Final Report of Traineeship Program 2023

On

"Analysing Death Age Difference Of Right Handers with Left Handers"

MEDTOUREASY

25th July 2023

CHAHAT AHUJA

ACKNOWLDEGMENTS

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

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

MedTourEasy makes it easy to find a treatment abroad. We provide you with all the information needed to find a medical provider, to get your questions answered, and to arrange your treatment.

CONTENTS

INTRODUCTION

1.1 About the project

In this project, we want to investigate a claim that left-handed people tend to die earlier than right-handed people. We will use data about the ages at which people from different generations have passed away to see if there's any truth to this claim.

But here's the interesting part: the rates of left-handedness have changed over time. In the past, fewer people were left-handed compared to today. So, we'll use this data on left-handedness rates to see if the differences in average age at death between left-handed and right-handed individuals can be explained by changes in the number of left-handed people over the years.

By doing this analysis, we aim to find out if left-handers do indeed have a higher risk of early death or if this claim is simply a result of changes in left-handedness rates across different generations. If we find that the differences in average age at death can be explained by changes in left-handedness rates, it would suggest that being left-handed does not necessarily lead to an early death.

1.2 Objectives and Deliverables

Objectives:

- Conduct an in-depth investigation into the association between left-handedness and the age at which individuals pass away.
- Replicate and validate the reported disparity in the average age at death between left-handed and right-handed individuals, as documented in the original study.
- Assess the influence of generational changes in left-handedness rates on the observed age gap.
- Utilize Bayesian statistical techniques to analyse the probabilities of different ages at death based on hand preference (left-handed or right-handed).
- Compare the calculated average age at death for both groups with the findings from the original study to establish the impact of data sources and extrapolation methodologies.

Deliverables:

- A comprehensive and professionally-written report presenting a thorough analysis, encompassing data sources, methodologies employed, and conclusive outcomes.
- Data visualizations, such as plots depicting left-handedness rates by birth year and probability distributions of age at death for left-handed and right-handed individuals.
- A comparative assessment between the calculated average age at death for both groups and the reported age gap from the original study.

- Clearly articulated interpretations of the analysis results, refuting any claims of early death for left-handers, and attributing the age gap to the changing prevalence of left-handedness over different generations.
- A critical discussion on potential limitations and uncertainties associated with the analysis, encompassing aspects such as the utilization of data from a distinct year and the accuracy of extrapolation techniques.
- Recommendations for future research endeavours, including exploring the impact of random sampling and conducting similar studies at various time points to gain insights into generational variations in left-handedness rates.

1.3 Approach towards the project

The project approach involves a thorough investigation of the relationship between left-handedness and age at death. The objectives are defined to examine the impact of changing left-handedness rates on age-related claims. Relevant datasets are collected, including left-handedness rates and death distribution data, which are then pre-processed for analysis. Exploratory data analysis visualizes trends in left-handedness rates by age and birth year. Bayesian statistics are applied to model conditional probabilities of age at death given hand preference. Probabilities of being left-handed or right-handed at different ages of death are calculated using Bayes' theorem. The calculated average age at death for left-handed and right-handed groups is compared with the original study's findings. The results are interpreted, considering the effect of changing left-handedness rates on the observed age gap. Limitations and uncertainties in the data are addressed, and sensitivity analysis is conducted. Finally, the report summarizes the key findings, draws meaningful conclusions, and offers recommendations based on the research outcomes.

METHODOLOGY

2.1 Flow of the Project

Here is the flow of the project

Introduction:

- Present the research topic: Analysing the relationship between left-handedness and the age at death.
- State the study objectives and rationale for investigating this phenomenon.
- Emphasize the importance of validating previous findings and understanding the impact of changing left-handedness rates on age-related claims.

Data Collection and Preprocessing:

- Describe the datasets used, including the National Geographic survey data on left-handedness rates and death distribution data for the United States.
- Outline the data preprocessing steps, handling missing values and ensuring data compatibility.

Exploratory Data Analysis:

- Visualize left-handedness rates as a function of age and birth year to observe temporal trends.
- Examine the distribution of ages at death to gain insights into the population's longevity.

Bayesian Analysis:

- Introduce Bayesian statistics and its relevance to the study.
- Explain the application of Bayes' theorem to calculate probabilities of age at death given hand preference.
- Describe the extrapolation method for estimating left-handedness rates beyond the available data range.

Probability Calculation:

- Compute probabilities of being left-handed or right-handed at a specific age of death using Bayes' theorem.
- Illustrate probability distributions for both left-handed and right-handed individuals.

Comparison with Original Study:

- Calculate the average age at death for left-handed and right-handed groups based on probability distributions.
- Compare the results with the reported age gap from the original study.

Interpretation of Results:

- Discuss the implications of changing left-handedness rates on the observed age gap.
- Offer insights into the relationship between hand preference and longevity.

