Ordinal Logit Models of Perceived Academic Freedom Among U.S. Academic Librarians

Tom Bareket

University of Minnesota

STAT4893W

Section 001

April 4, 2024

Ordinal Logit Models of Perceived Academic Freedom Among U.S. Academic Librarians

Introduction

There is a substantial body of evidence documenting the differential treatment of individuals with privileged and underprivileged characteristics. One common example is the persistent income gap between members of different races, and between men and women (e.g. Kochhar, 2023). Research suggests that many of these gaps exist in the workplace - for example, one study found differential rates of workplace discrimination among different ages, races and genders (Fekedulegn et al., 2019).

The existing literature on workplace discrimination has primarily focused on quantitative outcomes such as wages and productivity (e.g. Heiserman and Simpson, 2023). The objective of this research is to extend our understanding of the mechanisms of workplace discrimination. Specifically, we aim to understand if and how the gaps between privileged and underprivileged individuals extend to freedom of expression. In particular, we will examine academic librarians' perceptions of their own freedom, as reported through a qualitative survey. Academic freedom is an essential tenet of academia; thus, we are interested in identifying the circumstances under which academics' ideas are suppressed.

Academic librarians are a suitable population for this study because their profession requires various forms of expression, including research and publishing, collection development,and programming decisions such as extending invitations to speakers. By examining the statistical relationships between academic librarians' immutable and unchangeable characteristics and their perceived freedom of expression, we hope to gain insight on the groups whose voices are disproportionately silenced within this profession.

Methods and Materials

Dataset

In order to conduct our analysis of academic librarians' perceived academic freedom, we will use a dataset (Leebaw, 2021) collected by a pair of researchers at the University of Minnesota. This dataset was adapted from a survey which was distributed to academic librarians in the United States through listservs and professional organizations(Leebaw, 2019). This survey asks librarians about their personal characteristics and their workplace experiences. We will

utilize a subset of the variables contained in the original dataset.

Our independent variables are Gender, Sexual Orientation, Race and Age. All of these are expressed as categorical variables; Age is an ordinal variable with levels "18-34", "35-54" and "55+". Our dependent variable is the response to the question "I feel protected to express myself when questioning workplace policies and procedures", with levels "Never", "Sometimes", "Usually" and "Always". The independent variables were selected because they represent librarians' immutable and unchangeable characteristics. The dependent variable was selected because it represents one facet of librarians' perceived academic freedom, and because it had a relatively high response rate when compared to similar variables in the survey.

Data Segmentation

Some values of the identity questions have very few observations in this dataset. There are two reasons for this. The first is that some identities are rare; for example, there are very few Americans who identify as Pacific Islander. The second is that, because participants were permitted to select multiple identities for a single question, there are combinations of identities which are rare. For example, there are very few Americans who identify as both non-binary and a woman. When very few survey respondents identify with a particular identity label, it is impossible to make any meaningful inference with respect to that identity group; any inferences we make would say more about the small number of survey respondents who identify as members of the group than the group itself.

In addition, it does not make sense to count librarians who identify as members of multiple groups as an independent member of all of these groups; for example, one can imagine that the experiences of a librarian who identifies as genderqueer and as a man are very different from those of a librarian who only identifies as a man. It is also important to note that discriminatory agents[1] with whom the survey respondents interact may not care about, or even understand the nuances of the respondents' identities.

Since we cannot make inference with respect to rare identities, and since discriminatory agents may not discern between different rare identities, we assume that any academic freedom discrimination is homogeneous for identities with <10 observations. We combine such identities

---

[1] "agent" can refer to a coworker, student, institution, or any other person or entity with whom the librarian interacts.

into an "Other" category for each identity variable. This cutoff was chosen because most of our variables have a small number of observations with 100-400 observations and a large number of observations with 1-10 observations; a lower threshold such as 5 would be more arbitrary since there are several identities just above that threshold.

