**Exploring Empathy toward Others and Moderators on Prosocial Behavior: An Analysis on the 2004 GSS Dataset**

Lisette Horne

Department of Psychology, University of Minnesota

PSY 8712: Data Science in Psychology

Dr. Richard Landers

May 04, 2024

**Exploring Empathy toward Others and Moderators on Prosocial Behavior: An Analysis on the 2004 GSS Dataset**

This project aims to examine prosocial behaviors across various demographics and the role of empathy toward others. The demographics include race, age, sex, political party identification, and daily strength of connection to one’s religion or spiritual affiliation. The use of these demographics is because there is a variety of research in prosocial behavior literature that emphasizes that these demographic variables can influence one’s empathy reported and the extent in which they engage in prosocial behavior.

**Hypotheses and Research Questions**

*Research Question 1*: Is there a correlation between one’s reported empathy towards others and there reported prosocial behavior? *Hypothesis 1:* There will be a moderately strong and positive correlation between empathy and prosocial behavior.

*Research Question 2*: Do demographic variables moderate people’s reported empathy toward others and enacted prosocial behavior? *Hypothesis 2:* All demographic variables will, individually and in conjunction, moderate people’s reported empathy toward others and the reported enacted prosocial behavior.

*Research Question 3:* Can empathy and demographic variables be used to predict enacted prosocial behavior utilizing machine learning models? *Hypothesis 3*: The predictive power of empathy and demographic variables on enacted prosocial behavior using machine learning models is statistically significant.

**Methods**

**Open Science Materials**

To access this project using the version of R, Rstudio, the packages, the data, and the codebook used, a web binder of the project was created and can be accessed using the following link: https://mybinder.org/v2/gh/lisette-18/psy8712-final/HEAD?urlpath=rstudio. This binder serves as an online code repository to contain the code and all content related to the project and was created to allow for easily reproducible projects. When accessing, it will automatically begin a session in Rstudio with all material used in this project. In addition, the materials can be accessed using the web serve GitHub and by clicking on the following link: https://github.com/lisette-18/psy8712-final. When you have clicked the link, it will bring you to the respective GitHub page contain all of the code and related content to the project. A README file will appear at the bottom of the page, as well, and will provide a description of the project.

**Participants**

This project used pre-collected and publicly available data from the 2004 General Social Survey (GSS) dataset to explain trends in behaviors, demographics, and opinions of United States Americans. After cleaning the dataset, the number of participants was 561.

**Measures**

Empathy was measured using an aggregated measure of 7 empathy items from the GSS dataset. The scores were averaged across the rows and greater values indicate greater reported empathy toward others.Prosocial behavior was measured using an aggregated measure of 11-items from the GSS dataset. The scores were averaged across the rows and lower scores indicate greater enacted prosocial behavior. Sex was determined by asking participants for their current sex (i.e., “Male” or “Female”). Race was determined by asking participants for their racial identification (i.e., “White,” “Black,” or “Other”). Age was measured by asking participants for their current age at the time of the survey using a response scale of 18 – 89 years of age. The strength of an individual’s daily connection to their religion or spirituality was determined by asking “How often do you find strength in your connection to your religion or spirituality?” using a 6-option scale (1 *= many times a day*, 6 = *never/almost*). Participants political party identification was measured by asking which party they best identified as and for the sake of the project, the party identifications was limited to three options (i.e., Strong Democrat, Independent, or Strong Republican).

**Procedure**

561 participants were surveyed in 2004 about their behaviors, demographics, and opinions. Based on the current GSS design, a nationally representative sample of participants used a cross-section survey to respond to demographics, attitudes, and behaviors.

**Analyses**

**Descriptive Statistics and Static Visualizations**

**Table 1**

*Descriptive data by Sex*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sex** | **N** | **Mean Empathy Scores** | **SD Empathy Scores** | **Mean Prosocial Behavior Scores** | **SD Prosocial Behavior Score** |
| Male | 274 | 3.28 | 0.46 | 4.64 | 0.66 |
| Female | 287 | 3.36 | 0.47 | 4.68 | 0.63 |

*Note.* Descriptive data of individual’s sex on empathy and prosocial behaviors,

**Figure 1**

A graph of a graph of empathy scores

Description automatically generated

*Note*. Displaying the histogram of empathy scores and its distribution among participants.

A detailed table of descriptive data for each demographic group and the regressions ran can be found online using the GitHub repository.

**Interactive Visualization**

An online shiny app can be accessed through the following link: <https://lisette-18.shinyapps.io/final_shiny/>. The app was created with all three research questions in mind to explore the relationship between empathy and prosocial behavior based different demographic groups. It will allow users to explore the correlation between the two variables from five options: sex, race, age, political party identification, and daily strength of connection to religion/spirituality. Based on the choices by the user, the app will provide a scatterplot showing the correlational relationship between empathy and prosocial behavior, along with a linear regression line.