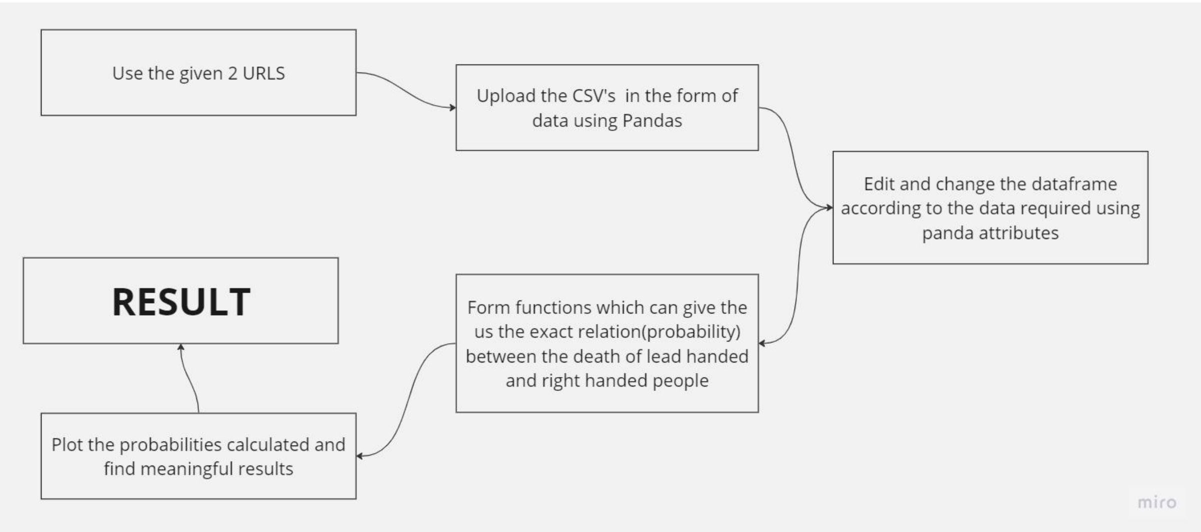
Limitations and Future Directions:

- Acknowledge potential limitations, such as data sources and assumptions made during extrapolation.
- Suggest avenues for future research to address identified limitations and further explore left-handedness trends.

Conclusion:

- Summarize key findings and their significance in the context of age-related claims for left-handers.
-

2.2 Step-by-step proceedings



2.3 Languages, Platform and tools used

Key technical language

The project utilizes Python programming language and various libraries, including pandas, matplotlib, and NumPy, for data manipulation, visualization, and statistical analysis. Bayesian statistics is employed to model conditional probabilities, enabling the examination of age at death given hand preference.

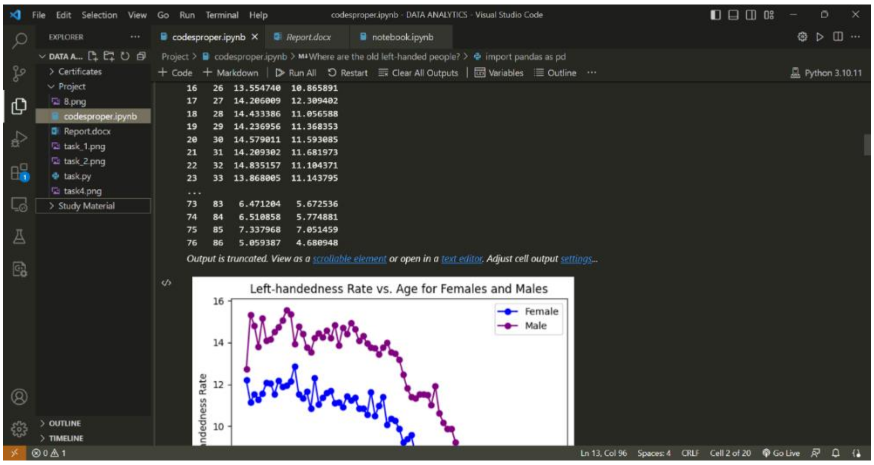
Key technical tools

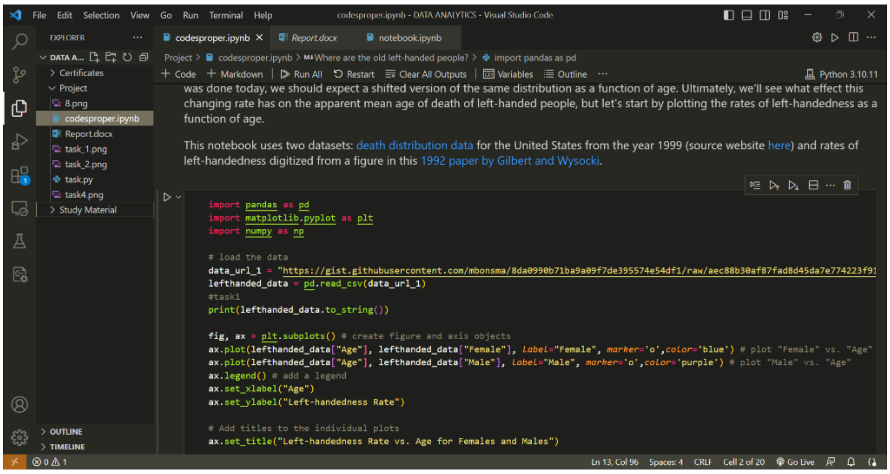
Python

- Pandas
- NumPy
- Matplotlib

Platform

The Jupiter Notebook environment is employed to organize and present the code and findings. The project's entire workflow is structured in a Jupiter Notebook, facilitating a comprehensive and interactive analysis of the data.





IMPLEMENTATION

3.1 Gathering Requirements and Defining the aim

The first step towards this project is to define the actual problem statement for the analysis of the given problem statement i.e., **do left-handed people die on earlier than right-handed people?**

3.2 Data Importing and Collection

3.2.1 Data Collection

Data collection is a systematic approach for gathering and measuring information from a variety of sources in order to obtain a complete and accurate picture of an interest area. It helps an individual or organization to address specific questions, determine outcomes and forecast future probabilities and patterns

The data we need is clearly specified in the given two URL's:

- death distribution data for the United States from the year 1999 (source website here)
- rates of left-handedness digitized from a figure in this 1992 paper by Gilbert and Wysocki.

