final_project

Emily Phillips

8/1/2021

Introduction

The idea and question of fertility has been a topic that I have been researching for some years now, especially being a young woman. It has become more of a prevalent issue, at least in the United States, for women to struggle with infertility and therefore not being able to have children naturally or maybe at all. In terms of birth rates, I would think that the issue of reproductive health in women (and maybe even men) has an impact on the number of kids that are born year-to-year in the country. Therefore, the problem that I want to research is the problem of infertility in America and how it has impacted birth rate. In order for any country's population to grow, people need to be born, and it could have a huge impact on the population if people experience issues with fertility and reproductive health. With the concept of infertility, I would also like to research how various factors influence one's fertility and whether we can predict the fertility of an individual based on their demographic and environmental variables.

Research questions

1. What is the weight of men and women's reproductive health in influencing a couple's ability to have children?
2. What is the current difference in birth rates from one country to another?
3. What is the average age for women to try to start having children?
4. How have non-traditional methods of having children influenced birth rate, such as adoption/IVF/etc?
5. What resources are provided to people who are experiencing issues with infertility?
6. What role does proper sex education play in fertility and reproductive health?
7. Does the current calculation of birth rate account for non-traditional methods of child delivery?
8. What are the key factors that play a role in one's fertility, men and women?

Approach

I plan to address this problem statement by first building regression models that will allow for me to determine what variables or factors allow for predicting one's fertility or at a more major level, which factors influence/predict the fertility rate of a country. From there, knowing the fertility rate of a region, which indicates the ability of someone to have a child, I want to visualize and compare the birth rates between people who experience fertility issues versus those who don't, and see how it has then impacted the birth rate in that region for the overall population growth of the area. I think both rates (fertility & birth) play into each other, and my approach will try to address how fertility influence both numerical values.

How your approach addresses (fully or partially) the problem.

In terms of partiality, I think it will be difficult to address this problem at a global level (at least within the context of this course) but my approach will partially address the problem by identifying how reproductive health plays into the overall birth rate of a certain area. It will identify how we can predict fertility and

draw a comparison to understand how fertility or lack of it affects birth rate and therefore also population growth.

Data (Minimum of 3 Datasets - but no requirement on number of fields or rows)

1. Kaggle Fertility Dataset

- 100 volunteers provided a semen sample analyzed according to the WHO 2010 criteria
- Sperm concentration are related to socio-demographic data, environmental factors, health status and life habits
- 100 rows
- 10 attributes/columns
- Date the data was donated: 2013-01-17
- No missing values
- Original dataset can be found at UCI Machine Learning Repository

2. World Bank Data

- Collected data from 1960 to 2016
- Data pertaining to countries' population, fertility rate and life expectancy
- Three different CSV files:
  - country_population.csv (61 columns)
  - fertility_rate.csv (61 columns)
  - life_expectancy.csv (61 columns)
- Collected from over the years but downloaded in 2018
- 264 rows in each CSV file

3. Female Respondent Data File
4. Female Pregnancy Data File
5. Male Respondent Data File

- All three above datasets:
  - Original source: National Survey of Family Growth
  - Collected from 2017-2019
  - Collection methods:
    - In-person interviews by trained female interviewers, in respondents' homes
  - Number of columns:
    - Female Respondent (3087)
    - Female Pregnancy (244)
    - Male Respondent (3059)
  - Handling of missing data
    - Only included completed cases
    - If any survey participants answered "don't know", "refused" or "not ascertained", they were given a default value .. i.e., 9, 8, 7, etc.
    - A case was defined as being complete if the respondent answered the last applicable question before ACASI
    - The small number of respondents who did not complete the ACASI section, partially or completely, will have "not ascertained" values assigned to all variables after their break-off point
  - Survey Info

Required Packages

- readxl
- tidyverse
- car
- QuantPsyc
- boot
- ggplot2

Plots and Table Needs

What types of plots and tables will help you to illustrate the findings to your research questions?

- Scatterplots – understanding relationship between variables
- Residual plots – assessing outliers
- Histograms – assess normality of the data sets
- Boxplots for comparing distributions among groups of study participants
- Tables:
  - Population of each country
  - Current birth rate in each country
  - Historical birth rates in each country
  - Current/historical fertility rates
  - Assessments of reproductive health for sample participants
  - Reproductive health comparison between men and women

Questions for future steps

- Assessing categorical variables when used in regression
- Generalization of analysis on one region to the greater regions – EX) USA to the world
- Merging of datasets, especially when bringing in data on men & women