Gender	Frequency
Woman	402
Man	130
Prefer not to respond	10
Genderqueer or gender nonconforming	8
Non-binary	7
Man, Genderqueer or gender nonconforming	3
Other	3
Transgender, Non-binary, Genderqueer or gender nonconforming	3
Woman, Genderqueer or gender nonconforming	3
Man, Other	2

Figure 1
Distribution of Gender identities in the dataset. Values such as "Man, Other" indicate that multiple identities were selected.

After this segmentation, the levels of each variable are listed below:

Variable Name	Levels
Gender	Woman, Man, Other
Orientation	Heterosexual, Bisexual, Queer, Gay, Lesbian, Other
Race	White or Caucasian, Hispanic or Latino, Asian American or Asian, African American or Black, Other
Age	18-34, 35-54, 55+
Question_Workplace	"Never", "Sometimes", "Usually", "Always"

Figure 2
Levels of each variable after segmentation

Regression Models

Given the ordered categorical structure of our response variable, as well as our inference-focused research question, an appropriate model is Ordinal Logistic Regression (OLR), also known as the Logit Proportional-Odds Model (Kleinbaum, 2003, p. 464). Given a dataset $\{(Y_1, X_{1,1}...X_{1,p}),...(Y_n, X_{n,1},,,X_{n,p})\}$, the OLR model describes the odds that a response variable Y with a discrete ordered domain $\{1...K\}$ falls within a subset of its domain given a set of predictor variables $X_1...X_p$ using the equation:

$$logit(\gamma_{i,k}) = \beta_{0,k} - \sum_{j=1}^{p} \beta_j * X_{i,j}$$

Where $logit(X) = \frac{X}{1-X}$ and $\gamma_{i,k} = P(Y_i \leq k)$ (Algeri, 2024b, p. 23).

This equation assumes that the effects of our predictors $X_1...X_j$ are the same across all levels of Y; in other words, $\beta_{1,j} = ... = \beta_{k,j} = \beta_j$. We also assume that our samples are independent and that our sampling is unbiased; finally, we assume no perfect multicollinearity among the predictor variables.

In this particular application, in which the predictor variables are categorical, it is useful to interpret the Odds Ratio (OR) between the reference level of a variable and a different level, ℓ. Because of the first assumption we mentioned, this odds ratio is the same for all levels of Y. In this report we discuss interpretations with respect to the OR at the level Y = 3. This OR represents the ratio of the probability that $Y \leq 3$ when $X = \ell$ to the probability that $Y \leq 3$ when $X =$ baseline. This OR is given by the equation:

$$OR_3(X_j = \ell) = \frac{odds(Y \leq 3|X_j = \ell)}{odds(Y \leq 3|X_j = baseline)} = e^{-\beta_{j,\ell}}$$

Where $\beta_{j,\ell}$ is the coefficient of X_j at the level ℓ.

This Odds Ratio is useful because it tells us whether the ℓ level or the baseline level is more likely to have $Y \leq 3$. In our case, $Y \leq 3$ represents the cases where a librarian "Never", "Sometimes" or "Usually" feels safe to express themselves while questioning workplace policies; in other words, the cases where their freedom of expression is inconsistently or never protected. Thus, if $OR_3(X_j = \ell) > 1 \rightarrow \beta_{j,\ell} < 0$, it would suggest that the ℓ level of X_j experiences less workplace academic freedom than the baseline level. We can also interpret $OR_3(X_j = \ell)$ as the factor by which the ℓ level is more likely to experience this form of workplace discrimination (Algeri, 2024a).

Variable Selection

We begin with our full model:

$$logit(\gamma_{i,k}) = \beta_{0,k} - \beta_1 * Gender_i Woman - \beta_2 * Gender_i Other - \beta_3 * Orientation_i Bisexual - \beta_4 * Orientation_i Queer - \beta_1 * Orientation_i Gay - \beta_5 * Orientation_i Lesbian - \beta_6 * Orientation_i Other - \beta_7 * Race_i Hispanic/Latino - \beta_8 * Race_i AsianAmerican/Asian - \beta_9 *$$

$Race_iAfricanAmerican/Black - \beta_{10} * Race_iOther - \beta_{11} * Age_i35 - 54 - \beta_{12} * Age_i55P^2$

Where terms such as $Gender_iWoman$ are dummy variables whose value is 1 if the variable (Gender in this case) is equal to the given level (Woman in this case) for the ith observation.

