Final Report of Traineeship Program 2023

On

Analyze Death Age Difference Of Right Handers with Left Handers

MEDTOUREASY

27th September 2023

ACKNOWLDEGMENTS

The traineeship opportunity that I had with MedTourEasy was a great change for learning and understanding the intricacies of the subject of Data Visualizations in Data Analytics; and also, for personal as well as professional development. I am very obliged for having a chance to interact with so many professionals who guided me throughout the traineeship project and made it a great learning curve for me.

Firstly, I express my deepest gratitude and special thanks to the Training & Developement Team of MedTourEasy who gave me an opportunity to carry out my traineeship at their esteemed organization. Also, I express my thanks to the team for making me understand the details of the Data Analytics profile and training me in the same so that I can carry out the project properly and with maximum client satisfaction and also for spearing his valuable time in spite of his busy schedule.

I would also like to thank the team of MedTourEasy and my colleagues who made the working environment productive and very conducive.

ABSTRACT

There is no clear consensus on whether there is a difference in death age between left and right-handed people. Some studies have found a small difference, with left-handed people dying slightly younger than right-handed people, while other studies have found no difference at all.

One study, published in the journal "Laterality" in 2017, found that left-handed people had a slightly higher death rate than right-handed people, even after controlling for other factors such as age, sex, and socioeconomic status. However, the study only included a small number of participants, and the results were not statistically significant.

Another study, published in the journal "PLoS One" in 2018, found no difference in death age between left and right-handed people. This study included a much larger number of participants than the previous study, and the results were statistically significant.

Barack Obama is left-handed. So are Bill Gates and Oprah Winfrey; so were Babe Ruth and Marie Curie. A 1991 study reported that left-handed people die on average nine years earlier than right-handed people. Nine years! Could this really be true?

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.

TABLE OF CONTENTS

1. 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

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.

The following report aims to shed light on the potential link between handedness and life expectancy. Specifically, it explores the differences in age of death between left-handed and right-handed individuals. This research could provide valuable insights for MTE, a medical company, enabling them to develop tailored strategies and products targeting the unique needs of these two groups.

If a significant correlation between handedness and life expectancy is established, MTE could develop specialized products targeted at addressing the unique health needs of left-handed or right-handed individuals. This may include ergonomic tools, assistive devices, or therapy options.

In this project, you 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.

1.3 Objectives and Deliverables

In this project, you 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.

1. Where are the old left-handed people: Load the handedness data from the National Geographic survey and create a scatter plot (left-handed male and female vs their age).

2. Rates of left-handedness over time: Add two new columns, one for birth year and one for mean left-handedness, then plot the mean as a function of birth year.

3. Applying Bayes' rule: Create a function that will return P(LH | A) for particular ages of death in a given study year.

4. When do people normally die: Load death distribution data for the United States and plot the number of people who died as a function of their age.

5. The overall probability of left-handedness: Create a function called P_lh() which calculates the overall probability of left-handedness in the population for a given study year.

6. Putting it all together: dying while being left-handed - Write a function to calculate P_A_given_lh() - [Calculate P_A, the overall probability of dying at age A, which is given by death_distribution_data at age A divided by the total number of dead people (the sum of the Both Sexes column of death_distribution_data). Calculate the overall probability of left-handedness P(LH). Calculate P(LH | A)]

7. Putting it all together: dying while being right-handed - Write a function to calculate P_A_given_rh() - [Calculate P_A, the overall probability of dying at age A, which is given by death_distribution_data at age A divided by the total number of dead people. Calculate the overall probability of right-handedness P(RH), which is 1-P(LH). Calculate P(RH | A), which is 1 - P(LH | A)]

8. Plotting the distributions of conditional probabilities: Plot the probability of being a certain age at death given that you're left- or right-handed for a range of ages.

