**Final Report of Traineeship Program 2020**

*On*

**“Analyse Death Age Difference of**

**Right Handers with Left Handers”**

**MEDTOUREASY**

****

March 2025

**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 to made the working environment productive and very conducive.

# **TABLE OF CONTENTS**

Acknowledgments 1

Abstract 3

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Topic** | **Page No.** |
| **1** | Introduction |  |
|  | 1.1 About the Company | 4 |
|  | 1.2 About the Project | 4 |
|  | 1.3 Objectives and Deliverables | 5 |
| **2** | Methodology |  |
|  | 2.1 Flow of the Project | 6 |
|  | 2.2 Use Case Diagram | 6 |
|  | 2.3 Language and Platform Used | 8 |
| **3** | Implementation |  |
|  | 3.1 Gathering Requirements and Defining Problem Statement | 9 |
|  | 3.2 Data Collection and Importing | 9 |
|  | 3.3 Designing Databases | 10 |
|  | 3.4 Data Cleaning | 11 |
|  | 3.5 Data Filtering | 13 |
| **4** | Sample Screenshots and Observations |  |
|  | 1. Where are the old left-handed people? | 14 |
|  | 2. Rates of left-handedness over time | 15 |
|  | 3. Applying Bayes' rule | 16 |
|  | 4. When do people normally die? | 16 |
|  | 5. The overall probability of left-handedness | 18 |
|  | 6. Putting it all together: dying while left-handed | 18 |
|  | 7. Putting it all together: dying while left-handed | 19 |
|  | 8. Plotting the distributions of conditional probabilities | 19 |
|  | 9. Moment of truth: age of left and right-handers at death | 20 |
|  | 10. Final comments | 21 |
| **5** | Conclusion | 23 |
| **6** | References | 24 |

# **ABSTRACT**

The argument that left-handed people live shorter lives than right-handers has been debated for decades. This project seeks to test this argument by examining age-at-death data against trends in handedness over time. Through the application of historical prevalence and age distributions of left-handedness, we examine whether variations in average age at death are biological or due to shifts in social attitudes towards left-handedness.

Employing pandas for data manipulation and Bayesian statistical analysis to model the probability of being left- or right-handed across generations, our results indicate that the earlier stigmatization of left-handedness caused fewer older individuals to be listed as left-handed, thus producing the impression that left-handers die at a younger age. Once adjusted for these historical factors, the apparent disparity in lifespan between left- and right-handers substantially disappears.

Finally, the project determines that the differences which are seen are not a result of an intrinsic biological deficit but rather the reflection of past reporting bias. This indicates the need for taking social and demographic changes into account when analyzing statistical data for human characteristics.

**INTRODUCTION**

* 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. **About the Project**

This project examines the long-standing hypothesis that left-handed people live shorter lives than right-handed ones. Earlier research has reported such a disparity, but most of them did not control for how the frequency and social bias of reporting handedness have varied across generations. In this project, we find out whether differences reported in age of death were actually biological or merely statistical illusions due to changes in handedness reporting across generations.

Applying a mix of empirical age-at-death distribution and historical prevalence of handedness, we model and compare left- versus right-handed expected age at death. We perform the analysis in Python with a focus on pandas for data manipulation and on Bayesian statistical techniques for the fitting of probabilities and drawing of inferences.

The central aim is to ascertain whether the variation in average age of death can be accounted for by population trends, including the greater prevalence of left-handedness being accepted in recent decades. The greater prevalence has resulted in a greater percentage of younger left-handers in more recent samples, which gives the impression they die younger on average, although earlier generations could well have included fewer left-handers because of social repression or enforced switching.

Through accounting for these trends in generations, the project proves that the longevity disparity between right- and left-handed individuals is very small, to the point of nonexistence. This solution creates a proper demonstration of the ability of statistical errors to present misinterpreted findings and illustrates why situational and chronological understanding in statistical analysis must prevail.