From this full model, we select variables using both Forward and Backward Stepwise Selection. In the following sections, we will describe the model selected by the Stepwise Selection algorithms and its implications.

<div align="center">

Results

</div>

The Forward and Backward Stepwise Selection algorithms both recommend the model which includes Gender, Race and Age and excludes Orientation (Appendix 2, figure 7) (computed using code provided by Faraway, 2016). Thus, we arrive at the reduced model:

$logit(\gamma_{i,k}) = \beta_{0,k} - 0.061 * Gender_iWoman + 1.091 * Gender_iOther - 0.431 *$
$Race_iHispanic/Latino + 0.284 * Race_iAsianAmerican/Asian + 0.546 *$
$Race_iAfricanAmerican/Black + 1.127 * Race_iOther - 0.213 * Age_i35 - 54 - 0.520 * Age_i55P$ [3],
where $(\beta_{0,1}, \beta_{0,2}, \beta_{0,3}) = (-2.468, -0.332, 1.676)$.

Since AIC was used as the optimization criterion, this means that the inclusion of Orientation does not lead to significant increases in log-likelihood. This is a reasonable result; Orientation is rarely physically perceptible to discriminatory agents, whereas Race is associated with visible characteristics such as skin pigmentation, Age is associated with visible characteristics such as wrinkled skin and Gender is associated with visible characteristics such as clothing.

We consider the baseline group to be individuals who are least likely to be be victims of discrimination - those who are young and identify as White/Caucasian men. With this specification in mind, we can compute the Odds Ratio of each non-baseline group. These are listed below:

---

[2] Age_i55P is a dummy variable whose value is 1 if the age of the ith observation is at least 55. This is used in place of the notation Age_i55+ to avoid the implication that the "+" indicates addition.

[3] Note that the βs in this reduced model are different from the βs in the full model.

Variable/baseline level	Level	Odds Ratio
Race/White or Caucasian	African American or Black	1.726
Race/White or Caucasian	Asian American or Asian	1.329
Race/White or Caucasian	Hispanic or Latino	0.650
Race/White or Caucasian	Other	3.086
Gender/Man	Woman	0.940
Gender/Man	Other	2.977
Age/18-34	35-54	0.808
Age/18-34	55+	0.596

Figure 3

The Odds Ratio of each variable in the reduced model. A value greater than 1 indicates that the group is less likely to report that they feel consistently protected to express their opinion while questioning workplace policies than the baseline group.

the following variable levels have statistically significant deviations from the baseline group at the $\alpha = 0.05$ level: Races in the "Other" category, Genders in the "Other" category, and Ages greater than or equal to 55. The p values of all the variables in this model are listed in Appendix A, Figure 6 (computed using code provided by University of California - Los Angeles, 2017).

<div align="center">

Discussion

</div>

Interpretation of Coefficients & Significance

The "Other" level of Race and Gender have Odds Ratio values of approximately 3 (Figure 4). This means that librarians with rare Gender identities are roughly 3 times as likely as librarians who identify as men to report that they feel inconsistently protected to express themselves while questioning workplace policies, controlling for Race and Age. Similarly, librarians with rare racial identities are roughly 3 times as likely as librarians who identify as White or Caucasian to report that they feel inconsistently protected to express themselves while questioning workplace policies, controlling for Gender and Age. This is consistent with our initial hypotheses.

On the other hand, librarians aged 18-34 are roughly 1.68[4] times as likely as librarians aged 55+ to report that they feel inconsistently protected to express themselves while questioning workplace policies, controlling for Gender and Race. This contradicts our initial hypothesis; in fact, although the coefficient for librarians aged 35-54 is not statistically significant, the Odds

---

[4] the reciprocal of 0.596.

