Final Report of Traineeship Program 2023

On

Analyze Death Age Difference of Right Handers with Left Handers

MEDTOUREASY

Submitted By:

Md Farhan Ashraf

ACKNOWLDEGMENTS

My experience with MedTourEasy during my traineeship was truly invaluable. It provided me with an excellent opportunity to delve into the world of Data Visualizations within Data Analytics, fostering both personal and professional growth. I'm immensely grateful for the chance to collaborate with seasoned professionals who generously shared their expertise and guided me through the intricacies of the field.

I extend my heartfelt appreciation to the Training & Development Team at MedTourEasy for entrusting me with this opportunity. Their support was instrumental in helping me grasp the nuances of the Data Analytics domain and equipping me with the necessary skills to execute the project effectively, ensuring maximum client satisfaction. Despite their busy schedules, they dedicated their time to mentor me, for which I am truly thankful.

Table of Content

Introduction

1.1 About the Company

MedTourEasy, a global healthcare company, provides you the informational resources needed to evaluate your global options. MedTourEasy provides analytical solutions to our partner healthcare providers globally

1.2 About the Project

In this project, we will explore this phenomenon using age distribution data to see if we can reproduce a difference in average age at death purely from the changing rates of left-handedness over time, refuting the claim of early death for left-handers. This notebook uses pandas and Bayesian statistics to analyze the probability of being a certain age at death given that you are reported as left-handed or right-handed.

A National Geographic survey in 1986 resulted in over a million responses that included age, sex, and hand preference for throwing and writing. Researchers Avery Gilbert and Charles Wysocki analyzed this data and noticed that rates of left-handedness were around 13% for people younger than 40 but decreased with age to about 5% by the age of 80. They concluded based on analysis of a subgroup of people who throw left-handed but write right-handed that this age-dependence was primarily due to changing social acceptability of left-handedness. This means that the rates aren't a factor of *age* specifically but rather of the *year you were born*, and if the same study was done today, we should expect a shifted version of the same distribution as a function of age. Ultimately, we'll see what effect this changing rate has on the apparent mean age of death of left-handed people, but let's start by plotting the rates of left-handedness as a function of age.

1.3 Objectives and Deliverables

Objectives:
1. Investigate the impact of changing rates of left-handedness on the observed age gap between left-handed and right-handed individuals.
2. Assess the implications of increasing rates of left-handedness over time for the longevity of left-handed individuals.
3. Evaluate potential factors contributing to the observed differences in age at death between left-handed and right-handed individuals, including societal attitudes and demographic shifts.

Deliverables:
1. Analysis of the relationship between changing rates of left-handedness and differences in age at death between left-handed and right-handed individuals.
2. Interpretation of findings regarding the potential impact of increasing left-handedness rates on the longevity of left-handed individuals, providing insights into health outcomes associated with handedness.
3. Discussion of the role of societal attitudes and demographic trends in shaping observed differences in age at death, highlighting the complex interplay of biological and social factors.

Methodology

2.1 Flow of the Project

Here's a description of the flow of the project for writing the project report which will be explained in detail in the following steps:
❖ Introduction
❖ Data Collection and Preparation
❖ Exploratory Data Analysis
❖ Bayesian Analysis
❖ Results
❖ Conclusion
❖ References
❖ Appendices

2.2 Language and Platform Used
Python:
- Data Cleaning and Preprocessing: Python was utilized for data cleaning and preprocessing tasks, such as handling missing values, converting data formats, and preparing the datasets for analysis. Libraries like pandas were used for efficient data manipulation and cleaning.
- Statistical Analysis: Python, along with libraries like NumPy was employed for statistical analysis, including exploratory data analysis (EDA) to understand the distributions and relationships in the data, as well as Bayesian analysis to estimate probabilities and assess the impact of changing rates of left-handedness.
- Data Visualization: Python's matplotlib weas utilized for data visualization, allowing for the creation of informative plots and visualizations to illustrate trends, patterns, and results derived from the analysis.