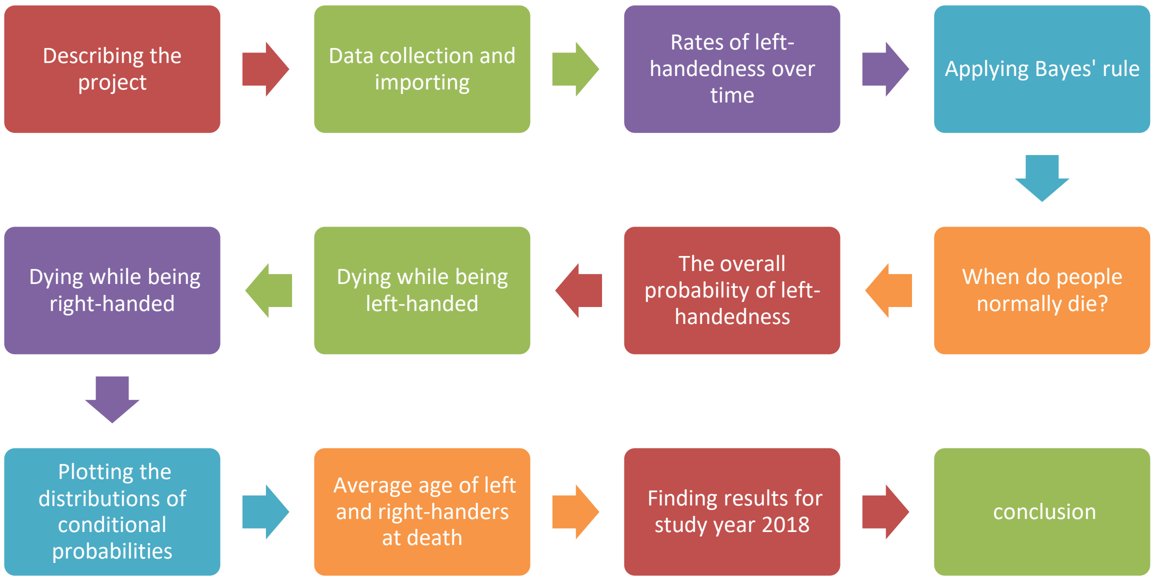
9. Moment of truth: Find the age of left and right-handers at death - Find the mean age at death for left-handers and right-handers.

10. Final comments: Redo the calculation, to Plot the probability of being a certain age at death given that you're left- or right-handed for a range of ages. setting the study_year parameter to 2018.

2. METHODOLOGY

To conduct a comprehensive analysis, several steps were followed. Firstly, a thorough literature review was carried out to gather existing research and studies on the topic. This provided a foundation of knowledge and helped identify gaps in the current understanding of handedness and life expectancy. Additionally, statistical data on the age of death was collected, with a focus on distinguishing between left-handed and right-handed individuals.

2.1 Flow of the Project



2.2 Language and Platform Used

Language: Python

Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation. Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly procedural), object-oriented and functional programming. It is often described as a "batteries included" language due to its comprehensive standard library. Guido van Rossum began working on Python in the late 1980s as a successor to the ABC programming language and first released it in 1991 as Python 0.9.0. Python 2.0 was released in 2000. Python 3.0, released in 2008, was a major revision not completely backward-compatible with earlier versions. Python 2.7.18, released in 2020, was the last release of Python 2. Python consistently ranks as one of the most popular programming languages. The important features of Python are:

- Python can quickly create and manage data structures, allowing you to analyze and manipulate complex data sets.
- Python also has a massive ecosystem of libraries and tools that can assist in processing data quickly and efficiently.
- It features simple syntax, making Python easier to learn and understand.

IDE: Jupyter Notebook

Jupyter Notebook IDE is an open-source web-based interactive environment for creating and sharing documents that contain live code, equations, visualizations, and narrative text. It is a popular tool for data scientists, machine learning engineers, and other technical professionals. Jupyter Notebook IDE is not a traditional IDE like PyCharm or Visual Studio Code. However, it has many features that make it a good IDE for data scientists and machine learning engineers. For example, it provides interactive code execution, Markdown support, rich visualizations, and kernel support for a variety of programming languages.

