# INVESTIGATING GENDER BIAS AMONG GRANT APPLICANTS

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Michael Thomas Heim

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Statistics

March 2020

Fargo, North Dakota

# North Dakota State University
## Graduate School

**Title**

Investigating Gender Bias Among Grant Applicants

**By**

Michael Thomas Heim

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

**MASTER OF SCIENCE**

SUPERVISORY COMMITTEE:

Dr. Rhonda Magel
Chair

Dr. Ronald Degges

Dr. Chanchai Tangpong

Approved:

03/31/2020
Date

Dr. Rhonda Magel
Department Chair

## ABSTRACT

An ongoing debate in society is about the existence of a wage gap between genders, and society's alleged preference to hire a man over an equally qualified woman. This debate extends from the commercial employment world into the funding of research grants. Given data collected at North Dakota State University between 2012 and 2018, have women who have sought federal funding for their research experienced a gender bias? To investigate, a logistic regression model is fit to determine whether gender affects funding probability. Other characteristics such as the primary investigator's college, requested amount, and the research team's make up of tenured and Caucasian members is also investigated. It was found that there is not a gender bias towards faculty at NDSU. Naturally, there was a bias towards researchers from different colleges and towards proposals requesting less funding. Surprisingly, a bias towards higher-proportion Caucasian research projects was found.

## ACKNOWLEDGEMENTS

Firstly, I would like to thank Dr. Rhonda Magel for being my advisor and mentor since the start of my undergraduate career, and for suggesting this research topic. Also, thanks to my other committee members, Dr. Ronald Degges and Dr. Chanchai Tangpong for serving on my committee and helping me through this big step in my graduate and professional career.

Most importantly, thanks to my parents, Dan and Sue Heim, for their continued financial and emotional support, and for standing beside me through the ups and downs of my educational career. You both taught me that I can do anything I set out to do, and you were right.

# TABLE OF CONTENTS

# LIST OF TABLES

# CHAPTER 1. INTRODUCTION

North Dakota State University and its students and faculty have always taken pride in their research. This is one aspect that draws students and faculty, as there is no lack of opportunities to conduct research. Faculty from almost every college, such as Education, Science and Math, Agriculture, and Health Professions, just to name a few, conduct multiple studies a year, many being interdisciplinary. Since NDSU operates on a limited research budget, research investigators must often seek external funding from federal or private grants. Since 2013, NDSU has tracked when its faculty members have sought funding outside of the university. With each occurrence, the college and department, budget, tenured status, gender, ethnicity, citizenship status, time as a faculty member at NDSU of all the principal and co-researchers with each proposal were recorded, and lastly if the request was funded.

Gender bias in all aspects of society has become an increasingly sensitive topic throughout recent years. The gender wage gap is an ongoing debate, along with society's alleged tendency to hire a man over an equally qualified woman for the same position. The funding of research through grants fits right into this subject. The primary research question to be investigated is: does the gender of grant applicants affect the proposal being funded? Since each proposal can consist of multiple researchers, does having a woman as the lead investigator on a proposal alter the chances of receiving funding? Can the same be said for a research team of entirely or primarily women?

Since the data collected by NDSU over the years includes other variables, these will also be used to see if any other biases exist. Some potential biases that could arise include budgetary bias, ethnicity bias, or tenure bias.

# CHAPTER 2. LITERATURE REVIEW

## 2.1. Other Studies

Other agencies have set out to investigate if there is a gender bias in federal grant programs. In 2002, an amendment to the National Science Foundation Authorization Act ordered a study to investigate gender bias in the distribution of Federal research and development funding (Hosek et al, 2005, p. xi). The study investigated funding of research within the National Institute of Health (NIH), the U.S. Department of Agriculture (USDA), and the National Science Foundation (NSF) itself from data collected between 2001-2003.

### *2.1.1. NSF*

The portion of this study conducted within the NSF shows that, at first glance, there are gender disparities across every category compared in Table 1. Women were statistically significantly less likely to request and receive funding and were statistically significantly more likely to request a smaller amount and receive a smaller amount than their male counterparts. Also, there was a substantial difference in the number of male and female applicants to begin with (Hosek et al, 2005, p. 21).

