# Longitudinal Review of Confidence in the Military

Toni Merhar

Department of Psychology, University of Minnesota – Twin Cities

PSY 8712: Data Science

Dr. Richard Landers

May 4, 2024

# Longitudinal Review of Confidence in the Military

This project examines confidence in the military compared to other institutions among the general population. The data utilized was pulled from The General Social Survey (GSS) over a span of 46 years from 1973 to 2018. The data was used to compare self-reported levels of confidence in the U.S. military with an individual’s demographics (age, gender, and years of education) as well as their confidence in ten other institutions.

## Hypotheses and Research Questions

H1: An individual’s level of confidence in the military can be adequately predicted (0.7 R-squared or higher) by their level of confidence in other institutions along with the year, their age, gender, and education.

H2: Confidence in the military has increased over the past 46 years.

H3: Confidence in the military increases with age.

H4: Confidence in the military increases with more years of education.

H5: Males will have a higher level of confidence in the military than females.

# Method

## Open Science Materials

A link to mybinder, a functional web-based binder where the code can be executed using the same libraries and R version I did, can be accessed here: [https://mybinder.org/v2/gh/merha013/psy8712-final.git/HEAD](https://mybinder.org/v2/gh/merha013/psy8712-final.git/HEAD?)

Materials can also be accessed via a public repository titled psy8712-final on github at this link: <https://github.com/merha013/psy8712-final>. Upon accessing this repository, start by reading the README file located in the main folder. It provides the order in which files should be viewed and explains where each is located.

## Participants

The participants consisted of 42,796 adults between the age of 18 and 89 who were surveyed by The General Social Survey (GSS).

## Measures

The demographic data utilized included year (1973-2018), age (from 18 to 89), gender (1 = male, 2 = female), and years of education (from 0 to 20).

The institutions judged by each participant were banks, major companies, organized religion, education, federal government, organized labor, press, medicine, U.S. Supreme Court, scientific community, congress, and the military. Each of these were ranked on a 3-point scale. The GSS assigned 1 = ‘a great deal’, 2 = ‘only some’, and 3 = ‘hardly any’. However, for this analysis I reversed this code so that 3 = ‘a great deal’ and 1 = ‘hardly any’ in order to make the graphs and charts more intuitive with higher values being better.

## Procedures

The GSS collected the data via personal-interview surveys designed to monitor changes in both social characteristics and attitudes in the United States. The data can be found at this website: <https://gss.norc.org/get-the-data/spss> in the first file under ‘Cumulative Data Set (cross-sectional samples from all years)’ titled ‘GSS 1972-2018 Cross-Sectional Cumulative Data (Release 3, May 27, 2020)’. The code book for the data can also be found there.

# Analyses

## Descriptive Statistics and Static Visualizations

The following figures and tables depict the differences in confidence levels between institutions over multiple years as well as confidence in the military compared to years, age, education level, and gender.

**Figure 1:** Average Confidence across Organizations by Year (H1)

A graph of a graph of a graph

Description automatically generated with medium confidence

**Figure 2:** Average Confidence in Military by Year (H2)

A graph of a graph showing the average confidence levels

Description automatically generated

**Figure 3:** Average Confidence in Military by Age (H3)

A graph of a graph showing the average confidence levels in the military by age

Description automatically generated

**Figure 4:** Average Confidence in Military by Education (H4)

A graph of a graph showing the average confidence levels

Description automatically generated

**Figure 5:** Average Confidence in Military by Gender (H5)

A graph with blue squares

Description automatically generated

**Table 1:** Correlation Matrix between Confidence in the Military and other Institutions (H1)

**A table of numbers and symbols

Description automatically generated with medium confidence**

**Table 2:** Correlation Matrix between Confidence in the Military and Demographics (H1)

A screenshot of a computer

Description automatically generated

## Interactive Visualization

I chose to produce a web application/dashboard to allow a closer look at the data in Figure 1 since there is a lot of information—more than can be properly displayed in paper format. An interactive view can enable readers to take a closer look at specific pieces of the data they are interested in. The shiny web app can be found here: <https://merha013.shinyapps.io/final/>

