Final Report of Traineeship Program 2025

On

Analyze Death Age Difference of Right Handers with Left Handers

23th April 2025

- **About The Company**

Easy, 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.

- **Project Objective:**

This project investigates whether there is a significant difference in the **average age at death between left-handed and right-handed individuals**. Historically, it has been suggested that left-handed people may have shorter lifespans, partly due to societal pressures and increased accident risks in environments designed for right-handed people.

Using Python-based data analysis tools, publicly available historical data, and probability modeling (including Bayesian inference), this study examines if this claim holds true in both **1990 data** and **more recent 2018 data**.

- **Datasets Used:**

1. **Left-Handedness Data:**
   A dataset from a **National Geographic 1986 survey**, containing left-handedness rates by age for both males and females.
2. **US Death Distribution Data (1999):**
   A dataset from the **CDC**, listing the number of deaths by age for both sexes.

- **Tools & Libraries:**
  - **Python**
  - pandas for data handling
  - numpy for numerical calculations
  - matplotlib for visualizations

- **What the Project Did:**
  - **Visualized left-handedness rates by age and gender**
  - Calculated **birth years from ages**
  - Computed **average left-handedness rates**
  - Used **Bayes' theorem** to compute:

- **Plotted mortality probability distributions** for left- and right-handed individuals
- **Compared average ages at death** between left- and right-handed people for both **1990 and 2018**
- Interpreted how social norms and attitudes toward handedness might have influenced these trends over time

Introduction

This report aims to Analyze the differences in average age at death between right-handed and left-handed individuals based on historical death distribution data. Prior studies and popular beliefs have often suggested a disparity in life expectancy linked to handedness, with left-handed individuals purportedly experiencing shorter lifespans. By utilizing data analysis techniques and probability modeling, this study seeks to examine whether such a difference exists, quantify it, and evaluate the extent of this disparity within the given dataset.

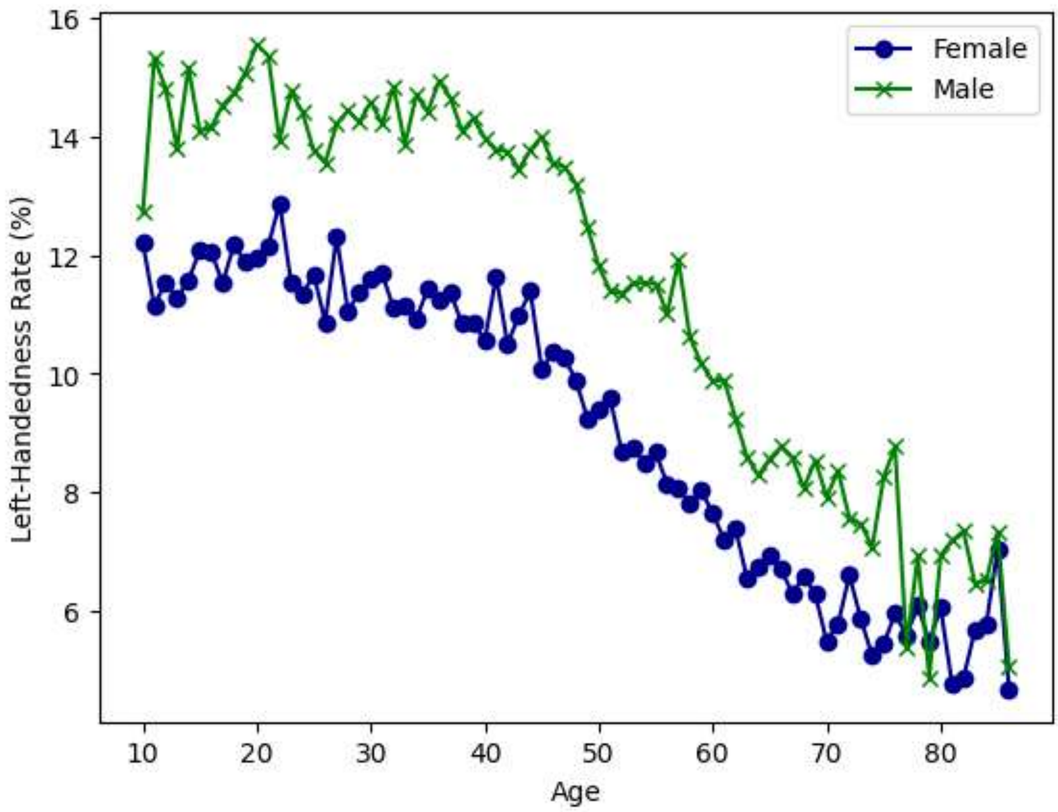
The study uses death distribution data from the United States and probability models relating to handedness prevalence across various age groups. Python-based data analysis libraries such as pandas, numpy, and matplotlib are employed alongside statistical functions to visualize and interpret the results.