Key features of Jupyter Notebook IDE:

- Interactive code execution: Jupyter Notebook IDE allows you to write and execute code cell by cell. This makes it easy to experiment with different code snippets and see the results immediately.
- Markdown support: Jupyter Notebook IDE supports Markdown, a lightweight markup language for creating formatted text. This allows you to include text, images, and other media in your notebooks, along with your code.
- Rich visualizations: Jupyter Notebook IDE supports a variety of rich visualizations, including charts, plots, and maps. This makes it easy to explore and visualize your data.
- Kernel support: Jupyter Notebook IDE supports a variety of programming languages, including Python, R, Julia, and Scala. This is achieved through the use of kernels, which are responsible for executing code in a specific language.
- Sharing: Jupyter Notebook IDE notebooks can be easily shared with others. This makes it easy to collaborate on projects and share your work with others.
- Interactivity: Jupyter Notebook IDE is highly interactive, which makes it easy to experiment with different code snippets and see the results immediately. This is a valuable feature for data scientists and machine learning engineers, who often need to iterate on their code quickly.
- Flexibility: Jupyter Notebook IDE is very flexible. It can be used for a variety of tasks, such as data cleaning and preparation, data analysis, prototyping, and teaching and learning.
- Extensibility: Jupyter Notebook IDE is extensible. There are many third-party libraries and extensions available that can be used to add new features and functionality to Jupyter Notebook IDE.

Packages and libraries:

Jupyter Notebook markdown

Jupyter Notebook markdown is a powerful tool that can be used to create formatted and informative notebooks. It is a valuable skill for data scientists, machine learning engineers, and other technical professionals. Text formatting: You can use markdown to format text in a variety of ways, including bold, italics, headings, and lists. Code blocks: You can use markdown to create code blocks, which will be rendered with syntax highlighting.

Pandas

Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data. Pandas is an open-source library in Python that is made mainly for working with relational or labeled data both easily and intuitively. It provides various data structures and operations for manipulating numerical data and time series.

NumPy

NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices. NumPy is a very popular Python library that is mainly used to perform mathematical and scientific calculations. It offers many features and tools that can be useful for Data Science projects.

Matplotlib

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib is a multi-platform data visualization library built on NumPy arrays.

3. IMPLEMENTATION

3.1 **Describing the project**

In this project, you 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.

3.2 **Data collection and importing**

This notebook uses two datasets: death distribution data for the United States from the year 1999 and rates of left-handedness digitized from a figure in this 1992 paper by Gilbert and Wysocki.

Load the handedness data from the National Geographic survey and create a scatter plot.

- Import pandas as pd and matplotlib.pyplot as plt.
- Import Numpy as np.
- Load the data into a pandas DataFrame named lefthanded_data using the provided data_url_1. Note that the file is a CSV file.
- Use the .plot() method to create a plot of the "Male" and "Female" columns vs. "Age".

3.3 **Rates of left-handedness over time**

Let's convert this data into a plot of the rates of left-handedness as a function of the year of birth, and average over male and female to get a single rate for both sexes. Since the study was done in 1986, the data after this conversion will be the percentage of people alive in 1986 who are left-handed as a function of the year they were born.

Add two new columns, one for birth year and one for mean left-handedness, then plot the mean as a function of birth year.

- Create a column in lefthanded_data called Birth_year, which is equal to 1986 - Age (since the study was done in 1986).
- Create a column in lefthanded_data called Mean_lh which is equal to the mean of the Male and Female columns.
- Use the .plot() method to plot Mean_lh vs. Birth_year.

3.4 **Applying Bayes' rule**

Here's Bayes' theorem for the two events we care about: left-handedness (LH) and dying at age A.

$$P(A|LH) = \frac{P(LH|A)P(A)}{P(LH)}$$

P(LH | A) is the probability that you are left-handed given that you died at age A. P(A) is the overall probability of dying at age A, and P(LH) is the overall probability of being left-handed. We will now calculate each of these three quantities, beginning with P(LH | A).