*Table 1.* Funding Requested and Awarded in NSF by Gender, 2001-2003 (Hosek et al, 2005, p. 22)

| | Women | Men |
|---|---|---|
| Number of Applicants | 24,860 | 47,339 |
| Percentage | (21.4%) | (78.6%) |
| Average Funding Requested | $483,003 | $494,228 |
| Average Funding Awarded | $80,508 | $84,970 |
| Average Award Size (among those who were awarded funding) | $227,720 | $232,462 |
| Percentage Receiving an Award | 35.4% | 36.6% |

One possible explanation for these disparities is the competitiveness for grant funding. Since it is highly competitive, the proposal submitted as well as the credentials of the researcher affect the outcome. Therefore, a researcher associated with a well-known research university that has more experience will likely have a better chance of receiving funding. Personal characteristics, such as ones' ability, motivation, or research interests may also influence the outcome, but are difficult to measure. After some predictive modeling on this data from the NSF, the researchers of this study discovered that when controlling for research discipline, academic degree, experience, and type of institution, all gender differences practically disappeared. The values in Table 2. below show that the predicted mean funding requested, predicted mean funding awarded, predicted probability of receiving funding, and predicted funding given the funding request was granted, is within the margin of error between males and females.

*Table 2.* Predicted Values in the NSF by Gender Controlling for Other Variables, 2001-2003 (Hosek et al, 2005, p. 24-26)

| | Women | Men |
|---|---|---|
| Probability of Receiving Funding | 36.5% | 36.3% |
| Average Funding Requested | $500,923 | $491,327 |
| Average Funding Awarded | $85,121 | $84,186 |
| Average Award Size (among those who were awarded funding) | $228,008 | $226,834 |

These values show that a woman and man of the same research discipline from the same institution with the same academic degree and experience are nearly equal in each of these categories, and well within the margin of error. Thus, the variables that were controlled for are likely the reason for the disparities between gender, and not a bias.

### *2.1.2. NIH*

This same study looked for gender differences in the NIH. Nearly 30% of primary investigators (PIs) applying for grants were women. This was a higher rate than in the NSF because women are more equally represented in medical and biological sciences than in physical sciences and engineering. As shown in Table 3., the study found that women were 11% less likely to receive funding, and the funding amount, when granted, was 30% smaller than their male counterparts. Initially, the gender disparities appear to be much larger in the NIH than in the NSF (Hosek et al, 2005, p. 28).

*Table 3.* Funding Requested and Awarded in NIH by Gender, 2001-2003 (Hosek et al, 2005, p. 28)

| | Women | Men |
|---|---|---|
| Number of Applicants | 18,571 | 47,339 |
| Percentage | (28.2%) | (71.8%) |
| Average Funding Requested | NA | NA |
| Average Funding Awarded | $367,842 | $582,091 |
| Average Award Size (among those who were awarded funding) | $1,281,679 | $1,807,736 |
| Percentage Receiving an Award | 26.3% | 29.2% |

After constructing some predictive models that control for the same potentially confounding variables discussed in the NSF, the difference in average funding between gender dropped, but remained statistically significant. Upon further investigation, the researchers found that there were a few explanations for this difference; one being the data was skewed. There was a large cluster of grants awarded that exceeded $7.39 million, and women were underrepresented in this grouping, receiving only 13% of the grants. Among this group, the women's average award size was only 17% smaller, compared to 30% smaller in grant awards less than $7.39 million (Hosek et al, 2005, p. 30). Also, the study suggested that women acquire less experience with age because they are less likely to be working full time. Long et al. discovered in 1973, approximately 23% of female doctorates in science and engineering were employed less than full time, versus only 3% of male doctorates. When reevaluated in 1995, this difference had decreased by half, but still remained (National Research Council, 2001). Secondly, the study hypothesized that women being less likely than men to be in the medical field contributes to the gender differences. Thus, grant applications would be sent through a different organization or agency. Lastly, a lack of data may have contributed to these results. Two major variables, the PIs' institution's research ranking, and average funding requested, were not provided (Hosek et al, 2005, p. 32-33). Thus, between the underrepresentation of women in the largest awards, women being less likely to work full time, and the same characteristics that were controlled for in the NSF research (research discipline, academic degree, experience, type of institution), it is difficult to come to a solid conclusion that these gender disparities are significant, and that a bias exists.

### *2.1.3. USDA*

Lastly, this study investigated gender differences within the USDA. Between 2001-2003, 23% of grant applicants were women. As seen in Table 4., the average amount requested and received were similar between genders, as well as the proportion of those receiving an award.