Ratio of these librarians is less than 1 and greater than that of librarians aged 55+, indicating that there is a negative relationship between age and perceived academic freedom discrimination which spans the entire range of ages considered in this study.

There are three possible explanations for the contradiction of our hypothesis about Age. The first is that older librarians are truly less likely to experience academic freedom discrimination; while older workers face discrimination in other fields, this may be a field which privileges them. The second is that perceptions of discrimination are different across age cohorts; younger librarians may be more likely to report that they feel unsafe to express their opinions. The third is that older librarians tend to hold higher-level positions within an organization; this can give them more power and therefore more protections to express controversial ideas. Of course, it is also possible that a combination of these possibilities is at play.

The non-significance of the other variable levels also contradicts our initial hypotheses. As in the case of the Age coefficient whose Odds Ratio contradicted our expectations, this may be because of the ways that academic librarians' field is different from other fields; for example, it is possible that librarians who identify as women do not experience high levels of discrimination because they are a majority group in this industry (e.g. 70% of librarians in our sample identify as Women). However, this would not explain the non-significance of several levels of the Race variable. As in the case of Age, it is possible that perceptions of discrimination are different across Genders and Races; it is possible, for example, that librarians who identify as Black face higher rates of discrimination but are less likely to report these incidents in a survey.

Conclusions & Future Research

Our findings contradict many aspects of the hypothesis we initially introduced. We did not find significant differences between men and women, nor between librarians who identify as White and those who identify as African American/Black, Asian American/Asian or Hispanic/Latino. This may be caused by a limitation of our methodology: we do not know whether the survey respondents' perception of their protection to self-express varies across Racial and Gender identities. Future research might repeat this methodology with a survey which asks participants if they have experienced specific incidents, which would reduce the role of participants' subjective perceptions.

One result which stands in stark contrast to our initial hypotheses is that older librarians

report feeling *more* protected while questioning workplace policies. Future research may further investigate the effects of age on the academic protections of librarians; as mentioned, it is possible that this industry is different from others; however, it is also possible that this result is the consequence of seniority or generational attitudes.

Our results do support some components of our initial hypothesis. We found that librarians with rare Gender and Racial identities experience a tripled likelihood of feeling inconsistently protected when questioning workplace policies when compared to librarians who identify as Men. This suggests that, while there is a relatively high degree of equality between Men and Women in regards to freedom of expression, this equality is not extended to individuals with other gender identities. It is more difficult to interpret this finding with respect to Race. Thus, future research may investigate the treatment of librarians with rare and common Racial identities.

References

Algeri, S. (2024a). *Personal communication (office hours)*.

Algeri, S. (2024b). Handout 6: GLM. University of Minnesota - Department of Statistics.

Faraway, J. J. (2016). Extending the linear model with r: Generalized linear, mixed effects and nonparametric regression models. *Technometrics*, 119.

Fekedulegn, D., Alterman, T., Charles, L. E., Kershaw, K. N., Safford, M. M., Howard, V. J., & MacDonald, L. A. (2019). Prevalence of workplace discrimination and mistreatment in a national sample of older u.s. workers: The regards cohort study. *SSM - Population Health*, *8*, 100444. https://doi.org/https://doi.org/10.1016/j.ssmph.2019.100444

Heiserman, N., & Simpson, B. (2023). Discrimination reduces work effort of those who are disadvantaged and those who are advantaged by it. *Nature Human Behavior*, *7*, 1980–1989. https://doi.org/https://doi.org/10.1038/s41562-023-01703-9

Kleinbaum, M., David G. Klein. (2003). Logistic regression: A self-learning text. *Technometrics*, *45*(1), 466. https://doi.org/10.1198/tech.2003.s30

Kochhar, R. (2023). The Enduring Grip of the Gender Pay Gap. https://doi.org/https://www.pewresearch.org/social-trends/2023/03/01/the-enduring-grip-of-the-gender-pay-gap/