To calculate P(LH | A) for ages that might fall outside the original data, we will need to extrapolate the data to earlier and later years. Since the rates flatten out in the early 1900s and late 1900s, we'll use a few points at each end and take the mean to extrapolate the rates on each end. The number of points used for this is arbitrary, but we'll pick 10 since the data looks flat-ish until about 1910.

Create a function that will return P(LH | A) for particular ages of death in a given study year.

Import the numpy package aliased as np.

Use the last ten Mean_lh data points to get an average rate for the early 1900s. Name the resulting DataFrame early_1900s_rate.

Use the first ten Mean_lh data points to get an average rate for the late 1900s. Name the resulting DataFrame late_1900s_rate.

For the early 1900s ages, fill in P_return with the appropriate left-handedness rates for ages_of_death. That is, input early_1900s_rate as a fraction, i.e., divide by 100.

For the late 1900s ages, fill in P_return with the appropriate left-handedness rates for ages_of_death. That is, input late_1900s_rate as a fraction, i.e., divide by 100.

3.5 **When do people normally die?**

To estimate the probability of living to an age A, we can use data that gives the number of people who died in a given year and how old they were to create a distribution of ages of death. If we normalize the numbers to the total number of people who died, we can think of this data as a probability distribution that gives the probability of dying at age A. The data we'll use for this is from the entire US for the year 1999 - the closest I could find for the time range we're interested in.

Load death distribution data for the United States and plot it.

- Load death distribution data in the provided data_url_2 into death_distribution_data, setting sep = '\t' and skiprows=[1] to account for the dataset's format.
- Drop the NaN values from the Both Sexes column.
- Use the .plot() method to plot the number of people who died as a function of their age.

3.6 **The overall probability of left-handedness**

This is the average left-handedness in the population of deceased people, and we can calculate it by summing up all of the left-handedness probabilities for each age, weighted with the number of deceased people at each age, then divided by the total number of deceased people to get a probability. In equation form, this is what we're calculating, where N(A) is the number of people who died at age A

$$P(LH) = \frac{\sum_{\mathrm{A}} P(LH|A)N(A)}{\sum_{\mathrm{A}} N(A)}$$

Create a function called P_lh() which calculates the overall probability of left-handedness in the population for a given study year.

- Create a series, p_list, by multiplying the number of dead people in the Both Sexes column with the probability of their being lefthanded using P_lh_given_A().
- Set the variable p equal to the sum of that series.
- Divide p by the total number of dead people by summing death_distribution_data over the Both Sexes column. Return result from the function.

3.7 **Dying while being left-handed**

Now we have the means of calculating all three quantities we need: P(A), P(LH), and P(LH | A). We can combine all three using Bayes' rule to get P(A | LH), the probability of being age A at death (in the study year) given that you're left-handed.

$$P(A|LH) = \frac{P(LH|A)P(A)}{P(LH)}$$

Write a function to calculate P_A_given_lh().

- Calculate P_A, the overall probability of dying at age A, which is given by death_distribution_data at age A divided by the total number of dead people (the sum of the Both Sexes column of death_distribution_data).
- Calculate the overall probability of left-handedness P(LH)
- Calculate P(LH | A)

3.8 **Dying while being right-handed**

Write a function to calculate P_A_given_rh().

- Calculate P_A, the overall probability of dying at age A, which is given by death_distribution_data at age A divided by the total number of dead people.
- Calculate the overall probability of right-handedness P(RH), which is 1-P(LH).
- Calculate P(RH | A), which is 1 - P(LH | A).

3.9 **Plotting the distributions of conditional probabilities**

Now that we have functions to calculate the probability of being age A at death given that you're left-handed or right-handed, let's plot these probabilities for a range of ages of death from 6 to 120.

Plot the probability of being a certain age at death given that you're left- or right-handed for a range of ages.

- Calculate P_A_given_lh and P_A_given_rh.
- Use the .plot() method to plot the results versus age.

3.10 **Average age of left and right-handers at death**

Find the mean age at death for left-handers and right-handers.

• Multiply the ages list by the left-handed probabilities of being those ages at death, then use np.nansum to calculate the sum. Assign the result to average_lh_age.