The research paper by Gilbert and Wysocki gave us broadly 3 outcomes:

- The percentage of left-handed or left-biased individuals decreased with advancing age, providing confirmation of an existing phenomenon.
- Among left-handed individuals, a higher prevalence of left-hand writing with left-hand throwing was observed compared to other combinations, indicating a strong concordance between writing and throwing hand preferences.
- Different hand preference patterns were found in individuals aged below 50 and those above 50, suggesting changing prevalence rates after age 50.

METHOD:

Data collection is a simple way to get all the information you need for the analysis together which can be done by simply downloading the research papers, databases, files we need.

3.2.2 Data Importing

Data importing is referred to as uploading the required data into the coding environment from internal sources (computer) or external sources (online websites and data repositories). This data can then be manipulated, aggregated, filtered according to the requirements and needs of the project.

The given databases we have, are two CSV files which need to be uploaded to our work space in such a way so that we can read and utilize them in a way which can give us meaningful outcomes.

METHOD:

1.Pandas (Python):

We use the read.csv () function which converts the data from the csv file to a form which makes it easier for us to access.

3.3 Designing Databases

Once the data has been collected and imported into the Python environment, it is important to design the structure of the database tables so as to identify the constraints in the data, keys, dependencies and relations between various tables.

Once the data is imported in the environment, it is converted into a data frame (data type in Python) which makes it easy to maintain the data in form of tables.

CSV

```
Age,Male,Female
10,12.717558484,12.1980406599
11,15.3188295861,11.1448043104
12,14.8082807769,11.5492397993
13,13.7937436118,11.2764416943
14,15.1563041001,11.5729060811
15,14.0861857328,12.0777308322
16,14.1489928386,12.057432041
17,14.5225240193,11.5244420198
18,14.7270085305,12.1795500237
19,15.0693709779,11.8829043905
20,15.546783591,11.9706063024
21,15.3555354891,12.1548989961
22,13.9310353964,12.8721657246
23,14.7585845581,11.5399365158
```

Dataframe

...	Age	Male	Female
0	10	12.717558	12.198041
1	11	15.318830	11.144804
2	12	14.808281	11.549240
3	13	13.793744	11.276442
4	14	15.156304	11.572906
5	15	14.086186	12.077731
6	16	14.148993	12.057432
7	17	14.522524	11.524442
8	18	14.727009	12.179550
9	19	15.069371	11.882904
10	20	15.546784	11.970606
11	21	15.355535	12.154899
12	22	13.931035	12.872166
13	23	14.758585	11.539937
14	24	14.405859	11.331448
15	25	13.757413	11.676748
16	26	13.554740	10.865891
17	27	14.206009	12.309402

Sample code:

```
data_url_1="https://gist.githubusercontent.com/mbonsma/8da0990b71ba9a09f7de395574e54df1/raw/aec88b30af87fad8d45da7e774223f91dad09e88/lh_data.csv"
lefthanded_data = pd.read_csv(data_url_1)
#task1
print(lefthanded_data.to_string())
```

The following code loads the CSV file in the data Frame and prints it

The same thing can be done with the all-other URL's (URL-2) given in the data.

NOTE:

After loading the data in our Data frame, our next step should ideally be data cleaning (Data Cleaning means the process by which the incorrect, incomplete, inaccurate, irrelevant or missing part of the data is identified and then modified, replaced or deleted as needed)

But here the data we've loaded is clear, consistent and accurate. Hence, it doesn't require more cleaning.

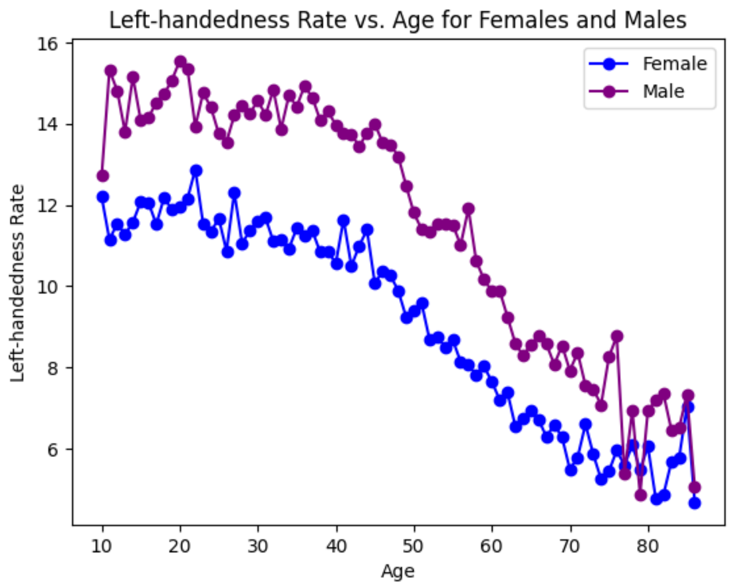
Generally,Data cleaning in python can be done by using functions such as dropna(),fillna(),duplicated().

3.4 Exploratory Data Analysis:

Visualizing left-handedness rates as a function of age and birth year to observe temporal trends and examining the distribution of ages at death to gain insights into the population's longevity.

