Final Report of Traineeship Program 2023

On

"ANALYZING THE DEATH AGE DIFFERENCE OF RIGHT HANDERS WITH LEFT HANDERS"

MEDTOUREASY

28th December 2023

ACKNOWLDEGMENTS

The traineeship opportunity that I had with MedTourEasy was a great change for learning and understanding the intricacies of the subject 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.

TABLE OF CONTENTS

ABSTRACT

A1991 study reported that left-handed people die on average nine years earlier than right-handed people.

To put this into testing, A National Geographic survey was conducted in 1986 which 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 .

Let's 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..

1.1 About the Company

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

1.2 About the Project

There is a common perception among the masses that people with left-handedness die early than their right-handed counterpart. To put this theory into test a study on this phenomenon was made in 1991, survey of 1,177,507 U.S. men and women between the ages of 10 and 86 included questions regarding hand preference for writing and throwing. Three effects were observed. Individuals with at least some left motoric bias comprised a smaller percent of the population with advancing age. This finding provides large-scale confirmation of a previously described phenomenon.

- Among sinistral, concordance for writing and throwing was 2.2 times as prevalent as left-writing with right-throwing, and 4.1 times as prevalent as right-writing with left-throwing.
- These sinistral subpopulations displayed distinct and stable prevalence prior to age 50 and changing patterns of prevalence subsequent to age 50. The results confirm a decrease with age in the prevalence of sinistrality, but indicate that age-specific rates of mixed- and left-handedness are distinct.
- A decrease with age in the proportion of left-handers in the population has been reported by many investigators.
- The implications for hypotheses regarding age-related change in the prevalence of sinistrality are discussed.

National Geographic Survey, The Smell Survey (a questionnaire and microencapsulated set of odorants) was inserted in each copy of the September 1986 issue of the National Geographic, and sent worldwide to 10.7 million members of the Society.

- Participants provided their age. sex, country of residence, and ethnicity (response options: black, white, Asian. American Indian, Hispanic, other, prefer not to answer). They also reported which hand (left or right) they used to write, and to throw. spanic, other, prefer not to answer). They also reported which hand (left or right) they used to write, and to throw.

- Only respondents who answered both hand-use questions were retained for analysis. Based on their answers, respondents were assigned to four behavioral phenotypes: left-writing and left-throwing (LwL,). left-writing and right-throwing (LwR,), right-writing and left-throwing (R-L,), and right-writing and right throwing (RwR,). A worldwide total.

- A worldwide total of 1,420,000 usable surveys were returned (a 13% response rate). The results presented here are based on 1,77, 507 U.S. respondents between 10 and 86 years of age.
  The age specific frequencies handedness for specific subjects(MALE AND FEMALE) data is used in determining the prevalence of left-handedness.

1.3 Objectives and Deliverables

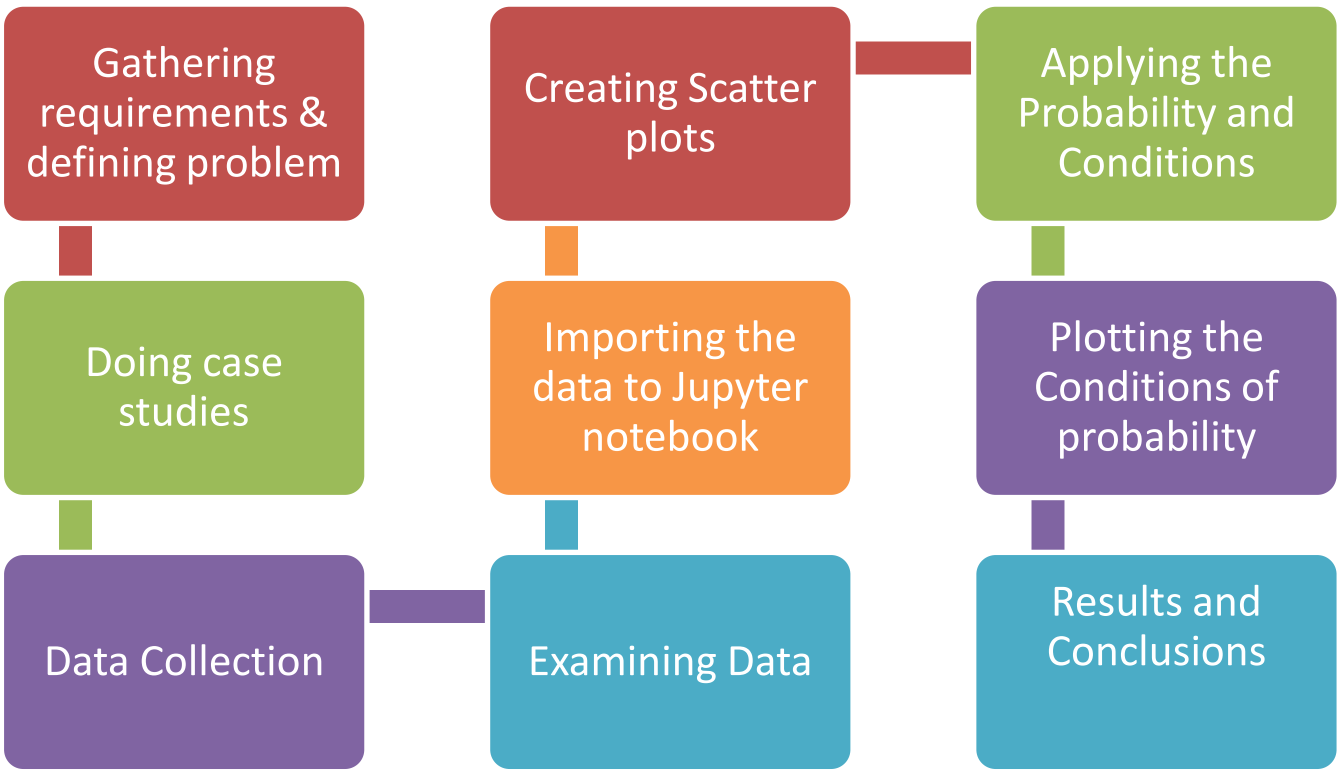
In this project we will explore the phenomena of left-handedness causing early death. This is done by using age distribution data to see if we reproduce a difference in average age at death purely from the changing rates of left-handedness overtime. Thus, reflecting the claim the early death of left hand over time. This notebook uses Panda and Bayesian statistics to analyse the probability of being a certain age at death given left-handed or right-handed.