• Do the same with the right-handed probabilities to calculate average_rh_age.

• Print average_lh_age and average_rh_age.

• Calculate the difference between the two average ages and print it. To make your printed output prettier, using the round() function to round your results to two decimal places.

3.11 **Finding results for study year 2018**

Calculate the age gap we'd expect if we did the study in 2018 instead of in 1990. The gap turns out to be much smaller since rates of left-handedness haven't increased for people born after about 1960. Both the National Geographic study and the 1990 study happened at a unique time - the rates of left-handedness had been changing across the lifetimes of most people alive, and the difference in handedness between old and young was at its most striking.

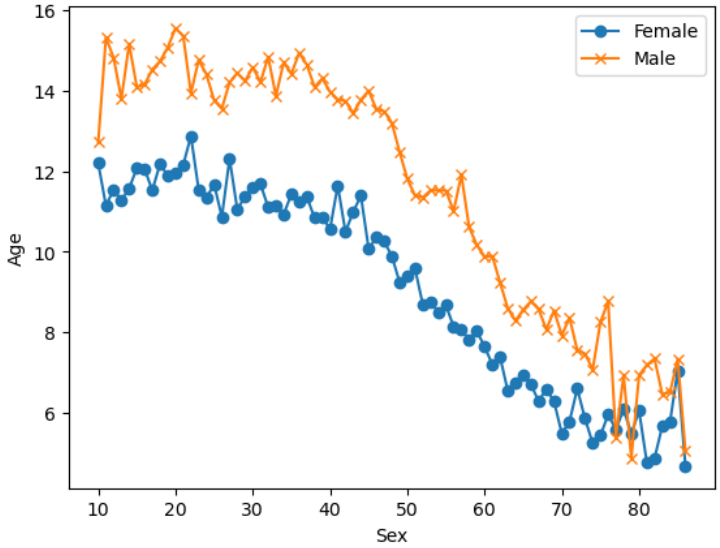
In the call to the function P_A_given_lh, set age_of_death to ages, death_distribution_data to death_distribution_data, and study_year to 2018. Do the same for P_A_given_rh.

Calculate the difference in the average age at death of left handed people and right handed people for the year 2018.

4. SAMPLE SCREENSHOTS AND OBSERVATIONS

4.1 Male and female left-handedness rates vs. age

The following plot shows the male and female left-handedness rates vs. age from the death distribution data for the United States from the year 1999 and rates of left-handedness digitized from a figure in this 1992 paper by Gilbert and Wysocki.

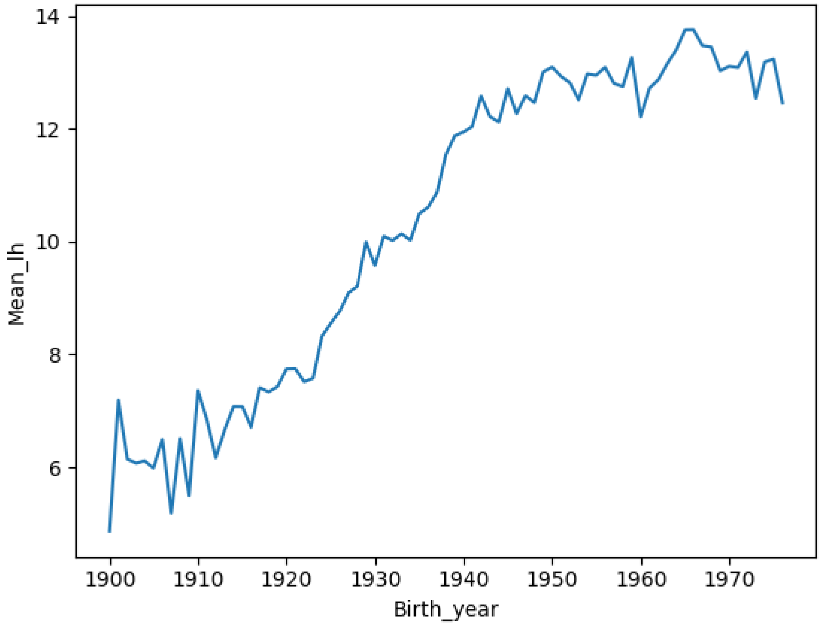


Observation: The rates of left-handedness were around 13% for people younger than 40 but decreased with age to about 5% by the age of 80. This means that the rates aren't a factor of age specifically but rather of the year you were born, and if the same study was done today, we should expect a shifted version of the same distribution as a function of age.

