

STAT3494 Proposal

Julia Andronowitz
University of Connecticut

October 10, 2022

1 Introduction

I will be researching the correlation between a community's access to safe drinking water and their life expectancy. Living in the United States, we often take this necessity for granted, but it is still a pressing issue in many parts of the world. I think it would be interesting to see how strong a possible correlation is. Current research shows that unsafe water is responsible for millions of deaths ([Ritchie and Roser, 2021](#)) and an increase in both access to and quality of safe drinking water can dramatically increase life expectancy ([Angelakis et al., 2021](#)).

2 Specific aims

I am expecting to see a positive correlation between a community's access to drinking water and their recorded life expectancy. Such a study could help emphasize the need for clean water worldwide, as it has numerous health benefits beyond longevity ([Popkin et al., 2010](#)).

3 Data description

I will be using two data sets from the World Health Organization and provided by Kaggle. The first contains information regarding basic drinking water services across various countries and in different years. There are 3455 rows with 4 columns: Location, Period, Indicator, and Tooltip. In this data set, we will be looking at the country, year, and percentage of population using at least basic drinking water services. The second data set contains information about life expectancy in various countries. It has 2197 rows and 5 columns: Location, Period, Indicator, Dim1, and Tooltip. We will be using the country and year columns to look at life expectancy.

4 Research design/methods/schedule

Using the data sets above, I will create one data frame in Python with correlating countries and years to examine the relationship between access to drinking water and average life expectancy. The countries and years will be narrowed down to those only in both data

sets. Additionally, we will only be using rows that include data for both sexes and omitting rows with NA values. The analysis will be done using linear regression (James et al., 2021). Linear regression would help in exploring the research hypothesis as we are looking to see if a correlation exists between two variables and how strong that relationship may be.

5 Discussion

I expect to find a positive correlation between access to drinking water and life expectancy, as clean water is essential for the body to function and many diseases can result from drinking contaminated water. My study can confirm existing notions of this relationship or can explore what other factors may affect life expectancy if a weak correlation is found. This work can impact how people view drinking water and help stress the need for worldwide access to clean water. If the results are not what I am expecting, I would have to look at what other factors may also impact longevity, as a person's health is very complex.

6 Conclusion

This research proposal focuses on investigating the effect of a population's access to clean drinking water on their life expectancy. Insights in this field can further promote the basic right to safe water and highlight just how important water is in human health.

References

- Angelakis, A., H. S. Vuorinen, C. Nikolaidis, P. S. Juuti, T. S. Katko, R. P. Juuti, J. Zhang, and G. Samonis (2021). Water quality and life expectancy: Parallel courses in time. *Water* 13(752).
- James, G., D. Witten, T. Hastie, and R. Tibshirani (2021). *An Introduction to Statistical Learning with Applications in R* (2nd ed.). Springer New York, New York: Springer Science + Business Media, LLC. ISBN 978-1-0716-1417-4.
- Popkin, B. M., K. E. D'Anci, and I. H. Rosenberg (2010). Water, hydration, and health. *Nutrition Reviews* 68(8), 439–458.
- Ritchie, H. and M. Roser (2021). Clean water and sanitation. *Our World in Data*.