**Final Project: An Examination of Misinformation Questions**

Mackenzie Nickle

Department of Psychology, University of Minnesota

PSY 8712 Data Science

Dr. Richard Landers

May 04, 2024

**Project Overview**

The purpose of my project is to examine accuracy on misinformation questions across various demographics. The demographics were chosen based on demographics that are often attributed to acceptance of misinformation, such as party affiliation and vote choice, as well as other basic demographic information, such as gender and income level.

**Hypotheses and Research Questions**

RQ1: How does vote choice in the 2020 election impact accuracy scores on various misinformation questions?

H1: There will be significant group differences in accuracy score by vote choice.

RQ2: How does gender impact accuracy scores on various misinformation questions?

H2: There will be significant group differences in accuracy score by gender.

RQ3: How does party identification impact accuracy scores on various misinformation questions?

H3: There will be significant group differences in accuracy score by party identification.

RQ4: How does state impact accuracy scores on various misinformation questions?

RQ5: How does income impact accuracy scores on various misinformation questions?

RQ6: How do these variables interact?

**Method**

**Open Science Materials**

To allow access to this project using the exact version of R, Rstudio, and packages, a web binder of this project was created and can be accessed using this link:

https://mybinder.org/v2/gh/nickl103/psy8712-final.git/HEAD

A web binder allows people to view a project using the exact same software versions that were used by the project creator when they created the binder, as almost a time capsule of the project. I created one for my project so if various functions I used become outdated by updates by code will still be usable through the binder.

All the materials for this project are also available on github which can be downloaded to one’s own computer. The link for the git repository is <https://github.com/nickl103/psy8712-final.git>. On the main page of the repository, there is a readme file which provides an overview of all of the folders and files within the repository.

**Participants**

My participants are people.

**Measures**

Party identification was measured by asking participants which party they identify with, with the response options being “Republican”, “Democrat”, “Independent”, or “something else”.

Gender was measured by asking participants what gender they were, with the response options being “male” or “female”.

Vote choice was measured by asking participants who they voted for in the 2020 election, with the response options being “Joe Biden”, “Donald Trump”, “Someone else”, or “Did not vote”.

Income was measured by asking participants their household income which was divided into seven response options ranging from “under $10,000” to “$150,000 or more”.

State was measured by asking participants what state they lived in, with the response options being the states and DC.

Accuracy score was measured using ten misinformation questions with “True”, “False”, and “I don’t know” options, where each participants correct answers were counted and then divided by the total number of questions to get an accuracy percentage score. Participants who did not answer or said “I don’t know” for more than 4 of the ten questions were excluded from analysis.

**Procedure**

1,111 adults were surveyed from December 2, 2021, to December 3, 2021. They were asked various question about conspiracy theories and misinformation.

**Analyses**

**Descriptive Statistics and Static Visualizations**

**Table 1.**

*Descriptive data by Gender.*

|  |  |  |  |
| --- | --- | --- | --- |
| **Gender** | ***N*** | **Mean Accuracy Score** | **SD**  **Accuracy Score** |
| Male | 365 | .80 | .26 |
| Female | 315 | .78 | .26 |

**Table 2.**

*Descriptive data by Party Affiliation.*

|  |  |  |  |
| --- | --- | --- | --- |
| **PID** | ***N*** | **Mean Accuracy Score** | **SD**  **Accuracy Score** |
| Republican | 197 | .55 | .27 |
| Democrat | 281 | .94 | .11 |
| Independent | n202 | .83 | .23 |

**Table 3.**

*Descriptive data by Vote Choice.*

|  |  |  |  |
| --- | --- | --- | --- |
| **Presidential Choice** | ***N*** | **Mean Accuracy Score** | **SD**  **Accuracy Score** |
| Joe Biden | 441 | .93 | .11 |
| Donald Trump | 239 | .53 | .26 |