- To create scatter plots and establish relationship between the relative frequency among Male and female vs Age.
- To create a graph by plotting Birth year vs Lefthanded data.
- Applying probability and conditions to find the Probability of dying at age A.
- Plotting death distribution for year 1999.
- Probability of being lefthanded and overall probability of being left and right-handed respectively.
- Plotting the distribution of conditional probabilities “Age at death vs Being Age A at death.
- Finding the difference in the mean age between left and right handers at death.
- Calculating the mean age difference of left and right handers at death in 2018.

II METHODOLOGY

2.1 Flow of the Project

The project followed the following steps to accomplish the desired objectives and deliverables. Each step has been explained in detail in the following section.



2.2 Language and Platform Used

2.2.1 Language: Python

Python is an easy to learn, high-level general purpose programming language. Developed by Van Rossum in the year 2000. It has efficient high-level data structures and a simple but effective approach to object-oriented programming. Python’s elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas on most platforms. Its core philosophy is summarized as flows:

- Beautiful is better than ugly.
- Explicit is better than implicit.
- Simple is better than complex.
- Complex is better than complicated.
- Readability counts.

Rather than building all of its functionality into its core, Python was designed to be highly extensible via modules. This compact modularity has made it particularly popular as a means of adding programmable interfaces to existing applications. Van Rossum's vision of a small core language with a large standard library and easily extensible interpreter stemmed from his frustrations with ABC, which espoused the opposite approach. Python strives for a simpler, less-cluttered syntax and grammar while giving developers a choice in their coding methodology.

2.2.2 Platform: Jupyter Notebooks

Jupyter Notebook previously known as iPython, allows users to create documents that include live code, equations, visualizations, and text. It's often used by programmers, data scientists, and students to document and demonstrate coding workflows or experiment with code.

Jupyter Notebook supports multiple languages, including Python, Ruby, R, and C++. It can also produce interactive outputs, such as HTML, images,videos, LaTeX, and custom MIME types

III IMPLEMENTATION

3.1 Gathering Requirements and Defining Problem Statement

This is the first step wherein the requirements are collected from the clients to understand the deliverables and goals to be achieved after which a problem statement is defined which has to be adhered to while development of the project.

3.2 Data Collection and Importing

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 for analyzing the age and left hand usage has been collected through various GitHub repositories, mentioned as follows:

- Death distribution data for the United States from the year 1999.
- Rates of left handedness-*Hand preference and age in the united states avery n. gilbertand charles j. wysxki*

Libraries Used in Jupyter Notebooks:

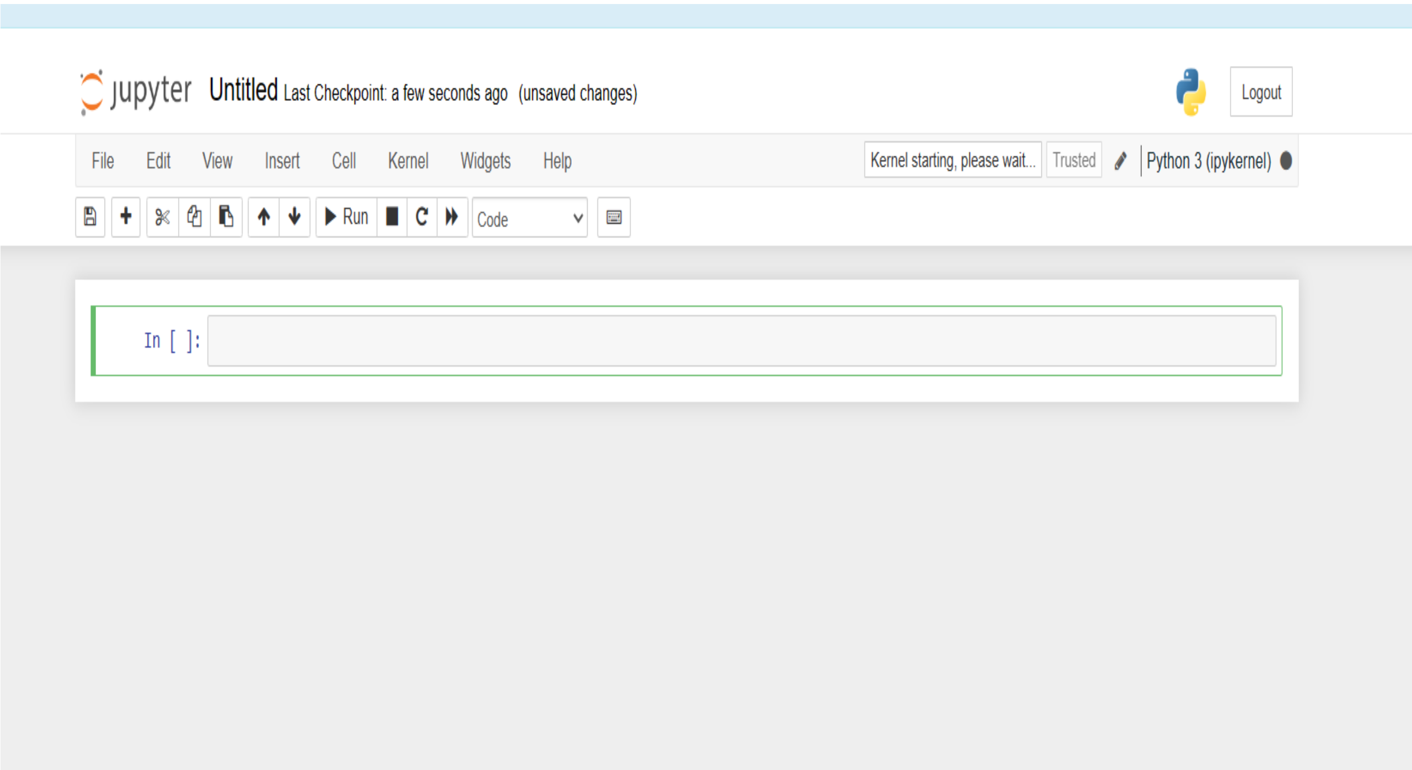
Pandas: Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data. The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" and was created by Wes McKinney in 2008. Pandas allows us to analyze big data and make conclusions based on statistical theories. Pandas can clean messy data sets, and make them readable and relevant. Relevant data is a very important aspect in data science.

Matplotlib: Matplotlib is a cross-platform, data visualization and graphical plotting library (histograms, scatter plots, bar charts, etc) for Python and its numerical extension NumPy. As such, it offers a viable open source alternative to MATLAB.

Some of the features of Matplotlib:

- It can create a variety of plots, including line plots, bar charts, histograms, scatter plots, and heatmaps.

- It is highly customizable, allowing you to control the appearance of your plots in great detail.



Jupyter Notebook layout

- It can export visualizations to a variety of formats, including PDF, SVG, PNG, and EPS.
- It is well-documented, with a large user community and many tutorials available online.

Numpy: NumPy is a Python library that adds support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. It is one of the most used Python packages for scientific computing with Python. NumPy arrays are called ndarrays and can be one-dimensional or multi-dimensional. They are much faster to process than Python lists, and they can be used to store a wide variety of data types, including numbers, strings, and objects.

some of the advantages of using NumPy:

- NumPy arrays are much faster to process than Python lists.
- NumPy arrays can be used to store a wide variety of data types, including numbers, strings, and objects.
- NumPy provides a variety of functions for working with ndarrays, including functions for creating, manipulating, and analyzing arrays.
- NumPy also provides functions for linear algebra, Fourier transforms, and random number generation.

Functions Used:

read_csv (): It is a wrapper function for read.table() that mandates a comma as separator and uses the input file's first line as header that specifies the table's column names. Thus, it is an ideal candidate to read CSV files. It has an additional parameter of url() which is used to pull live data directly from GitHub repository.

Read_tsv (): TSV stands for Tab-Separated Values, which is a text-based file format for storing tabular data. In a TSV file, records are separated by newlines, and values within a record are separated by tab characters. TSV is a delimiter-separated values format, similar to comma-separated values (CSV)

Actual Code:

```
lefthanded_data = pd.read_csv(r'C:\Users\pooji\OneDrive\Documents\Left handed data\data.csv')
```

```
death_distribution_data = pd.read_csv(r'C:\Users\pooji\OneDrive\Documents\Left handed data\data2.tsv', sep='\t', skiprows=[1])
death_distribution_data = death_distribution_data.dropna(subset=['Both Sexes'])
```

3.3 Designing Databases

Once the data has been collected and imported into the Jupyter notebook 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. A table called lefthanded_data is created by running the csv file form Github.

Relative frequency of left handedness among 0-76

```
In [40]: lefthanded_data = pd.read_csv(r'C:\Users\pooji\OneDrive\Documents\Left handed data\data.csv')

In [41]: lefthanded_data
```

Out[41]:

	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
...	...	...	...
72	82	7.350204	4.874899
73	83	6.471204	5.672536
74	84	6.510858	5.774881
75	85	7.337968	7.051459

The columns represent the relative frequency of left-handedness among male and female participants across ages 10-86. From the journal it is found that the survey. A survey of 1,177,507 U.S. men and women between the ages of 10 and 86 included questions regarding hand preference for writing and throwing. Three effects were observed. Individuals with at least some left motoric bias comprised a smaller percent of the population with advancing age

Deaths by Age, Sexes in the year 1999 in the United States

Preview Code Blame 127 lines (127 loc) · 2.18 KB

Search this file

Age	Both Sexes	Male	Female
ALL	2391399	1175460	1215939
0	27937	15646	12291
1	1989	1103	886
2	1376	797	579
3	1046	601	445
4	838	474	364
5	763	446	317
6	696	384	312
7	683	386	297
8	692	389	303
9	640	359	281

