CS-GY 6053 Final Project - Report

Group Members

Zhongyu Zhang - zz46xx

Kevin Tan - st52xx

Runze Li - rl50xx

1 Introduction

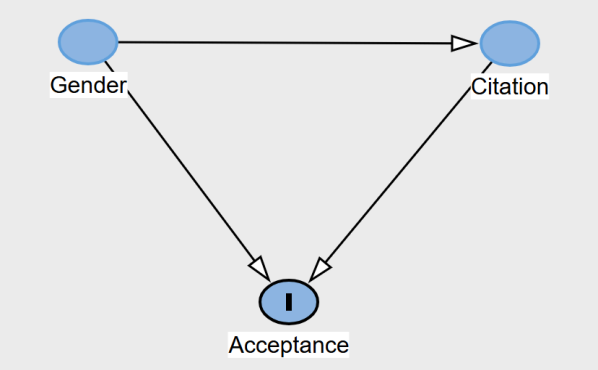
This report presents an in-depth analysis of the ICLR 2017-2022 dataset with a focus on understanding the influence of gender and citation counts on paper acceptance decisions. By conducting quantitative analysis, we aim to uncover potential biases in the acceptance process and provide insights into how gender and citation metrics may impact the likelihood of a paper being accepted for publication. Our study investigates whether these factors are contributing to systemic biases in the peer review process, which could perpetuate inequalities in the academic community. In addition to identifying these patterns, this research will serve to highlight the broader implications for academic publishing. By shedding light on any existing biases, we hope to contribute to the development of more equitable review processes, ultimately fostering an environment where all researchers, regardless of gender, are provided equal opportunities based on the quality and merit of their work. The findings from this analysis can not only inform the academic community but also guide the creation of policies aimed at promoting fairness and transparency in academic publishing.

2 Dataset Description

The ICLR 2017-2022 dataset provides open-access data on the peer-review process for papers submitted to the International Conference on Learning Representations (ICLR) over these years. This dataset includes detailed information on authors, institutions, submissions, reviews, and decisions, which serves as a valuable resource for examining patterns and potential biases in the review and acceptance process within a leading AI and machine learning conference.

3 Causal Model

A causal model is proposed to analyze the relationships between the variables in our study. The model assumes that the gender of authors, citations of authors, and acceptance of papers are interconnected, potentially influencing each other. We visualize these relationships using a Directed Acyclic Graph (DAG), which provides a clear representation of the causal dependencies. In this DAG, we assume that Gender influences both Citations and Acceptance, and Citations influence Acceptance directly.

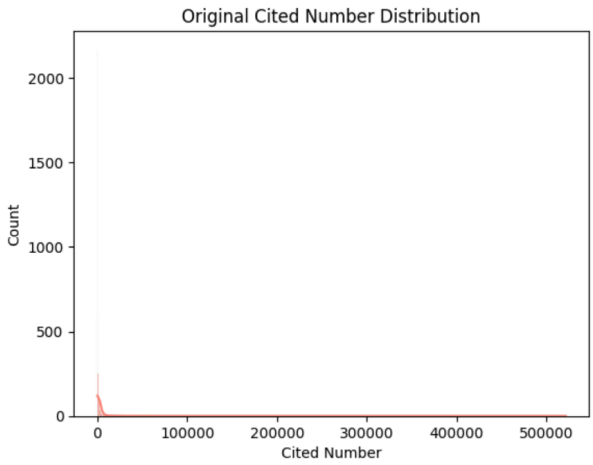


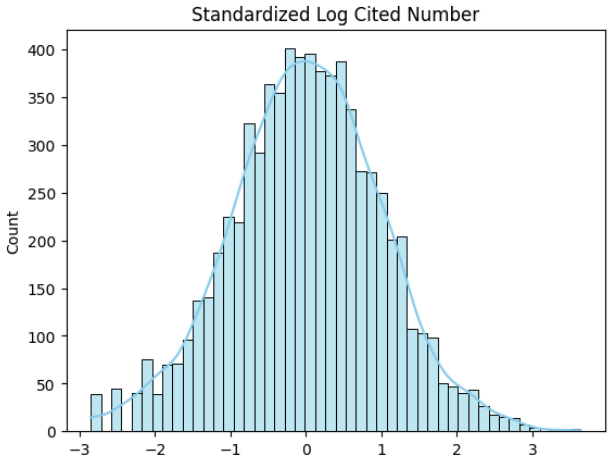
Each variable in the causal model is clearly labeled:

- The gender of authors is categorized into “Male,” “Female,” “Unspecified,” and “Non-Binary.” Here we take only the first author of each paper for a single paper since there are several authors for each paper, thus avoiding confusion caused by non-first authors.
- Citation numbers refer to the total number of citations of the authors. It is presumed that higher citation counts could lead to higher chances of paper acceptance, as papers authored by highly cited researchers may be perceived as more credible or impactful.
- Acceptance of papers is categorized into five states: “Oral,” “Spotlight,” “Poster,” “Workshop,” and “Reject.” These categories represent the level of acceptance, ranging from the highest (Oral) to the lowest (Reject). It is assumed that both gender and citations play a role in determining the level of paper acceptance.

4 Data Processing

All the data was saved in a .db file, and we extracted information that we needed using SQL query. The original distribution of authors' cited numbers has a really large scale and some of the authors have extremely high cited numbers. So we treat it logarithmically and then standardize it.





Besides, we encode the five different states of acceptance result to 0~4, from low acceptance to high.

5 Statistical Model

5.1 Direct Effect of Gender on Acceptance

$$Acceptance \sim Bernoulli(p_k)$$

$$logit(p_k) = b_{G[k]}$$

$$b_{G[k]} \sim Normal(0, 1)$$

Here p_k is the probability that a paper will be accepted, given the gender condition. With the logit transformation model, the effect of gender on acceptance can be captured by the coefficients corresponding to different genders. If the effect of gender on acceptance is significant, the coefficient $b_{G[k]}$ will not be zero, indicating that gender has a statistically significant effect on the probability of acceptance. To simplify this model, we divided the acceptance level into binary state-rejected (contains Reject and Workshop) and accepted (all other situations).

5.2 Total Effect

In addition to the direct effect, we also considered the total effect of gender on acceptance. Here, we use ordered Logit models to model acceptance.

$$Acceptance \sim OrderedLogit(\varphi_i, \alpha)$$

$$\varphi_i = \beta_{C,G[i]} \times C$$

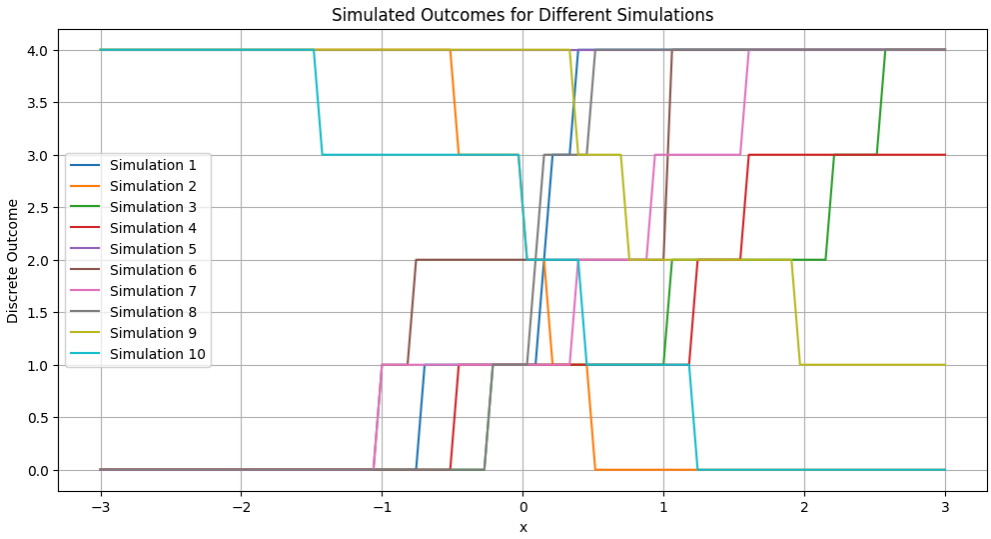
$$\beta_{C,G[i]} \sim Normal(0,\ 1)$$

$$\alpha_k \sim Normal(0, 1)$$

Here, we used a variable $\beta_{C,G[i]}$ to capture the interaction effect between gender and citations, which denotes the prior distribution of the coefficient. By constructing such a statistical model, we are able to reveal the role that gender and author-cited numbers play in dissertation acceptance decisions.

