



Module 3 - Session 2 - Data Analysis

Working effectively with data

CivicDataLab

2021/09/16 (updated: 2021-09-17)

Data Analysis Methodologies

What to use when

It can be hard to know which methodology to use when designing a data project. The most important thing for the success of your selection (and equity) is to make sure that your chosen methodology matches the kinds of research questions you have.

Read more [here](#)

Access the matrix [here](#)

 WE ALL COUNT
project for equity in data science

We All Count Methodology Matrix

	WHO	WHAT	WHERE	WHEN	WHY
Select an option in each column -->	-	-	-	-	-
Bayesian Latent Variables	A, B, E	A, B	A, B	A, B, C	C
Bayesian Net Analysis	A, B, C, D	A, B	A, B	A, B, C	C
Collaborative Yarning	A, B, C, D, E	A, B, C	A, B	A, B, C	A, B, C, E, F
Convergent Parallel Design	B, D, E	A, B	A	B, C	C, D
Difference in Difference	B, D, E	A	A	B	D, E
Directed Acyclic Graph Analysis (DAG)	A, B, C, D	A, B	A	A	C, E, F
Focus Groups	A, B, C, D, E	A, B, C	A, B	A, B	A, B, C, F
Key Informant Interviews	A, B, C, D, E	A, B, C	A, B	A, B	A, B, C, F
Latency-Constructed Counterfactual	B, E	A, B	A	A, B, C	C, D

Methodology Matrix - Link

Reading empirical legal research

Structure of a journal article

Structure of a Journal Article

- 1) Title
- 2) Keywords
- 3) Abstract
- 4) Introduction
- 5) Experimental
- 6) Results and Discussion
 - Tables, Figures
- 7) Summary/Conclusions
- 8) References

Surveying a paper

One of the ways

1. Read the title and keyword

Surveying a paper

One of the ways

1. Read the title and keyword
2. Read the abstract

Surveying a paper

One of the ways

1. Read the title and keyword
2. Read the abstract
3. Jump to the conclusions

The first 3 steps will help you decide, if reading the paper in detail is worth it or not!!

Reading the paper

1. Look at the tables and figures (including captions). *Might be helpful before we dig into the details which will slow down the reading.*

Reading the paper

1. Look at the tables and figures (including captions). *Might be helpful before we dig into the details which will slow down the reading.*
2. Read the Introduction section *Background information, why the study was done..*

Reading the paper

1. Look at the tables and figures (including captions). *Might be helpful before we dig into the details which will slow down the reading.*
2. Read the Introduction section *Background information, why the study was done..*
3. Read the Results and Discussion section

Reading the paper

1. Look at the tables and figures (including captions). *Might be helpful before we dig into the details which will slow down the reading.*
2. Read the Introduction section *Background information, why the study was done..*
3. Read the Results and Discussion section

If the research is really relevant then

1. Read the experimental section - *The How's, to understand what was done to better understand the meaning of data and its interpretation..*

Other tips



1. Develop a system of note-taking that is easy to access.

Other tips

1. Develop a system of note-taking that is easy to access.
2. Learn about [Notion](#) , [Obsidian](#) or a bunch of other documentation tools that suit your needs.

Other tips

1. Develop a system of note-taking that is easy to access.
2. Learn about [Notion](#) , [Obsidian](#) or a bunch of other documentation tools that suit your needs.

The faintest writing is better than the best memory

Analysis - Deep Dive

SocioEconomic Indicators and Case Pendency



Does economic growth reduces case pendency ?

SocioEconomic Indicators and Case Pendency

Does economic growth reduces case pendency ?

Objective

Explore how *socioeconomic indicators* affect

1. Proportion of cases pending for more than 10 years
2. Proportion of pending cases classified as criminal
3. Proportion of pending cases filed by women
4. Proportion of pending cases filed by senior citizens

SocioEconomic Indicators and Case Pendency



Does economic growth reduces case pendency ?