Implementation

3.1 Gathering Requirements and Defining Problem Statement

1. Define the Objective:
The objective is to investigate the phenomenon of early death among left-handed individuals and assess the impact of changing rates of left-handedness over time on the average age at death.
Determine whether the reported nine-year difference in average age at death between left-handed and right-handed individuals is supported by the data.

2. Data Requirements:
Left-handedness survey data: Information on age, sex, and hand preference for throwing and writing, covering a wide range of birth years.
Death distribution data: Age-specific mortality rates for the population, preferably from a period close to the study year.

3. Data Sources:
National Geographic survey data from 1986: Provides insights into the prevalence of left-handedness across different age groups.
Death distribution data for the United States from 1999: Offers age-specific mortality rates necessary for Bayesian analysis.

4. Tools and Technologies:

Python programming language: For data manipulation, statistical analysis, and visualization.
Libraries such as pandas, matplotlib, and numpy: To handle data, create visualizations, and perform Bayesian analysis.

5. Problem Statement:

The problem is to determine whether there is a significant difference in the average age at death between left-handed and right-handed individuals, considering the changing rates of left-handedness over time.

3.2 Data Collection and Importing

The data for the project was sourced from:

- National Geographic survey data from 1986: Provides insights into the prevalence of left-handedness across different age groups.
- Death distribution data for the United States from 1999: Offers age-specific mortality rates necessary for Bayesian analysis.

3.3 Designing Databases

1. Tables:

LeftHandednessSurvey:
Columns: ParticipantID (Primary Key), Age, Sex, HandPreferenceThrowing, HandPreferenceWriting
DeathDistribution:
Columns: Age, NumberDeceased (Both Sexes)

2. Data Types:

ParticipantID: Integer (Primary Key)
Age: Integer
Sex: Text (Male, Female)
HandPreferenceThrowing: Text (Left, Right)
HandPreferenceWriting: Text (Left, Right)
NumberDeceased: Integer

3. Relationships:

LeftHandednessSurvey.ParticipantID → DeathDistribution.Age (Foreign Key)
By designing databases with these considerations, we ensure efficient data storage, retrieval, and management, supporting the analytical requirements of the code provided.

3.4 Data Cleaning

- Loading Data: The code loads two datasets:

Left-handedness survey data (lefthanded_data): Contains information about age, sex, and hand preference for throwing and writing.

Death distribution data (death_distribution_data): Provides the number of deceased individuals at different ages.

- Handling Missing Values:

NaN values are dropped from the Both Sexes column in the death distribution data using the dropna() function.

- Creating new columns:

Birth year column (Birth_year) is created in the left-handedness survey data by subtracting the age from 1986.

Mean left-handedness column (Mean_lh) is calculated as the average of left-handedness rates for male and female participants.

3.5 Data Filtering

- Left-handedness Survey Data Filtering:

Any missing or invalid values in the dataset have been filtered out during the data loading process.

- Death Distribution Data Filtering:

NaN values in the Both Sexes column are explicitly filtered out using the dropna() function, ensuring that only valid data points are considered for further analysis.