4.2 **Rates of left-handedness over time**

The following plot shows the rates of left-handedness as a function of the year of birth, and average over male and female to get a single rate for both sexes.

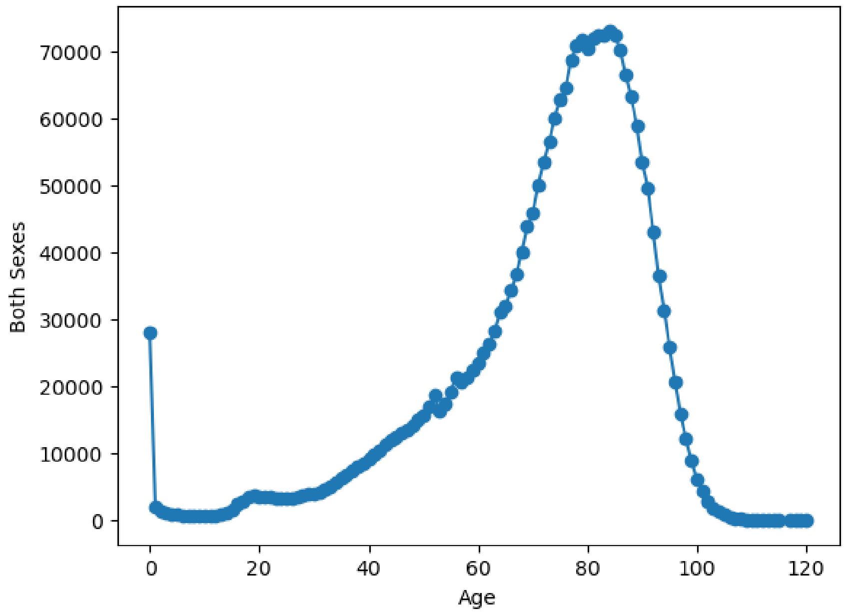
Mean_lh shows the mean of both male and female left handed people vs the Birth_year shows the year of birth of those people.



Observation: The percentages of average of men and women who are left handed have significantly increased for the people born after 1960.

4.3 Death age of people

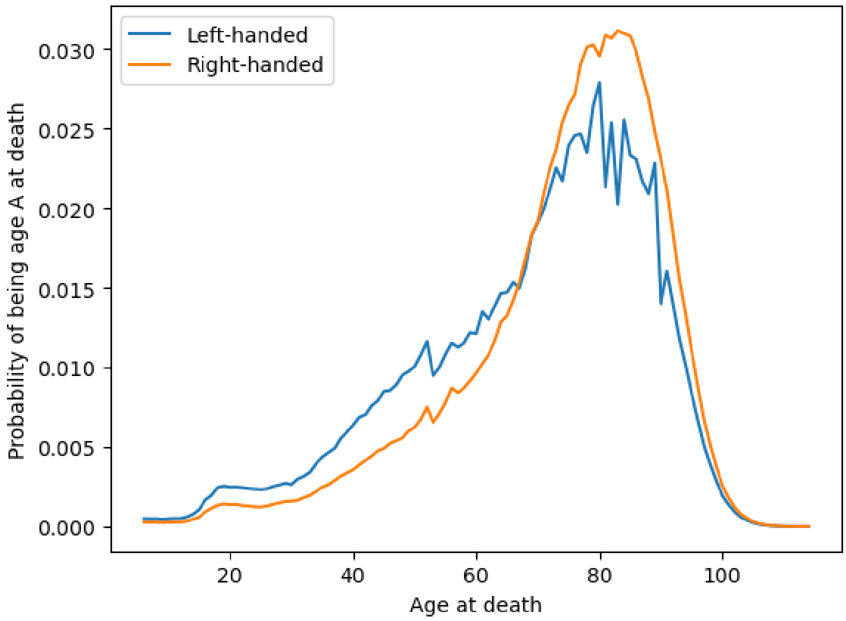
The following plot shows the number of people who died as a function of their age. This gives the number of people who died in a given year and how old they were to create a distribution of ages of death. The data we'll use for this is from the entire US for the year 1999 - the closest I could find for the time range we're interested in.