*Table 4.* Funding Requested and Awarded in USDA by Gender, 2001-2003 (Hosek et al, 2005, p. 34)

| | Women | Men |
|---|---|---|
| Number of Applicants | 2,452 | 8,104 |
| Percentage | (23.2%) | (76.8%) |
| Average Funding Requested | $176,260 | $175,285 |
| Average Funding Awarded | $28,896 | $28,222 |
| Average Award Size (among those who were awarded funding) | $110,542 | $109,260 |
| Percentage Receiving an Award | 26.1% | 25.8% |

The largest difference was in the proportion of women submitting grant applications. This is likely due to female applicants being more likely to submit proposals within other fields, such as biomedical or social sciences, and less in traditional agriculture. Using predictive modeling within the USDA data, the researchers' prediction of mean funding requested and awarded, as well as probability of receiving funding were within a margin of error as seen in Table 5.

*Table 5.* Predicted Values in USDA by Gender Controlling for Other Variables, 2001-2003 (Hosek et al, 2005, p. 35-36)

| | Women | Men |
|---|---|---|
| Probability of Receiving Funding | 25.8% | 25.9% |
| Average Funding Requested | $173,593 | $176,000 |
| Average Funding Awarded | $27,563 | $28,297 |
| Average Award Size (among those who were awarded funding) | $107,927 | $110,669 |

These values were calculated controlling for fiscal year, type of research institute, department, and USDA grant program. The differences between men and women are minute and not statistically significant.

### *2.1.4. Prior Study Results*

When examining this research conducted in three major federal agencies, the results seem clear. Not enough evidence has been shown to make a strong case that a gender bias exists in these cases and during these times. Within each of these federal organizations there initially appeared to be differences between men and women's amounts when requesting and receiving funding, but predictive modeling conducted in the study has shown that when controlling for other variables such as research discipline, academic degree, experience, and type of institution, the differences were within the margin of error. The portion of the study conducted in the NIH arguably shows a more noticeable difference between the funding amounts requested and dispersed among men and women. However, these findings are likely not valid due to skewed data, violating the assumption of normality, and women being less likely to work full time.

# CHAPTER 3. METHODS

## 3.1. The Data

The data obtained was collected by North Dakota State University between the years 2012-2018 and consisted of 6,249 observations post data cleansing. It recorded the following variables for each investigator on a proposal:

- Fiscal year
- Role (principal or co-investigator)
- Department
- College
- Investigator’s budget
- Proportion investigator’s budget was of total proposal
- Investigator’s tenure status (Yes/No)
- Gender (M/F)
- Citizenship status (citizen, permanent resident, international)
- Ethnicity (White, Asian, Hispanic, American Indian, Black, a combination of, or not specified)
- Years spent at NDSU
- Status of proposal (Funded, not funded, submitted).

## 3.2. Data Transformation, Cleaning, and Variables Used

The original dataset took a unique form. It was multivariate with multiple observations in a sense. One proposal often consisted of multiple investigators, and each investigator corresponded to a single observation. Therefore, to use this data in a statistical model, the data must be transformed, and some new variables defined. The variable that was most logical to use as a dependent variable was the status of a proposal. Proposals whose status was simply submitted must be dropped because they provide no information to answer the research question. Thus, proposal is now a binary variable, and a logistic regression model is the best method to use. The reasons for this will be discussed more in depth later. Other variables created to answer the primary research question include an indicator variable of whether the PI was male or female, and a variable to represent the proportion of female investigators on a proposal.

The secondary research question asks if other biases exist in this dataset. Another typical bias aside from gender is ethnicity. Using the data provided, another variable is created to represent the proportion of "Caucasian" investigators on each proposal. Since the data allowed investigators to declare mixed ethnicities, Caucasian was defined as those declaring strictly white. All other observations that were not strictly white were declared as non-Caucasian. Some investigators chose not to specify so those observations are removed from the data. A similar variable to represent the proportion of tenured members on a proposal is also created.

The other bias that could be investigated from this data is how federal agencies tended to fund proposals of NDSU faculty based on the college or department the PI was from, and the total cost of the proposal. Since there were hundreds of possible departments an investigator could be from, there would be no way to detect any significance of a variable using such information in the model. The initial data also consisted of twelve colleges/institutions at NDSU. It would also be unlikely to discover any statistical significance when using a set of indicator variables with this many levels. Thus, this variable was reorganized into a smaller number of groups based on similarity in discipline. The groupings are displayed below in Table 6.