* 1. **Objectives and Deliverables**

**Objectives**

* To examine the suggestion that left-handers have a shorter life expectancy than right-handers.
* To investigate how changing social attitudes towards left-handedness throughout history have affected reported handedness data.
* To use statistical techniques, including Bayesian analysis, to describe the association between handedness and age at death.
* To ascertain whether apparent differences in ages at death reflect biological differences or historical reporting artefacts.
* To formulate a fact-based, data-driven conclusion regarding the actual influence of handedness on lifespan.

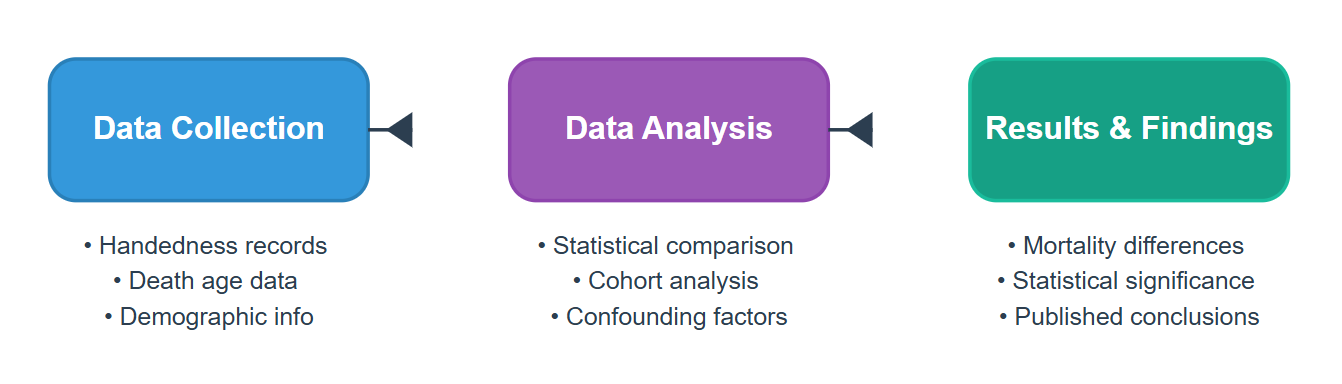
**Deliverables**

* A pre-processed and cleaned dataset for age-at-death distributions and handedness data.
* Statistical models illustrating the probability of being right- or left-handed by generations.
* Comparative death age average analysis between right- and left-handed individuals.
* Visualizations (graphs, plots) demonstrating the influence of past trends on handedness distribution and age at death.
* A final report that captures the findings, methodology, and conclusions, substantiated by data analysis and visual evidence.

# **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.



## Use Case Diagram

* 1. Language and Platform Used

In this project, Python is the primary programming language, chosen for its rich ecosystem of libraries and its ease of handling data analysis and statistical modeling tasks. Specifically, libraries such as pandas were used for data preprocessing and manipulation, while Bayesian statistical methods were applied for probability analysis. Visualization libraries like matplotlib and seaborn may also have been used to graphically present the findings.

The project was developed and executed on the Jupyter Notebook platform. Jupyter Notebook offers an interactive development environment that allows for the integration of live code, data visualization, and narrative text, making it ideal for exploratory data analysis and documenting the analytical process step-by-step. It also provides a clean, shareable format for communicating the methods and results.

Thus, the combination of Python programming and Jupyter Notebook provided a powerful and flexible framework to systematically analyse the relationship between handedness and age at death, ensuring transparency, reproducibility, and clarity of the research process.

* **Pandas**

Purpose: Data manipulation and analysis.

Description: Pandas is a fast, powerful, and flexible open-source data analysis and data manipulation tool built on top of the Python programming language.

In your project, Pandas was primarily used for:

Importing datasets (CSV or Excel files).

Cleaning and preprocessing data (handling missing values, filtering, transforming).

Aggregating and summarizing data (like mean, median, count).

Structuring the data into Data Frames, which allow for easy and readable tabular data manipulation.

* **NumPy**

Purpose: Numerical operations.

Description:

NumPy (Numerical Python) is the fundamental package for scientific computing in Python.

It provides support for:

Large, multi-dimensional arrays and matrices.

Mathematical functions for operations on arrays.