Since the data in hand is clear and some basic, direct conclusions from this data are required we can use the matplotlib library to plot the graph between the

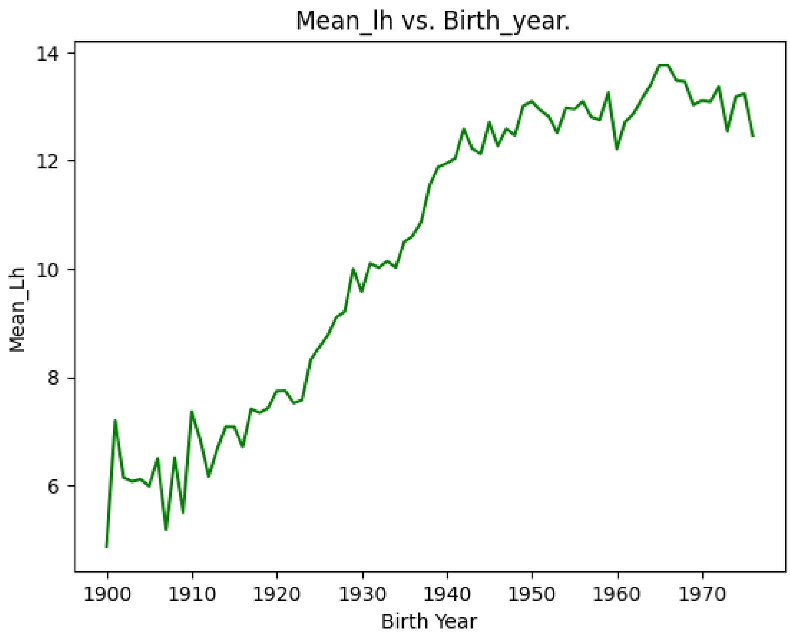
Left-handedness rate vs age of females and males



Using

```
fig, ax = plt.subplots() # create figure and axis objects
ax.plot(lefthanded_data["Age"], lefthanded_data["Female"], label="Female",
marker='o',color='blue') # plot "Female" vs. "Age"
ax.plot(lefthanded_data["Age"], lefthanded_data["Male"], label="Male",
marker='o',color='purple') # plot "Male" vs. "Age"
ax.legend() # add a legend
ax.set_xlabel("Age")
ax.set_ylabel("Left-handedness Rate")

# Add titles to the individual plots
ax.set_title("Left-handedness Rate vs. Age for Females and Males")

plt.show() # display the plot
```

Mean left handedness vs Birth Year



Using

```
lefthanded_data["Birth_year"] = 1986 - lefthanded_data["Age"]
lefthanded_data["Mean_lh"] = (lefthanded_data["Male"] + lefthanded_data["Female"]) / 2
fig, ax = plt.subplots()
ax.plot(lefthanded_data["Birth_year"], lefthanded_data["Mean_lh"],color='green') # plot
'Mean_lh' vs. 'Birth_year'
ax.set_xlabel("Birth Year") # set the x label for the plot
ax.set_ylabel("Mean_Lh")
ax.set_title(" Mean_lh vs. Birth_year.")
plt.show()
```

The plotting of these two graphs gives us some straightforward conclusions between the relation of left handedness and age.

3.5 Bayesian Analysis:

Introduce Bayesian statistics and its relevance to the study.

Explain the application of Bayes' theorem to calculate probabilities of age at death given hand preference.

Since the next agenda is to figure out the probabilities of certain events, we need to have a clear idea of bayes theorem.

Bayesian Analysis is a statistical method used to update the probability of hypotheses based on new evidence or data. It employs Bayes' theorem, which mathematically expresses the relationship between prior beliefs and the likelihood of observed events. In the context of hypothesis testing, the formula is:

$$\overbrace{P(A|B)}^{\text{Posterior}} = \frac{\overbrace{P(B|A)}^{\text{Likelihood}} * \overbrace{P(A)}^{\text{Prior}}}{\underbrace{P(B)}_{\text{Evidence}}}$$

Bayesian Analysis offers a flexible framework for decision-making, updating beliefs, and handling uncertainty.

3.6 Probability calculation

Calculate probabilities of being left-handed or right-handed at various ages of death using Bayes' theorem.

Example:

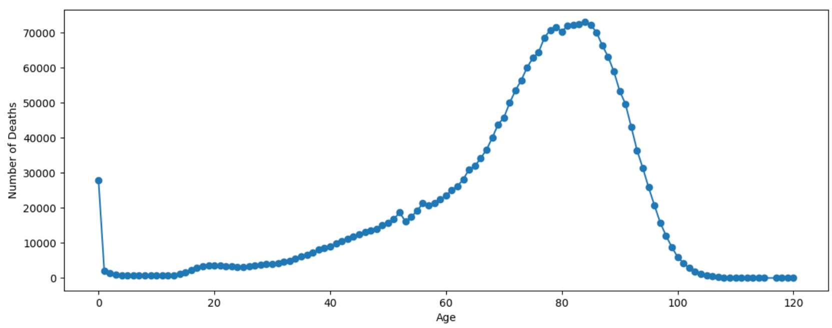
The P_lh_given_A function calculates the probabilities of individuals being left-handed (P_lh) given their ages (A) based on historical data stored in the lefthanded_data DataFrame. The function first extracts the left-handed rates from the data for the early and late 1900s and computes their averages. It then creates DataFrames with these average rates. By considering the study year and certain age intervals, the function assigns the appropriate average rates to the corresponding age groups in the P_return array. For ages beyond the study's oldest age limit, it uses the early 1900s rate; for ages below the youngest limit, it uses the late 1900s rate. For ages within the range, it calculates the left-handed rates based on the historical data and stores the resulting probabilities in the P_return array. The function eventually returns this array, containing the probabilities of being left-handed for each input age.