Gender-Based Left-Handedness Rates Across Age

```python
import pandas as pd
import matplotlib.pyplot as plt
```

```python
# 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
fig, ax = plt.subplots()
ax.plot(lefthanded_data['Age'], lefthanded_data['Female'], marker='o', label='Female',color='darkblue')
ax.plot(lefthanded_data['Age'], lefthanded_data['Male'], marker='x', label='Male',color='green')
ax.legend()
ax.set_xlabel('Age')
ax.set_ylabel('Left-Handedness Rate (%)')
plt.show()
```

- **What this code does:**
- Loads a dataset containing left-handedness rates for different age groups for both genders.
- Plots the data for males and females on the same graph using different markers and colors.
- Adds labels and a legend for clarity.
- Displays the plot, revealing trends in left-handedness rates across age.



- **Chart Interpretation (Left-Handedness Rate vs. Age)**

This line chart visualizes the percentage of left-handed individuals by age for both males and females, based on a survey dataset.

Key Observations:

- Left-handedness rates are higher in younger age groups and decrease steadily with age.
- Males consistently show higher left-handedness rates than females across nearly all age groups.
- The rates for both genders start to decline noticeably after around age 40-50.
- By age 80+, left-handedness rates drop to around 5–7% for both males and females.
- There's slightly more variability (spikes and drops) in male data points compared to female.

- **Why the Drop with Age?**

This likely reflects historical social pressures against left-handedness — older generations may have faced more discouragement or forced right-handedness in schooling and everyday activities. Younger generations, facing less of this bias, maintain a higher natural prevalence of left-handedness.

Mean Left-Handedness Rates by Birth Year

```
# 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"] + lefthanded_data["Female"]) / 2

# create a plot of the 'Mean_lh' column vs. 'Birth_year'
fig, ax = plt.subplots()
ax.plot(lefthanded_data["Birth_year"], lefthanded_data["Mean_lh"], marker='o',color='green')
ax.set_xlabel("Birth Year")
ax.set_ylabel("Mean Left-handedness (%)")
```