**Data Cleaning**

Data cleaning for this project relied on a variety of ‘dplyr’ functions from the ‘tidyverse’ package. I only kept participants who had responses to all of the items. Participants who did not complete all variables were excluded from the analysis. I selected my variables of interest and converted all respective variables to factors in order for the shiny web page to function and to rename the levels and labels. Last, I created an aggregated and averaged measure of empathy scores and of prosocial behavior scores for participants.

**Analysis**

*Hypothesis 1.* To test the first hypothesis, I ran a correlation test on the average scores of empathy toward others and the average scores of prosocial behavior. Among the participants, the average empathy experienced for other and enacted prosocial behaviors were not correlated, *r*(559) = -.54, *p* = 0.59. See Figure 2 for scatterplot of correlation.

**Figure 2**

A graph of a number of black dots

Description automatically generated with medium confidence

*Hypothesis 2.* To test the second hypothesis, I ran a multiple linear regression model to predict prosocial behavior from empathy toward others, age, race, sex, religious/spirituality strength, and political party identification. This resulted in a significant model *F*(12, 548) = 5.71, *p* < .001, *R2*= 0.09. Based on further analysis, it was found that sex, race, political party identification, and age were not a significant predictor, but religious/spiritual strength was (*t* = 5.14, *p* < .001). See Table 2 for regression table.

**Table 2**

*Regression results using avg\_probehav as the criterion*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | *b* | | | *b*  95% CI  [LL, UL] | *sr2* | *sr2*  95% CI  [LL, UL] | | Fit | |
| (Intercept) | | | 4.04\*\* | [3.61, 4.47] |  | |  | |  |
| Female | | | 0.10 | [-0.01, 0.21] | .01 | | [-.01, .02] | |  |
| Race: Black | | | 0.17\* | [0.01, 0.34] | .01 | | [-.01, .02] | |  |
| Race: Other | | | -0.05 | [-0.28, 0.18] | .00 | | [-.00, .00] | |  |
| Age | | | 0.01\*\* | [0.00, 0.01] | .03 | | [.00, .05] | |  |
| PartyID:Independent | | | 0.10 | [-0.03, 0.24] | .00 | | [-.01, .01] | |  |
| PartyID: Strong Republican | | | 0.02 | [-0.12, 0.16] | .00 | | [-.00, .00] | |  |
| Rel\_strength: Every day | | | 0.17\* | [0.02, 0.31] | .01 | | [-.01, .02] | |  |
| Rel\_strength: Most days | | | 0.32\*\* | [0.15, 0.49] | .02 | | [-.00, .05] | |  |
| Rel\_strength: Some days | | | 0.35\*\* | [0.14, 0.56] | .02 | | [-.00, .04] | |  |
| Rel\_strength: Once in a while | | | 0.62\*\* | [0.43, 0.82] | .06 | | [.02, .10] | |  |
| Rel\_strength: Never/Almost Never | | | 0.53\*\* | [0.33, 0.72] | .05 | | [.01, .08] | |  |
| avg\_empathy | | | -0.02 | [-0.14, 0.09] | .00 | | [-.00, .00] | |  |
|  | |  | |  |  |  | | *R2*  = .111\*\* | |
|  | |  | |  |  |  | | 95% CI[.05,.14] | |
|  | |  | |  |  |  | |  | |

*Note.* \* indicates p < .05. \*\* indicates p < .01.

*Hypothesis 3.* To test the third hypothesis, I ran machine learning models to compare against the OLS models. Based on the respective R-squared, we can determine that the OLS regression model is sufficient in explaining variation within the data.

**Table 3**

*Machine Learning Results*

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **cv\_R2** | **ho\_R2** |
| OLS regression | 0.16 | 0.05 |
| Elastic Net | 0.15 | 0.05 |
| Random Forest | 0.81 | 0.00 |
| xgbLinear | 0.86 | 0.00 |

All other analyses and respective tables and figures can be found in the online GitHub repository.

**Reflection**

In my reflection during my time in Data Science, I have learned a lot, especially in the context of reproducible research. I have learned the importance of not copying and pasting code in order to run my data, but also to be mindful to look at the most appropriate packages and functions to use, to explore the R Studio Help page, to keep my R most up to date, as well as knowing why I am using certain functions over others. These concepts I have learned are ones I will bring back to my own research and have because it has allowed for clearer code, more understandable code, and code that can be more easily reproduced. In addition, I will be adding the new procedures of creating separate folders in my R root project in order to keep my figures and outputs more organized because it has helped tremendously with the class and I know will help my own research a lot. I will be changing my mindless copying and pasting from online or ChatGPT and I found learning better ways of keeping my research and code more organized as the most valuable tool from this class.