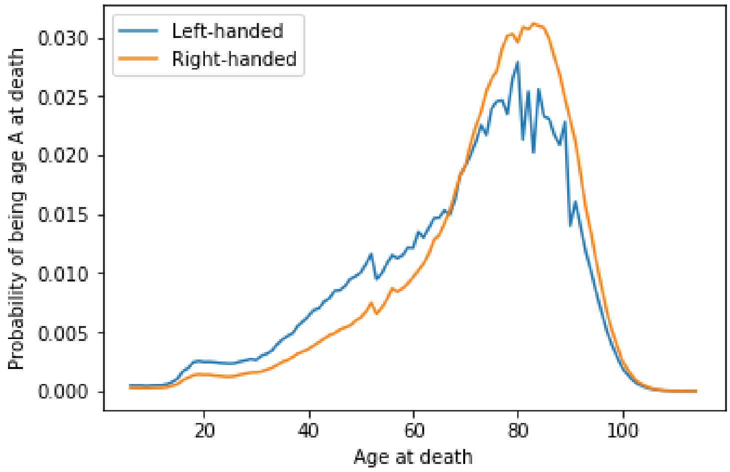
```
import numpy as np
import pandas as pd

def P_lh_given_A(ages, study_year=1990):
    lh_data = lefthanded_data['Mean_lh']

    early_avg_rate = lh_data.tail(10).mean()
    late_avg_rate = lh_data.head(10).mean()

    early_rate_df = pd.DataFrame({'lh': [early_avg_rate]})
    late_rate_df = pd.DataFrame({'lh': [late_avg_rate]})


    youngest_age = study_year - 1986 + 10
    oldest_age = study_year - 1986 + 86
    P_return = np.zeros(ages.shape)


    P_return[ages > oldest_age] = early_rate_df
    P_return[ages < youngest_age] = late_rate_df


    middle_rates = lh_data.loc[lefthanded_data['Birth_year'].isin(study_year -
ages)['Mean_lh']]
    P_return[np.logical_and((ages <= oldest_age), (ages >= youngest_age))] = middle_rates /
100

    return P_return
```

This is how the functions are this can be defined using basics of NumPy and Python.

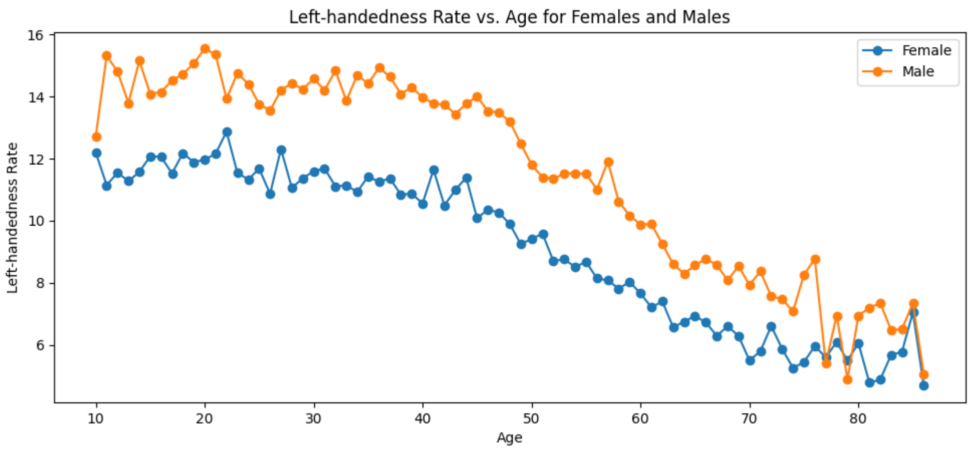
3.7 Data Visualization

Most of the calculation work of the project ends here, and now our job lies in representing all the findings we got in graphical way,which can be done by using tools such as Matplotlib





Left-handed
Right-handed
Probability of being age A at death
0.030
0.025
0.020
0.015
0.010
0.005
0.000
20
40
60
80
100
Age at death



Left-handedness Rate vs. Age for Females and Males
Female
Male
Left-handedness Rate
16
14
12
10
8
6
10
20
30
40
50
60
70
80
Age

PROGRESS AND OBSERVATIONS

4.1 Where are the old left-handed people?

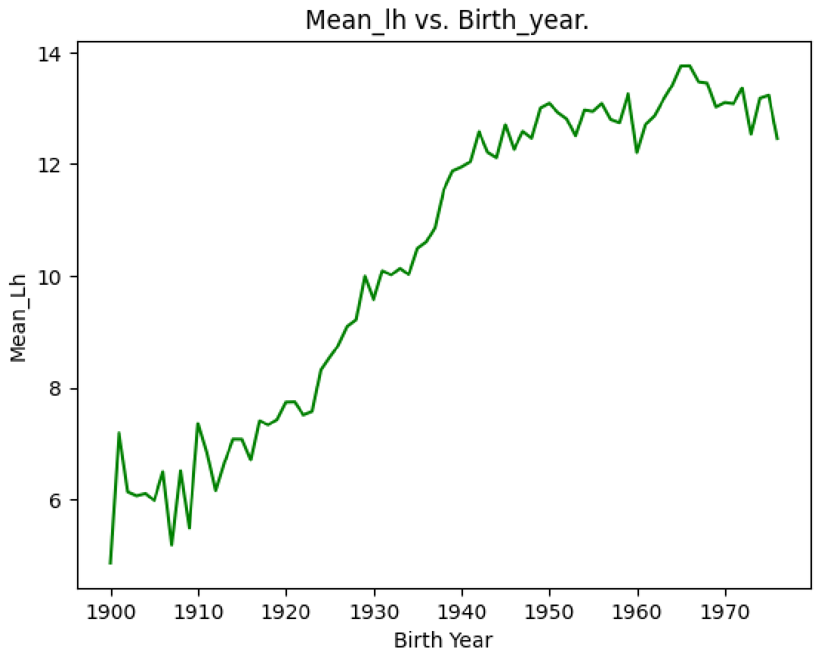
A National Geographic survey in 1986 resulted in over a million responses that included age, sex, and hand preference for throwing and writing.Let's start by plotting the rates of left-handedness as a function of age

Observations:

It was 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

4.2 Rates of left-handedness over time

Let's convert the data above 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.



