Survey Research Methodology II Assignment - Summary

Analyzing discrimination against transgender people across the EU

In this study we had two main objectives: to examine cross-country differences between European countries' support for transgender individuals obtaining official documents reflecting their gender identity, and to develop a predictive model in order to forecast support levels in other countries based on observed trends. Our analysis revealed that support was linked to being a woman, knowing a transgender person, being non-religious, and frequent internet use, while opposition was associated with anti-LGBTQ+ views, broader discriminatory attitudes, lower life satisfaction, and right-wing ideology. Our predictive models achieved highly accurate forecasts at the country level, though concerns about overfitting and limited generalizability beyond Europe pose the need for a larger and more diverse dataset.

What we already know

Existing research into support for and opposition to transgender rights helped give us a starting point for what themes may be important in our model. A study by Kanamori and Xu (2020) found that religious fundamentalism has a significant positive effect on transphobia, both directly and indirectly through reduced contact with transgender individuals. Additionally, this study identified that increased contact with transgender people is associated with lower levels of transphobia. Individual factors, such as religiosity, gender (where men tend to exhibit higher levels of transphobia), age, and right-wing authoritarianism, have also been identified as relevant. Moreover, adherence to other prejudices, such as homophobia, sexism, and beliefs in traditional gender roles, is related to transphobia. Aguirre-Sánchez-Beato (2020), in their review of theoretical approaches, summarizes how cognitive perspectives highlight these individual characteristics, external factors such as contact and information about trans people, and the connection with other prejudices as key elements in explaining transphobia. All of these variables have been captured in our model.

Data preprocessing: selecting, recoding, and feature engineering

Given the large dataset (coming from the Eurobarometer May 2019), we initially filtered out variables irrelevant to our analysis, such as those related to trade, globalization, and EU energy policies. For many sociodemographic variables, multiple versions were available (ex. different categorization of occupation), we selected only one of their versions by making informed decisions based on visualizations and logical relationships.

Creating new measures to reduce dimensionality

The dataset contained numerous variables on discrimination at both individual and country levels, posing the challenge of capturing attitudinal differences without introducing multicollinearity or losing information. To address this, questions about perceptions of discrimination at the country-level were not included in our dataset to prevent conflating societal (captured by our country-level additional variables) and personal perceptions (captured by the Eurobarometer survey). A composite personal discrimination score was constructed by aggregating attitudes toward various groups, condensing highly correlated variables. Additionally, minority socialization scores were developed, including an index of friendships with minorities, a count of reported anti-discrimination actions, and a social alienation index based on perceived societal influence, all constructed by aggregating results of several questions.

Inclusion of country level-data

To complement the individual level data and build an explanatory mixed model, additional country level economic and demographic indicators were included. We sourced reputable aggregated country level measures, including:

- GDP per capita (adjusted PPP) as a macroeconomic indicator (World Bank).
- LGBT+ rights index (Our World in Data) that scores on 18 different policy measures.
- Gender Development and Gender Inequality Indices (UNDP), which assess gender disparities in health, education, and economic participation.
- Democracy Index (The Economist) to score on factors like electoral process, civil liberties and political participation.

For completeness in coverage and simplicity in modelling, where multiple disaggregated scores existed, the overall measure was selected for each index.

Dealing with missing data

Our dataset didn't have explicit NAs, as most variables were encoded as factor variables including levels such as "Don't Know" or "Refusal to answer". We replaced those responses, as well as certain ambiguous categories, with NAs, in order to enhance the quality of our analysis after exploratory testing. We did have one key exception. We hard coded refusal to answer the question about having transgender friends (sd1_7) for direct relevance and this ended up being a significant variable.

We also conducted analysis into NA responses to the target variable to try identify potentially concerning patterns. Univariate analysis and logistic regression were undertaken with non-response to qc17 as the binary target. Women, less frequent internet users and working class people were some of the groups to exhibit higher non-response rates. Cross-country disparities—ranging from 1.4% in Belgium to 28.5% in Bulgaria—highlight cultural reluctance to engage with sensitive topics.

When considering paradata, survey cooperation emerged as a significant factor, with uncooperative respondents disproportionately answering "Don't Know" to our target question. "Bad" cooperation was 3.5 times more likely to result in non-response than "excellent" cooperation. This was the most extreme odds-ratio observed, while other components were not so worrying. This allowed us to perform imputation of the handful of missing observations for our predictor variables.

Explanatory model

Selecting our independent variables

To develop a robust explanatory model for public support of transgender rights, we first focused on selecting significant individual-level variables and minimizing multicollinearity among them. The key challenge was managing numerous predictors to balance model complexity with interpretability. We employed a simple logistic regression, stepwise selection, and Lasso regularization to refine our understanding of the most relevant variables. Using these results, our previous descriptive analysis and existing literature, we finalized our selection. To enhance model simplicity we retained only the most relevant levels for certain categorical variables, resulting in a final set of about 15 individual-level variables including demographic characteristics (such as age, gender, social-economic status), other individual attitudes question (political orientation, religious beliefs, happiness etc.), but also many variables related to being discriminated against or being discriminatory against others.