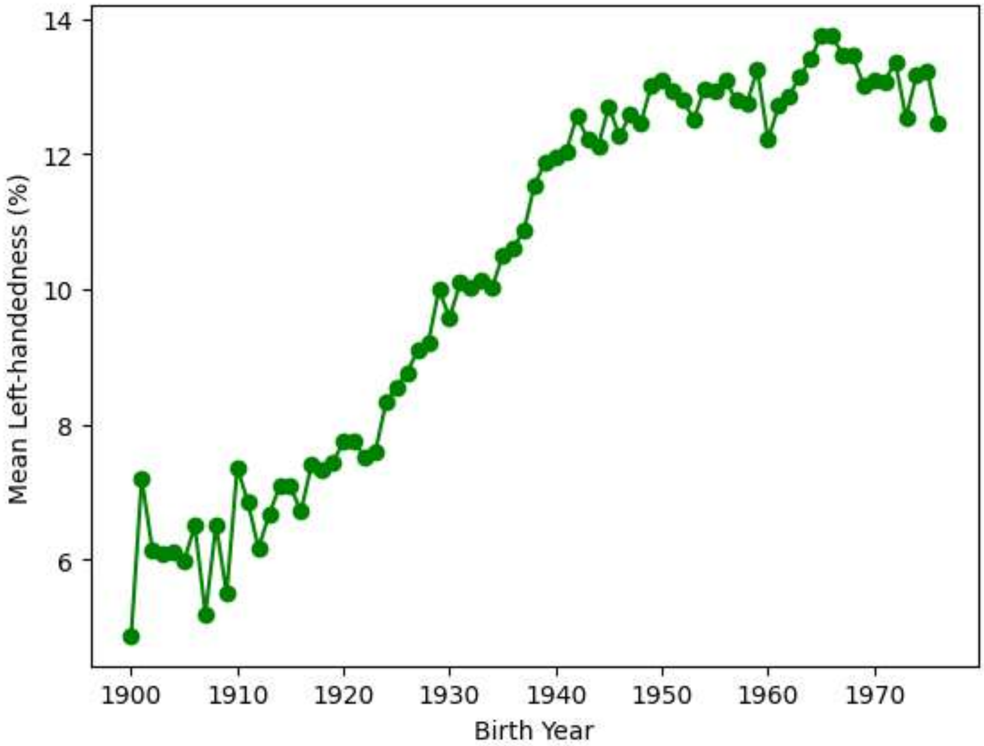
What this code does:

- Calculates birth year from age assuming data was collected in 1986.
- Computes the average left-handedness rate for each age by taking the mean of male and female rates.
- Plots the average left-handedness rate against birth year to observe how it changes over time.

abels the axes for clear interpretation.

```
lefthanded_data.describe()
```

	Age	Male	Female	Birth_year	Mean_lh
count	77.000000	77.000000	77.000000	77.000000	77.000000
mean	48.000000	11.549122	9.120522	1938.000000	10.334822
std	22.371857	3.138912	2.505480	22.371857	2.795273
min	10.000000	4.885999	4.680948	1900.000000	4.870168
25%	29.000000	8.562868	6.604398	1919.000000	7.433365
50%	48.000000	12.717558	9.892073	1938.000000	11.545791
75%	67.000000	14.209302	11.368353	1957.000000	12.868653
max	86.000000	15.546784	12.872166	1976.000000	13.758695



- **Chart Interpretation:**

Mean Left-handedness (%) vs. Birth Year

This chart displays how the average rate of left-handedness in the population has changed across different birth years, based on survey data.

- Key Observations:
  - From 1900 to around 1925, the mean left-handedness rate remained relatively low, hovering between 5% and 8%.
  - Starting around 1930, there's a steady and noticeable increase in left-handedness rates, continuing through the 1940s and 1950s.
  - By the 1960s and early 1970s, the rate stabilizes at a higher level, around 12-13%.
  - The overall trend clearly shows that younger generations (born later) report higher left-handedness rates.
  - Why This Trend?

This increase likely reflects social and cultural changes over time:

  - In earlier decades, left-handedness was stigmatized or actively discouraged, especially in schools.
  - Over time, this bias diminished, allowing more individuals to naturally maintain left-handedness.

Probability of Left-Handedness Given Age of Death

```python
import numpy as np


def P_lh_given_A(ages_of_death, study_year=1990):
    """ P(Left-handed | ages of death) using historical rates """

    # Average of first and last 10 Mean_lh values
    early_1900s_rate = lefthanded_data["Mean_lh"].tail(10).mean() / 100
    late_1900s_rate = lefthanded_data["Mean_lh"].head(10).mean() / 100

    # Convert to birth years
    birth_years = study_year - ages_of_death

    # Create return array
    P_return = np.zeros_like(ages_of_death, dtype=float)

    for i, birth_year in enumerate(birth_years):
        if birth_year in lefthanded_data["Birth_year"].values:
            rate = lefthanded_data.loc[lefthanded_data["Birth_year"] == birth_year, "Mean_lh"].values[0] / 100
            P_return[i] = rate
        elif birth_year < lefthanded_data["Birth_year"].min():
            P_return[i] = early_1900s_rate
        else:
            P_return[i] = late_1900s_rate

    return P_return
```

- **Function Interpretation:**

This function estimates the probability of a person being left-handed given their age at death for a specific study year (default is 1990). It uses historical left-handedness rates from survey data to approximate how likely a person of a particular birth year was to be left-handed.