*Table 6.* Groupings of Colleges for Indicator Variable

| College Grouping Indicator | Colleges Grouped |
|---|---|
| Grouping 1 | Ag Exp Station & Ext Services; Agriculture, Food Systems & Natural Resources |
| Grouping 2 | Arts, Humanities & Social Science |
| Grouping 3 | Business; UGPTI |
| Grouping 4 | Engineering |
| Grouping 5 | Information Technology; Miscellaneous Offices; Graduate School |
| Grouping 6 | Health Professions; Human Development & Education |
| Grouping 7 | Science and Math |

Lastly, to include the proposal's budget in the model it is transformed into a categorical variable. Originally the budget was a continuous variable ranging from zero dollars to twenty million. This large range would likely make the results of the variable difficult to interpret, so the variable is broken down into categorical variable with three levels as seen in the table below.

*Table 7.* Groupings of Proposal Total Budget

| Budget Grouping Indicator | Budget Amounts |
|---|---|
| Budget Grouping 1 | ≤$50,000 |
| Budget Grouping 2 | $50,001 to $500,000 |
| Budget Grouping 3 | ≥$500,001 |

### 3.3. The Model

Since the dependent variable chosen (whether the proposal is funded or not) is binary, a logistic regression model is the best option. In linear regression, model parameters are derived to be consistent and minimum-variance unbiased estimators. When the dependent variable is binary these properties no longer hold. Maximum likelihood estimation is used to estimate model parameters (Hosmer et al, 2000, p. 7-8). The logistic regression model takes the form

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$$

where $\pi(x)$ denotes the probability of an event occurring. This equation uses the link function to relate a probability to various explanatory variables (Agresti, 2007, p. 71). The parameter $\beta_0$ is the model intercept and can be used to estimate the odds or probability of success regardless of the values of the other variables. The other parameters $\beta_i$ corresponding to the variable $X_i$ represent how the probability of the event occurring changes when $X_i$ changes. The form of this model is beneficial because it allows the user to take the exponential of the parameter estimates to make inferences on their corresponding variables.

For this model, there are $k = 12$ variables and thirteen parameters estimated including the model intercept: six indicator variables for the college group, two indicator variables for the budget group, percentage of tenured, female, and Caucasian members, and an indicator variable for the principal investigator being female. These variables are as follows:

- $X_1$: Indicator for College Grouping 1 vs. 7
- $X_2$: Indicator for College Grouping 2 vs. 7
- $X_3$: Indicator for College Grouping 3 vs. 7
- $X_4$: Indicator for College Grouping 4 vs. 7
- $X_5$: Indicator for College Grouping 5 vs. 7
- $X_6$: Indicator for College Grouping 6 vs. 7
- $X_7$: Indicator for Budget Grouping 0 vs. 2
- $X_8$: Indicator for Budget Grouping 1 vs. 2
- $X_9$: Percentage of Tenured Members
- $X_{10}$: Percentage of Women Members
- $X_{11}$: Percentage of Caucasian Members
- $X_{12}$: Indicator for the Lead PI being Female

After an initial model is constructed and the parameters are examined for significance, often a selection method is used to determine the “best” model for predicting the outcome. In this case, the model was primarily fitted to see what variables are statistically significant predictors in determining a proposal's funding status, not necessarily attempt to predict the status. Thus, no selection method was used to determine the best model for predicting the status of a proposal.

# CHAPTER 4. RESULTS

## 4.1. Model Significance

Once the model parameters are estimated and the model is fit, a hypothesis test is run on the model's usefulness (Hosmer et al, 2000, p. 11). The null hypothesis $H_0: \beta_i = 0$ is tested against the alternative hypothesis $H_a: some\ \beta_i \neq 0$ for $i = 1, \ldots, 12$. If at least one parameter is not equal to 0 then the model is considered useful. The test statistic from the Likelihood Ratio Test is 1096.6783 (P-value<0.0001). Under the null hypothesis this test statistic follows a Chi-Square distribution with 12 degrees of freedom (Hosmer et al, 2000, p. 146). The null hypothesis will be rejected at $\alpha = 0.05$. At least one model parameter $\beta_i$ is different from 0, and the model is significant and useful.

## 4.2. Variable and Group Significance

Since the overall model is useful, the next step is to test each parameter within the model for significance (Hosmer et al, 2000, p. 92). Based on the results of this information displayed in Table 8., it is possible to conduct a variable selection method; however, these results will only be used to determine which variables and groups of indicator variables affect a research proposal receiving funding.