4. **Sample Screenshots and Observations**
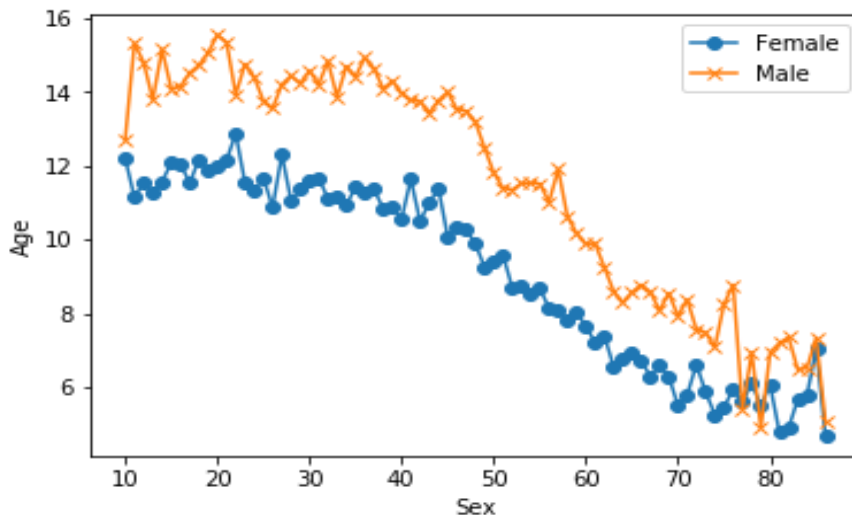
- The graph shows the left-handedness rates for males and females across different ages where the x-axis represents the age while the y axis represents the percentage of left handed people.
- By observing the graph, you can see how left-handedness rates vary with age and whether there are any differences between males and females in terms of left-handedness prevalence across different age groups.
- The observations clearly shows the increase in percentage of left handed people with decrease in age which could be due to Varity of reason such as acceptance in society.
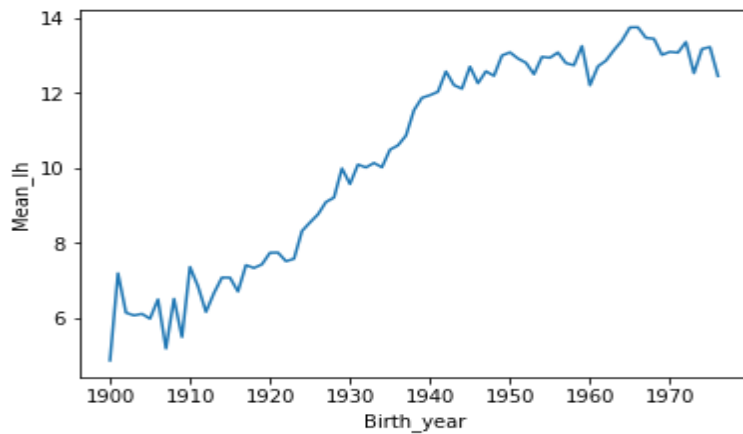
- The x-axis represents birth year and y axis representing the mean left-handedness rate for each birth year group.
- By observing the graph it clearly solidifies the previous graphs conclusion of increase in the number of left-handed people with increase in every consecutive year
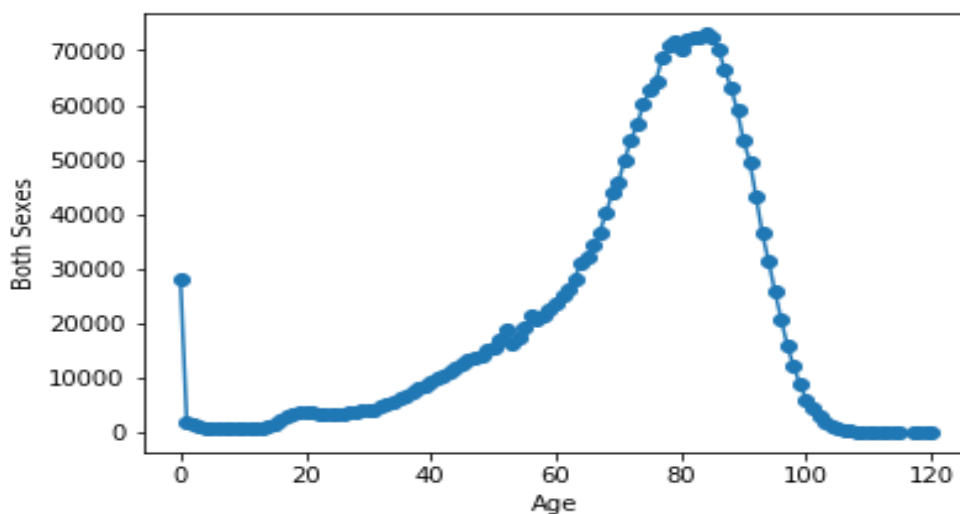
- This graph provides insights into mortality rates, showing how the number of deaths varies with age
- Peaks and troughs in the plot indicate age groups with higher or lower mortality rates, respectively. Analysing the shape of the curve can help identify trends such as infant mortality, age-related diseases, and life expectancy
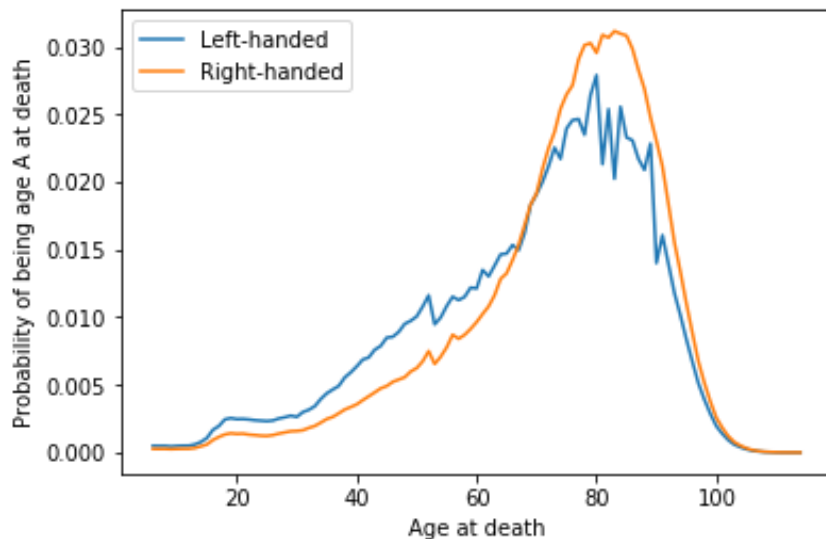
- The x-axis represents the age at death, ranging from 6 to 115 years while the y-axis indicates the probability of being a certain age at death. Two lines are plotted: one for left-handed individuals and another for right-handed individuals.
- Peaks and troughs in the curves indicate age groups with higher or lower probabilities of death for left-handed and right-handed individuals, respectively. Comparing the two curves allows for insights into any differences in mortality patterns between left-handed and right-handed individuals across different age groups.
- This graph indicates on a higher probability of dying earlier for a left handed people when compared to right handed. Which is further proved in subsequent codes showing an average difference of 5.5 years.