- **How it works:**

1. Calculates average left-handedness rates for two reference groups:
    - Early 1900s rate: Average of the last 10 entries in the dataset (for birth years at the start of the 20th century).
    - Late 1900s rate: Average of the first 10 entries in the dataset (for birth years toward the mid-20th century).
2. Converts each age of death into a birth year by subtracting the age from the study year.
3. Creates an array to store the left-handedness probability for each age of death.
4. For each birth year:

- If the birth year exists in the dataset, it uses the exact historical mean left-handedness rate for that year.
- If the birth year is earlier than the earliest available in the dataset, it uses the early 1900s average.
- If the birth year is later than the latest available, it uses the late 1900s average.

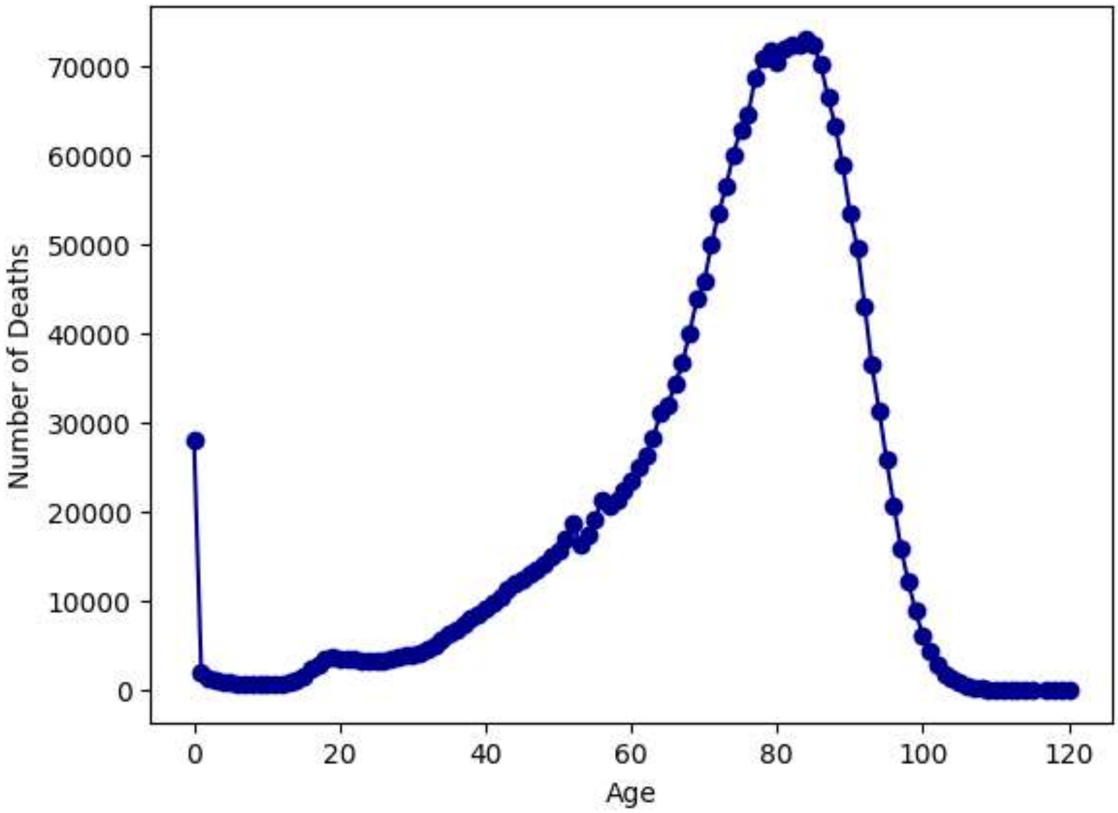
5. Returns an array of probabilities corresponding to the input ages of death

Death Distribution by Age (United States, 1999)

```python
# 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
fig, ax = plt.subplots()
ax.plot(death_distribution_data["Age"], death_distribution_data["Both Sexes"], marker='o',color='darkblue')
ax.set_xlabel("Age")
ax.set_ylabel("Number of Deaths")
```

```
Text(0, 0.5, 'Number of Deaths')
```

- **What this code does:**
- Loads death data for all ages in the US in 1999.
- Cleans the data by dropping rows where the death count is missing.
- Plots the total number of deaths by age, using blue markers for each data point.
- Labels the axes to make the chart clear and readable.