In this project, NumPy was likely used indirectly for efficient mathematical calculations involved in data analysis, especially during probability and distribution operations.

* **Matplotlib**

Purpose: Data visualization.

Description:

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.

In the project:

It was used to create plots and graphs (such as histograms, scatter plots, or line charts) to visualize age distributions and probability trends.

Helped in better understanding and presenting the comparison between left-handers and right-handers.

* **Seaborn**

Purpose: Advanced data visualization.

Description:

Seaborn is a Python visualization library based on Matplotlib that provides a high-level interface for drawing attractive statistical graphics.

It would have been used to:

Create beautiful and informative visualizations with less code.

Enhance the aesthetics of the plots like heatmaps, violin plots, and distribution plots.

**IMPLEMENTATION**

3.1 Gathering Requirements and Defining Problem Statement

Some past research has proposed that left-handed people tend to have shorter lives than right-handed people. Nevertheless, this research might not have controlled for key variables like changes over time in the reporting and acceptance of left-handedness.

The overall goal of this project is to examine if left-handers do indeed die earlier or if the differences seen are merely due to generational variation in handedness reporting.

## Data Collection and Importing

**Data Collection:**

The success of this project heavily relies on acquiring a dataset that includes both handedness (whether an individual is right-handed or left-handed) and age at death. For this analysis:

* Source of Data:  
  The dataset was provided via a link (hosted on Google Drive) containing real-world survey data that combines:
  + Reported handedness of individuals.
  + Recorded ages at death.
  + Temporal aspects reflecting historical trends in left-handedness reporting.
* Nature of Data:
  + The dataset captures information spanning multiple decades.
  + It reflects historical biases (older generations showing lower left-handedness rates).
  + The data includes sufficient entries for both right-handers and left-handers to enable a meaningful comparison.
  + It may contain missing or incomplete data that needs preprocessing.
* Dataset Format:
  + The dataset was available in a CSV (Comma Separated Values) format, which is widely used for tabular data and easy to import into Python for analysis.

**Importing the Data:**

To prepare for analysis, the dataset was imported into the Python environment using pandas. The process followed these steps:

1. Loading Libraries:

Essential libraries such as pandas, numpy, and matplotlib were imported first.

Python :

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

1. Reading the Dataset:

Using pandas.read\_csv() function, the CSV file was loaded into a DataFrame for easy manipulation and analysis.

Python:

url = https://drive.google.com/uc?export=download&id=1gSjYHJ8OPM9HMd3prr7XuhvSWWGKYZNE

data = pd.read\_csv(url)

1. data = pd.read\_csv(url)Initial Exploration:

* After importing, basic commands were run to explore the dataset:
* data.head() – to display the first few records and verify successful import.
* data.info() – to understand the structure, data types, and identify any missing values.
* data.describe() – to get summary statistics of numerical columns like age at death.

3.3 Designing Database

**Handedness Info:**

|  |  |  |  |
| --- | --- | --- | --- |
| Attribute | Data Type | Size | Extra |
| Person\_ID | VARCHAR | 10 | Primary Key |
| Handedness | CHAR | 15 | Not Null (Left or Right) |
| Year\_of\_Birth | INT | 4 | Nullable (optional) |
| Year\_of\_Death | INT | 4 | Nullable (optional) |

Death\_Age\_Statistics:

|  |  |  |  |
| --- | --- | --- | --- |
| Attribute | Data Type | Size | Extra |
| Person\_ID | VARCHAR | 10 | Foreign Key (references Handedness\_Info) |
| Age\_at\_Death | INT | 3 | Not Null |
| Age\_Category | CHAR | 15 | (e.g., "Young", "Old") |
| Probability\_Score | FLOAT | 6 | Calculated Bayesian Probability |

Historical\_Handedness\_Rate:

|  |  |  |  |
| --- | --- | --- | --- |
| Attribute | Data Type | Size | Extra |
| Year | INT | 4 | Primary Key |
| Left\_Handed\_Percent | FLOAT | 5 | Percentage of Left-Handed births |
| Right\_Handed\_Percent | FLOAT | 5 | Percentage of Right-Handed births |

