

Understanding Reader Satisfaction Through the Lens of Story Elements in Detective Novels

For Novels Written Between the Early 1800s - Early 1900s

Team Avengers: Aditya Khan, Yu Jin Shim, Tina Kim

December 3, 2021

Table of Contents

Section	Slide Number(s)
Introduction	3
Objectives and Questions	4
Question 1	5-8
Question 2	9-12
Question 3	13-16
Conclusions	17
Limitations	18
References	19

Introduction

- A key objective of a detective novel is to hook the audience into its narrative, and satisfy them with intrigue and mystery.
- One question that arises from observing detective novels is whether there are certain characteristics these books have that may influence reader satisfaction (our focus).
- We intend to garner insights from the “detective_data” dataset which contains information about a sample of 352 detective novels published between the early 1800s and early 1900s.
- We hope to use the data from this sample to make conclusions about early modern detective novels and reader satisfaction in a way that is accessible particularly for Professors Hammond and Stern.

Objectives and Questions

We consider three different perspectives, each leading to a research question. They may allow us to make conclusions about all detective novels published between the early 1800s - early 1900s (the **population**):

- 1 Publishing Perspective: find out whether there is a year range in which most satisfactory early modern detective novels were made.
 - **Question:** What are the range of plausible values that the publish year of a book can take for books that have at least a satisfaction level of 4?
- 2 Reader's Perspective: assess whether gender bias is present in significant events, and whether it may impact reader satisfaction.
 - **Question:** Is the proportion of male and female detectives the same in the stories of male and female authors?
- 3 Cultural Perspective: consider whether analysing reader satisfaction provides support for immense cultural impact of Conan Doyle's work (Armstrong, 2016).
 - **Question:** Is the mean reader satisfaction similar between novels written by Arthur Conan Doyle, and those which are not?

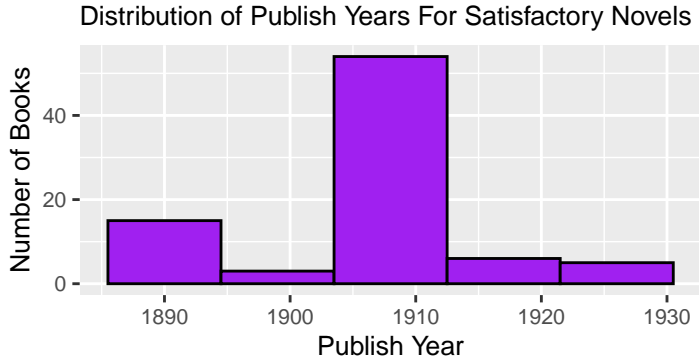
Q1: Introduction to the First Question

What are the range of plausible values that the publish year of a book can take for books that have at least a satisfaction level of 4?

- We define a satisfying novel to have a reader satisfaction of at least 4 out of 5.
 - If we consider 3 out of 5 to indicate indifference (as it is the middle of the scale), 4 appears to be the lower bound to consider a book satisfying.
 - The variable used to find reader satisfaction for each novel is “how_satisfying_is_this_story_as_a_piece_of_detective_fiction”.
 - The variable that indicates publish year of each book is “year_published”.
- Our goal is to infer a specific range of years between the 19th and 20th centuries when it was likely for a satisfying novel to have been published.

Q1: Visualising the Data

- This histogram shows the publish years of satisfactory books in this sample.
- The centre of the distribution is concentrated strongly around the years of 1905 to 1910, and there is only one peak in the distribution. This suggests that the majority of satisfactory books in this sample were published around those dates.



Q1: Statistical Analysis

- Using a bootstrap confidence interval test, we are able to use the data from the sample, to:
 - 1 Find a distribution of years in which the **true year** when the most satisfactory books were published lies (by a process called resampling - done 1000 times).
 - We use the **mean** publish year of satisfactory books to build this distribution.
 - 2 Narrow down in the new distribution a plausible range of years where the true year lies.
 - We exclude the 2.5% most extreme values on both ends of the distribution.
- As we exclude those extreme values, we are left with 95% confidence that the true year is in the remaining interval of years the distribution covers:

2.5%	97.5%
1904.481	1906.197

Q1: Results and Discussion

- Based on the test, we can claim with 95% confidence that the true year when the most satisfactory books were published between the early 1800s - early 1900s, between 1904.5 and 1906.2.
 - That is, approximately 1905.
- Furthermore, this may provide credence to a claim that 1905 can be considered a “golden year” for detective novels.
- However, this conclusion does not determine why 1905 was a successful year.
- Thus, we recommend further research into considering why 1905 may have yielded such a relatively high number of satisfying novels.

Q2: Introduction to the Second Question

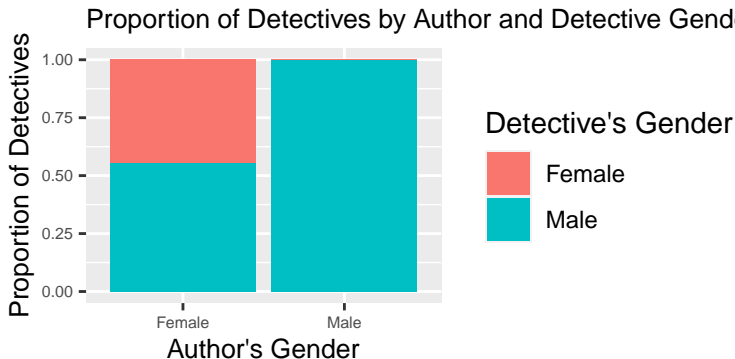
Is the proportion of male and female detectives the same in the stories of male and female authors?