*Table 8.* Hypothesis Test Results for the Variables

| | P-value (compare at $\alpha = 0.05$) | Rejection Decision |
|---|---|---|
| College Grouping | <.0001 | Reject |
| Budget Grouping | <.0001 | Reject |
| Percent Tenured Investigators | 0.5706 | Fail to Reject |
| Percent Female Investigators | 0.9391 | Fail to Reject |
| Percent Caucasian Investigators | <.0001 | Reject |
| Women Principal Investigator | 0.5916 | Fail to Reject |

A Partial F-Test was conducted for the College Grouping and Budget Grouping variables, respectively. The P-values for these groups displayed in Table 8., compared at $\alpha = 0.05$, shows these groups of variables are significant predictors in the status of a proposal. The other variable significant in the model is the percentage of Caucasian investigators. The percentage of tenured investigators on a proposal was not a statistically significant predictor in the odds of the proposal receiving funding along with the percentage of female investigators. Also, whether the primary investigator was female did not impact the odds. Further interpretations and conclusions about these results will be discussed more in length later.

For the variables that are determined to be statistically significant predictors, it is easy to examine how they affect the dependent variable by taking the exponential of the model parameters. This results in the odds ratio, and an odds ratio confidence interval can also be calculated and examined. For a categorical variable such as the college and budget groupings, a baseline is selected, and each odds ratio within the variable is in comparison to this baseline group. Group seven is selected as the baseline for the college grouping and group three for the budget amount. The results of the odds ratios for the college groupings and budget groupings are displayed in Table 9. and Table 10., respectively.

*Table 9.* Confidence Intervals for Odds Ratios of College Group Comparisons

| | Odds Ratio Point Estimate | 95% Confidence Lower Bound | 95% Confidence Upper Bound |
|---|---|---|---|
| College Group 1 vs 7 | 2.880 | 2.412 | 3.440 |
| College Group 2 vs 7 | 2.377 | 1.729 | 3.270 |
| College Group 3 vs 7 | 7.786 | 5.182 | 11.698 |
| College Group 4 vs 7 | 0.911 | 0.717 | 1.158 |
| College Group 5 vs 7 | 3.260 | 2.412 | 4.405 |
| College Group 6 vs 7 | 2.025 | 1.600 | 2.563 |

*Table 10.* Confidence Intervals for Odds Ratio of Budget Group Comparisons

| | Odds Ratio Point Estimate | 95% Confidence Lower Bound | 95% Confidence Upper Bound |
|---|---|---|---|
| Budget Group 1 vs 3 | 5.180 | 4.039 | 6.644 |
| Budget Group 2 vs 3 | 2.250 | 1.756 | 2.882 |

The general interpretation for the odds ratios from these tables is that a proposal from college group one, for example, was 2.88 times more likely to receive funding that that from college group seven. Further interpretations and discussion of these values will be in Chapter 5.

For the other significant variables that are considered continuous, i.e. the percentage of Caucasian investigators, there is a single value for the odds ratio estimate. These values are displayed in Table 11.

*Table 11.* Confidence Interval for Odds Ratio of the Percentage of Caucasian Investigators

| | Odds Ratio Point Estimate | 95% Confidence Lower Bound | 95% Confidence Upper Bound |
|---|---|---|---|
| Percent Caucasian Investigators | 2.109 | 1.833 | 2.427 |

The general interpretation for this odds ratio value is that as the percentage of Caucasian investigators on a proposal increase one percent, the odds of receiving funding increases by 2.109 times. Further discussion of this will also be found in Chapter 5.

### 4.3. Limitations

Unlike the research studies mentioned previously in this paper, what agency or type of agency each research proposal was submitted to is unknown. Thus, it is not possible to further investigate if a certain local or federal organization that disperses funds is biased towards faculty at NDSU or in general.

The grouping of colleges at NDSU that is applied in the model restricts the interpretations. If it is found that one college grouping is statistically significantly more likely to receive funding than another college grouping, one cannot extend the inference to say that a single college within group one is statistically significantly more likely to receive funding than a single college within group two, for example.

