

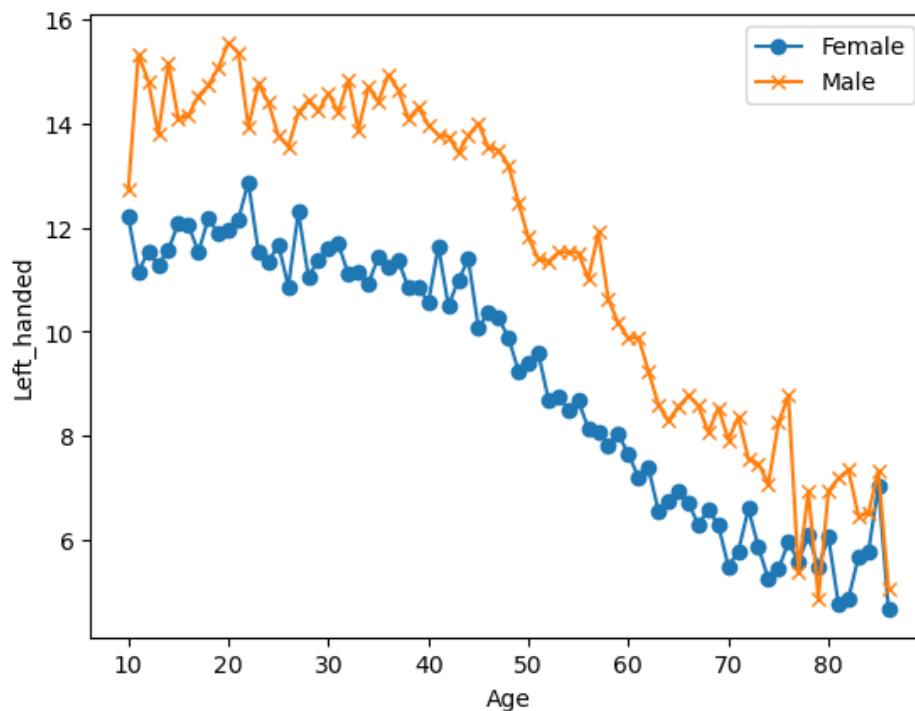
Analyze Death Age Difference of Right Handers with Left Handers

Project Description

In this project, you will explore this phenomenon using age distribution data to see if we can reproduce a difference in average age at death purely from the changing rates of left-handedness over time, refuting the claim of early death for left-handers. This notebook uses pandas and Bayesian statistics to analyze the probability of being a certain age at death given that you are reported as left-handed or right-handed.