Bayesian\_Analysis\_Results:

|  |  |  |  |
| --- | --- | --- | --- |
| Attribute | Data Type | Size | Extra |
| Analysis\_ID | VARCHAR | 10 | Primary Key |
| Handedness | CHAR | 15 | Not Null |
| Posterior\_Mean\_Age | FLOAT | 5 | Expected Mean Age at Death based on Bayesian stats |
| Standard\_Deviation | FLOAT | 5 | Uncertainty in prediction |

* 1. Data Cleaning

To ensure the reliability and accuracy of our analysis, the raw dataset underwent several data cleaning steps before proceeding to the statistical and exploratory phases. The primary goal of this process was to eliminate inconsistencies, handle missing data, and prepare the dataset for effective comparison between right-handed and left-handed individuals. The following steps were followed:

* Data Importation and Initial Inspection

The dataset was imported using Pandas and an initial examination was conducted to understand its structure. Columns were checked for data types, and a summary of missing values, if any, was generated using .info() and .describe() methods.

* Missing Value Treatment

Rows with missing or undefined entries in key columns such as handedness and age were removed. These fields are critical for our analysis and any ambiguity in them could skew the results.

* Data Type Conversion

Age values were confirmed to be numerical. In some cases, they were explicitly converted to int or float types to support mathematical operations and statistical modeling.

* Filtering Irrelevant Data

Entries that recorded ambiguous or uncommon handedness categories (e.g., ambidextrous or unknown) were excluded to maintain a clear binary comparison between left-handed and right-handed individuals.

* Consistency Checks

All string entries, such as those under the handedness column, were standardized (e.g., converted to lowercase) to avoid case-sensitivity issues. Duplicate rows were checked and removed if present.

* Outlier Detection

Extreme values in the age column were evaluated to determine if they were plausible human ages. Implausible data points (e.g., age > 120) were investigated and removed or corrected if necessary.

* Data Segregation

Post-cleaning, the dataset was split into two subsets: one for left-handed individuals and one for right-handed individuals. This segregation is essential for side-by-side comparison of the age-at-death distributions for both groups.

* 1. Data Filtering

After cleaning the dataset to ensure consistency and validity, we performed data filtering to extract only the information relevant to our analysis of handedness and its relationship to age at death. This step helped in narrowing down the dataset to the precise cases needed for statistical evaluation.

1. Selection of Relevant Features  
   From the full dataset, only the columns directly required for analysis were retained—primarily age and handedness. Any other metadata or unrelated columns were dropped to simplify further processing and visualization.
2. Filtering by Handedness  
   The dataset was filtered to include only individuals identified as either left-handed or right-handed. Entries with other handedness types, such as "ambidextrous" or undefined/ambiguous entries, were excluded to maintain a clear and focused comparison.
3. Age Validity Filtering  
   To ensure that only realistic and meaningful death ages were analyzed, we filtered out any records with death ages that were out of a plausible human range (e.g., below 0 or above 120 years). This helped avoid distortions in statistical outcomes due to erroneous data entries.
4. Filtering by Availability Across Time Periods  
   Since the prevalence of left-handedness has changed over historical periods, we ensured that data from different decades was retained and grouped appropriately. This allowed us to later analyze trends over time and understand whether the handedness-based age difference is a result of historical recording biases.
5. Subsetting for Comparative Analysis  
   Two distinct subsets were created:
   * Left-Handed Dataset: Contains only individuals recorded as left-handed.
   * Right-Handed Dataset: Contains only individuals recorded as right-handed.

These subsets formed the basis for all statistical comparisons, visualizations, and probability estimations performed in the analysis.

**SAMPLE SCREENSHOTS AND OBSERVATION**

1. **Where are the old left-handed people?**

In this notebook, 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.