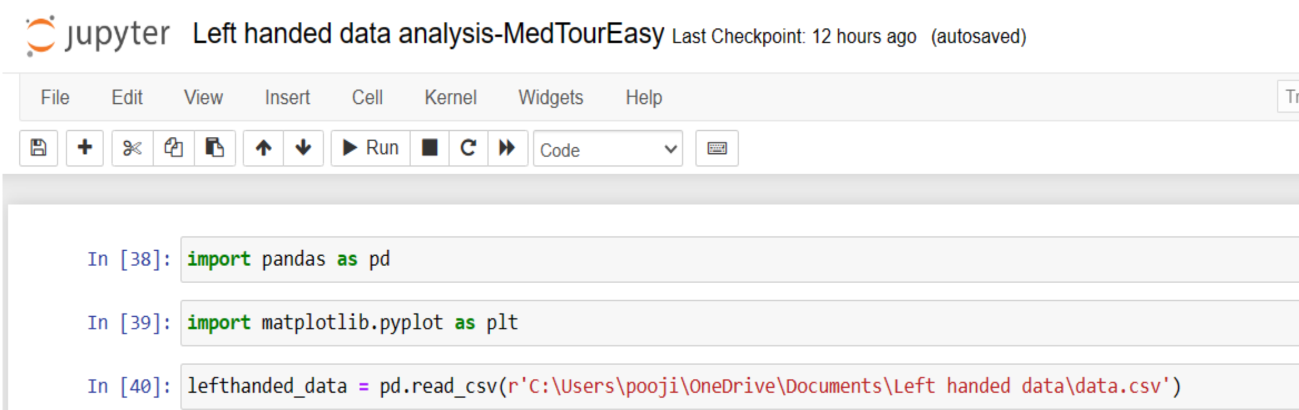
The columns represent the relative frequency of left-handedness among male and female participants across ages 10-86. From the journal it is found that the survey. A survey of 1,177,507 U.S. men and women between the ages of 10 and 86 included questions regarding hand preference for writing and throwing. Three effects were observed. Individuals with at least some left motoric bias comprised a smaller percent of the population with advancing age

3.4 Importing the data to Jupyter notebooks

There are a few ways to import data into a Jupyter Notebook. One way is to use the %csv magic command. This command allows you to read a CSV file directly into the notebook. To use the %csv magic command, simply type %csv filename in a code cell, where filename is the name of the CSV file you want to read. Another way to import data into a Jupyter Notebook is to use the pandas library. Pandas is a Python library that provides data structures and operations forworking withdata.

To use Pandas to import data,you can use the pandas.read_csv() function. This function takes the name of the CSV file you want to read as an argument and returns a Pandas DataFrame object.

Actual code:



Load the handedness data from the National Geographic survey and create a scatter plot.
Import the pandas library as pd and matplotlib.pyplot as plt. Pandas will be useful in importing .csv file into the jupyter environment and the matplotlib library will be helpful in plotting graphs. Once you have imported your data into a Jupyter Notebook, you can use it for analysis, visualization, and other tasks

3.5 Data Visualization by Creating Scatter plots

Data visualization is presenting data in a graphical or pictorial format. It allows decision-makers to see visually presented analytics, so that they can grasp difficult concepts or identify new patterns. In interactive visualizations, technology can be used to dig in charts and graphs for more detail, interactively modifying what data one can see and how it works.

A scatter plot identifies a possible relationship between changes observed in two different sets of variables. It provides a visual and statistical means to test the strength of a relationship between two variables.

3.5.1 Male and Female vs Age

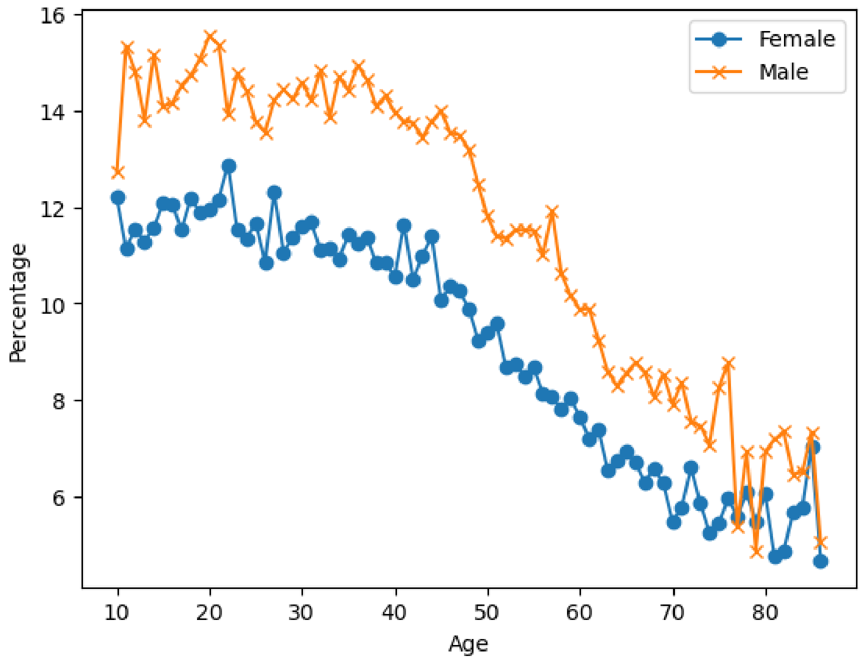
A National Geographic survey in 1986 resulted in over a million responses that included age, sex, and hand preference for throwing and writing. Researchers Avery Gilbert and Charles Wysocki analyzed this data and noticed that rates of left-handedness were around 13% for people younger than 40 but decreased 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.

