

Datathon Writeup

Linda Cao, Georgie Stammer, Enzo Moraes Mescall, Erik Mendes Novak

2021-11-07

Introduction

Immediately after going over the data set our group was interested in section P of the questionnaire where individuals were asked a series of intense political questions and responded with their opinions. This provides a unique subjective insight of a population's perspective on their civic space. The data set provides information about the material conditions, and their perceived material conditions, of the surveyed and thus allows us to attempt to estimate the impact of an individual's material conditions on their political leanings. Notably, all the questions in section P are on a scale of 1 - 4 where answering 1 indicates support of a larger authoritarian government while answering 4 indicates support for a more democratic form of government.

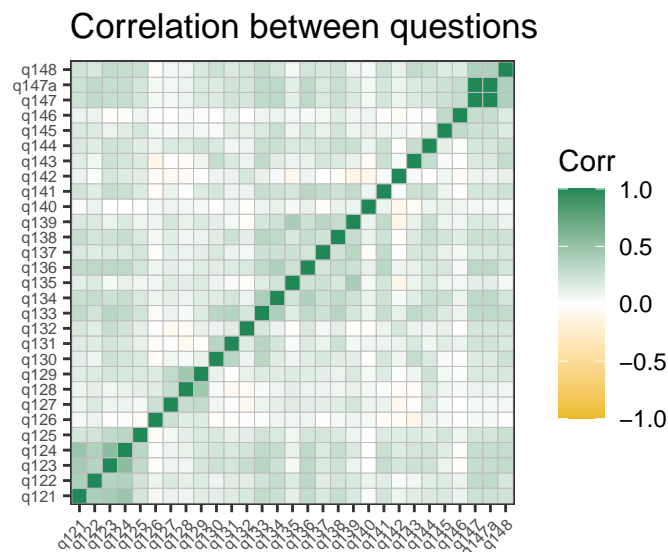
Research Question *To what extent do an individual's material conditions affect their leanings toward authoritarian form of government?*

Methodology

To measure an individual's political leanings based off of their answers from the questionnaire we took their responses from questions Q121 to Q148 which scale from 1 to 4 and took the arithmetic mean across selected questions, which we refer to as the political index.

//space for erik

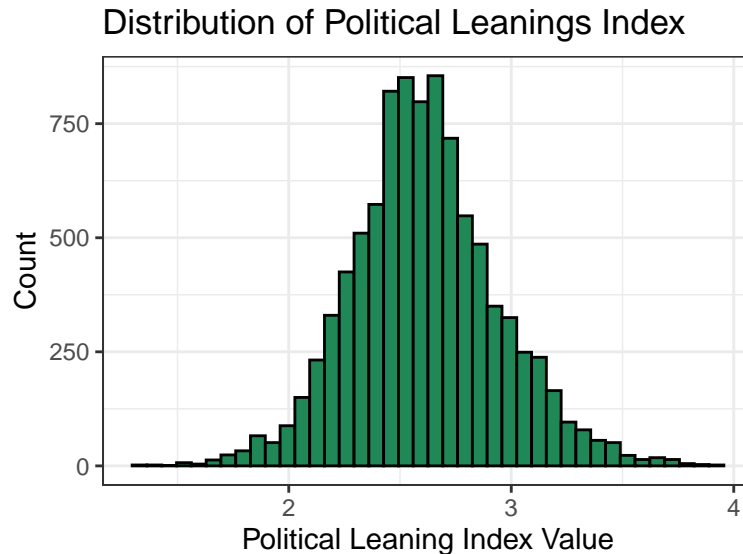
```
ggcorrplot(corr_ideology, ggtheme = theme_bw(), tl.cex = 6,  
           colors = c("#ebbb2a", "white", "#208756")) +  
  labs(title = "Correlation between questions")
```



Next we look at the aforementioned questions in section P.

Exploratory Data Analysis

```
ggplot(data = final_data, aes(x = avg)) +  
  geom_histogram(fill = "#208756", color = "black", bins = 40) +  
  labs(title = "Distribution of Political Leanings Index",  
       x = "Political Leaning Index Value",  
       y = "Count") +  
  theme_bw()
```



```
kable(learning_index, digits = 3,  
      caption = "Summary Statistics for Learning Index",  
      col.names = c("Mean", "Max", "Median", "Min", "IQR", "SD"))
```

Table 1: Summary Statistics for Learning Index

| Mean | Max | Median | Min | IQR | SD |
|-------|-------|--------|-------|-------|------|
| 2.613 | 3.926 | 2.593 | 1.333 | 0.415 | 0.33 |

The distribution is unimodal, looks vaguely normal and doesn't have any apparent outliers.

Data Analysis

After cleaning up the data we created a linear model using the responses to questions SE001 through SE017 and then selected the most relevant predictor variables to keep in the model based off of their p-values with the threshold being $p > 0.1$.

```
model <- lm(avg ~ country + gender + education, data = final_data) %>%  
  tidy(conf.int = TRUE) %>%  
  kable(digits = 3, title = "Linear Regression Model for Political Index")  
model
```

| term | estimate | std.error | statistic | p.value | conf.low | conf.high |
|------------------|----------|-----------|-----------|---------|----------|-----------|
| (Intercept) | 2.767 | 0.014 | 202.361 | 0.000 | 2.740 | 2.794 |
| countryHong Kong | -0.296 | 0.014 | -21.311 | 0.000 | -0.324 | -0.269 |
| countryKorea | -0.204 | 0.011 | -18.378 | 0.000 | -0.225 | -0.182 |

| term | estimate | std.error | statistic | p.value | conf.low | conf.high |
|--------------------|----------|-----------|-----------|---------|----------|-----------|
| countryChina | -0.473 | 0.012 | -38.357 | 0.000 | -0.497 | -0.449 |
| countryMongolia | -0.430 | 0.012 | -35.948 | 0.000 | -0.454 | -0.407 |
| countryPhilippines | -0.327 | 0.012 | -27.324 | 0.000 | -0.350 | -0.304 |
| countryTaiwan | -0.268 | 0.012 | -22.565 | 0.000 | -0.291 | -0.245 |
| countryThailand | -0.436 | 0.012 | -36.208 | 0.000 | -0.460 | -0.412 |
| gender | -0.014 | 0.006 | -2.396 | 0.017 | -0.025 | -0.003 |
| education | 0.016 | 0.001 | 19.557 | 0.000 | 0.014 | 0.017 |

Which we may represent in the equation:

$$\begin{aligned}
\hat{PolIndex} = & 2.767 - 0.296 \text{ countryHK} - 0.204 \text{ countryKO} - 0.473 \text{ countryCN} - 0.430 \text{ countryMN} \\
& - 0.327 \text{ countryPH} - 0.268 \text{ countryTW} - 0.014 \text{ gender} + 0.016 \text{ education} + \epsilon_i, \quad \epsilon \sim N(, \sigma_\epsilon^2)
\end{aligned}$$

The y-intercept represents a female individual from Japan with 0 years of formal education. All the countries are indicator variables and these carry the most weight on the response variable. Gender is also categorical but having a small impact of only a -0.014 difference in male political indices, on average and holding all other variables constant. The only numerical is years of education where, on average, for every additional year of education we would expect an individual's political index to raise by 0.017, holding all other variables constant.

Conclusion

These findings would match our intuition about education where increasing years of schooling dissuades authoritarian ideology. Another point to consider is that belonging to countries generally associated to authoritarian regimes, like the PRC, Mongolia and Thailand, generally have a much greater negative effect on an individual's political ideology

ANOVA conclusion analysis