Mixed model testing at multiple levels

Following the multilevel modelling approach outlined by Hox (2010), we tested models at progressive levels to account for both individual and country-level effects:

1. Baseline Testing: A null model was run with only random intercepts per country, capturing country-level variance.
2. Level 1: Added individual-level fixed effects to examine demographic and attitudinal predictors.
3. Level 2: Included country-level fixed effects (e.g., GDP per capita, gender inequality index, LGBTQ+ policy index) to assess national influences.
4. Level 3: Introduced random slopes and cross-level interactions to analyze how national contexts influence individual attitudes.

Key findings

Stronger support for transgenders' rights was found among women, non-believers, those who know somebody that identifies as transgender, and those who use the internet daily. In contrast, opposition was more common among men, right-wing supporters, ethnic minorities, unhappy people and those who are more discriminatory in general. None of our country-level fixed effects was significant, although some of our cross level interactions were significant, such as gender or right-wing ideology and the LGBTQ score of a country, as well as being non-religious and the democracy index or the gender inequality index. These interactions, for example, suggest that non-religious people are less likely to support transgenders' rights in countries where there is higher gender inequality, but also that in more democratic countries being non-religious is not such a determining factor in explaining support for transgenders' rights. They also suggest that right-wing ideology has a stronger negative effect on support in countries with weaker LGBTQ+ protections.

In our models, the percentage of variance explained by the grouping structure (i.e. by the differences between countries, random effects) is very low, indicating that our explanatory power comes from individual-level fixed effects. Indeed, incorporating individual-level fixed effects increased our models' performances by a lot, while incorporating country-level fixed effects did not increase it and incorporating cross-level interactions improved the performance of the model, but only slightly, suggesting that again most of the variance is explained at an individual level.

Our findings at the individual level align (at least partially) with existing research and seem robust across all our model specifications. We do have some concern that the ‘no’ votes are underreported due to the overlap in who is more likely to respond ‘don’t know’ and ‘no’ to our target. This may indicate some social desirability bias hidden in the missing responses.

Predictive model of country level support

To develop a predictive model for assessing support for transgender individuals' rights to change their civil documents across different countries, we aggregated individual-level survey data to a country level. The dependent variable is therefore defined as the percentage of individuals in each country supporting this right. Explanatory variables included country-level means for all our numerical variables and country level proportions of different levels for all our categorical variables. Due to the limited number of observations (28 countries), we used Leave-One-Out Cross-Validation (LOOCV) to assess model performance instead of a traditional train-test split. We applied multiple modeling approaches, namely random forest, gradient boosting, and elastic-net linear regression, with LOOCV for hyperparameter tuning and model selection.

Model performance

Elastic Net and Random Forest already perform extremely well, however, Gradient Boosting performs even better with almost perfect predictions. This extreme predictive power is due to the small number of observations of which our dataset is built off, which allows us to train the models using LOOCV yielding impressive results. However, we should be skeptical of these results. Given the small number of observations and their shared characteristics, our models can predict extremely well within this dataset, but their reliability outside of the European context is questionable. A larger and more diverse dataset would be necessary to improve generalizability and mitigate overfitting.

Key findings

Some of the most important predictors of support for transgender change of documents at the country level were: the proportion of population against the inclusion of a third gender option in official documents, and against LGBTQ rights (aggregated from survey questions); roma population, manual workers, right-wingers, the average educational level, and the country’s score on the LGBT Policy Index. While most of them align with the observations from our explanatory models, the educational level variable had no relevance when analyzing mixed-effects, which is indicative of a more complex or non-linear relationship with the target.

Overall thoughts

Overall, the individual level variables are central to our mixed-level models. While country-level data on indices like LGBTI+ policy progressiveness are useful, predictive models of attitudes are better suited to individual level data. With our models, the attitude to our target is around 84% explainable. Our models agree with existing literature on social inclusion and are a useful tool to understand what influences individuals' attitudes.

References:

Aguirre-Sánchez-Beato, S. (2020). EXPLAINING TRANSPHOBIA AND DISCRIMINATION AGAINST TRANS PEOPLE: A REVIEW OF THEORETICAL APPROACHES. *Sexualidad y Salud*. https://doi.org/10.1590/1807-0310/2020v32n190274

Hox, J.J. (2010). Multilevel analysis: Techniques and applications. Second edition. New York: Routledge.

Kanamori, Y., & Xu, Y. J. (2020). Factors Associated with Transphobia: A Structural Equation Modeling Approach. *Journal of Homosexuality*. Publicado en línea: 15 de diciembre de 2020.