Let's start by plotting the rates of left-handedness as a function of age.Create a scatter graph with Age along the x-axis and percentage along y-axis.

Actual Code:

```
In [86]: # scatter plot MALE and FEMALE vs Age
fig, ax = plt.subplots() # create figure and axis objects
ax.plot(lefthanded_data.Age, lefthanded_data.Female,label='Female', marker = 'o') # plot "Female" vs. "Age"
ax.plot(lefthanded_data.Age, lefthanded_data.Male, label='Male', marker = 'x') # plot "Male" vs. "Age"

plt.xlabel('Age')
plt.ylabel('Percentage')
plt.legend(loc="upper right")
plt.show()
```

Scatter plot Age vs Percentage



It can be seen from the above scatter plot that the prevalence of left handedness is higher in males than females. Also after the age of 50 there is a constant fall in the percentage of left handedness among both men and women. Thus we can conclude that age plays a role in the hand usage prevalences.

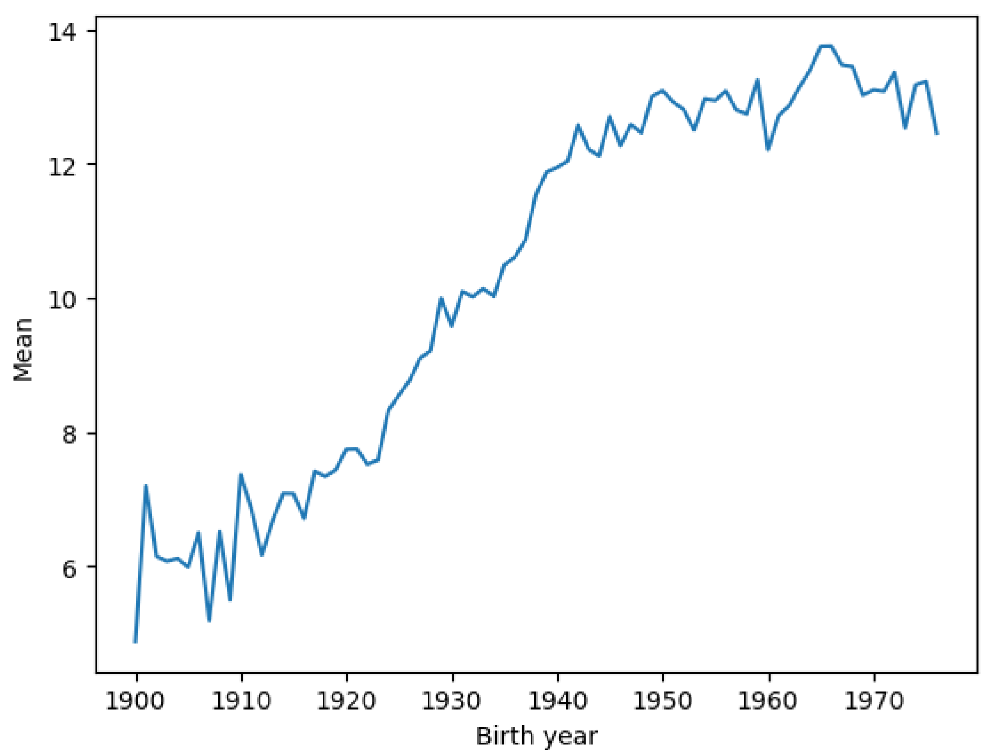
3.5.2 Birth year vs Mean left hand data

By creating this graph we can convert the 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.

Code:

In [88]:
```
# scatter plot Birth year vs Mean left hand data
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) # plot 'Mean_lh' vs. 'Birth_year'

plt.xlabel('Birth year')
plt.ylabel('Mean')


plt.show()
```

Graph showing birth year vs Mean left-handedness



It can be seen from the above scatter plot that the prevalence of left handedness is constantly increasing over time in 1900s. With the mean being around 7 at the start of the 1900s to 13 in the 1970s. We can conclude that as societal norms faded many people began using left hand for writing and throwing activities in the later 1900s.

IV APPLYING THE PROBABILITY AND CONDITIONS

4.1 APPLYING BAYES RULE

Bayes' theorem is a mathematical formula that calculates the probability of an event occurring given that another event has already occurred. In the analysis we want to calculate the **probability of dying at age A given that you're left-handed**. 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) the probability that you are left-handed given that you died at age A
- P(A) overall probability of dying at age A
- P(LH) overall probability of being left-handed

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. Create a function that will return P(LH | A) for particular ages of death in a given study year. Import the numpy package as np. Using last 10 Mean_lh data points to get an average rate and naming the resulting data frame early_1900s_rate. Using last 10 Mean_lh data points to get an average rate and naming the resulting data frame early_1900s_rate. Using the first 10 Mean_lh data points to get an average rate and naming the resulting data frame late_1900s_rate.