5.3 Prior Simulation

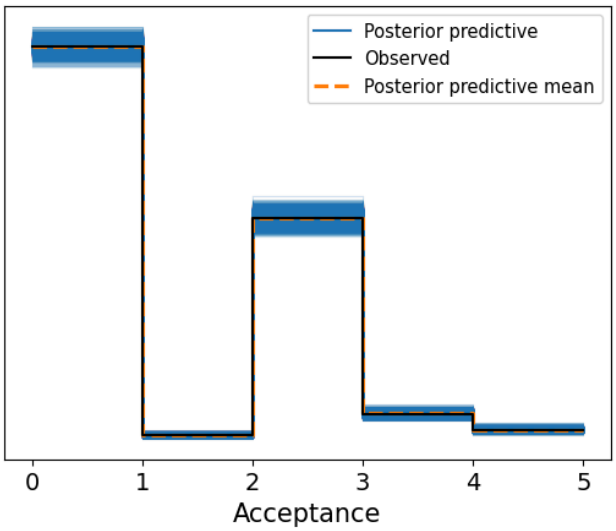
We provided prior predictive simulation for the total effect model and proved that for different treatment values (cited number), the outcome (acceptance) was effectively distributed in categories from 0 to 4.



The figure shows the results of 10 simulations. The results are slightly different for each simulation, reflecting the fact that due to the randomization α and *b*, the acceptance decision level of the paper jumps differently with the x-value.

6 Posterior Predictive Checks

Posterior predictive checks (PPC) are an important step in model evaluation that determines the fit of a model by comparing the observed data with the predicted data generated by the model. By using the az.plot_ppc() function, we obtained the following figure:



In the figure, the blue shaded area indicates the posterior predictive distribution, which is the distribution range of the predicted values generated from the posterior distribution. By plotting multiple model predictions sampled from the posterior distribution, the shaded area shows the width of the distribution of these predictions. The orange dashed line indicates the mean of the posterior predictive distribution. It reflects that the average model prediction of acceptance fits the original data distribution of paper submission results.

We can see that the model's posterior predictive distribution covers the actual data reasonably well, and the means are closer to the observed data, indicating that the model's predictions are plausible at these data points.

7 Discussion

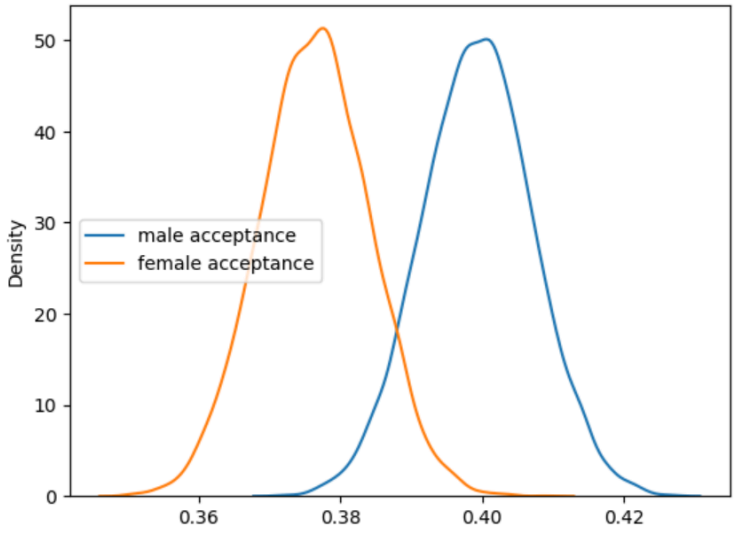
Since the number of people with non-binary gender is too few in the dataset and may not represent the real fact. We currently only focus on the identified male and female authors.

For the result of the direct effect of gender on acceptance, the coefficient b is shown as follows.