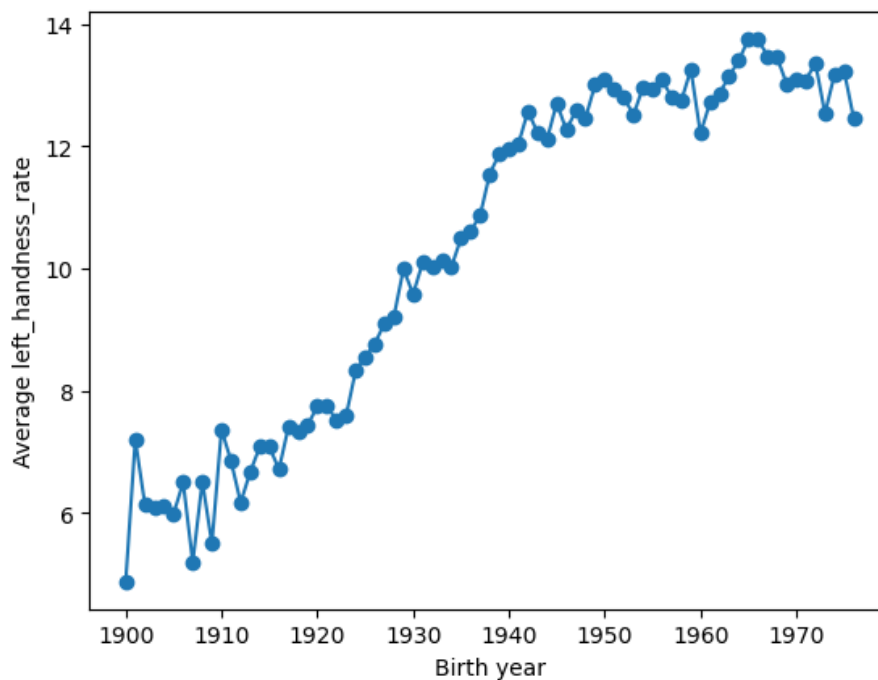
EXPERIMENT AND RESULT

1. Male and Female Left-Handedness Rates vs. Age:



- In this visual representation, we observe the scatter plot, which demonstrates the left-handedness rates for both males and females across different age groups.
- The plot reveals the distribution of left-handedness within the population across age brackets. It allows us to discern whether there are any age-related patterns in handedness.
- By labelling the axes and incorporating a legend, the plot becomes more interpretable, enabling us to distinguish between the male and female data points. This distinction is valuable for understanding potential gender-based differences in handedness.

2. Rates of Left-Handedness Over Time:



- This plot showcases the temporal evolution of left-handedness rates by computing the mean of male and female left-handedness rates across various birth years.
- It provides insights into how the prevalence of left-handedness has potentially changed over time, offering a historical perspective on handedness trends.
- The plot's trend line enables us to identify whether left-handedness has shown any significant shifts or fluctuations across different birth cohorts.

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.

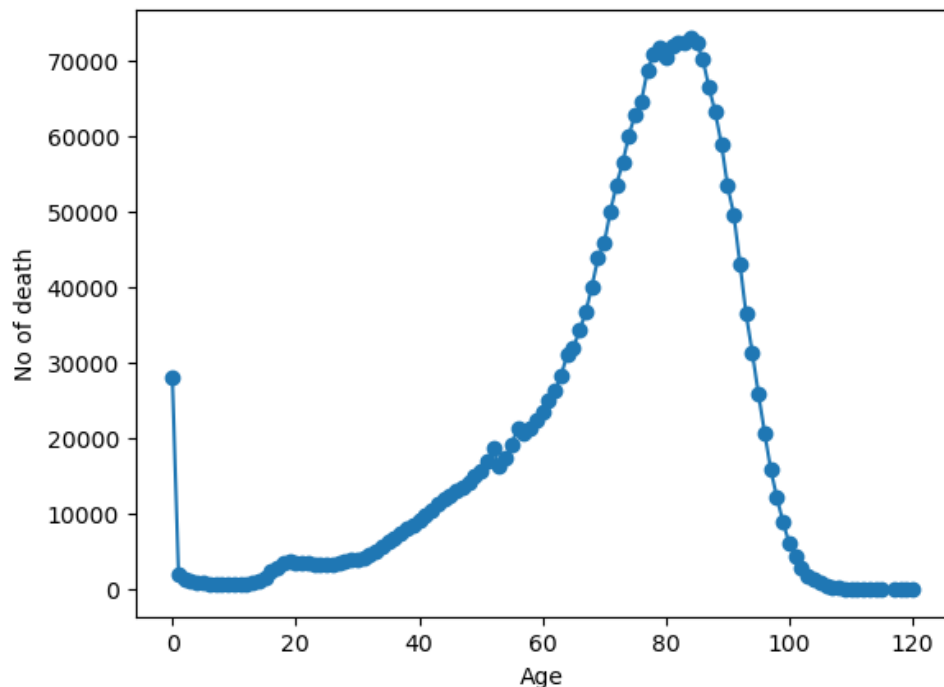
We want to calculate the probability of dying at age A given that you're left-handed. Let's write this in shorthand as $P(A | LH)$. We also want the same quantity for right-handers: $P(A | RH)$.

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

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

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

4. Death Distribution Data for the United States in 1999



This data set in 1999 offers an overview of the ages at which individuals passed away during that year.

- It provides a visual representation of the distribution of ages at death, shedding light on the demographic aspects of mortality in that specific time frame.
- This analysis is foundational for understanding the typical age at which people passed away during the year 1999, offering context for assessing the age distribution within the population.