# CHAPTER 5. DISCUSSION

Previously, the college group, budget amount, and proportion of Caucasian members were found to be significant predictors in a proposal's funding status. Now, how these variables affect the funding status will be examined by looking at odds ratio estimates and the corresponding confidence intervals. First, the groupings of the colleges at NDSU will be broken down and examined from above. Table 9. shows the odds ratio estimate and the 95% confidence upper and lower bounds for odds ratio of each college group's odds of being funded in comparison to group seven (proposals from the College of Science and Mathematics). Every comparison had statistically significant results except for group four. Since its confidence limit includes on odds of 1.00, it is possible that a proposal from group four is no more or less likely to receive funding than a proposal from group seven. The other groups not only produced significant results but were all more likely to receive funding than research conducted primarily from the College of Science and Math. Proposals from the third group have the highest odds in comparison to group seven, being about nine times more likely to receive funding. Since group seven contains UGPTI (Upper Great Plains Transportation Institute), a unique department at NDSU organized to specifically conduct research regarding all modes of transportation, it is logical that this group is significantly more likely than any other to receive funding into comparison to group seven.

Likely contributing to the significance in the college grouping variable is different disciplines requesting funding through different agencies. As demonstrated in the literature review section, proposals from the Department of Agriculture are submitted through the USDA, Department of Health Professions through the NIH, and so on. If possible, an interdisciplinary research project could hypothetically be arranged so that funding can be requested from a different organization than originally planned to increase the likelihood of receiving funding.

The other two significant variables in the model were the budget amount and the percentage of Caucasian members. Table 10. shows that a proposal with a budget less than $50,000 (belonging to group one), is 5.18 times more likely to receive funding than a proposal with a budget in excess of $500,001. A budget request from a proposal between $50,001 and $500,000 was 2.25 times more likely to receive funding than budget group 3. From this, it is evident that a larger budget is less likely to be funded. This result makes sense, as agencies must wisely disperse the funding to research that is deserving, yet not exorbitant.

The last significant predictor, the proportion of Caucasian investigators, has a 95% odds ratio confidence interval of all values above 1.00, as seen in Table 11. This indicates that a higher proportion of Caucasian members on a proposal is more likely to receive funding than a proposal comprised of a lower proportion Caucasian members. If this is truly the case, it suggests that there is some bias occurring, whether this significance comes primarily from a certain agency, or within a certain department at NDSU, or a bias towards Caucasian faculty in general at NDSU.

Lastly, the primary purpose of this statistical model was to determine if there was bias between genders in grant funding requested by faculty at NDSU. The variables that are set up to provide insight into this included the proportion of female investigators and an indicator variable of the lead investigator being female or male. Both of these variables are insignificant in the model. This provides evidence that there is no gender bias among NDSU grant applicants after accounting for other variables in the statistical model.

# CHAPTER 6. CONCLUSION

After model parameter estimation, it was found that the college of the principal investigator, the budget, and the proportion of Caucasian members on the research team influenced a proposal's likelihood of receiving grant funding. With this information, adjustments could be made to increase the odds of receiving funding. If an investigator wanted to conduct a certain research project, naturally they should attempt to set up the research so that it will be as inexpensive as possible. If the research will be quite costly, the investigator could first propose to conduct a small portion of the research, gather results, and resubmit another proposal to finish the research if the results are interesting. Since the college of the principal investigator was also a significant predictor, an interdisciplinary group could define the principal investigator to be a colleague within the research team from another college, or request funding from another agency in an attempt to increase the likelihood of receiving funding.

Although the proportion of Caucasian members on a research team influences funding probability, it would be unethical to use this information to select a purely or primarily white group of members to conduct a research project. However, a potential research topic going forward could be connecting which funding agency a proposal was submitted to. With this information, one could investigate if a certain agency or agencies largely contributed to this bias. This model could also be broken down into several models by college group or each distinct college. This racial bias may be explained by an unbalanced number of primarily white faculty within a certain college.

The main purpose of this model, to find if there was a gender bias among grant applicants comprised of NDSU faculty, showed that there was not a gender bias present. This was shown by two variables in the model, the proportion of female investigators and if the lead investigator was female, being insignificant predictors in the likelihood of the proposal receiving a grant. Also, the model does not show a bias towards tenured investigators exist. Overall, the results of this research align with the other studies examined, which were conducted in three large federal funding agencies, and show no evidence of a gender bias.

# REFERENCES

Agresti, A. (2007). An introduction to categorical data analysis. Hoboken, NJ: Wiley.

Hosmer, D.W., Lemeshow, S., & Sturdivant, R.X. (2000). Applied logistic regression. Hoboken, NJ: Wiley.