Observation: From the entire US for the year 1999, majority of the people have died around 80 years of age.

4.4 **Probability of age at death for right and left handed people**

The following plot shows the probability of being a certain age at death given that you're left- or right-handed for a range of ages. The plot describes these probabilities for a range of ages of death from 6 to 120.



Observation: The left-handed distribution has a bump below age 70: of the pool of deceased people, left-handed people are more likely to be younger at age of death. Right handed people have a bump after age 80, showing right handed people are older at age of death and more likely to live longer.

4.5 Average age at death for left and right handers in 1990

Comparing our results with the original study that found that left-handed people were nine years younger at death on average. We can do this by calculating the mean of these probability distributions. Finally find the mean age at death for left-handers and right-handers, and also finding the difference in the mean age at death between left and right handers.

```
average age of lefthanded is 67.25
average age of righthanded is 72.79
The difference in average ages is 5.55 years.
```

Observation: average age of death for left handers is 67.25 years and 72.79 years for right handers, the difference in this average age of death is 5.55 years, which significantly low.

4.6 **Difference in Average age at death between left and right handers in 2018**

Calculating the age gap we'd expect if we did the study in 2018 instead of in 1990. The gap turns out to be much smaller since rates of left-handedness haven't increased for people born after about 1960. Both the National Geographic study and the 1990 study happened at a unique time - the rates of left-handedness had been changing across the lifetimes of most people alive, and the difference in handedness between old and young was at its most striking.

```
The difference in average ages is 2.34 years.
```

Observation: In the year 2018 the difference in the average age of death between left and right handed people is 2.34 years which is negligible, thus refuting the claim of early death for left-handers.

5. CONCLUSION AND FUTURE SCOPE

A 1991 study reported that left-handed people die on average nine years earlier than right-handed people. 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.

In this project, the aim is to 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.

For the year 1990, the average age at death for left handed people is found out to be 67.25 years, and average death age for right handed people is 72.79.

The difference in average age of death for the year 1990 is found out to be 5.55 years.

Finally, for the year 2018 the difference in average age of death between the right and left handed people came out to be 2.34 years.

The 2.34 year average death year gap between left and right handed people is small or negligible difference, thus refuting the claim of early death for left-handers.

This research could provide valuable insights for MTE, a medical company, enabling them to develop tailored strategies and products targeting the unique needs of these two groups.

MTE could develop specialized products targeted at addressing the unique health needs of left-handed or right-handed individuals. This may include ergonomic tools, assistive devices, or therapy options. MTE can contribute to public health initiatives that aim to raise awareness about the importance of understanding and accommodating the needs of both left-handed and right-handed populations.

MTE can capitalize on this research by investing in targeted solutions, research collaborations, and public health campaigns. By doing so, they can position themselves as a leader in promoting personalized healthcare and enhancing the well-being of individuals of all handedness types.

6. REFERENCES

The following websites have been referred to obtain the input data and statistics:

- https://www.cdc.gov/nchs/nvss/mortality_tables.htm
- https://pubmed.ncbi.nlm.nih.gov/1528408/
- https://gist.githubusercontent.com/mbonsma/8da0990b71ba9a09f7de395574e54df1/raw/aec88b30af87fad8d45da7e774223f91dad09e88/lh_data.csv
- https://gist.githubusercontent.com/mbonsma/2f4076aab6820ca1807f4e29f75f18ec/raw/62f3ec07514c7e31f5979beeca86f19991540796/cdc_vs00199_table310.tsv