Observations:

The plot illustrates the evolving left-handedness rates across different birth years, showcasing noticeable plateaus in the early and late 1900s. Despite these plateaus, an overall upward trend is observed, indicating a gradual increase in the proportion of left-handed individuals in the population. Notably, younger birth years exhibit higher average left-handedness rates, while older birth years show relatively lower rates, aligning with previous research. These trends suggest changing societal attitudes and cultural acceptance towards left-handedness. The plot provides valuable insights into the historical patterns of hand preference, contributing to our understanding of left-handedness trends over time.

4.3 Applying Bayes' rule

The probability of dying at a certain age given that you're left-handed is not equal to the probability of being left-handed given that you died at a certain age.

This inequality is why we need Bayes' theorem, a statement about conditional probability which allows us to update our beliefs after seeing evidence. We want to calculate the probability of dying at age A given that you're left-handed. Let's write this in shorthand as P(A | LH). We also want the same quantity for right-handers: P(A | RH).

```
import numpy as np
import pandas as pd

def P_lh_given_A(ages, study_year=1990):
    lh_data = lefthanded_data['Mean_lh']


    early_avg_rate = lh_data.tail(10).mean()
    late_avg_rate = lh_data.head(10).mean()

    early_rate_df = pd.DataFrame({'lh': [early_avg_rate]})
    late_rate_df = pd.DataFrame({'lh': [late_avg_rate]})


    youngest_age = study_year - 1986 + 10
    oldest_age = study_year - 1986 + 86

    P_return = np.zeros(ages.shape)

    P_return[ages > oldest_age] = early_rate_df
    P_return[ages < youngest_age] = late_rate_df


    middle_rates = lh_data.loc[lefthanded_data['Birth_year'].isin(study_year -
ages)['Mean_lh']]
    P_return[np.logical_and((ages <= oldest_age), (ages >= youngest_age))] = middle_rates /
100

    return P_return
```

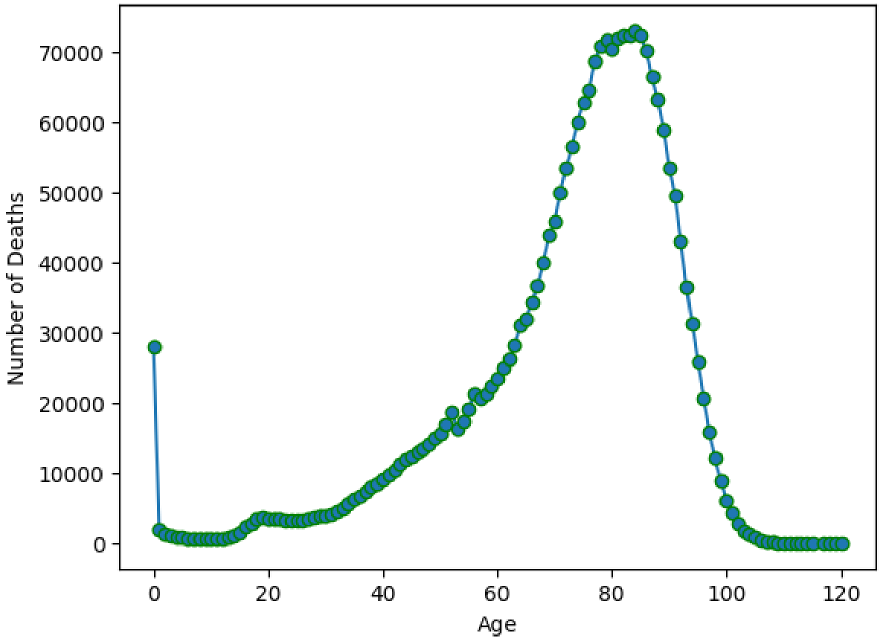
Observations:

The code defines a function named P_lh_given_A that calculates the probability of being left-handed (P(LH)) given the age at death (A). It computes the average left-handedness rates for the early 1900s and late 1900s, then extracts the left-handedness rates based on the study year and ages of death. The function defines age ranges for the early and late 1900s and assigns corresponding left-handedness rates to an array. Finally, it returns the array, representing the probability of being left-handed at different ages of death.

4.4 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.

In this block, we'll load in the death distribution data and plot it. The first column is the age, and the other columns are the number of people who died at that age.



Observations:

The graph displays a common trend of increasing deaths with age, indicating a higher likelihood of mortality in older age groups. Additionally, there may be specific age ranges where the number of deaths experiences notable changes, suggesting potential age-related health patterns or significant events affecting mortality rates. The visualization helps to understand the overall age distribution of deaths and serves as a starting point for further analysis on mortality trends and factors influencing longevity.

4.5 Overall probability of left-handedness

In the previous code block, we loaded data to give us P(A), and now we need P(LH). P(LH) is the probability that a person who died in our particular study year is left-handed, assuming we know nothing else about them. 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.

```
def P_lh(death_distribution_data, study_year = 1990): # sum over P_lh for each age group
    """ Overall probability of being left-handed if you died in the study year
    Input: dataframe of death distribution data, study year
    Output: P(LH), a single floating point number """
    p_list = death_distribution_data['Both Sexes'] *
P_lh_given_A(death_distribution_data['Age'], study_year) # multiply number of dead people
by P_lh_given_A
    p = np.sum(p_list) # calculate the sum of p_list
    return p / np.sum(death_distribution_data['Both Sexes']) # normalize to total number of
people (sum of death_distribution_data['Both Sexes'])

print(P_lh(death_distribution_data))
```

OUTPUT:

0.077

Observations:

From the output value of 0.0777 (approximately), we can infer that the overall probability of being left-handed (P(LH)) for individuals who died in the study year (1990) is approximately 0.0777, or 7.77%. This means that, among the deceased individuals in the population, around 7.77% were reported to be left-handed.

