

POLI 170A - Problem Set 1

Gonzalo Rocha-Vazquez

7/3/2022

Question 1. What are the 5 components that your research project will be asked to address?

1. Question

- *This is the central focus of the project, a question with important implications (e.g., how does political affiliation affect voting turnouts in non-swing states?). Traditionally in quantitative analysis it's a phenomenon, puzzle, or variation that isn't explained by current theory; however, it could also be something that hasn't received attention before and you're attempting to explain it using literature on adjacent topics (e.g., using Political Violence literature to explain Criminal Violence & State-Cartel Conflict).*

2. Answer

- *This is your theoretical framework, the causal mechanism that gives your hypothesized explanation for how the Independent Variable affects the Dependent Variable. You build this by using the existing literature on the variables associated with your project.*

3. Data & Methods

- *This is (1) The data sources that you utilize to answer the question and (2) the statistical methods that you utilize to test these methods. The data source should include the unit of analysis & temporal information (number of observations optional), and the statistical methods should be detailed (e.g., specify covariates) and appropriate for testing the hypothesis.*

4. Findings

- *This is where you detail the results of your analysis: what's the answer to your question? Was your theory correct? The results could be null or inconclusive (i.e., not statistically significant), and ideally should note size of the effect and any interesting heterogeneity between groups (e.g., democracies have never gone to war with one another).*

5. Contributions (& room for improvement)

- *This should be where you talk about the implications of your results on social science research overall, what additional research needs to be done? Surprising (or null) results can elicit new questions, new avenues for future research—this is where you note that.*

Question 2. Identify the 5 components in the following abstract.

1. Question

- *“Despite the increasing globalization of housing markets, little is known about its political implications. This study investigates whether rising Chinese investments in U.S. homes influenced local voting in recent U.S. presidential elections.”*

2. Answer

- *“Building on pocketbook/sociotropic voting and nativism theories, I develop hypotheses on the electoral consequences of foreign real estate investment through greater home demand and equity, improved local economies, and changing neighborhoods.”*

3. Data & Methods

- *“Using difference-in-differences designs that combine a unique shock to Chinese capital outflows in 2013 with county-level measures of local attractiveness to Chinese investments...”*

4. Findings

- *“...I find that greater exposure reduced Democratic vote shares in both the 2016 and 2020 presidential elections. Furthermore, an initially larger white population strengthened this effect while a larger college-educated population weakened it. In contrast, local equity gains, housing competition, or economic strength did not systematically influence the effect.”*

5. Contributions (& room for improvement)

- *“Together, the results appear more consistent with the pro-conservative effects of nativism.”*

Question 3. How do we get from important but vague topics to specific, testable research questions?

1. Give an example of a broad concept that is of interest to social scientists. (2 points)

- *The impact of Corruption and Organized Crime on the development (economic, institutional, violence, etc.) of countries.*

2. Give an example of a property/variable that you could measure about that concept. (2 points)

- *You could measure the violence associated with Organized Criminal Groups (OCGs) and how this varies based on a government's Drug Trafficking Regulatory Arrangements (e.g., Tacit Coexistence or Particularistic Confrontation vs. Protection-extraction rackets or Particularistic negotiation).*

3. Give an example of a way you could operationalize that property/variable so that other researchers could measure it the same way. (5 points)

- *You could count the deaths in a region resulting from Inter-Cartel & State-Cartel conflicts, and categorize the region's Drug Trafficking regulatory arrangement (state-level in a federal country).*

4. What is the unit of analysis in the operationalization of the variable you specified? (2 points)
 - *The unit of analysis would be individual deaths attributed to Inter-Cartel & State-Cartel conflicts.*
5. Give an example of values that the operationalized variable could take on. (2 points)
 - *This could either be the total number of deaths or a rate (e.g., number of deaths per 100,000 people); it could either be binary (1 = caused by Inter-Cartel & State-Cartel conflicts, or 0 = not caused by Inter-Cartel & State-Cartel conflicts), or you could differentiate between deaths caused by Inter-Cartel conflicts & deaths caused by State-Cartel conflicts to see how regulatory arrangements affect not only violence but also the type of violence.*
6. What type of measurement is your hypothesized variable? (Nominal, Ordinal, Continuous) (2 points)
 - *I believe this would be a continuous (ratio) variable: we can say that State X has a rate of 50% & State Y has a rate of 25%, and that's comparable to State A having 25% and State B having 12.5%; in each pair the former is twice the latter.*

Question 4. Loading data into R

```
# 1) Open a new R Script and set your working directory to be the folder where
#    you have saved the data. (2 points)
```

```
getwd() #Checks our current Directory
```

```
## [1] "/Users/gonzalorochoa-vazquez/Desktop/POLI 170A/Problem Set 1"
```

```
setwd("/Users/gonzalorochoa-vazquez/Desktop/POLI 170A/Problem Set 1")
#Changes directory if needed
```

```
# 2) Install the readxl package if you have not already.
#    Load the readxl package. (2 points)
```

```
#install.packages("readxl")
library(readxl) #Only need to install once, so comment the installation,
                #but leave the load uncommented or else we'll get an error
                #when running this script
```

```
# 3) Load the dataset into R. When doing so, change the variable type of
#    "Rural population (% of total population)" from character to numeric.
#    This will automatically convert missing data (which the World Bank format
#    records as ..) into NA (3 points)
```

```
dataset <- read_xlsx("RuralPopData_POLI170A.xlsx") #loads our dataset
dataset$`Rural population (% of total population)` <-
  as.numeric(dataset$`Rural population (% of total population)`)
#rewrites our selected variable from chr to num
```

```
# 4) Make a new version of the dataset called "rural pop" (2 points)
```

```

rural_pop <- dataset #Creates a second object identical to our first object

# 5) Remove the old version of the dataset (2 points)

rm(dataset) #removes only our first object, leaving us with object 2

# 6) Rename the variables to be: (a) country; (b) iso; (c) rural pop percent (3 points)

#install.packages("dplyr")
library(dplyr) #dplyr is your best friend for wrangling data

