

STAT 251 - Project

Joshua Carpenter, Cecilia Fu

3/24/2021

Introduction

Between the years of 2000 and 2019, the World Health Organization (WHO) has collected data on causes of death in 183 countries. We are interested particularly in Cardiovascular Disease. The purpose of this analysis will be to determine if cardiovascular disease is a more prevalent cause of death in the United States than elsewhere, that is if the proportion of deaths due to cardiovascular disease is greater in the US than in other countries.

To do this, we will model the proportion of deaths due to cardiovascular disease in the United States as the probability of success of a binomial random variable, where the population consists of all deaths in the United States and a trial consists of sampling one death and determining whether or not the cause was cardiovascular disease. We will consider a success to be that the death was caused by cardiovascular disease and a failure that it was not. We will similarly model the proportion of deaths outside the United States caused by cardiovascular disease as the probability of success of a binomial random variable where the population is all deaths that occurred outside of the United States.

We will then determine an appropriate gamma prior distribution, which we will use for both data distributions; we will run a Bayesian update based on data from WHO; and we will compare the posterior distributions using Monte-Carlo methods. Based on the Monte-Carlo estimated posterior distribution for the difference in proportions, we will determine a 95% confidence interval and conclude whether the proportions are significantly different.

Data

Below is a summary of the data to be used. All values are deaths per thousand.

| ID | Country | Cardio_Disease | Total_Deaths |
|-----|---------|----------------|--------------|
| 1 | AFG | 0.01062 | 254.8099 |
| 2 | ALB | 0.00044 | 31.1542 |
| 3 | DZA | 0.01072 | 203.3004 |
| ... | ... | ... | ... |
| 175 | USA | 1.91123 | 2949.2139 |
| ... | ... | ... | ... |
| 182 | ZMB | 0.0052 | 121.1049 |
| 183 | ZWE | 6e-04 | 117.7098 |

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|----------------|--------|---------|---------|----------|----------|------------|
| Cardio_Disease | 0.0000 | 0.0005 | 0.0027 | 0.0243 | 0.0088 | 1.9112 |
| Total_Deaths | 0.6186 | 20.5604 | 72.8564 | 302.8184 | 187.4312 | 10105.5956 |

Appendix A: Data Source

Global health estimates: Leading causes of death

Cause-specific mortality, 2000–2019

See Global summary estimates

<https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghle-leading-causes-of-death>

Appendix B: Code

```
library(knitr)
opts_chunk$set(echo = FALSE, comment=NA)

library(readxl)
library(tidyverse)
set.seed(3812)
# Read in the data
country_codes <- read_xlsx("deaths2019.xlsx",
  range = "'Deaths All ages'!H8:GH8",
  col_names = FALSE) %>%
  pivot_longer(everything(), names_to = "Names", values_to = "Country") %>%
  select(-Names)
cardio_disease_vals <- read_xlsx("deaths2019.xlsx",
  range = "'Deaths All ages'!H128:GH128",
  col_names = FALSE) %>%
  pivot_longer(everything(), names_to = "Names", values_to = "Cardio_Disease") %>%
  select(-Names)
total_deaths_vals <- read_xlsx("deaths2019.xlsx",
  range = "'Deaths All ages'!H11:GH11",
  col_names = FALSE) %>%
  pivot_longer(everything(), names_to = "Names", values_to = "Total_Deaths") %>%
  select(-Names)
cardio <- bind_cols(country_codes, cardio_disease_vals, total_deaths_vals)
cardio_tail <- cardio %>%
  rownames_to_column("ID") %>%
  tail(2) %>%
  mutate(Cardio_Disease = as.character(round(Cardio_Disease, 4)),
    Total_Deaths = as.character(round(Total_Deaths, 4)))
row_USA <- cardio %>%
  rownames_to_column("ID") %>%
  filter(Country == "USA") %>%
  mutate(Cardio_Disease = as.character(round(Cardio_Disease, 5)),
    Total_Deaths = as.character(round(Total_Deaths, 4)))
cardio_head <- cardio %>%
  rownames_to_column("ID") %>%
  head(3) %>%
  mutate(Cardio_Disease = as.character(round(Cardio_Disease, 5)),
    Total_Deaths = as.character(round(Total_Deaths, 4))) %>%
  add_row(ID = "...", Country = "...",
    Cardio_Disease = "...", Total_Deaths = "...") %>%
  add_row(row_USA) %>%
```

```

    add_row(ID = "...", Country = "...",
            Cardio_Disease = "...", Total_Deaths = "...") %>%
    bind_rows(cardio_tail)
kable(cardio_head, align = "c")

cardio_summ <- cardio$Cardio_Disease %>%
  summary() %>%
  as.matrix() %>%
  t()
death_summ <- cardio$Total_Deaths %>%
  summary() %>%
  as.matrix() %>%
  t()
overall_summ <- rbind(cardio_summ, death_summ) %>%
  round(4)
row.names(overall_summ) <- c("Cardio_Disease", "Total_Deaths")
kable(overall_summ, align = "c")

```