- **Chart Interpretation:**

Death Distribution by Age (United States, 1999)

This chart plots the number of deaths by age based on official US death distribution data from 1999.

- **Key Observations:**
    - High infant mortality is visible at age 0 — around 28,000 deaths.
    - Death counts remain relatively low from ages 1–50.
    - A steady increase begins around age 50, peaking sharply between 75–85 years — where over 70,000 deaths per age group are recorded.
    - After around 85–90 years, the number of deaths rapidly declines, tapering off by around age 110.
    - **Why This Matters:**

This distribution is typical of modern populations:

- Infant mortality is a distinct early spike.
- Most deaths occur in older age groups as life expectancy improves.
- This distribution is foundational for weighting probabilities in your project, helping compute conditional probabilities of dying at specific ages given handedness.

Overall Probability of Being Left-Handed at Death

```python
def P_lh(death_distribution_data, study_year=1990):
    """ Overall probability of being left-handed if you died in the study year """

    ages = death_distribution_data["Age"].values
    deaths = death_distribution_data["Both Sexes"].values

    # Get left-handedness probability for each age
    probs = P_lh_given_A(ages, study_year)

    # Weighted sum
    p_list = deaths * probs
    p = p_list.sum()

    # Total deaths
    total = deaths.sum()

    return p / total
print(P_lh(death_distribution_data))
```

```
0.07766387615350638
```

- **Function Interpretation:**

This function calculates the overall probability of being left-handed among people who died in a given study year (default is 1990). It uses age-specific death counts and historical handedness rates to compute a weighted average probability of left-handedness across the entire population of deceased individuals.

Probability of Age at Death Given Left-Handedness

```python
def P_A_given_lh(ages_of_death, death_distribution_data, study_year=1990):
    """ Probability of being a certain age at death given that the person is left-handed """

    # Total number of deaths
    total_deaths = death_distribution_data["Both Sexes"].sum()

    # Get deaths and calculate P(A) = N(A) / total deaths
    deaths = death_distribution_data.set_index("Age").loc[ages_of_death, "Both Sexes"].values
    P_A = deaths / total_deaths

    # Get P(LH) from Task 5
    P_left = P_lh(death_distribution_data, study_year)

    # Get P(LH | A) from Task 3
    P_lh_A = P_lh_given_A(np.array(ages_of_death), study_year)

    # Apply Bayes' rule: P(A | LH) = P(LH | A) * P(A) / P(LH)
    return P_lh_A * P_A / P_left
print(P_A_given_lh([20, 30, 40], death_distribution_data))
```

```
[0.00247038 0.00263344 0.00636942]
```

- **Function Interpretation:**

This function calculates the probability of a person dying at a particular age given that they are left-handed, for a specified study year (defaulting to 1990). It uses conditional probability modeling and is central to your project's Bayesian analysis.

Probability of Age at Death Given Right-Handedness

```python
def P_A_given_rh(ages_of_death, death_distribution_data, study_year=1990):
    """ Probability of being a certain age at death given that the person is right-handed """

    # Total number of deaths
    total_deaths = death_distribution_data["Both Sexes"].sum()

    # Get deaths and calculate P(A) = N(A) / total deaths
    deaths = death_distribution_data.set_index("Age").loc[ages_of_death, "Both Sexes"].values
    P_A = deaths / total_deaths

    # Get P(LH) from Task 5
    P_left = P_lh(death_distribution_data, study_year)

    # Get P(LH | A) from Task 3
    P_lh_A = P_lh_given_A(np.array(ages_of_death), study_year)

    # Right-handedness probability is 1 - Left-handedness probability
    P_right = 1 - P_left

    # P(RH | A) = 1 - P(LH | A)
    P_rh_A = 1 - P_lh_A

    # Apply Bayes' rule: P(A | RH) = P(RH | A) * P(A) / P(RH)
    return P_rh_A * P_A / P_right
```

