Public Trust in Banks and Financial Institutions: A Knowledge Mining Approach

Soha Arian, Kasra Akbari, Eman Ajmal, Joseph Martinez

Introduction

Public trust in banks shapes how people see the financial system. This project looks at trust as a self-reported feeling, not as actual financial actions. It does not track whether people open accounts, take out loans, or invest. The main focus is on how confident people say they feel about banks. Earlier studies show that trust and access to financial services can differ by income, education, and age. This project does not claim that trust leads to certain financial actions. Instead, it looks for patterns in reported confidence among different groups. The study uses national survey data to see how income, education, and age relate to confidence in banks. It also checks if artificial intelligence systems can accurately summarize these patterns. The project uses both statistical analysis and a review of AI-generated summaries to study differences in trust and how they are described. In this study, income, education, and age are used as basic life factors that may shape how safe or stable people feel about money. These factors are studied as related to confidence, but they are not treated as direct causes.

Research Question(s)

This project answers three clearly defined questions.

1. How does confidence in banks differ across income, education, and age groups in the United States?
2. When income, education, and age are analyzed together in a statistical model, which of these factors remain related to higher or lower confidence, and how large are those differences?
3. When artificial intelligence systems summarize the statistical findings, do their summaries accurately reflect the statistical evidence without exaggeration or unsupported cause-and-effect claims?

Background

Trust in financial institutions depends on economic experiences, financial stability, and education. People with lower incomes might feel less secure and trust banks less. Education can influence trust by shaping financial understanding. Age is important because each generation has faced different economic events. The General Social Survey (GSS) gathers data on confidence in banks, along with demographic details and the year of the survey. This information makes it possible to study patterns across groups and over time. In this project, knowledge mining means using statistical tools to find patterns in large datasets, rather than just looking at simple averages.

Methodology

This study uses the General Social Survey cumulative dataset from 1972 to 2024. The dataset has responses about confidence in banks, income, education, age, and the year of the survey. All analyses will be done in R. Every step of data cleaning and modeling will be written in code so the analysis can be repeated exactly. The main outcome is confidence in banks. Respondents say if they have "a great deal," "only some," or "hardly any" confidence. These answers will be coded as high, moderate, or low confidence. The categories are ranked from high to low but are not treated as evenly spaced numbers. Income categories change across survey years, so income will be grouped into three brackets (low, middle, and high) based on the income distribution for each year. This keeps income categories comparable over time. This method compares people based on their position in the income ranking for each year, not their exact dollar amount. To make sure this choice does not change the results, an extra model using inflation-adjusted income will also be tested. Education will be grouped by the highest degree completed. Age will be measured in years as a continuous variable. The survey year will be included in all models to account for long-term changes in trust. The models will also test whether the relationship between income and confidence changes across different time periods. Survey weights from the GSS will be used so results represent the U.S. adult population. Cases missing income, education, or age will first be removed from the regression analysis, and the number of removed cases will be reported. An additional test will be done to check whether missing answers change the results.

The first part of the analysis uses descriptive statistics. This means calculating the percentage of people reporting each level of confidence and making tables to show how confidence varies by income, education, and age. Graphs will help show these patterns. This step answers the first research question. The second part uses ordered logistic regression, a method for analyzing ranked categories. This model estimates how income, education, and age relate to higher or lower confidence, while keeping other factors constant. The survey year is included to control for changes in trust over time. The proportional odds assumption for ordered logistic regression will be tested in R. If this assumption does not hold, multinomial logistic regression will be used instead. After the model is estimated, predicted probabilities will show how likely each group is to report high confidence. These results will be shown in graphs to highlight group differences. This step answers the second research question. The last part reviews artificial intelligence summaries.

Once all statistical models are finished, a set written prompt will be used to give the same model results to an AI system. The same prompt will be used for every summary to keep things consistent. The evaluation focuses only on how the AI describes the specific results from this study, not on reviewing what other researchers have written about AI. The AI will create one summary for each model, and a separate summary will be written by a person. Both summaries will be checked using a scoring guide with four criteria: unsupported causal claims, exaggeration of group differences, missing statistical uncertainty, and biased language. Two different reviewers will score each summary using the same guide. Their scores will be compared to make sure the evaluation is consistent. The total score will show how closely the AI summary matches the statistical results and whether it avoids overstating the findings. This step answers the third research question.

Knowledge Mining Contribution

This project uses knowledge mining to find patterns in how people feel about banks, using a large national survey. In this project, knowledge mining means using clear statistical models and step-by-step analysis to find patterns in survey data in a structured way. Instead of just reporting an overall average, this study looks at how trust changes by income, education, and age and examines these factors together. This helps find clear patterns in the data. The project also checks if artificial intelligence systems explain the statistical results accurately. In short, the study finds patterns in the data and checks how well they are described.

Policy and Governance Relevance

Learning about differences in confidence in banks shows how different groups see financial institutions. If some income, education, or age groups report lower confidence, it could point to concerns about fairness or stability in the financial system. These differences do not prove there is a problem, but they highlight areas where institutions should pay attention to public opinion. The project also matters for governance because artificial intelligence systems are now used to summarize research. It is important that AI summaries are accurate and do not exaggerate differences or make unsupported claims. By checking AI summaries, this study helps make sure research findings are shared clearly and responsibly.

Limitations

This study finds associations, not cause and effect. Confidence in banks is self-reported and might not match actual behavior. Grouping income across years could make some differences less clear. The AI evaluation checks how accurately results are interpreted but does not measure bias in the AI training data. Because the dataset covers many decades, it may hide short-term changes in trust that happened during major financial events.

Conclusion

This project studies demographic differences in public confidence in banks using national survey data and reproducible statistical methods in R. The study combines structured statistical modeling with a careful review of AI-generated summaries to produce findings that are clear, transparent, and responsibly shared. All AI prompts, outputs, and changes will be included in an appendix, along with a short explanation of how AI was used in the research.

References

Grable, John E., Eun Jin Kwak, and Kristy L. Archuleta. "Distrust of Banks among the Unbanked and Banked." *International Journal of Bank Marketing*, vol. 41, no. 6, 2023, pp. 1498–1520. https://doi.org/10.1108/IJBM-10-2022-0441.

Crosignani, Matteo, et al. "Financial Inclusion in the United States: Measurement, Determinants, and Recent Developments." *SSRN Electronic Journal*, 2025, https://doi.org/10.2139/ssrn.5534239.

General Social Survey. "Get the Data in Stata Format." *GSS – NORC at the University of Chicago*, https://gss.norc.org/get-the-data/stata.html. Accessed 26 Feb. 2026.