Conclusion & Future Scope

- Mortality rate: The analysis clearly showed a difference with left handed people peaking below 70 which is significantly less than their right handed counterpart.
- Effect of Changing Rates of Left-Handedness: By applying Bayes' theorem and considering the changing rates of left-handedness over time, we were able to show that the reported nine-year age gap between left-handed and right-handed individuals is likely influenced by the changing rates of left-handedness in the population. However, our calculated age gap of 5.5 years is still lower than the reported gap, suggesting that some approximations or factors not considered in the analysis may contribute to the observed difference.

Future Scope:
- Refinement of Data Analysis: Further research could involve refining the data analysis techniques and addressing potential limitations such as the extrapolation of left-handedness rates for older and younger age groups. Using more precise methods for estimating age distributions and handedness rates could provide more accurate results.
- Comparative Studies: Comparing handedness trends and mortality patterns across different populations and geographical regions could offer valuable comparative insights. Understanding how cultural, environmental, and demographic factors influence handedness and age-related health outcomes could contribute to more comprehensive conclusions.
- This analysis can help companies prepare their stocks, analyse their target audience for more target specific ads also

understanding handedness opens up the path for companies to bring out more products which can help left-handed people improve their lifestyle and make it easier for them.

Reference:
https://www.cdc.gov/nchs/data/statab/vs00199_table310.pdf
https://www.cdc.gov/nchs/nvss/mortality_tables.htm
https://pubmed.ncbi.nlm.nih.gov/1528408/