Actual Code:

```
In [93]: import numpy as np
def P_lh_given_A(ages_of_death, study_year = 1990):
""" P(Left-handed | age of death), calculated based on the reported rates of left-handedness.
Inputs: age of death, study_year
Returns: probability of left-handedness given that a subject died in `study_year` at age `age_of_death` """

# Use the mean of the 10 neighbouring points for rates before and after the start
early_1900s_rate = lefthanded_data['Mean_lh'][-10:].mean()
late_1900s_rate = lefthanded_data['Mean_lh'][:10].mean()
middle_rates = lefthanded_data.loc[lefthanded_data['Birth_year'].isin(study_year - ages_of_death)]['Mean_lh']

youngest_age = study_year - 1986 + 10 # the youngest age in the NatGeo dataset is 10
oldest_age = study_year - 1986 + 86 # the oldest age in the NatGeo dataset is 86
```

For the early_1900s_ ages, we fill in P_return with appropriate left_handedness rates for the ages_of_death i.e by dividing the early_1900s_rate by 100. Similarly, for the late_1900s_ages, we fill in the p_return with appropriate left handedness rates for the ages_of_death i.e by dividing the late_1900s_rate by 100.

Code:

```
P_return = np.zeros(ages_of_death.shape)  # create an empty array to store the results
# extract rate of left-handedness for people of age age_of_death
P_return[ages_of_death > oldest_age] = early_1900s_rate / 100
P_return[ages_of_death < youngest_age] = late_1900s_rate / 100
P_return[np.logical_and((ages_of_death <= oldest_age), (ages_of_death >= youngest_age))] = middle_rates / 100

return P_return
```

4.2 DEATH DISTRIBUTION DATA FOR UNITED STATES

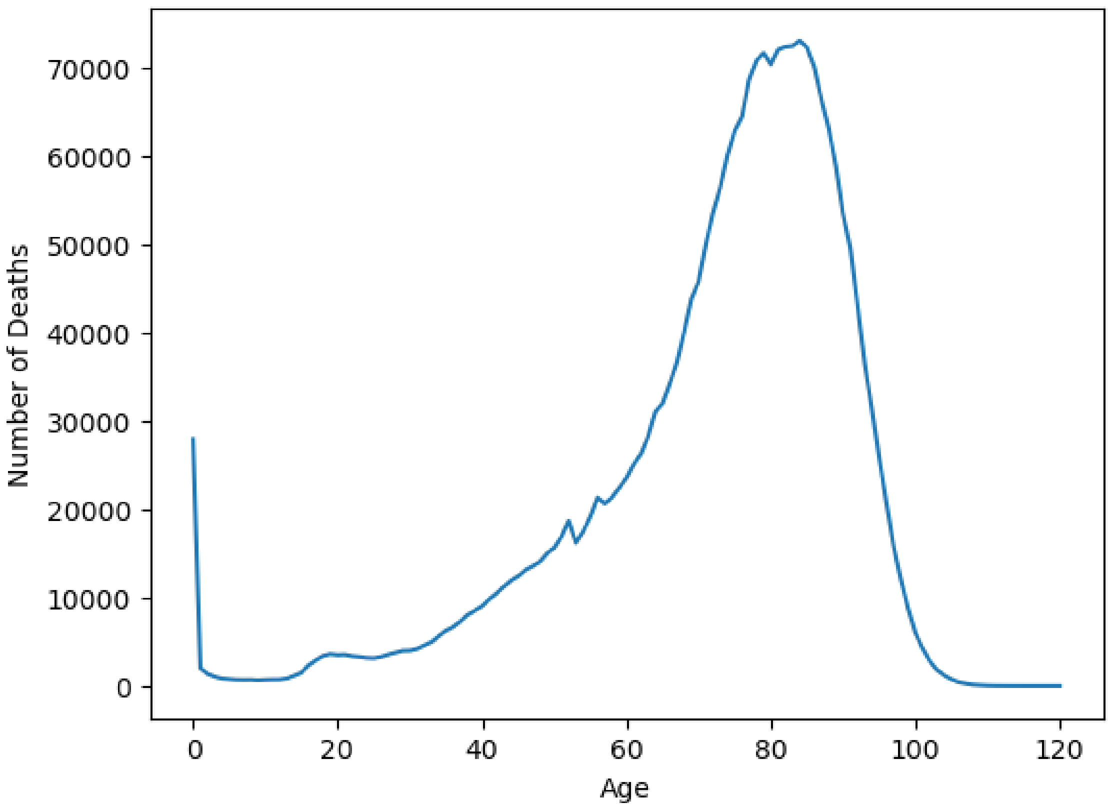
Load the aeth distribution data for the United Stated (1999) and plot it to find P(A) which gives us theprobability of age at which people normally died.

Actual Code:

```
In [91]: # Death distribution data for the United States in 1999

death_distribution_data = pd.read_csv(r'C:\Users\pooji\OneDrive\Documents\Left handed data\data2.tsv', sep='\t', skiprows=[1])
death_distribution_data = death_distribution_data.dropna(subset=['Both Sexes'])

fig, ax = plt.subplots()
ax.plot(death_distribution_data['Age'], death_distribution_data['Both Sexes']) # plot 'Both Sexes' vs. 'Age'
plt.xlabel('Age')
plt.ylabel('Number of Deaths')

plt.show()
```

Load the death distribution data from the .tsv file in Github url given. Name the new data table as death_distribution data. To remove the NA values from both the sexes column we use .dropna('Both Sexes'). By using the plot method we create a grph with age along x-axis and number of deaths along y-axis.