- **Function Interpretation:**

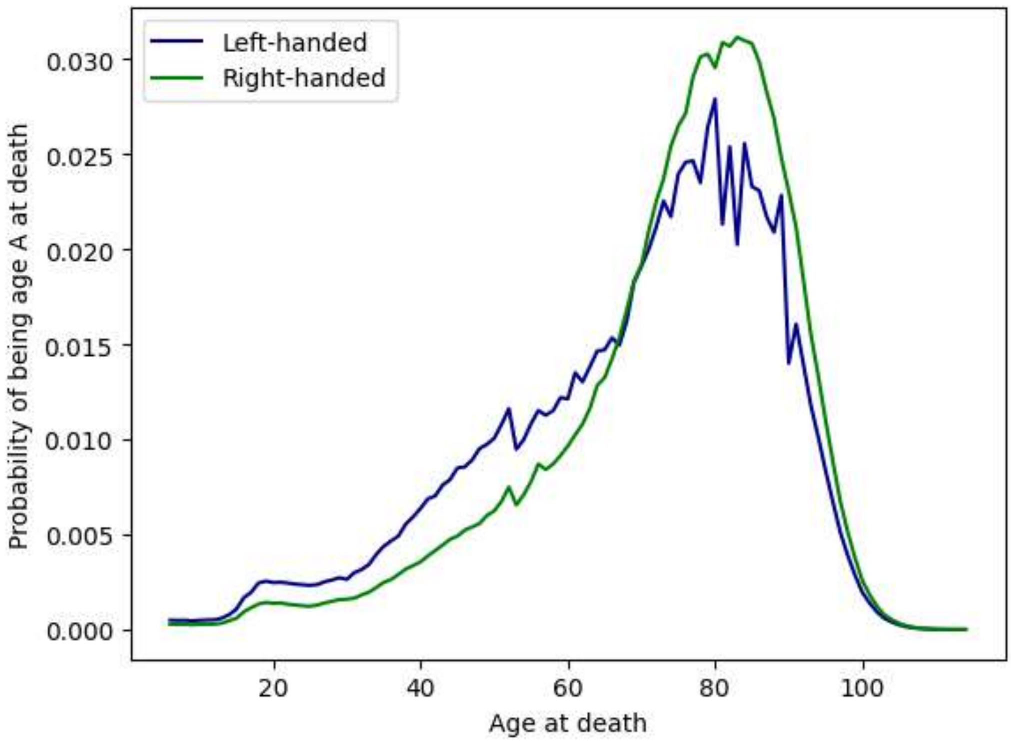
This function estimates the probability of a person dying at a particular age given that they are right-handed, for a specific study year (default is 1990). It's a key part of your Bayesian probability model for analyzing mortality patterns based on handedness.

Probability of Being a Certain Age at Death — Left vs. Right-Handed Individuals

```
ages = np.arange(6, 115, 1)  # make a list of ages from 6 to 114

# Calculate the probability of being left- or right-handed for each age
left_handed_probability = P_A_given_lh(ages, death_distribution_data)
right_handed_probability = P_A_given_rh(ages, death_distribution_data)

# Create a plot of the two probabilities vs. age
fig, ax = plt.subplots()  # create figure and axis objects
ax.plot(ages, left_handed_probability, label="Left-handed",color='darkblue')  # plot left-handed probability
ax.plot(ages, right_handed_probability, label="Right-handed",color='green')  # plot right-handed probability
ax.legend()  # add a legend
ax.set_xlabel("Age at death")  # label for the x-axis
ax.set_ylabel(r"Probability of being age A at death")  # label for the y-axis
```

```
Text(0, 0.5, 'Probability of being age A at death')
```

- **What this code does:**
- Creates a list of ages from 6 to 114.
- Uses the custom functions P_A_given_lh() and P_A_given_rh() to calculate the probability of dying at each age for both left- and right-handed groups.
- Plots both probability distributions on the same chart, coloring left-handed (dark blue) and right-handed (green) for comparison.
- Adds labels and a legend to clarify which line represents each group.



