Finding & Using Data & Statistics

A brief introduction to Empirical Legal Research

For further assistance with ELR, contact:
Sarah Ryan, YLS Empirical Research Librarian
sarah.ryan@yale.edu

For the birds... Images of birds in art are taken from the Yale University image collection, see Bibliography
Contents

1. Data, data... everywhere (and statistics too)
2. Data versus statistics
3. How to find legal data and statistics, quickly
4. How to find data sets, for long-term study
5. How to calculate statistics
6. Data, statistics, and empirical legal research (ELR)
7. How to get help with ELR

Throughout the show, click suns or blue text for more information!
Data, data... everywhere (and statistics too)

Data and statistics are all around us...

1. We all analyze data and perform statistical operations (e.g., calculating grades).

2. Statistics play a key role in the law, from average billable hours per client to parts per million of PCBs to percentage of a certain group denied employment.
Data, data... everywhere (and statistics too)

The problem with the ubiquity of data and statistics is that they are often used in misleading and irresponsible ways... Consider the following:

**Claim:** African Americans favor lower taxes

**Data:** Survey of 1,500 taxpayers in Pittsburgh, who were called at home at 5pm. 3 participants were African American. 2 of those 3 “strongly favor lower taxes.” (2 out of 3 = 66%)

**Statistic:** 66% of African Americans reported that they “strongly favor lower taxes”
Data, data... everywhere (and statistics too)

As the obviously faulty “African Americans favor lower taxes...” example illustrates, data and statistics can be used in mathematically precise but illogical and unethical ways. Social scientists are therefore vigilant about issues such as sample size (how many members of the population are surveyed), sample composition (how many people of each race, age, etc. participate), survey research methods (e.g., what time people are called), and more. Concomitantly, a trustworthy researcher always discloses how data were collected and statistics were calculated.
Data are numbers that have not been analyzed. If you download a “data set” it will contain rows and columns of data. In social science research, each row typically represents a person. So, row 9 might be Carl’s survey answers, row 10 Claribel’s. Columns represent survey questions or items, etc. Survey answers are transformed into numeric variables in 3 steps...
3. I find the cafeteria’s food appetizing.
   a. Strongly agree
   b. Agree
   c. Neutral
   d. Disagree
   e. Strongly disagree

Step 1: A question or statement is presented in a survey or other instrument.

Step 2: It is given a variable name (e.g., CAFFOOD), which is recorded in a codebook, and the answer choices are numbered (e.g., Agree=4).

CAFFOOD
3. I find the cafeteria’s food appetizing.
   5= Strongly agree
   4= Agree
   3= Neutral
   2= Disagree
   1= Strongly disagree

Step 3: Individual answers are recorded as numbers (e.g., Adele “agrees”=4), in the variable column (CAFFOOD). (Participant names are often removed to protect privacy and minimize bias.)
Statistics are processed data. That is, someone has used mathematics to process or analyze the data. For instance, an average (i.e., mean) is a statistic that involves two mathematical steps (add up the numbers, divide by how many numbers there are).
Oftentimes, rather simple statistical calculations will yield the answer to a question. Medians (e.g., median, or mid-point, income) can tell compelling stories about what life is like for those “in the middle”... Basic statistics are called “descriptive statistics.”

Correlation and causality claims require advanced statistics that permit inferences or projections...
How to Find Legal Data and Statistics, Quickly

If you want to make a quick argument, you typically want to find a statistic (e.g., 47% of incarcerated people are...). Some great places for law-related statistics are...

1. Department of Justice/Bureau of Justice Statistics. Click here for general statistics and here for economic data.
3. U.S. Census Bureau. Click here for data and here to calculate statistics online using the Bureau’s “Data Ferrett.”
How to Find Data Sets, for Long-term Study

Thousands of data sets are available online...The trick is finding the data you need amidst the numerical “haystack.” The following process can help...

1. Articulate a hypothesis or research question
2. Underline the key variables (e.g., race, incarceration rates)
3. Specify your research aim (e.g., test a relationship), which statistical tests match your aim, and what sorts of data you’ll need
4. Search for data in a data clearinghouse (e.g., ICPSR), then...
5. Determine which agencies, organizations, and/or individuals might have been motivated to collect the data you’re seeking
6. Ask for help from a data librarian
How to Find Data Sets, for Long-term Study

Example

1. RQ: Do Native Americans receive longer sentences for non-violent crimes than others...

1. Native Americans sentences for non-violent crimes...? I am seeking sentencing data... on nonviolent crimes... by race, ethnicity, or group (e.g., a ‘race’ column)

3. I want to compare ethnic groups to each other. I might like to conduct an ANOVA using the ☀ variables race and sentencing.