Distribution Graph (Both Sexes) Age vs Number of Death



From the graph it is evident that most of the left handers lived upto their 80s , which is a good longevity as far as the last century was considered. The mean, median and mode which lies around 75-85 suggests that 70000 left handed persons lived to that age range which was considered normal.

4.3 OVERALL PROBABILITY OF LEFT HANDEDNESS

We now create a function called P_lh() which is used to calculate the overall probability of left_handedness in the population for a given study year. We create a series, p_list, by multiplying the number of dead people in the Both_sexes column with the probability of them being lefthanded using P_lh_given_A(). After this we set the variable p equal to the sum of the series. Divide p by the total number of dead people by summing death_distribution_data over the Both_Sexes column, now we return result from the function.

P(LH | A) from the previous steps is used to calculate N(A) is the value of both sexes. DataFrame where the Age column is equal to the denominatior is the total number of dead people sum of Both_Sexes column.

Actual Code:

```
In [65]: 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
             P_lh = P(LH | Age of death) P(Age of death) + P(LH | not A) P(not A) = sum over ages
             Input: dataframe of death distribution data
             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)
             p = np.sum(p_list)
             return p/np.sum(death_distribution_data['Both Sexes']) # normalize to total number of people in distribution
         P_lh(death_distribution_data)

Out[65]: 0.07766387615350638
```

We get an output of 0.7766 which is equal to the probability of being left handed if someone died in the study year 1990. Similarly we can find the probability for other years. Since the value is non negative and >0 we can conclude that the probability of non left handers is also > 0.

4.3.1 THE OVERALL PROBABILITY OF HANDEDNESS P(LH),P(RH)

P(LH) is the probability that a person who died in our particular study year is left-handed. 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. N(A) is the number of people who died at age A

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

We write a function to calculate P_A_given_lh(). Calculate P_A, the overall probability of dying at age A , which ia given by death_distribution_data at age A divided by the total number of dead people. Now we calculate the overall probability of right_handedness P(RH) which is 1 – P(LH).

Actual Code:

```
In [66]: #for 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['Both Sexes'][ages_of_death] / np.sum(death_distribution_data['Both Sexes'])
    P_left = P_lh(death_distribution_data, study_year) # use P_lh function to get probability of left-handedness overall
    P_lh_A = P_lh_given_A(ages_of_death, study_year) # use P_lh_given_A to get probability of left-handedness for a certain age
    return P_lh_A*P_A/P_left
```

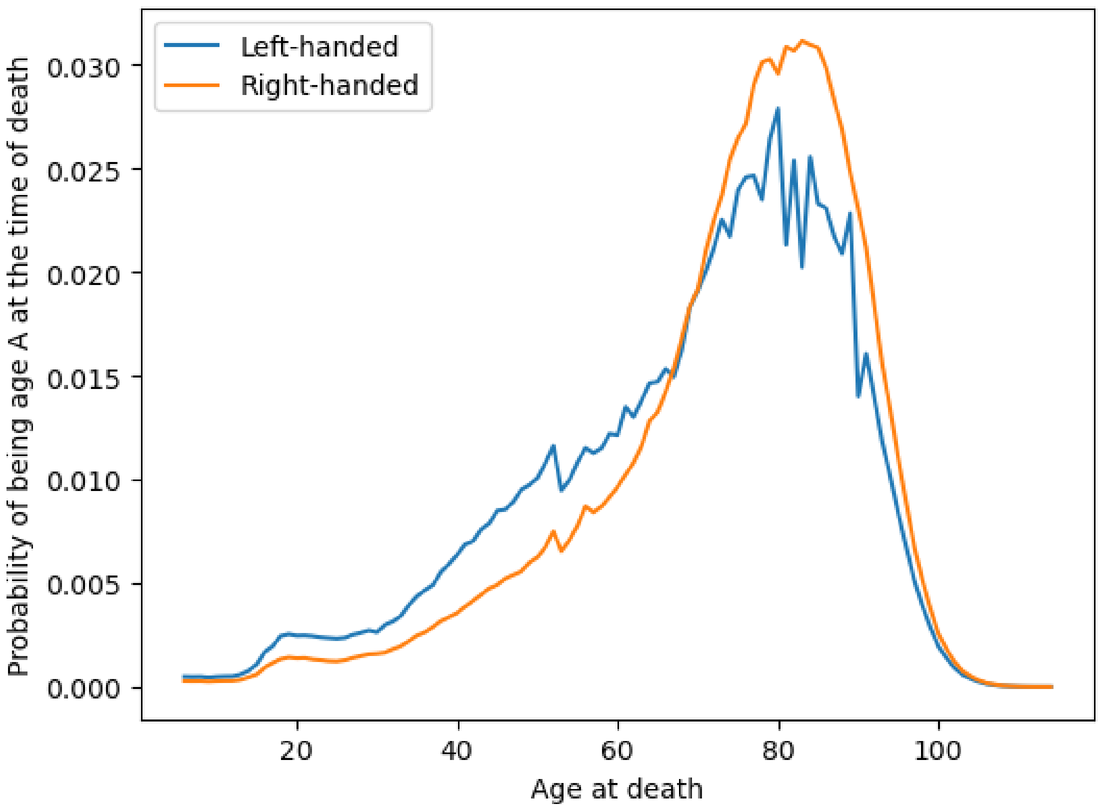
```
In [67]: #for right handers
def P_A_given_rh(ages_of_death, death_distribution_data, study_year = 1990):
    """ The overall probability of being a particular `age_of_death` given that you're right-handed """
    P_A = death_distribution_data['Both Sexes'][ages_of_death] / np.sum(death_distribution_data['Both Sexes'])
    P_right = 1- P_lh(death_distribution_data, study_year) # either you're left-handed or right-handed, so these sum to 1
    P_rh_A = 1-P_lh_given_A(ages_of_death, study_year) # these also sum to 1
    return P_rh_A*P_A/P_right
```