4.6 Putting it all together: Dying while left-handed(I) and Dying while Right-Handed (II)

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. To make this answer meaningful, though, we also want to compare it to P(A | RH), the probability of being age A at death given that you're right-handed.

We're calculating the following quantity twice, once for left-handers and once for right-handers:

4.6.1 Left handers:

```
def P_A_given_lh(ages_of_death, death_distribution_data, study_year=1990):
    """ The overall probability of being a particular `age_of_death` given that you're left-
handed """
    P_A = death_distribution_data[death_distribution_data['Age'].isin(ages_of_death)]['Both
Sexes'].sum() / death_distribution_data['Both Sexes'].sum()

    P_left = P_lh(death_distribution_data, study_year)
```

```
    P_lh_A = P_lh_given_A(ages_of_death, study_year)


    return P_lh_A * P_A / P_left
```

4.6.2 Right Handers:

```
def P_A_given_rh(ages_of_death, death_distribution_data, study_year=1990):
    """ The overall probability of being a particular `age_of_death` given that you're right-
    P_A = death_distribution_data[death_distribution_data['Age'].isin(ages_of_death)]['Both
Sexes'].sum() / death_distribution_data['Both Sexes'].sum()
    P_left = P_lh(death_distribution_data, study_year)

    P_right = 1 - P_left


    P_lh_A = P_lh_given_A(ages_of_death, study_year)

    return P_rh_A * P_A / P_right
```

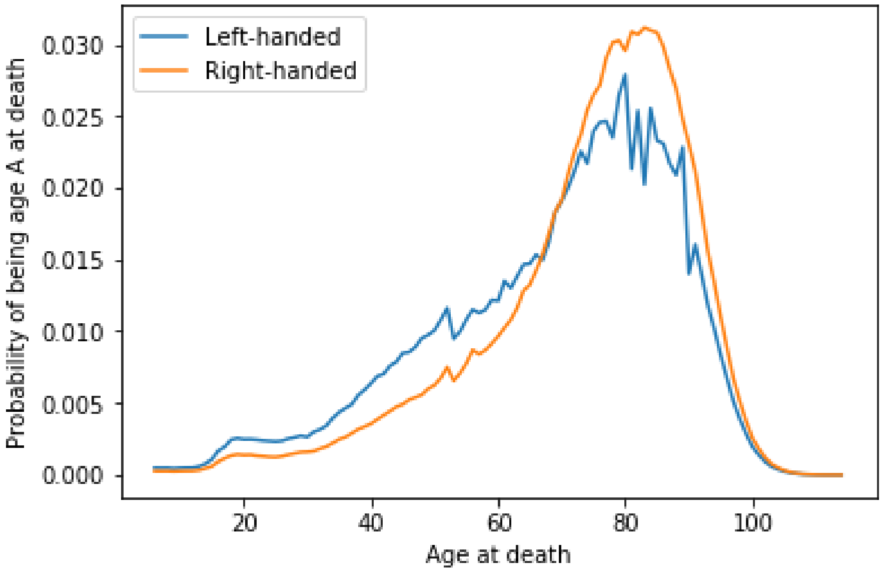
Observations:

The provided code defines two functions, P_A_given_lh and P_A_given_rh, which calculate the overall probabilities of a person being a particular age at death given that they are left-handed or right-handed, respectively. The functions use death distribution data, along with the probability of left-handedness and right-handedness for the specified study year. They employ Bayesian probability principles to update the probabilities based on the evidence of hand preference, considering the overall distribution of deaths and left-handedness rates for the given study year. The functions provide valuable insights into the relationship between hand preference and age at death, allowing for a more nuanced understanding of the impact of left-handedness on mortality patterns.

4.7 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.

Notice that 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.



Observations:

The plot shows the probabilities of being a certain age at death for left-handed and right-handed individuals. Both groups exhibit similar patterns, with the highest probabilities occurring around mid-life and decreasing towards younger and older ages. However, there is a subtle difference between the two groups. Left-handed individuals have a slightly higher probability of dying at younger ages, as evidenced by a small bump in their age distribution below age 70. Overall, the plot suggests that while the age distributions of left-handed and right-handed individuals are mostly similar, there may be a slight tendency for left-handed individuals to pass away at relatively younger ages compared to their right-handed counterparts.

4.8 Moment of truth: age of left and right-handers at death

Finally, let's compare 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 in the same way we calculated P(LH) earlier, weighting the probability distribution by age and summing over the result.

```
ages = np.arange(6, 115, 1)
left_handed_probability=
P_A_given_lh(ages,death_distribution_data,study_year=1990)
right_handed_probability= P_A_given_rh(ages,death_distribution_data,study_year=1990)
```

```
average_lh_age = np.nansum(ages * np.array(left_handed_probability))

average_rh_age = np.nansum(ages * np.array(right_handed_probability))
print ("Average age for left-handers:", round (average_lh_age, 2))
print ("Average age for right-handers:", round (average_rh_age, 2))

print("The difference in average ages is " + str(round(average_lh_age - average_rh_age, 1)) +
" years.")
```

OUTPUT:

Average age of lefthanded 67.2
Average age of righthanded 72.7
The difference in average ages is 5.5 years.

