

Part I: Stat 404 Group Project

Chris Faires, Jakin Kolars, Jeremy Miranda

1. Give a detailed description for the source of your data.

The data came from the The Global Health Observatory (GHO) data repository under the World Health Organization (WHO) that keeps track of the health status as well as many other related factors for all countries.

2. Give a detailed description of the variables in your data set; including why the data were collected.

From the original data set we took the variables country, year, life expectancy, alcohol consumption, percentage expenditure (of GDP), bmi, and population. Since each country was in the data set 15 times (the years in the data set were from 2000 to 2015) we took out 14 of the years to focus on just 1 year, which will be 2013.

These variables were collected because understanding people's health is extremely important to the wellbeing of everyone all over the world.

- Alcohol: Per capita consumption of alcohol in liters
- Health Expenditure Per Capita: Expenditure on health as a percentage of GDP, per capita
- BMI: Average Body Mass Index of entire country population
- Population: Population of the country

3. Describe your sample. (e.g., How large is it, how were the data collected...).

The data set is provided by the World Health Organization, which tracks health statistics for all countries of the world. The original data set contains data for 193 countries for years between 2000 and 2015. We will analyze a sample of 2013 data for 193 countries. The data were collected from economic data and the WHO data repository website.

4. Define the population your sample was drawn from.

The population the sample was drawn from is meant to include all humans in the world. The country-level variables, such as life expectancy and BMI, represent mean values of individual-level measures. Therefore, the analysis from this sample is meant to be used to make inferences for citizens in all countries of the world.

5. What is your research question(s)? One research question is sufficient, but you should have no more than 3 research questions. Make sure your research “questions” are worded as questions with a ? at the end.

What factors contribute most to increasing life expectancy?

Is life expectancy affected more by physical traits or environmental conditions?

6. Describe your dependent variable (e.g., what is it, what are the units...). Remember this variable needs to be quantitative.

The dependent variable for the data set is life expectancy. It is the average life expectancy for each country in years.

7. Describe each of your independent variables (e.g., what are they, what are the units, are they categorical or quantitative, if it is categorical how was it coded in the .csv file...).

Alcohol: Per capita consumption of alcohol in liters. Quantitative.

Health Expenditure Per Capita: Expenditure on health as a percentage of GDP, per capita. Quantitative.

BMI: Average Body Mass Index of entire country population. Quantitative.

Population: Population of the country. Quantitative.

8. Do you have missing data? If so, how is it coded in the .csv file (e.g., blank, ., 0, 999)?

There are a few countries that have some missing data. In the data set, the missing data is coded with a 0 or a blank.