Hosek, Susan D., Amy G. Cox, Bonnie Ghosh-Dastidar, Aaron Kofner, Nishal Ray Ramphal, Jon Scott, and Sandra H. Berry, Gender Differences in Major Federal External Grant Programs. Santa Monica, CA: RAND Corporation, 2005. https://www.rand.org/pubs/technical_reports/TR307.html

National Research Council. 2001. From Scarcity to Visibility: Gender Differences in the Careers of Doctoral Scientists and Engineers. Washington, DC: The National Academies Press. https://doi.org/10.17226/5363.

# APPENDIX. SAS CODE

```
data GrantData (Drop=Dept Department College Tenure DeptID Citizenship
SeniorityDate Years TenuredDescrpt Ethnicity PI_Budget Total_Proposal);
    infile 'Final_Data_NonConfidential.txt' delimiter='09'x firstobs=3
missover dsd;
    input
    Proposal : $10.
    FY
    Role $
    Dept
    Department : $40.
    College : $50.
    PI_Budget
    Total_Proposal
    Tenure $
    DeptID
    Gender $
    Citizenship : $20.
    Ethnicity : $12.
    SeniorityDate : $10.
    Years
    TenuredDescrpt $
    Status : $10.;
      if FY<0 Then delete;

*Selectively delete proposals categorized as submitted and create binary
      variable;
      if Status='Submitted' then delete;
      if Status='Funded' then Status=1;
      else Status=0;

*Delete variables that had certain missing values;
      if Proposal = 'FAR0020657' then delete; if Proposal = 'FAR0020834' then
      delete; if Proposal = 'FAR0021108' then delete; if Proposal =
      'FAR0021145' then delete;
      if Proposal = 'FAR0021315' then delete; if Proposal = 'FAR0021507' then
      delete; if Proposal = 'FAR0021700' then delete; if Proposal =
      'FAR0021783' then delete;
      if Proposal = 'FAR0021821' then delete; if Proposal = 'FAR0021899' then
      delete; if Proposal = 'FAR0022709' then delete; if Proposal =
      'FAR0022913' then delete;
      if Proposal = 'FAR0023050' then delete; if Proposal = 'FAR0023051' then
      delete; if Proposal = 'FAR0023053' then delete; if Proposal =
      'FAR0023054' then delete;
      if Proposal = 'FAR0023110' then delete; if Proposal = 'FAR0023115' then
      delete; if Proposal = 'FAR0023170' then delete; if Proposal =
      'FAR0023219' then delete;
      if Proposal = 'FAR0023369' then delete; if Proposal = 'FAR0023542' then
      delete; if Proposal = 'FAR0023543' then delete; if Proposal =
      'FAR0023624' then delete;
      if Proposal = 'FAR0023665' then delete; if Proposal = 'FAR0023669' then
      delete; if Proposal = 'FAR0023753' then delete; if Proposal =
      'FAR0024289' then delete;
```

```
      if Proposal = 'FAR0024291' then delete; if Proposal = 'FAR0024391' then
      delete; if Proposal = 'FAR0024576' then delete; if Proposal =
      'FAR0024600' then delete;
      if Proposal = 'FAR0024617' then delete; if Proposal = 'FAR0024898' then
      delete; if Proposal = 'FAR0024958' then delete; if Proposal =
      'FAR0024959' then delete;
      if Proposal = 'FAR0025002' then delete; if Proposal = 'FAR0025087' then
      delete; if Proposal = 'FAR0025293' then delete; if Proposal =
      'FAR0025362' then delete;
      if Proposal = 'FAR0025700' then delete; if Proposal = 'FAR0025758' then
      delete; if Proposal = 'FAR0026006' then delete; if Proposal =
      'FAR0026057' then delete;
      if Proposal = 'FAR0026217' then delete; if Proposal = 'FAR0026454' then
      delete; if Proposal = 'FAR0026819' then delete; if Proposal =
      'FAR0026873' then delete;
      if Proposal = 'FAR0026919' then delete; if Proposal = 'FAR0026920' then
      delete; if Proposal = 'FAR0026924' then delete; if Proposal =
      'FAR0027443' then delete;
      if Proposal = 'FAR0027631' then delete; if Proposal = 'FAR0027739' then
      delete; if Proposal = 'FAR0027983' then delete; if Proposal =
      'FAR0028196' then delete;
      if Proposal = 'FAR0028214' then delete; if Proposal = 'FAR0028381' then
