in the future).

On the other hand, the reported estimates for the US population reaching age 65 today refer to *life expectancy at age 65.* The intuition is that in this case we are asking about expected lifespan for a group of people who were born in 1953 (= 2018-65). In this case the available information is obviously incomplete: we cannot compute the average lifespan, because (hopefully!) many of them are still alive. With some approximation, most of the deceased WBG retirees were born around year 1937 (= 2018-81). Since we know that the life expectancy has been growing over time, it should not surprise that for a “younger” cohort, there are better prospects.

1. The second reason has to do with **sampling**, that is how to select a subset of population for which I am able to collect information that I wish to generalize to a larger population of interest. Let’s consider a sort of *reductio ad absurdum* argument. Imagine we are studying the life expectancy of the US population as a whole and we learn that the World Bank collects rigorous and complete information about the mortality and life expectancy of its employees. *Is this a* ***representative*** *sample, i.e. one providing measures of central tendency (mean, median, etc.) and dispersion (variance) that you can confidently* ***generalize*** *to the US population?*

My answer is “No!” because the key to obtain a credible (~ representative) sample is to draw it in such a way that all variables affecting the outcome of interest, have the same probability of being observed as in the target population.

This brings me to examine closer the **outcome of interest**: life expectancy at birth. We know that this indicator is closely connected with health conditions, which are in turn an integral part of a country’s level of development, in terms of living standards, access to healthcare, education, etc. In practice, I could capture most of the variation due to these dimensions by using one single proxy variable: “country of birth”. Thus, a reasonable way to address the task at hand could be to select a sample in which the distribution of individuals by “country of birth” is exactly the same as per the US census results. Looking back to our initial question, the WBG staff population is likely much more diverse (by country of birth) than the US population – hence the remarkably high variance of the age of death of WBG retirees. So we have at least one good reason to challenge a credible comparison between the two distributions.

Following the same reasoning, we should not forget that there are other important variables that likely predictors of the life expectancy outcome: such as level of education, health-affecting lifestyle (frequent travel, high stress jobs, etc.). All the more reasons to refrain from quick and inaccurate comparisons.