Objective

Explore how *socioeconomic indicators* affect

1. Proportion of cases pending for more than 10 years
2. Proportion of pending cases classified as criminal
3. Proportion of pending cases filed by women
4. Proportion of pending cases filed by senior citizens

Data Sources

1. National Judicial Data Grid (NJDG) - Civil and criminal cases filed and pending
2. Demographic information from the annual population Census of India, 2011
3. GDP data from Planning Commission and MOSPI

Features



1. Final dataset covers a total of 496 districts over 25 States

Features

1. Final dataset covers a total of 496 districts over 25 States
2. Socioeconomic and structural factors include:
 1. Judicial factors - number of judges per court and courts per district
 2. Economic factors - per capita GDP, degree of urbanisation, household assets ownership, proportion of working population
 3. Social factors - age, education, literacy levels, social identities based on caste, religion and gender of litigants.

Features

1. Final dataset covers a total of 496 districts over 25 States
2. Socioeconomic and structural factors include:
 1. Judicial factors - number of judges per court and courts per district
 2. Economic factors - per capita GDP, degree of urbanisation, household assets ownership, proportion of working population
 3. Social factors - age, education, literacy levels, social identities based on caste, religion and gender of litigants.
3. The study only traces the association between socioeconomic factors and proportion of pendency and does not indicate any causal relationships between them.

Features

1. Final dataset covers a total of 496 districts over 25 States
2. Socioeconomic and structural factors include:
 1. Judicial factors - number of judges per court and courts per district
 2. Economic factors - per capita GDP, degree of urbanisation, household assets ownership, proportion of working population
 3. Social factors - age, education, literacy levels, social identities based on caste, religion and gender of litigants.
3. The study only traces the association between socioeconomic factors and proportion of pendency and does not indicate any causal relationships between them.
4. Uses Cross sectional data (*Data collected at one point in time*) instead of panel data (*Data collected over time*)

Findings



Indicator	Demography	Economy	Judicial
Proportion of cases pending over 10 years	x	-	-
Proportion of criminal cases	x	+	+
Proportion of cases filed by women	+	x	-
Proportion of cases filed by senior citizen	+	-	x

Methodology

Primary Method - **Ordinary least square regressions**

For checking the robustness of selected methodology:

1. Logarithmic Transformations
2. Heteroskedasticity
3. Cooks distance for treating outliers

Dataset and Other Resources



1. Research Paper - [Link](#)
2. Datasets - [Link](#)
3. Author Details and Paper Summary - [Link](#)
4. Journal of Indian Law and Society - [Link](#)

Jobs for Justice



Whether judges respond to **pandering incentives** by ruling in favour of the government in the hope of receiving jobs after retiring from the Court ?_

Jobs for Justice



Whether judges respond to **pandering incentives** by ruling in favour of the government in the hope of receiving jobs after retiring from the Court ?_

Variables that influence incentives

1. The importance of a case
2. Whether the judge retires with enough time left in a government's term (at-least 47 weeks) to be rewarded with a prestigious job
3. Authoring judgements in favour of the government

Data Sources

1. All reported **Supreme Court of India cases** involving the government from 1999 till 2014
 1. Total 2,605 cases
 2. 941 cases with judgements
 3. 742 cases decided by a 2 judge bench (random allocation)
 4. Both judges retire before March 2015 - 687 cases
 5. One of the two judges write a judgement - 681 cases
 6. Decision was unambiguously for or against the government - 652 cases

Data Sources

1. All reported **Supreme Court of India cases** involving the government from 1999 till 2014

1. Total 2,605 cases
2. 941 cases with judgements
3. 742 cases decided by a 2 judge bench (random allocation)
4. Both judges retire before March 2015 - 687 cases
5. One of the two judges write a judgement - 681 cases
6. Decision was unambiguously for or against the government - 652 cases

2. **Data on Judges**

1. Date of appointment
2. Date of retirement
3. Date of elevation to CJI
4. Date of birth

Data Sources

1. All reported **Supreme Court of India cases** involving the government from 1999 till 2014

1. Total 2,605 cases
2. 941 cases with judgements
3. 742 cases decided by a 2 judge bench (random allocation)
4. Both judges retire before March 2015 - 687 cases
5. One of the two judges write a judgement - 681 cases
6. Decision was unambiguously for or against the government - 652 cases