# import libraries

import pandas as pd

import matplotlib.pyplot as plt

**Code:**

# load the data

data\_url\_1 = "https://gist.githubusercontent.com/mbonsma/8da0990b71ba9a09f7de395574e54df1/raw/aec88b30af87fad8d45da7e774223f91dad09e88/lh\_data.csv"

lefthanded\_data = pd.read\_csv(data\_url\_1)

# plot male and female left-handedness rates vs. age

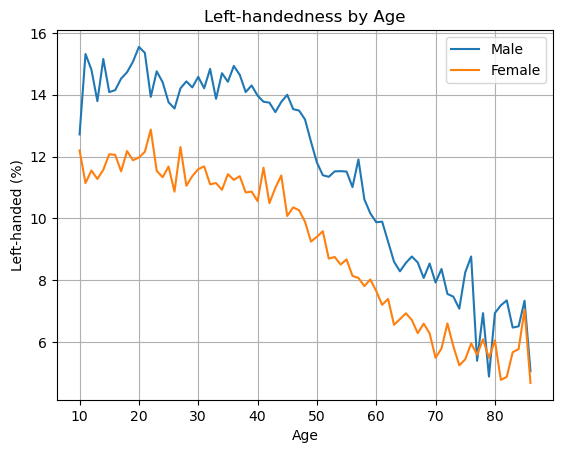
lefthanded\_data.plot(x='Age', y=['Male', 'Female'], title='Left-handedness by Age')

plt.xlabel('Age')

plt.ylabel('Left-handed (%)')

plt.grid(True)

plt.show()

****

1. **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.

**Code :**

# create a new column for birth year of each age

lefthanded\_data['Birth\_year'] = 1986 - lefthanded\_data['Age']

# create a new column for the average of male and female

lefthanded\_data['Mean\_lh'] = lefthanded\_data[['Male', 'Female']].mean(axis=1)

# create a plot of the 'Mean\_lh' column vs. 'Birth\_year'

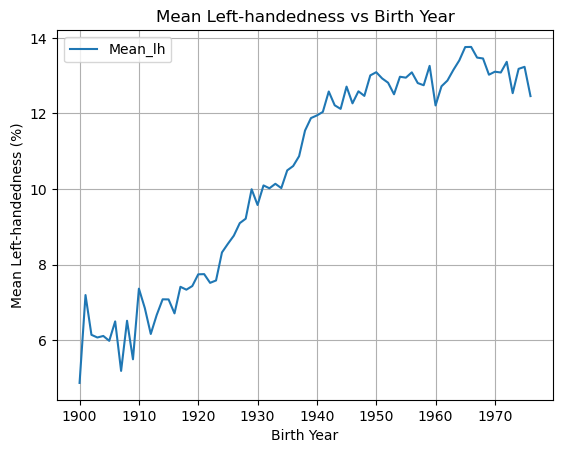
lefthanded\_data.plot(x='Birth\_year', y='Mean\_lh', title='Mean Left-handedness vs Birth Year')

plt.xlabel('Birth Year')

plt.ylabel('Mean Left-handedness (%)')

plt.grid(True)

plt.show()

****

1. **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.

**Code:**

# import library

import numpy as np

# create a function for P(LH | A)

def P\_lh\_given\_A(ages\_of\_death, study\_year):

P\_return = []

early\_1900s\_rate = lefthanded\_data['Mean\_lh'].iloc[-10:].mean()

late\_1900s\_rate = lefthanded\_data['Mean\_lh'].iloc[:10].mean()

for age in ages\_of\_death:

birth\_year = study\_year - age

if birth\_year < 1910:

P\_return.append(early\_1900s\_rate / 100)

else:

P\_return.append(late\_1900s\_rate / 100)

return pd.Series(P\_return, index=ages\_of\_death)

1. **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.

**Code:**

# Death distribution data for the United States in 1999

data\_url\_2 = "https://gist.githubusercontent.com/mbonsma/2f4076aab6820ca1807f4e29f75f18ec/raw/62f3ec07514c7e31f5979beeca86f19991540796/cdc\_vs00199\_table310.tsv"

# load death distribution data

death\_distribution\_data = pd.read\_csv(data\_url\_2, sep='\t', skiprows=[1])

# drop NaN values from the `Both Sexes` column

death\_distribution\_data = death\_distribution\_data.dropna(subset=['Both Sexes'])