5. 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`):

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

The overall probability of left-handedness is 0.06891656945882148

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

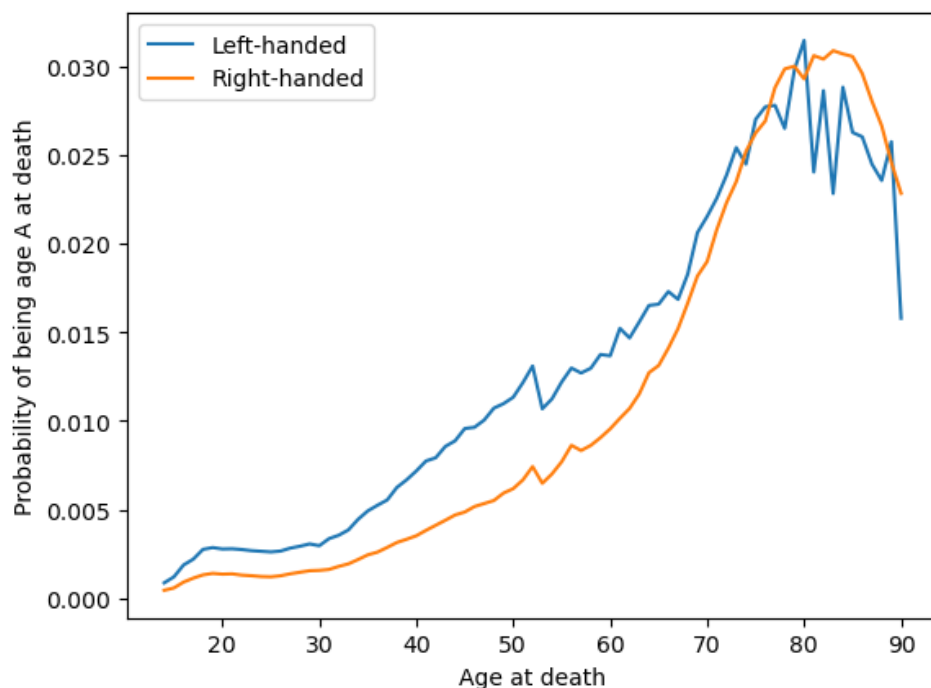
Now we have the means of calculating all three quantities we need: $P(A)$, $P(LH)$, and $P(LH | A)$. We can combine all three using Bayes' rule to get $P(A | LH)$, the probability of being age A at death (in the study year) given that you're left-handed. To make this answer meaningful, though, we also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

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

7. Putting it all together: dying while left-handed (ii)

8. Probability of Age at Death Given Handedness:



- The probability plots reveal the distributions of ages at death for left-handed and right-handed individuals. They offer insights into the likelihood of reaching specific ages at death based on handedness.
- By contrasting the probability distributions for the two groups, the plots facilitate the identification of potential disparities in age at death between left-handed and right-handed individuals.

- This analysis underscores the relevance of handedness as a factor in mortality studies, potentially challenging or confirming prevailing assumptions about its impact on the lifespan.

These visual representations and analyses are essential for extracting meaningful insights from the data, enabling us to draw conclusions and contribute to the overarching research question of whether there is a significant age difference between left-handers and right-handers.

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.

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

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

Av age of left-handed is : 66.73784037679073
 Av age of right-handed is : 61.66446645150637
 The difference in average ages is : 5.1 years.

10. 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 sinister ness. 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 more 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:

- 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).
- 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.

The difference in average ages is 4.1 years.

CONCLUSION

From the analysis conducted, we can see that left-handers experience an earlier average age of death than right – handers.