2. **Data on Judges**

1. Date of appointment
2. Date of retirement
3. Date of elevation to CJI
4. Date of birth

3. **Jobs Data** - Collected using Gazette of India, Newspaper reports

1. Position
2. Date of appointment

Features

1. Examining how judicial decisions are affected by factors unrelated to legal reasoning

Features

1. Examining how judicial decisions are affected by factors unrelated to legal reasoning
2. Identify and measure corruption at an aggregate institutional level

Features

1. Examining how judicial decisions are affected by factors unrelated to legal reasoning
2. Identify and measure corruption at an aggregate institutional level
3. Identify the **causal effect** of career-concern incentives on judicial decision-making

Findings

We observe that the government has a **14-27% chance of winning a case** that is one standard deviation higher than mean importance that is decided by a bench with both judges retiring shortly before an election. Our estimates imply that the probability of the government **winning such a case more than doubles** when it is instead decided by a bench with both judges retiring long before an election.

Methodology

1. Statistically identify (using OLS) the presence of pandering by comparing benches composed of judges who have stronger incentives to pander to those who have weaker incentives to pander

Methodology

1. Statistically identify (using OLS) the presence of pandering by comparing benches composed of judges who have stronger incentives to pander to those who have weaker incentives to pander
2. Construct an index for measuring the importance of a case by identifying the number of:
 1. Attorneys General
 2. Solicitors General
 3. Senior Advocates, and
 4. Advocates that appeared in the case

Methodology

1. Statistically identify (using OLS) the presence of pandering by comparing benches composed of judges who have stronger incentives to pander to those who have weaker incentives to pander
2. Construct an index for measuring the importance of a case by identifying the number of:
 1. Attorneys General
 2. Solicitors General
 3. Senior Advocates, and
 4. Advocates that appeared in the case
3. Treatment and Control groups
 1. Benches with two judges retiring long before an election as the **high treatment group**"
 2. Those with just one judge retiring long before an election as the **low treatment group**
 3. Those with both retiring shortly before an election as the **control group**

Dataset and other resources

1. Research Report - [Link](#)
2. Authors website - [Link](#)
3. Paper summary - [IdeasofIndia](#)
4. Media Coverage - [ThePrint](#)

Tools

Documenting Qualitative Info



Features

1. Open Source
2. Data collection, exploration and analysis at one place
3. Can be used to store much more rich information, like reports, articles (full-text), audio/video files for every entry
4. Securely store, organise, and publish document collection to make the stories within it more accessible and understandable to a wider audience
5. Establish relationships between responses/entities

UWAZI Use-Cases



1. Using Uwazi to improve access to decisions issued by human rights courts and commissions - [Link](#)

UWAZI Use-Cases



1. Using Uwazi to improve access to decisions issued by human rights courts and commissions - [Link](#)
2. Database and commentary on jurisprudence of the African Court on Human and Peoples' Rights - [Link](#)

UWAZI Use-Cases



1. Using Uwazi to improve access to decisions issued by human rights courts and commissions - [Link](#)
2. Database and commentary on jurisprudence of the African Court on Human and Peoples' Rights - [Link](#)
3. Pakistan's Death Penalty Database - [Link](#)

UWAZI Use-Cases



1. Using Uwazi to improve access to decisions issued by human rights courts and commissions - [Link](#)
2. Database and commentary on jurisprudence of the African Court on Human and Peoples' Rights - [Link](#)
3. Pakistan's Death Penalty Database - [Link](#)

More info [here](#)

Resources

1. [Putting data back into context](#)
2. Learning How to Learn: Powerful mental tools to help you master tough subjects - [Link](#)
3. [How to Read a Paper Efficiently \(By Prof. Pete Carr\)](#)
4. [How to lie with indices](#)

Queries and Feedback