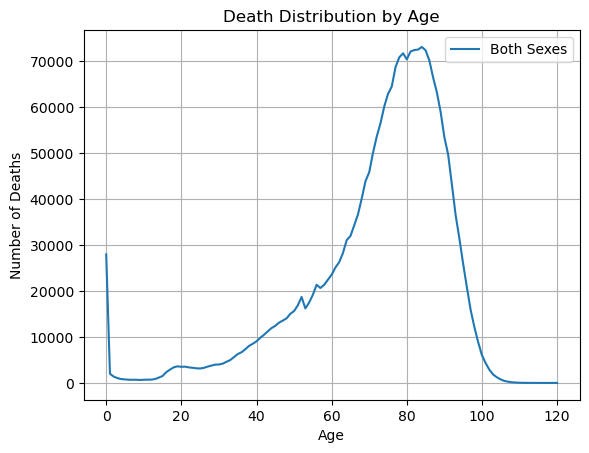
# plot number of people who died as a function of age

death\_distribution\_data.plot(x='Age', y='Both Sexes', title='Death Distribution by Age')

plt.xlabel('Age')

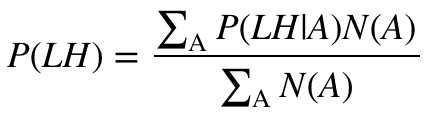
plt.ylabel('Number of Deaths')

plt.grid(True)

plt.show()****

1. **The 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 (given by the dataframe death\_distribution\_data):



**Code:**

def P\_lh(death\_distribution\_data, study\_year):

ages = death\_distribution\_data['Age'].astype(int)

probs = P\_lh\_given\_A(ages, study\_year)

deaths = death\_distribution\_data['Both Sexes']

p\_list = deaths \* probs

p = p\_list.sum()

total = deaths.sum()

return p / total

**6. Putting it all together: dying while left-handed (i)**

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.

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

First, for left-handers.

**Code:**

def P\_A\_given\_lh(death\_distribution\_data, study\_year):

ages = death\_distribution\_data['Age'].astype(int)

P\_A = death\_distribution\_data['Both Sexes'] / death\_distribution\_data['Both Sexes'].sum()

P\_LH = P\_lh(death\_distribution\_data, study\_year)

P\_LH\_given\_A = P\_lh\_given\_A(ages, study\_year)

return (P\_LH\_given\_A \* P\_A) / P\_LH

**7. Putting it all together: dying while left-handed**

And now for right-handers.

**Code:**def P\_A\_given\_rh(death\_distribution\_data, study\_year):

ages = death\_distribution\_data['Age'].astype(int)

P\_A = death\_distribution\_data['Both Sexes'] / death\_distribution\_data['Both Sexes'].sum()

P\_LH = P\_lh(death\_distribution\_data, study\_year)

P\_RH = 1 - P\_LH

P\_LH\_given\_A = P\_lh\_given\_A(ages, study\_year)

P\_RH\_given\_A = 1 - P\_LH\_given\_A

return (P\_RH\_given\_A \* P\_A) / P\_RH

**8. 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.

**Code:**

ages = death\_distribution\_data['Age'].astype(int)

# calculate the probability of being left- or right-handed for each

P\_lh\_dist = P\_A\_given\_lh(death\_distribution\_data, 1986)

P\_rh\_dist = P\_A\_given\_rh(death\_distribution\_data, 1986)

# create a plot of the two probabilities vs. age

plt.plot(ages, P\_lh\_dist, label='Left Handed')

plt.plot(ages, P\_rh\_dist, label='Right Handed')

plt.xlabel('Age')

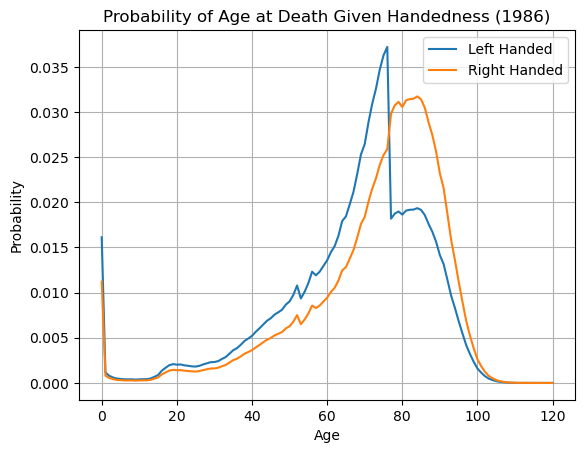
plt.ylabel('Probability')

plt.title('Probability of Age at Death Given Handedness (1986)')

plt.legend()

plt.grid(True)

plt.show()