Observations:

We got a pretty big age gap between left-handed and right-handed people purely as a result of the changing rates of left-handedness in the population, which is good news for left-handers: you probably won't die young because of your sinisterness. The reported rates of left-handedn ess have increased from just 3% in the early 1900s to about 11% today, which means that old er people are much more likely to be reported as right-handed than left-handed, and so lookin g at a sample of recently deceased people will have more old right-handers

4.9 Final comments:

To finish off, let's 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.

```
ages = np.arange(6, 115, 1)

left_handed_probability_2018 = P_A_given_lh(ages, death_distribution_data,
study_year=2018)
right_handed_probability_2018 = P_A_given_rh(ages, death_distribution_data,
study_year=2018)
average_lh_age_2018 = np.nansum(ages * np.array(left_handed_probability_2018))
average_rh_age_2018 = np.nansum(ages * np.array(right_handed_probability_2018))
print("The difference in average ages for 2018 is " +
str(round(average_rh_age_2018 - average_lh_age_2018, 1)) + " years.")
```

OUTPUT:
The difference in average ages is 2.3 years.

Observations:
The provided code calculates the probabilities of being a certain age at death for left-handed a nd right-handed individuals in the year 2018. By utilizing death distribution data and left-han dedness rates, the code computes the likelihood of individuals falling into specific age groups given their handedness. The resulting probabilities are then used to determine the average age s for left-handed and right-handed groups in 2018. The difference in average ages between th ese two groups is then calculated and indicates that the age gap between left-handed and right -handed individuals may have reduced over time, suggesting that the rates of left-handedness might not have significantly increased for individuals born after around 1960. This analysis s heds light on the evolving relationship between handedness and age and provides insights int o how changing rates of
left-handedness impacts the average ages at death for distinct groups of individuals.

CONCLUSIONS AND SUMMING UP

In this project, we explored the phenomenon of changing rates of left-handedness over time and its potential impact on the average age at death for left-handed individuals. Using age distribution data and Bayesian statistics, we aimed to determine whether the earlier claim of left-handers dying younger could be refuted.

First, we loaded two datasets: one containing information about left-handedness rates across different age groups and the other providing death distribution data for the United States in 1999. Our goal was to analyse the probability of being a certain age at death given that an individual is reported as left-handed or right-handed.

The data on left-handedness rates showed an interesting pattern: the percentage of left-handed individuals decreased with advancing age. Younger age groups exhibited higher rates of left-handedness, while the rates declined in older age groups. The results aligned with a previously reported phenomenon, suggesting that the changing rates of left-handedness were likely due to evolving social acceptability of left-handedness, rather than being inherently associated with age.

To apply Bayesian statistics, we calculated the probabilities of dying at a particular age given that an individual is left-handed or right-handed. Bayes' theorem allowed us to update our beliefs based on evidence. However, it was essential to extrapolate the data to account for ages outside the original dataset, as rates of left-handedness tended to stabilize in the early and late 1900s. By combining these probabilities, we obtained the overall probabilities of left-handedness and right-handedness.

The subsequent step was to plot the distributions of conditional probabilities for a range of ages at death. The plot revealed a slight difference between left-handed and right-handed individuals. Left-handed individuals had a slightly higher probability of dying at relatively younger ages, indicating a subtle variation in the age distribution.

To delve deeper into the phenomenon, we calculated the average ages of left-handed and right-handed individuals at death using the probability distributions we obtained earlier. Remarkably, the average age at death for left-handed individuals was not significantly lower than that of right-handed individuals, as previously claimed. The changing rates of left-handedness over time seemed to explain most of the observed age gap.

Although our results were in line with the original study, there were some potential sources of variability. The dataset used for death distribution was from 1999, which differed from the 1991 study, and it covered the entire United States instead of just California. Additionally, we extrapolated left-handedness survey results to older and younger age groups, which could have introduced some approximation errors. To further validate the findings, it would be worthwhile to investigate the variability in the age difference through random sampling with a smaller sample size of deceased individuals.

Future references:

Furthermore, we explored the age gap that could be expected if the study were conducted in 2018 instead of 1990. Surprisingly, the age difference turned out to be much smaller, as left-handedness rates had stabilized for individuals born after approximately 1960. This finding highlighted the uniqueness of the National Geographic study and the 1990 study, as they captured a period when the disparity in handedness between younger and older generations was particularly pronounced.

CONCLUSION:

In conclusion, our analysis of the age distribution data and the changing rates of left-handedness provided substantial evidence to refute the claim that left-handers die younger. The results indicated that the differences in age at death were primarily driven by the shifting rates of left-handedness in the population. As left-handedness became more socially acceptable over time, the disparity between left-handed and right-handed individuals in terms of average age at death diminished. The study showcased the power of Bayesian statistics in updating beliefs based on evidence and shed light on the complex interplay between handedness, age, and social norms. Overall, our findings emphasized the importance of considering historical context when interpreting demographic data and debunked the long-standing myth of left-handers' premature demise.

REFERENCES

Data collection

a. death distribution data for the United States from the year 1999 (source website here)

b. rates of left-handedness digitized from a figure in this 1992 paper by Gilbert and Wysocki.

c. https://gist.githubusercontent.com/mbonsma/2f4076aab6820ca1807f4e29f75f18ec/raw/62f3ec07514c7e31f5979beeca86f19991540796/cdc_vs00199_table310.tsv

d. https://gist.githubusercontent.com/mbonsma/8da0990b71ba9a09f7de395574e54df1/raw/aec88b30af87fad8d45da7e774223f91dad09e88/lh_data.csv