      delete; if Proposal = 'FAR0028403' then delete; if Proposal =
      'FAR0028538' then delete;
      if Proposal = 'FAR0028582' then delete; if Proposal = 'FAR0028771' then
      delete; if Proposal = 'FAR0028871' then delete; if Proposal =
      'FAR0028872' then delete;
      if Proposal = 'FAR0028926' then delete;

*Create tenure and ethnicity binary variables;
      if Tenure='yes' then TenureIND=1;
      else TenureIND=0;

      if Ethnicity='WHITE' then Caucasian=1;
      else Caucasian=0;

*Categorize budget amount;
      TotalBudget=PI_Budget/Total_Proposal;
      if TotalBudget=<50000 then BudgetGroup=0;
      else if (TotalBudget>50000 & TotalBudget=<500000) then BudgetGroup=1;
      else if TotalBudget>500000 then BudgetGroup=2;

*Group colleges;
      if College = "Ag Exp Station & Ext Service" or College = "Agriculture,
      Food Systems & Natural Resources" Then CollegeInd=1;
      Else if College = "Arts, Humanities & Social Science" Then
      CollegeInd=2;
      Else if College = "Business" or College = "UGPTI" Then CollegeInd=3;
      Else if College = "Engineering" Then CollegeInd=4;
      Else if College = "Information Technology" or College = "Miscellaneous
      Offices" or College = "Graduate School" Then CollegeInd=5;
      Else if College = "Health Professions" or College= "Human Development &
      Education" Then CollegeInd=6;
      Else if College = "Science & Mathematics" Then CollegeInd=7;
    run;

*Number of Members on each Proposal*;
```

```
proc freq data=GrantData;
      tables Proposal / out=NumberMembers (rename=(Count=MemberCount)
      drop=Percent) noprint;
      run;

proc sort data=NumberMembers;
      by Proposal;
      run;

*Number of Tenured Members on each Proposal*;
proc freq data=GrantData;
      tables Proposal / out=NumberTenured (rename=(Count=TenureCount)
      drop=Percent) noprint;
      where TenureIND=1;
      run;

proc sort data=NumberTenured;
      by Proposal;
      run;

*Number of Women Members on each Proposal*;
proc freq data=GrantData;
      tables Proposal / out=NumberWomen (rename=(Count=WomenCount)
      drop=Percent) noprint;
      where Gender='F';
      run;

proc sort data=NumberWomen;
      by Proposal;
      run;

*Number of 'Purely' Caucasian Members on each Proposal*;
proc freq data=GrantData;
      tables Proposal / out=NumberCaucasian (rename=(Count=CaucasianCount)
      drop=Percent) noprint;
      where Caucasian=1;
      run;

proc sort data=NumberCaucasian;
      by Proposal;
      run;

*If lead PI is women, create indicator variable*;
proc freq data=GrantData;
      tables Proposal / out=MainPIWomen (rename=(Count=WomenLeadPI)
      drop=Percent) noprint;
      where Role='PI' and Gender='F';
      run;

proc sort data=MainPIWomen;
      by Proposal;
      run;

proc sort data=GrantData;
      by Proposal Role;
      run;
```

```
*Merge and View all data*;
data GrantData (Drop= Gender Caucasian MemberCount TenureCount WomenCount
      CaucasianCount);
      merge GrantData NumberMembers NumberTenured NumberWomen NumberCaucasian
      MainPIWomen;
      by Proposal;
      if Role='CPI' then delete;
      If TenureCount='.' Then TenureCount=0;
      If WomenCount='.' Then WomenCount=0;
      If CaucasianCount='.' Then CaucasianCount=0;
      If WomenLeadPI='.' Then WomenLeadPI=0;
      TenurePercent = TenureCount / MemberCount;
      WomenPercent= WomenCount / MemberCount;
      CaucasianPercent = CaucasianCount / MemberCount;
      If CollegeInd='.' Then Delete;
      If Proposal=lag1(Proposal) Then Delete;
      run;

ods rtf file='F:\Thesis\Research\FinalOutput.rtf';

proc logistic data=GrantData;
      title 'Final Logistic Model Output with Categorical Budget';
      class CollegeInd WomenLeadPI BudgetGroup;
      model Status(event='1') = CollegeInd BudgetGroup TenurePercent
      WomenPercent CaucasianPercent WomenLeadPI / lackfit rsquare;
      run;

ods rtf close;
```