The Secerts of Mordern Detective Story

Gender distribution of victims, proportion of testimonial clues before and after 1905, average age that authors made their first publications.

Group YEL: Eric Li, Lisa Cheng, Yinuo Zhao

December 3, 2021

Introduction

Include here a few sentences to introduce the problem and provide context. You might want to briefly summarize the data in words (what is the data and what is it used for). You can present the questions you are investigating here.

Data Summary (optional)

Here you can explain how you cleaned the data and created variables suitable for answering your questions. You can also include graphical displays that either motivated or address the questions.

Research Question 1 - Statement of the Question

Research Question 1 - Relevant Visualization

Research Question 1 - Set-up of Statistical Model/Method

Research Question 1 - Results and Interpretation

Research Question 2 - Statement of the Question

Research Question 2 - Relevant Visualization

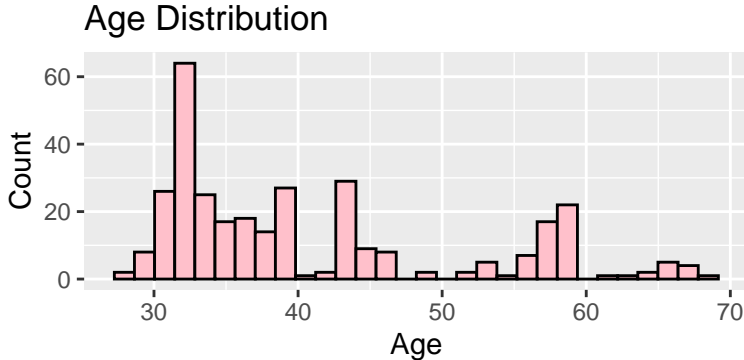# Research Question 2 - Set-up of Statistical Model/Method

Research Question 2 - Results and Interpretation

The research question is: **What is the range of plausible values for the average age of authors when the detective story is first published?**

This research will allow us to explore the age of which authors have the most inspiration that many of them wrote detective stories at that age.

Research Question 3 - Relevant Visualization

Age Distribution

The above figure is a histogram that shows the distribution of the authors' ages when the story is first published. We can see that all age ranges from around 27 to around 68. It is more common for authors to publish detective stories in their 30s to 40s than later in their life.

Research Question 3 - Set-up of Statistical Model/Method

Data Wrangling

- Selected to include the only two variables needed for this specific research question: age of the author and date of the publication of the story.
- Removed all observations that are missing any of these two variables.
- Created a new variable that represents the age of the author when the story is published. This is done by subtract these two variables.
- There exist several unreasonable values in this new variable we created(negative ages, and extremely small ages), we exclude these observations from the data since this is probably due to the mistakes during data collection.

Statistical Method

- The research question is completed using the bootstrap method. This method will give us the range of plausible values for the parameter we are trying to find, which is this case is the average age of authors when the detective story is first published.

Research Question 3 - Results and Interpretation

- From the result of bootstrapping, the 2.5% percentile is 39.59 and the 97.5% percentile is 41.91.
- This means that we are 95% confident that the average age of authors when the detective story is first published is between 39.59 to 41.91.
- If we repeated this procedure many times, 95% of those confidence intervals will include the true average age of authors when the detective story is first published for all detective stories from the early 1800s to the early 1900s.

Limitations

Overall Conclusions

Overall Conclusions

References and Acknowledgements (optional)