Leebaw, A., Danya Logsdon. (2019). The Cost of Speaking Out: Do Librarians Truly Experience Academic Freedom? https://doi.org/https://hdl.handle.net/11299/203282

Leebaw, A., Danya Logsdon. (2021). Academic Librarians and Academic Freedom Survey Data. https://doi.org/https://doi.org/10.13020/fk5t-6h30

University of California - Los Angeles. (2017, February). *Ordinal logistic regression | r data analysis examples*. https://stats.oarc.ucla.edu/r/dae/ordinal-logistic-regression/

Appendices

Appendix 1: Regression output

```
Call:
polr(formula = factor(Question_workplace) ~ relevel(factor(Race),
    ref = "White or Caucasian") + relevel(factor(Gender), ref = "Man") +
    Age, data = dfnew, Hess = TRUE)

Coefficients:
                                Value Std. Error t value
RaceAfrican American or Black -0.54600    0.6360 -0.8584
RaceAsian American or Asian   -0.28411    0.5010 -0.5670
RaceHispanic or Latino         0.43089    0.4551  0.9467
RaceOther                     -1.12676    0.3619 -3.1139
GenderOther                   -1.09077    0.3784 -2.8828
GenderWoman                    0.06175    0.1962  0.3148
Age35-54                       0.21317    0.1889  1.1286
Age55+                         0.51987    0.2540  2.0468

Intercepts:
    Value   Std. Error t value
1|2 -2.4683  0.2742    -9.0006
2|3 -0.3317  0.2334    -1.4210
3|4  1.6763  0.2470     6.7860

Residual Deviance: 1273.241
AIC: 1295.241
```

Figure 4

Summary of the reduced model specified in the Results section

```
                                Value Std. Error     t value      p value
RaceAfrican American or Black -0.54600262  0.6360386 -0.8584426 3.906481e-01
RaceAsian American or Asian   -0.28411453  0.5010404 -0.5670491 5.706808e-01
RaceHispanic or Latino         0.43088962  0.4551370  0.9467251 3.437789e-01
RaceOther                     -1.12676218  0.3618528 -3.1138685 1.846518e-03
GenderOther                   -1.09077095  0.3783727 -2.8827950 3.941639e-03
GenderWoman                    0.06174912  0.1961702  0.3147732 7.529339e-01
Age35-54                       0.21316597  0.1888792  1.1285834 2.590736e-01
Age55+                         0.51987216  0.2539900  2.0468213 4.067563e-02
1|2                           -2.46829680  0.2742368 -9.0006045 2.244781e-19
2|3                           -0.33171827  0.2334468 -1.4209589 1.553287e-01
3|4                            1.67631147  0.2470252  6.7859939 1.152900e-11
```

Figure 5

p values of the variables in the reduced model specified in the Results section(computed using code provided by University of California - Los Angeles, 2017

Appendix 2: Variable Selection

```
Start:  AIC=1299.78
factor(Question_workplace) ~ relevel(factor(Race), ref = "White or Caucasian") +
    relevel(factor(Gender), ref = "Man") + Age + Orientation

                                                    Df    AIC
- Orientation                                        5 1295.2
- Age                                                2 1298.8
<none>                                                 1299.8
- relevel(factor(Race), ref = "White or Caucasian")  4 1303.4
- relevel(factor(Gender), ref = "Man")               2 1303.6

Step:  AIC=1295.24
factor(Question_workplace) ~ relevel(factor(Race), ref = "White or Caucasian") +
    relevel(factor(Gender), ref = "Man") + Age

                                                    Df    AIC
<none>                                                 1295.2
- Age                                                2 1295.5
- relevel(factor(Race), ref = "White or Caucasian")  4 1298.9
- relevel(factor(Gender), ref = "Man")               2 1302.3
```

Figure 6

Stepwise variable selection with AIC(computed using code provided by Faraway, 2016