**Table 4.**

*Descriptive Information by State*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| State | N | Mean  Accuracy Score | SD  Accuracy Score | State | N | Mean  Accuracy Score | SD  Accuracy Score |
| **ME** | 3 | .80 | .17 | **SC** | 11 | .56 | .34 |
| **NH** | 3 | .47 | .50 | **GA** | 19 | .81 | .22 |
| **MA** | 16 | .81 | .23 | **FL** | 27 | .76 | .27 |
| **RI** | 5 | .84 | .30 | **KY** | 8 | .75 | .27 |
| **CT** | 5 | .98 | .05 | **TN** | 12 | .7 | .34 |
| **NY** | 51 | .78 | .25 | **AL** | 5 | .54 | .26 |
| **NJ** | 24 | .81 | .25 | **MS** | 3 | .97 | .06 |
| **PA** | 30 | .81 | .24 | **AR** | 5 | .48 | .33 |
| **OH** | 22 | .77 | .32 | **LA** | 9 | .82 | .23 |
| **IN** | 7 | .77 | .32 | **OK** | 8 | .84 | .23 |
| **IL** | 20 | .80 | .27 | **TX** | 44 | .75 | .24 |
| **MI** | 20 | .86 | .26 | **MT** | 5 | .72 | .28 |
| **WI** | 8 | .91 | .17 | **ID** | 7 | .74 | .32 |
| **MN** | 17 | .89 | .21 | **WY** | 1 | 1 | NA |
| **IA** | 13 | .72 | .36 | **CO** | 12 | .73 | .26 |
| **MO** | 10 | .73 | .34 | **NM** | 3 | .93 | .12 |
| **ND** | 4 | .9 | .12 | **AZ** | 28 | .67 | .28 |
| **SD** | 3 | .7 | .3 | **UT** | 4 | .88 | .19 |
| **NE** | 4 | .83 | .15 | **NV** | 8 | .79 | .25 |
| **KS** | 12 | .85 | .22 | **WA** | 23 | .89 | 0.22 |
| **DE** | 2 | .7 | .42 | **OR** | 6 | 1 | 0 |
| **MD** | 19 | .9 | .19 | **CA** | 79 | .81 | .25 |
| **DC** | 4 | 1 | 0 | **HI** | 3 | .87 | .15 |
| **VA** | 21 | .79 | .29 |  |  |  |  |
| **WV** | 5 | .56 | .31 |  |  |  |  |
| **NC** | 22 | .86 | .20 |  |  |  |  |

**Table 5.**

*Descriptive data by Income.*

|  |  |  |  |
| --- | --- | --- | --- |
| **Household**  **Income** | ***N*** | **Mean Accuracy Score** | **SD**  **Accuracy Score** |
| Under $10,000 | 7 | .5 | .32 |
| $10,000 to $24,999 | 37 | .77 | .32 |
| $25,000 to $49,999 | 80 | .75 | .29 |
| $50,000 to $74,999 | 118 | .77 | .23 |
| $75,000 to $99,999 | 95 | .81 | .25 |
| $100,000 to $149,999 | 144 | .78 | .27 |
| $150,000 or more | 199 | .84 | .24 |

**A graph of a bar graph

Description automatically generated**

**A graph showing a box plot

Description automatically generated with medium confidence**

**A graph of a box plot

Description automatically generated with medium confidence**

**Interactive Visualizations**

My web app can be accessed at this link :<https://nickl103.shinyapps.io/final_shiny/>

I chose to make a web application that displays a scatterplot, specifically a jitterplot, of accuracy score by case identification number. There are five different options on my web page, allowing users to select what gender, what party affiliation, which presidential candidate was voted for, what income bracket, and what state are displayed on the plot.

**Data Cleaning**

For data cleaning, I chose to use dplyr functions as those seemed liked the easiest method for cleaning my data. My first task was to select the columns I would use for my analysis which was determined by my research questions and hypotheses. My next step was to convert skips and “I don’t know”s to NAs for the columns I was doing analysis with as I did not want them to be counted as answers for later analysis. Next, I filtered so only respondents who answered at least six of the ten accuracy questions were kept. I converted various variables to numeric variables for later analysis. Next, I converted all of the demographic variables I would be using for analysis to factor variables so I could include their levels and labels as well as rename the columns to make it easier. I then removed the non-factored variables from the dataset. My last step was to create the accuracy scores that would be used for analysis which I did by counting the number of correct answers each participant had and then creating an accuracy score by doing the count over total number of questions.

**Analysis**

For H1, based on the OLS analysis, there are significant group differences by vote choice, with an Rsquared of .5351. For H2, based on the OLS analysis, accuracy score was significant for males but not for females, with a very small Rsquared of .002.For H3, Based on the OLS analysis, accuracy score was significantly different for all three party identification with an Rsquared of .39.

I ran an interaction OLS model to see how the different predictors interacted. Based on the OLS output, I wanted to see if other forms of modeling would better explain the data so I used machine learning. Based on the Rsquared for each model, the OLS regression is sufficient in explaining the variation in the data.

**Table 6.**

*Machine Learning Results*

|  |  |  |
| --- | --- | --- |
| **Algo** | **CV\_Rqs** | **Ho\_Rqs** |
| Regression | .59 | .58 |
| Elastic Net | .60 | .58 |
| Random Forests | .56 | .58 |
| Xgboost | .60 | .58 |

**Reflection**

In Data Science, I have learned better coding skills that will allow my code to be more legible and reproducible. One of the most useful skills I have learned is strategies on how to import and clean different types of data using tidyverse packages. Previously, I used the not-so great strategy of copying others’ code without truly understanding what each line did. Going forward, my plan is to no longer aimlessly copy code but to work my way through various coding tasks using different strategies I have learned in this course. Throughout the course, the skills I learned in trying to engineer solutions and problem solve will also be extremely helpful moving forward. I am also grateful for the greater understanding of coding and technology that this course has provided me as it has grown my understanding of R. As a proponent of open science, this course has also taught me various strategies for making my own research practices more open and reproducible, such as the use of a standard path structure and use of GitHub.