- To consider this question, we mutate the data in the following manner:
 - New variables, `sum_male` and `sum_female` are created to sum the number of detectives by gender, excluding missing values.
 - `mean_prop` is the proportion of male detectives calculated using the mentioned variables.
- This lets us observe the mean proportion of male detectives by author gender:

Sex	sum_male	sum_female	mean_prop
Female Author	35	28	0.556
Male Author	289	0	1.000

Q2: Visualising the Data

- We further explore the data in the sample to observe the difference in the number of female and male detectives in the novels of male and female authors in a bar plot.
- All male authors used male detectives, whereas female authors showed about equal usage of male and female detectives, motivating this research question.



Q2: Statistical Analysis

- Referring to the table in slide 9, there is a 0.443 difference in the proportion of male detectives in the novels of male and female authors.
- We conduct a randomisation test to check how significant this difference is, with the following hypotheses:
 - H_0 : there is no difference in the proportion of male detectives in the stories of male and female authors.
 - H_1 : there is a difference in the proportion of male detectives in the stories of male and female authors.
- This test simulates (1000 times) a hypothetical where if each of the novels in the sample were randomly written by a male or female author, how much of a difference there would still be in the proportion of male detectives.
- The resulting p-value calculated for this test was 0, which suggests very strong evidence against H_0 , as it is less than 0.001.
- So, male and female authors do indeed use different proportions of male detectives.

Q2: Results and Discussion

- The findings from the visualisation and the statistical analysis suggest a bias towards male detectives in the novels of male authors, as compared to novels by females.
- There is conflicting interpretations on whether this gender bias can be said to have a sufficient effect on reader satisfaction.
 - Claim 1: Readers tend to like to relate to the characters in stories that match their gender role (Jose, 1989). A dearth of female detectives in male authored stories may alienate female readers.
 - Claim 2: the existence of bias may be driven by a general preference for male protagonists, irrespective of the reader's gender (Jose, 1989).
- Further research is recommended to investigate why female authors during this time used a more balanced gender representation, than male authors.

Q3: Introduction to the Third Question

Is the mean reader satisfaction similar between novels written by Arthur Conan Doyle, and those which are not?

- For this question, we create a new variable called “conan” which returns a value of True if the novel was written by Arthur Conan Doyle, and False otherwise.
- This allows us to split the novels in the sample into the aforementioned groups, letting us compare the mean satisfaction of novels:

Mean Satisfaction (Conan Doyle)	Mean Satisfaction (Not Conan Doyle)
3.23	2.76

- Conan Doyle averages a 0.47 higher satisfaction rating than the rest of the authors in this sample.

Q3: Visualising the Data

- Observing the boxplot below, the higher whiskers, and the larger box indicate that Conan Doyle's novels have higher ratings than others' works.
- This shows that Conan Doyle's satisfaction rating does tend to be higher than other authors' in this sample.



Q3: Statistical Analysis

- We will use a randomisation test to check whether the statistic of 0.47 indicates a significant difference between the two means.
- The randomisation test will use the following hypotheses:
 - H_0 : there is no difference in the mean reader satisfaction for novels written by Conan Doyle, and novels not written by Conan Doyle.
 - H_1 : there is a difference in the mean reader satisfaction for novels written by Conan Doyle, and novels not written by Conan Doyle.
- After running the randomisation test, the p-value calculated was 0.
- As this p-value is less than 0.001, there is strong evidence against H_0 , which means we can reject it.
- This suggests that there is a significant difference between the mean reader satisfaction for novels written by Conan Doyle and other authors.

Q3: Results and Discussion

- Based on the visualisation and the results from the randomisation test, it suggests that Conan Doyle does have a significantly higher mean satisfaction rating.
- This result suggests that there is backing to the common belief that Conan Doyle and his famous Sherlock Holmes novels do lead to higher reader satisfaction, on average.
- Based on these findings, we suggest further research could be done to understand what exactly makes Conan Doyle's novels much more enjoyable.
- Understanding what stylistic elements make Conan Doyle's novels entertaining may lead to a greater understanding of what makes a detective novel great.

Conclusion

- Based on the findings from each of the research questions, we received insights into each of the three perspectives.
- From the perspective of publishing, we claim with 95% confidence that 1905 was the year when the most satisfying early modern detective novels were published.
- From the perspective of the reader, the presence of gender bias favouring male detectives in the male authors' novels, taps into a debate about whether gender bias does impact reader satisfaction.
- From the cultural perspective, we found backing for the belief that Conan Doyle is one of the great detective novelists, leading to questions about what makes him successful.
- These findings help elucidate the relationship between reader satisfaction and detective novel characteristics, while still leaving open avenues for further research.

Limitations

- A limitation for question 1 is that the variable that we are using in the bootstrapping process (mean of publish year), represents year as a decimal value.
 - Technically, years are not supposed to be considered in non-whole values.
 - A decimal year value still has meaning (as decimals represent partway through a year).
- A limitation for the randomisation test in question 2 was that reader satisfaction was treated as a numerical variable, it is a categorical variable.
 - This may still give useful results, as mean reader satisfaction is a valid metric.
 - However, this test is usually supposed to take numerical variables instead.
- The boxplot visualisation in question 3 is also supposed to take a numerical variable, instead of the categorical reader satisfaction.
 - Reader satisfaction still works as the values it takes have numerical meaning.
- Our questions were from a pool of many questions. Not all were significant. We don't wish to provide an impression that this dataset only provides positive results.

References

- Armstrong, J. K. (2016, January 6). How Sherlock Holmes changed the world. Retrieved December 3, 2021, from <https://www.bbc.com/culture/article/20160106-how-sherlock-holmes-changed-the-world>
- Jose, P. (1989). The role of gender and gender role similarity in readers' identification with story characters [Abstract]. *Sex Roles*, 21, 697-713. doi:<https://doi.org/10.1007/BF00289178>