## Data Cleaning

I pulled the entire longitudinal file from the GSS website. I then removed all columns except those identified for my research questions. I removed any data that had an N/A response for confidence in the military since it would not be useful for my analysis. I then re-coded the responses as described in the measures section and changed the column titles to be more intuitive. (The GSS assigned 1 = ‘a great deal’, 2 = ‘only some’, and 3 = ‘hardly any’ for responses in confidence level. However, for this analysis I reversed this code so that 3 = ‘a great deal’ and 1 = ‘hardly any’ in order to make the graphs and charts more intuitive with higher values being better.) I then saved the cleaned data for future use.

## Analysis

To test **H1** (prediction), I looked at the correlations depicted in Tables 1 and 2 and then ran machine learning to see if a sufficient model could be created to predict confidence in the military from the other variables. The correlations between the other variables were all small (less than 0.26), and the Random Forrest model (using *train* and *ranger* as the method in R) created the most accurate prediction model. It was able to get an R-squared of 0.83 using the training table. However, it was likely overfitting the training data, capturing noise rather than the underlying patterns since running it on the holdout table resulted in an R-squared of only 0.23. The other models tested via machine learning all did worse.

**Table 3:** R-squared values for Machine Learning Models

A screenshot of a computer

Description automatically generated

**H1: Rejected.** An individual’s level of confidence in the military can **NOT** be adequately predicted by their level of confidence in other institutions along with their demographics.

To test **H2** (years), I conducted a linear regression (via the *lm* function in R). The results showed a very small but statistically significant positive increase in the level of confidence in the military as years increased (F(1, 42794) = 861.8, R-squared = 0.0197, p < .001). However, this only explains approximately 1.974% of the variation in military confidence.

A screenshot of a computer

Description automatically generated

**H2: Accepted, but only slightly.** While year is a significant predictor of confidence in the military, it explains only a small portion of the variance in confidence.

To test **H3** (age), I conducted a linear regression (via the *lm* function in R). The results showed a small but statistically significant positive increase in the level of confidence in the military as age increases (F(1, 42653) = 254.6, R-squared = 0.006, p < .001). However, this only explains approximately 0.589% of the variation in military confidence. Additionally, looking at Figure 3, you can see that confidence tends to start off higher at age 18 then declines before starting to rise again around age 30, and it doesn’t pass above the confidence at 18 until individuals are in their mid-80s.

A screenshot of a computer code

Description automatically generated

**H3: Accepted (but only slightly).** While age is a significant predictor of confidence in the military, it explains only a very small portion of the variance in confidence.

To test **H4** (education), I conducted a linear regression (via the *lm* function in R). The results showed a very small but statistically significant decrease in the level of confidence in the military as years of education increased (F(1, 42696) = 253.1, R-squared = 0.006, p < .001). However, this only explains approximately 0.589% of the variation in military confidence.

A screenshot of a computer

Description automatically generated

**H4: Rejected.** While total years of education is a significant predictor of military service, it shows a **decrease rather than an increase** with more years of education. Additionally, it explains only a small portion of the variance in confidence.

To test **H5** (gender), I conducted a Welch Two Sample t-test (via the *t.test* function in R) to compare mean scores between male and female participants. The results showed a small but significant difference in mean scores (t(40355) = 9.476, p < .001). The true difference in mean level of confidence in the military between male and female participants is likely within the 0.049 and 0.075 range.

A computer code with black text

Description automatically generated

**H5: Accepted.** Males tend to have a slightly higher level of confidence in the military than females.

# Reflection

Overall, this class has forced me to get comfortable with portions of R and data science in general that I was wholly unaware of previously. Though I don’t foresee the data that I collect needing analysis via a supercomputer, web scraping, or text mining, the basic awareness of them and how they function is valuable. Meanwhile, I will be utilizing all of the R coding skills (and perhaps shiny web apps) this summer when I start coding my first dataset for analysis. Finger crossed!