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
b[0]	-0.409	0.029	-0.461	-0.353	0.000	0.000	6372.0	3334.0	1.0
b[1]	-0.505	0.087	-0.668	-0.342	0.001	0.001	6304.0	2767.0	1.0

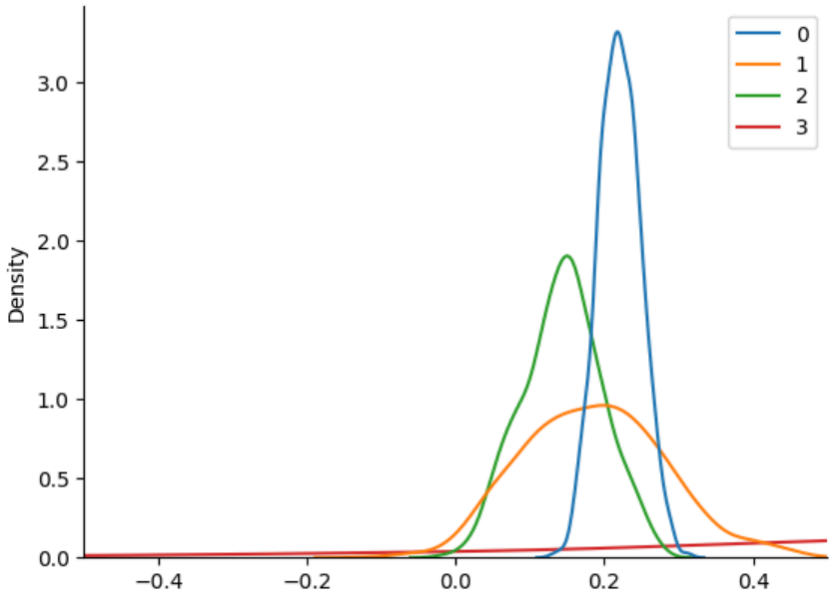
The mean of b0 and b1 indicates that the acceptance rate of both male and female authors is less than 50%. However, male authors have a higher acceptance rate.

We also used counterfactual simulation to reveal differences in acceptance rates of articles by authors of different genders. The figure below shows that overall male authors have a higher probability of being accepted. But the difference is rather small. The difference in the average acceptance rate is only about 3 percent. Although the number of confirmed male authors is much larger than females, which causes most of the accepted papers' 1st authors to be males, the peer review system is generally fair to males and females.



For the total effect of gender and citation number on acceptance rate. The result is shown below:

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
bC[0]	0.22	0.03	0.16	0.27	0.00	0.00	3657.97	2527.89	1.0
bC[1]	0.19	0.09	0.01	0.37	0.00	0.00	3521.64	2614.51	1.0
bC[2]	0.14	0.05	0.04	0.23	0.00	0.00	3702.03	2947.19	1.0
bC[3]	1.06	0.72	-0.32	2.35	0.01	0.01	3997.38	2626.92	1.0
alpha[0]	0.36	0.02	0.32	0.41	0.00	0.00	1677.86	2185.28	1.0
alpha[1]	0.42	0.02	0.37	0.46	0.00	0.00	1676.46	2250.66	1.0
alpha[2]	2.70	0.05	2.60	2.78	0.00	0.00	2417.36	2528.19	1.0
alpha[3]	3.91	0.09	3.74	4.07	0.00	0.00	2819.99	2718.25	1.0



The total effect model indicated that the number of citations has a positive influence on paper acceptance for both genders. The coefficients for bC are positive for all categories, indicating that higher citation counts increase the probability of paper acceptance. However, the impact is relatively modest. The effect of citations on paper acceptance is slightly more pronounced for male authors, as indicated by the larger coefficient for male authors (*bC[0]*). The lower standard deviation for male authors suggests that this relationship is more stable and confident compared to female authors. This may be due to more samples of males.

We also noticed that the difference between *alpha[0]* and *alpha[1]* is much smaller, which indicates that very few papers actually get the result of "Workshop," and most of them get the result of "Reject" and "Poster."

8 Future Work

In this study, we mainly analyzed the impact of gender on thesis acceptance decisions. However, the decision to accept a thesis may be influenced by a combination of factors, and a number of other academically relevant variables besides gender may also have a significant impact on the acceptance rate of a thesis, such as nationality, reputation of the research institution, and so on. Future research could further extend the model to consider the following potential factors with the aim of making a more comprehensive prediction of acceptance.