APPENDICES

APPENDIX 1

```
# import libraries
import pandas as pd
import matplotlib.pyplot as plt

# load the data
data_url_1 =
"https://gist.githubusercontent.com/mbonsma/8da0990b71ba9a09f7de395574
e54df1/raw/aec88b30af87fad8d45da7e774223f91dad09e88/lh_data.csv"
lefthanded_data = pd.read_csv(data_url_1)

# plot male and female left-handedness rates vs. age
%matplotlib inline
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"
ax.legend() # add a legend
ax.set_xlabel('Age')
ax.set_ylabel('Left_handed')
plt.show()
```

APPENDIX 2

```
# 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['Female'] +
lefthanded_data['Male'])/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') #
plot 'Mean_lh' vs. 'Birth_year'
ax.set_xlabel('Birth year') # set the x label for the plot
ax.set_ylabel('Average left_handness_rate') # set the y label for the plot
plt.show()
```

APPENDIX 3

```
# import library
import numpy as np

# create a function for P(LH | A)
def P_lh_given_A(ages_of_death, study_year = 1990):
    """ P(Left-handed | ages of death), calculated based on the reported rates of
    left-handedness.
    Inputs: numpy array of ages of death, study_year
    Returns: probability of left-handedness given that subjects died in
    `study_year` at ages `ages_of_death` """

    # Use the mean of the 10 last and 10 first points for left-handedness rates
    before and after the start
    early_1900s_rate =
    lefthanded_data.loc[lefthanded_data['Birth_year'].between(study_year - 1900 -
    10, study_year - 1900)]['Mean_lh'].mean()
    late_1900s_rate =
    lefthanded_data.loc[lefthanded_data['Birth_year'].between(study_year - 1900 ,
    study_year - 1900 + 10)]['Mean_lh'].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 is 10
    oldest_age = study_year - 1986 + 86 # the oldest age is 86

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

    return P_return
```

APPENDIX 4

```
import pandas as pd

# Death distribution data for the United States in 1999
data_url_2 =
"https://gist.githubusercontent.com/mbonsma/2f4076aab6820ca1807f4e29f75
f18ec/raw/62f3ec07514c7e31f5979beeca86f19991540796/cdc_vs00199_tabl
e310.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
=["2391399"])
```

```
# plot number of people who died as a function of age
fig, ax = plt.subplots()
ax.plot(death_distribution_data['ALL'], death_distribution_data["2391399"],
marker='o') # plot 'Both Sexes' vs. 'Age'
ax.set_xlabel('Age')
ax.set_ylabel('No of death')
plt.show()
```

APPENDIX 5

```
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 = P_Lh_given_A(death_distribution_data['ALL'], study_year) *
death_distribution_data["2391399"] # multiply number of dead people by
P_Lh_given_A
    p = p_list.sum() # calculate the sum of p_list
    P_Lh = p/ np.sum(death_distribution_data["2391399"])
    return P_Lh # normalize to total number of people (sum of
death_distribution_data['Both Sexes'])

print("The overall probability of left-handedness is ",
P_Lh(death_distribution_data))
```

APPENDIX 6

```
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["2391399"][ages_of_death] /
np.sum(death_distribution_data["2391399"])
    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
```

APPENDIX 7

```
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['2391399'][ages_of_death] /
np.sum(death_distribution_data['2391399'])
    P_right = 1 - P_lh(death_distribution_data, study_year) # either you're left-
handed or right-handed, so P_right = 1 - P_left
    P_rh_A = 1 - P_lh_given_A(ages_of_death, study_year) # P_rh_A = 1 -
P_lh_A
    return P_rh_A*P_A/P_right
```

APPENDIX 8

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

# calculate the probability of being left- or right-handed for each
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")
ax.plot(ages, right_handed_probability, label = "Right-handed")
ax.legend() # add a legend
ax.set_xlabel("Age at death")
ax.set_ylabel(r"Probability of being age A at death")
plt.show()
```

APPENDIX 9

```
# 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*np.array(left_handed_probability))
average_rh_age = np.nansum(ages*np.array(right_handed_probability))

# print the average ages for each group
print("Av age of left-handed is : " + str(average_lh_age))
print("Av age of right-handed is : " + str(average_rh_age))

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

APPENDIX 10

```
# Calculate the probability of being left- or right-handed for all ages
left_handed_probability_2018 = P_A_given_lh(ages, death_distribution_data,
2018)
right_handed_probability_2018 = P_A_given_rh(ages,
death_distribution_data, 2018)

# calculate average ages for left-handed and right-handed groups
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 is " +
      str(round(average_lh_age_2018 - average_rh_age_2018, 1)) + " years.")
```