```

```

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

```

```

rural_pop <- rename(rural_pop, country = 'Country Name', #rewrite original object
                    iso = 'Country Code', #format: new_name = old_name
                    rural_pop_percent =
                    'Rural population (% of total population)')

# 7) Report the mean of rural pop percent for all observations (3 points)

mean(rural_pop$rural_pop_percent, na.rm = FALSE) #mean of all observations

```

```
## [1] NA
```

```
mean(rural_pop$rural_pop_percent, na.rm = TRUE) #removes the NA observation(s)
```

```
## [1] 39.39201
```

```
# 8) What type of variable is rural pop percent? (3 points)
```

```
print("rural pop percent is a continuous (ratio) variable.")
```

```
## [1] "rural pop percent is a continuous (ratio) variable."
```

Question 5. Are you working with a partner on the final project, or by yourself?

As of right now, I will be working by myself on the final project.

Question 6. What is the research question you will answer for your final project?

Currently my question is *How does the role of local police in crime (Protect against vs. Participate in) affect citizen's threats to security?*. This is significant because one of the main priorities for regions dealing with Drug Trafficking Organizations (DTOs) is minimizing the violence (i.e., deaths) that results from these conflicts, and the regulatory arrangements that the state (i.e., police) has with DTOs greatly impacts the levels of violence in the area.

Note: I'm trying to balance attempting a more advanced project, but also accounting for feasibility given the timeframe and pace of the course. I might stop by office hours to talk through options for turning this into a slightly advanced project that should still be feasible in the remaining weeks.

Question 7. What is the theory that you will argue answers your research question?

According to Flom (2019), when the police coordinate with Drug Trafficking Organizations through Protection rackets, criminal violence is reduced because stability and predictability are reduced. This extends to refraining from “visible” violence, which is violence aimed at civilians uninvolved in the DTOs activities.

Note: The citation is a link to the article, clicking on Flom (2019) should take you to the article.

Question 8. What is the testable hypothesis your final project will assess to provide evidence that supports or does not support your theory?

Civilians whose local police participate in crime will be less likely to view Organized crime and drug trafficking groups as the biggest threat to their safety than civilians whose local police protect against crime.

Question 9. What is the independent variable in your hypothesis? The dependent variable? What types of measurements are these variables?

The independent variable here will be whether the civilian believe the local police protect against crime or participate in it (*AOJ18: Police Protect vs Police Participate in Crime*). The dependent variable is the group that pose the greatest threat to the civilian's safety (*AOJ21: Groups that Pose the Greatest Threat to Your Security*). The type of measurement for both of these variables is categorical, the discrete values in the dataset for each variable is just a factor representing a group.

Question 10. What data will you use to test your hypothesis?

I plan on using data from the *Latin American Public Opinion Project (LAPOP), 2004-2015 [28 COUNTRIES]* (ICPSR 36562) dataset available through the Inter-university Consortium for Political and Social Research (ICPSR) at the following link: <https://www.icpsr.umich.edu/web/ICPSR/studies/36562/summary>

Potential plan for Analysis: At the moment I could see myself doing a crosstab, then a scatterplot with a regression line, and then including a regression table with different models controlling for variables such as gang presence in neighborhood (*AOJ17: Gang Presence in Neighborhood*). I'd like to do something a bit more advanced, but unsure how to proceed.

```
getwd()
```

```
## [1] "/Users/gonzalorochoa-vazquez/Desktop/POLI 170A/Problem Set 1"
```

```
setwd("/Users/gonzalorochoa-vazquez/Desktop/POLI 170A/ICPSR_36562/DS0001")
load(file = "36562-0001-Data.rda")
```

The screenshot shows the RStudio interface. The main window displays a data table with the following columns: PAIS (Country), WAVE (Survey Wave), YEAR (Year), IDNUM (Questionnaire Number), IDNUM_14 (idnum_14 Questionnaire Number), ESTRATOPRI (Primary Strata (estratopri)), and ESTRATSEC (Secondary Strata (estratsec)). The table shows 23 rows of data for Argentina in 2008, with various questionnaire numbers and strata assignments.

The Environment pane on the right shows the loaded data object: da36562.0001, which contains 236,754 observations and 266 variables. The variables listed include PAIS, WAVE, YEAR, IDNUM, IDNUM_14, ESTRATOPRI, ESTRATSEC, STRATA, UPM, CLUSTER, UR, TAMANO, IDIOMAQ, FECHA, WT, WEIGHT1500, ODD, X_OR_Y, Q1, and Q2.

Figure 1: Extra Credit