PLSC 473: American Judicial Behavior

Christopher Zorn

September 15, 2015

Where Do Data Come From?

- Pre-Collected / Other Researchers
- Government
- Non-Profits / NGOs / etc.
- Corporations / Private Sector
- Collect it yourself
    - Experiments
    - "By hand"
    - Surveys (in-person/phone/web/MTurk)
    - Automated Methods (web scraping, APIs, etc.)

How Much Data?

- E&M: "as much as resources and time allow."
  - Best: *All of it.*
  - Second-best: A lot.
- Data Collection & *Selection Bias*
  - Entire "population" = no selection bias
  - Anything less = "sample"
  - **The only sampling strategy that doesn't lead to bias is random sampling**
  - Can be *simple random*, *stratified*, *clustered*

Coding

- Turning values into numbers...
- Key considerations:
    - Level of Measurement ("NOIR")
    - Categories:
        1. Mutually Exclusive
        2. Exhaustive
        3. Err on the side of *more information* (more categories, etc.)

The Codebook

- Description of sampling frame / universe of cases, sampling process, etc.
- Lists each variable's possible codes and their associated categories
- Also includes notes...
- **Goal: Source material + codebook = recreated data**

Missing Data

- When values aren't present...
- Why?
  - Impossible
  - Data unavailable
  - Nonresponse...
- Types of Nonresponse
  - Item (that question / measure)
  - Unit (that individual / unit)
  - Cluster / group
- Recorded in various forms.. (e.g., `NA`)
- Analysis: Often use *listwise deletion*

Data Structure

- *Cross-Sectional*
- *Time Series*
- *Time-Series Cross-Sectional* ("Panel")
- *Multilevel* ("Nested")
- *"Relational"*
  - "Dyads," "Triads," etc.
  - Units may be the same or different
  - Network structure

Example: Baseball Data (1997)

```
> Baseball <- read.csv("Class Data/Baseball.csv")
> head(Baseball,10)
   respon DH_appr age PID5 female followbaseball
1       1      NA  65    2 Female              0
2       2       1  63    1   Male              1
3       3      NA  56    1 Female              1
4       4      NA  24   NA Female              0
5       5      NA  47    5   Male              0
6       6      NA  81    5 Female              1
7       7       1  28    1   Male              1
8       8       0  76    1   Male              1
9       9      NA  22    2 Female              0
10     10      NA  39    1 Female              0
```

Example: SCOTUS Clerk Data (1953-2004)

```
> Clerks <- read.csv("Class Data/AnnualClerks.csv")
> head(Clerks,10)
   Term female white  top5law   lcclerk
1  1953      0   100 44.44445 12.500000
2  1954      0   100 64.70589 44.444450
3  1955      0   100 76.47059 41.666660
4  1956      0   100 55.55556 20.000000
5  1957      0   100 58.82353 30.000000
6  1958      0   100 57.89474 27.272730
7  1959      0   100 61.11111 44.444450
8  1960      0   100 66.66667  7.142858
9  1961      0   100 55.55556 21.428570
10 1962      0   100 71.42857 21.428570
```

Example: Country-Year (TSCS) Data

```
> TSCS <- read.csv("Class Data/CountryTSCS.csv")
> head(TSCS,10)
       country year ccode polity gdppc coldwar region
1  AFGHANISTAN 1946   700    -10    NA       1      6
2  AFGHANISTAN 1947   700    -10    NA       1      6
3  AFGHANISTAN 1948   700    -10    NA       1      6
4  AFGHANISTAN 1949   700    -10    NA       1      6
5  AFGHANISTAN 1950   700    -10   119       1      6
6  AFGHANISTAN 1951   700    -10   143       1      6
7  AFGHANISTAN 1952   700    -10   149       1      6
8  AFGHANISTAN 1953   700    -10   157       1      6
9  AFGHANISTAN 1954   700    -10   156       1      6
10 AFGHANISTAN 1955   700    -10   154       1      6
```

Example: Country-Year (TSCS) Data (cont'd)

```
> tail(TSCS,12)
      country year ccode polity gdppc coldwar region
7305 ZIMBABWE 1988   552     -6  2170       1      4
7306 ZIMBABWE 1989   552     -6  2357       1      4
7307 ZIMBABWE 1990   552     -6  2581       0      4
7308 ZIMBABWE 1991   552     -6  2811       0      4
7309 ZIMBABWE 1992   552     -6  2662       0      4
7310 ZIMBABWE 1993   552     -6  2518       0      4
7311 ZIMBABWE 1994   552     -6  2663       0      4
7312 ZIMBABWE 1995   552     -6  2695       0      4
7313 ZIMBABWE 1996   552     -6  2891       0      4
7314 ZIMBABWE 1997   552     -6  3153       0      4
7315 ZIMBABWE 1998   552     -6  3089       0      4
7316 ZIMBABWE 1999   552     -6    NA       0      4
```

Example: "Dyadic" Country Data (1968)

```
> Dyads <- read.csv("Class Data/Dyads1968.csv")
> head(Dyads,150)
    dyadid ccode1 ccode2 dem1 dem2 distance allies
1     2020      2     20   10   10        0      1
2     2040      2     40   10   -7     1135      0
3     2041      2     41   10   -9     1437      1
4     2042      2     42   10   -3     1477      1
5     2051      2     51   10   10     1446      0
.
.
.
126   2840      2    840   10    5     8570      1
127   2850      2    850   10   -7    10172      0
128   2900      2    900   10   10     9916      1
129   2920      2    920   10   10     8759      1
130  20040     20     40   10   -7     1586      0
131  20041     20     41   10   -9     1869      0
132  20042     20     42   10   -3     1893      0
133  20051     20     51   10   10     1897      0
134  20052     20     52   10    8     2547      0
.
.
.
```

Codebook: Supreme Court Judicial Database

Supreme Court Database Code Book
brick_2015_01

CONTRIBUTORS

Harold Spaeth
Michigan State University College of Law

Lee Epstein
Northwestern University

Ted Ruger
University of Pennsylvania School of Law

Codebook: Supreme Court Judicial Database

The Supreme Court Database Codebook

17 Manner in which the Court takes Jurisdiction

Variable Name	Spaeth Name	Normalizations
jurisdiction	JUR	varJurisdiction (13)

The Court uses a variety of means whereby it undertakes to consider cases that it has been petitioned to review. These are listed below. The most important ones are the writ of certiorari, the writ of appeal, and for legacy cases the writ of error, appeal, and certification.

- End of Content for Variable 17. Manner in which the Court takes Jurisdiction -

Codebook: Supreme Court Judicial Database

A15 varJurisdiction

13 Distinct Values

varJurisdiction is used in conjunction with:
jurisdiction

Values:

1	cert
2	appeal
3	bail
4	certification
5	docketing fee
6	rehearing or restored to calendar for reargument
7	injunction
8	mandamus
9	original
12	stay
13	writ of error
14	writ of habeas corpus
15	unspecified, other