Characterizing Political Opinions

Vincent Pun

Abstract

Political views have historically been divided between the Right and Left. Given both the complexity and variety of values, beliefs, and political attitudes present in society, a deeper understanding of the nation's changing political landscape should be achieved through the observation of factors and key issues that leave these political groups most deeply divided. Statistical methods such as principal component analysis and factor analysis are used to assess relationships between questions sourced from Pew Research Center's March 2019 Political Survey data, and this is used to characterize the political opinions in the United States.

Keywords Political survey, correlation plot, principal component analysis, data complexity, min-max normalization, factor analysis, path diagram, heatmap

Introduction

Conducting survey research on thousands of unique participants can be both time-consuming and expensive. Risks such as survey fatigue, incomplete responses, and uninteresting results may jeopardize the quality of the research. Thus, it is important to understand which key categories and characteristics create the distinct separations between political mindsets when developing surveys, so consultants may eliminate redundant questions and efficiently gather meaningful insights.

Literature review

The visualization of item-factor relationships is clearly displayed through the use of colored heatmaps; creating path diagrams is also a popular method for viewing the strength of loading between factors and their corresponding manifest variables (Chapman and Feit 2015).

Methods

Pew Research Center's March 2019 Political Survey data was used as the starting point for this research. The original dataset contains survey data from 1,503 participants, and there are a total of 129 variables that consist of questions covering areas such as basic demographics, politics, the federal government, foreign policy, and social issues.

This research focuses specifically on the variables where questions were asked to all participants. These questions are separated and coded numerically as part of the preparation process. A graphical display of a correlation matrix is applied to determine basic relationships between a total of 30 unique questions, and this is reordered to help identify groups of highly correlated data.

Figure 1: Correlation plot of survey research data asked to all participants

To formulate a method of reducing complexity in this dataset, traditional statistical methods such as principal component analysis (PCA) and factor analysis (FA) are calculated and visualized. The goal of PCA is to maximize the retained amount of interesting information by finding linear relationships between groups of variables, which are referred to as components. We see that the first four principal components account for 76.5% of the explainable linear variance in this dataset. To visualize the PCA solution, a scree plot is used to see that the cumulative proportion of variance levels out at the fourth principal component, which suggests this is the appropriate cutoff to reduce data complexity.

Figure 2: Scree plot with parallel analysis

Furthermore, we review the variable loadings in the rotation matrix, an output of the PCA computation, to confirm observed relationships between variables found on the correlation plot. Interpreting the direction (positive or negative) of each variable in the first four principal components helps us identify questions that function similarly in terms of how they explain the dataset.

For example, questions 20 and 25 in the survey ask participants if they are content with the federal government and if they can trust the government, respectively; both variables have high positive values in component 3 (PC3), which confirms that there is overlapping shared variance. This process is repeated for all 30 variables.

To reduce the dimensionality of the data, we eliminate 8 variables that did not contribute to explaining linear relationships in the first four principal components, and the remaining variables are combined based on similarities; the mean of each group is taken. Prior to these calculations, the data is brought to a common format through the use of min-max normalization, a process that transforms minimum and maximum values into decimals between 0 and 1. In result, the data is consolidated down to 10 variables that explain key characteristics found within the survey such as politics, the U.S. economy, tax regulation, foreign policy, race, and social discrimination.

Results

Performing a PCA computation on the new reduced dataset now shows that the first four principal components explains 90.74% of what is interesting in the data–an improvement from 76.5%. To assess the relationship between these ten variables, we use exploratory factor analysis to identify factors, which are regarded as latent variables that cannot be observed directly, but are understood through their relationship with other variables (Chapman and Feit 2015). Again, a scree plot is used, and it suggests three factors. To visualize the loadings for each key political topic, we use a heat map to show separation into these three factors. Additionally, a path diagram is produced, where significant relationships are traced from latent trait to significant variable loadings.

Figure 3: Heatmap - Factor loadings for political opinions

Figure 4: Path Diagram - Key survey characteristics traced to three factors

We see that the first factor in assessing the construct of political opinion find correlation across individual participant responses on views of Congress, activity in world affairs, same-sex marriage, and fairness in the economic system. The second factor is related to having trust in the current federal government, administration, and tax regulations. Finally, the third factor pertains to awareness on foreign policy and public affairs.

Conclusions

Effective survey research is highly dependent on good quality data. To reach this goal, it is optimal for researchers to ask questions that will produce aid in segmenting participants based on their views of the current political landscape. Methods such as PCA and FA have helped reduce the complexity and length of the overall survey by identifying the questions and key issues that are most important in differentiating political groups, which should reduce the costs and time needed to conduct future studies.

References

"How the Political Typology Groups Compare." Pew Research Center - U.S. Politics & Policy. Pew Research Center, July 2, 2020. https://www.pewresearch.org/politics/interactives/political-typology-comparison-2017/.

Chapman, Chris. & Feit, E. R For Marketing Research and Analytics. Cham: Springer, 2015.

Hatley, Nick. "How to Analyze Pew Research Center Survey Data in R." Medium. Pew Research Center: Decoded, January 4, 2020. https://medium.com/pew-research-center-decoded/how-to-analyze-pew-research-center-survey-data-in-r-f326df360713.

Hatley, Nick. "Using Tidyverse Tools with Pew Research Center Survey Data in R." Medium. Pew Research Center: Decoded, January 4, 2020. https://medium.com/pew-research-center-decoded/using-tidyverse-tools-with-pew-research-center-survey-data-in-r-bdfe61de0909.

About the Author

Vincent Pun is a Master's candidate in the data science program at Northwestern University, experienced in financial services process automation, tax reporting, and strategy consulting.