V PLOTTING THE CONDITIONS OF PROBABILITY

5.1 PLOTTING THE DISTRIBUTIONS OF CONDITIONAL PROBABILITIES

Plotting the probability of being certain age at the time of death given that you're right or left handed for a range of ages. For finding this this we calculate,
P_A_given_lh and P_A_given_rh using the functions defined earlier. After finding the probabilities we use .plot() method to plot the results vs Age.
Now Let us plot these probabilities for a range of ages of death from 6 to 120.

Distribution Graph Age at death vs Probability of being age A at death



From the above graph we can see that there is a small gap of 0.010 in probabilities of left and right handed people at the time of death. 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.

Actual Code:

```
In [90]: #plotting the distribution of conditional probablities
ages = np.arange(6,115,1)

left_handed_probability = P_A_given_lh(ages, death_distribution_data)
right_handed_probability = P_A_given_rh(ages, death_distribution_data)

fig, ax = plt.subplots() # create figure and axis objects
ax.plot(ages, left_handed_probability, label = "Left-handed")
ax.plot(ages, right_handed_probability, label = "Right-handed")
plt.xlabel('Age at death')
plt.ylabel('Probability of being age A at the time of death')
plt.legend(loc="upper left")

plt.show()
```

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.

5.2 AGE OF LEFT AND RIGHT HANDERS AT DEATH

In order to arrive at the final conclusion we calculate the mean age at death for left-handers an right-handers. To find this we multiply the ages list by left-handed probabilities of being those ages at death, then we use npnsum to calculate the sum. No, we assign the results to average_lh_age, We repeat the same steps with right-handed probabilities to calculate average-rh-age. Print average_lh_age and average_rh_age followed by calculating the difference between the two average ages and print, to make the printed output uniform we use round() to round off to two decimal places.

Actual Code:

```
In [73]: #Average age of left handers and right handers at death
         average_lh_age =  np.nansum(ages*np.array(left_handed_probability))
         average_rh_age =  np.nansum(ages*np.array(right_handed_probability))

In [74]: >>> print(round(average_lh_age,1))

         67.2

In [75]: >>> print(round(average_rh_age,1))

         72.8

In [76]: >>> print("The difference in average ages is " + str(round(average_rh_age - average_lh_age, 1)) + " years.")

         The difference in average ages is 5.5 years.
```

From the results we can see that the average age when left-handers pass away is 67.2 years and the average age when right- handers pass away is 72.8 years. The difference between the ages is 5.5 years for this particular data we analysed. Thus we can say that left-handers died 5.5 earlier than their right- hander counterparts in 1990.

5.3 CALCULATING THE AGE GAP FOR THE YEAR 2018

Setting the study year parameter to 2018 we calculate the average age gap between left and right handers during the time of death. In thre call to P_A_given_lh , we set age_of_death to ages, death_distribution_data to death_distribution_data, an d study year to 2018. We repeat the steps for P-A_given_rh.

Actual Code:

In [83]:
```
#Age gap expected if the test was done in 2018
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))
```

In [84]:
```
>>> print("The difference in average ages is " + str(round(average_rh_age_2018 - average_lh_age_2018, 1)) + " years.")
```

```
The difference in average ages is 2.3 years.
```

From the results we find the difference in average ages between left and right handers is 2.3 years.

VI CONCLUSION

The analysis throws light upon the independency of left-handedness and age of death. The prevalence of societal restraints in early 1900s can be the cause of less left-handeness among the population, as the years progressed the population of left-handedness also gradually increased. The male population has higher left-handed prevalence than the female population. Also, the normal distribution curve of the age of death for both left and right-handed people is concentrated around 70-80 years. The mean difference in the age of death between the left and right-handers for the year 1999 was 5.5 years. This can be due to other factors such as health issues, ethnicity, sex etc., Finally when the mean age difference was found for the year 2018 it was around 2.3 years which is very less compared to 5.5 years 20 years ago.

This difference is expected to further diminish and hence we can rule out that left-handedness has no effect in causing early death among individuals.

VII REFERENCES

Data Collection

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

a.https://www.cdc.gov/nchs/data/statab/vs00199_table310.pdf

b. Hand preference and age in the united states avery- n. gilbert* and charles j. wysxki~ *givaudan-roure corporation, teaneck, new jersey, u.s.a.; and tmonell chemical senses center, philadelphia, pennsylvania, u.s.a. (for calculating rates of left handeness)

Programming References

The following websites have been referred for Jupyter coding tutorials:

a. https://python.land/data-science/jupyter-notebook

b. https://realpython.com/jupyter-notebook-introduction/

c. https://in.video.search.yahoo.com/search/video?fr=mcafee&ei=UTF-8&p=jupyter+notebook+tutorial&vm=r&type=E210IN826G0#id=1&vid=8ac90795e07000075a7630f98c67674c&action=click