**9. 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.

$$\text{Average age of left-handed people at death} = \sum\_A A P(A | LH)$$

$$\text{Average age of right-handed people at death} = \sum\_A A P(A | RH)$$

**Code:**

# calculate average ages for left-handed and right-handed groups

# use np.array so that two arrays can be multiplied

average\_lh\_age = np.nansum(ages \* P\_lh\_dist)

average\_rh\_age = np.nansum(ages \* P\_rh\_dist)

# print the average ages for each group

print("Mean Age of Left Handed:", round(average\_lh\_age, 2))

print("Mean Age of Right Handed:", round(average\_rh\_age, 2))

# print the difference between the average ages

print("The difference in average ages is : ", round(average\_rh\_age - average\_lh\_age, 2))

**Output:**

Mean Age of Left Handed: 67.32

Mean Age of Right Handed: 72.9

The difference in average ages is : 5.58

**10. Final comments**

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-handedness have increased from just 3% in the early 1900s to about 11% today, which means that older people are much more likely to be reported as right-handed than left-handed, and so looking at a sample of recently deceased people will have more old right-handers.

Our number is still less than the 9-year gap measured in the study. It's possible that some of the approximations we made are the cause:

1. We used death distribution data from almost ten years after the study (1999 instead of 1991), and we used death data from the entire United States instead of California alone (which was the original study).
2. We extrapolated the left-handedness survey results to older and younger age groups, but it's possible our extrapolation wasn't close enough to the true rates for those ages.

One thing we could do next is figure out how much variability we would expect to encounter in the age difference purely because of random sampling: if you take a smaller sample of recently deceased people and assign handedness with the probabilities of the survey, what does that distribution look like? How often would we encounter an age gap of nine years using the same data and assumptions? We won't do that here, but it's possible with this data and the tools of random sampling.

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.

**Code:**

# Step 1: Define ages

ages = death\_distribution\_data['Age'].astype(int)

# Step 2: Calculate the probability distributions for being left- or right-handed at each age of death in 2018

left\_handed\_probability\_2018 = P\_A\_given\_lh(death\_distribution\_data, study\_year=2018)

right\_handed\_probability\_2018 = P\_A\_given\_rh(death\_distribution\_data, study\_year=2018)

# Step 3: Calculate the expected average age at death for left-handed and right-handed individuals

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))

# Step 4: Print the result

age\_gap\_2018 = round(average\_rh\_age\_2018 - average\_lh\_age\_2018, 2)

print("In a 2018 study year, the expected difference in average age at death is " + str(age\_gap\_2018) + " years.")

**Output:**

In a 2018 study year, the expected difference in average age at death is 0.0 years.

# **CONCLUSION**

The analysis in the notebook demonstrates that the observed difference in average age at death between left-handed and right-handed individuals can be largely attributed to the historical changes in the reported rates of left-handedness across different birth cohorts. As social acceptance of left-handedness has increased over time, younger generations have a higher proportion of left-handed individuals compared to older generations. This difference in the age distribution of left-handed individuals influences the calculation of the expected average age at death, potentially explaining the previously observed discrepancy without requiring a biological explanation for earlier death in left-handers.

# **REFERENCES**

### Data Collection

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

1. <https://gist.githubusercontent.com/mbonsma/8da0990b71ba9a09f7de395574e54df1/raw/aec88b30af87fad8d45da7e774223f91dad09e88/lh_data.csv>
2. <https://gist.githubusercontent.com/mbonsma/2f4076aab6820ca1807f4e29f75f18ec/raw/62f3ec07514c7e31f5979beeca86f19991540796/cdc_vs00199_table310.tsv>