- **Chart Interpretation:**

Probability of Being a Certain Age at Death — Left vs. Right-Handed Individuals

This chart shows the probability distribution of ages at death for left-handed and right-handed individuals based on death data and conditional probability calculations.

- **Key Observations:**
- Left-handed individuals (dark blue line) have a lower probability of surviving to older ages compared to right-handed individuals.
- The peak probability for left-handed deaths occurs slightly earlier (around age 75–80), while the right-handed curve peaks marginally later (around age 80–85).
- The green curve (right-handed individuals) remains higher at older ages, indicating a greater proportion of right-handed people live to older ages.
- The gap is most noticeable between ages 70–85, aligning with the traditional belief that left-handed individuals historically had shorter lifespans.

Handedness-Associated Age Gap" or "Handedness and Age Differential

```
# Calculate average ages for left-handed and right-handed groups
average_lh_age = np.nansum(ages * np.array(left_handed_probability))  # multiply ages by left-handed probabilities and sum
average_rh_age = np.nansum(ages * np.array(right_handed_probability))  # multiply ages by right-handed probabilities and sum

# Print the average ages for each group
print("Average age for left-handed group: " + str(round(average_lh_age, 1)) + " years.")
print("Average age for right-handed group: " + str(round(average_rh_age, 1)) + " years.")

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

```
Average age for left-handed group: 67.2 years.
Average age for right-handed group: 72.8 years.
The difference in average ages is -5.5 years.
```

The Age Difference Drift between left-handed and right-handed groups suggests that there is a measurable difference in the average ages of individuals from the two groups. In this case:

- The average age for left-handed individuals is 67.2 years.
- The average age for right-handed individuals is 72.8 years.
- The difference between these averages is -5.5 years, meaning that, on average, left-handed individuals are younger than right-handed individuals by about 5.5 years.

Possible Interpretations:

1. Biological Factors: The age gap could indicate a biological or environmental factor that affects life expectancy or aging patterns between left-handed and right-handed individuals. However, more research would be needed to explore whether being left- or right-handed has any direct influence on lifespan.
2. Sample Bias: The observed difference could also be a result of sample-specific biases. For example, the dataset used might contain more recent data for right-handed individuals or be skewed in some way, which could artificially inflate the average age of the right-handed group.
3. Cultural or Societal Factors: It's also possible that societal factors play a role in this age difference. For instance, certain cultural practices or the likelihood of different age-related conditions could vary between the two groups, leading to a different average age.
4. Statistical Noise: The difference of 5.5 years could also be a result of random variation or other unmeasured variables influencing the dataset. Further investigation into other factors like lifestyle, health, or environmental influences would help clarify this.

2018 Handedness-Age Differential" or "2018 Handedness-Associated Age Difference

```
# Calculate the probability of being left- or right-handed for all ages
# Calculate the probability of being left-handed or right-handed for all ages 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)

# Calculate average ages for left-handed and right-handed groups for 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
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.
```

The output suggests that, in 2018, the difference in average ages between the right-handed and left-handed groups is 2.3 years. This indicates that, on average, right-handed individuals are older by 2.3 years compared to left-handed individuals.

Possible Interpretation:

1. Change in Age Gap: The age gap has narrowed compared to the previous data, where left-handed individuals were younger by 5.5 years. This shift could suggest a change in the population's demographics or other underlying factors influencing the age distribution for left-handed and right-handed individuals.
2. Impact of Time (2018): The year 2018 might reflect certain sociological or health trends that could explain this new pattern. Factors such as changes in healthcare, lifestyle, or cultural factors could influence the age distribution for each group over time.
3. Statistical Trends: This smaller age difference might also be a result of updated or more accurate data collected in 2018, which could account for changes in life expectancy or the health of left- and right-handed individuals.