4. Search ICPSR

Tip: Perform very simple searches of variable names in the “Search/Compare Variables” page/box and then use the Compare button to compare data sets. Try it now! – Click
How to Find Data Sets, for Long-term Study

4. Brainstorm (or Google search) who might have collected data:
   - U.S. Sentencing Commission
   - U.S. Department of Justice
   - The Sentencing Project
   - The ACLU
   - State Orgs., Agencies
   - Legal Scholars

5. Ask a librarian... “I’ve found a description of data on racial disparities in sentencing collected by the ACLU, but I can’t locate the actual data set...”
How to calculate statistics

Calculating statistics involves 5 steps:

1. Articulate a hypothesis (H) or research question (RQ)
2. Determine what kind of data you need/have
3. Match-make your H/RQ and the kind of data you need/have with the appropriate statistical operation
4. Perform the statistical operation
5. Analyze the results

H: There is an inverse relationship between income and % of income tax favored among CT residents.
How to calculate statistics

2. Determine what kind of data you need/have

Data come in three basic forms or “levels of measurement”

1. **Nominal** – data whose numbers don’t mean anything in terms of “more or less,” or data whose numbers are just place-holders. For example, race can be recorded: 0=African American, 1=Asian, 2=Caucasian, 3=Latino... Latinos aren’t “worth” 3 times as much as Asians. Rather, a 3 just tells the researcher that the person is Latino; “Latino” is alphabetically after “Asian” and thus receives a higher number.

2. **Ordinal** – data whose numbers signal more or less but not an exact amount of more or less. For example, 5=strongly agree, 4=agree, 3=Neutral, 2=Disagree, 1=Strongly Disagree. Strongly agree is stronger than neutral and much stronger than strongly disagree, but we wouldn’t say strongly agree is 5 times more strong than strongly disagree.

3. **Interval-ratio** – data whose numbers actually indicate how much more or less. For example, an income of $50,000 is exactly $10,000 more than an income of $40,000.
How to calculate statistics

2. Determine what kind of data you need/have

Survey on Attitudes Regarding Connecticut Income Tax

1. Do you reside in Connecticut?
   ___ no   ___ yes

2. Are you employed in Connecticut?
   ___ no   ___ yes

If you answered no to questions 1 and 2, you are finished!

3. How would you feel about a 1% income tax?
   ___ strongly oppose   ___ oppose   ___ neutral   ___ favor   ___ strongly favor

4. How would you feel about a 2% income tax?
   ___ strongly oppose   ___ oppose   ___ neutral   ___ favor   ___ strongly favor

5. How would you feel about a 5% income tax?
   ___ strongly oppose   ___ oppose   ___ neutral   ___ favor   ___ strongly favor

6. What is your individual annual income? $_____________________
How to Calculate Statistics

3. Match-make your H/RQ and the kind of data you have with the right statistical operations

Different “levels” of data permit different statistical operations. For example, nominal data permit only basic sorts of counting/calculating. It would be nonsensical to take an average of nominal data. If you had a room with 9 women and 1 man, you wouldn’t say that the average person is 10% female.

So, the trick is to figure out what sorts of data you have (e.g., race=nominal, years of criminal sentence=interval-ratio), what you want to achieve or prove (e.g., describe the data, test a causal claim), and then match-make the data and goal with an operation. Go here for help: UCLA “What statistical analysis should I use?”
How to Calculate Statistics

4. Perform the statistical operations

Few researchers perform calculations by hand these days. Just as most folks write documents in Word, etc., *most statisticians use a program to calculate*. The four most popular programs among empirical legal scholars seem to be: Excel, R, SPSS, and Stata. Excel is the easiest to learn but has the fewest functions (e.g., least math, fewer graphing options); the rest are harder to learn (e.g., because they operate best using shorthand commands such as “regress” for “perform a regression”) but offer greater functionality.

**All require the researcher to already know what math s/he wants to use**... data must be matched with operations first.
How to Calculate Statistics

5. Analyze the results

Statistical programs generate mathematical output that must be analyzed. For example, if a researcher wanted to know if the brain weight of an animal was related to its body weight (i.e., prediction: bigger creatures have bigger brains), she might perform a regression. ☀ She would pay particular attention to the R-squared statistic, which suggests how strong the prediction is...

Learn more: http://people.duke.edu/~rnau/rsquared.htm
While the practice of law has always required the use of data (e.g., acres, heads of cattle, debt:assets), lawyers have often outsourced more complicated statistical work to social scientists (e.g., economists). Increasingly, lawyers and legal scholars are taking control/ownership of statistical calculation.

As YLS professors such as Ian Ayres, and others, have argued, data is a commodity and statistical “crunching” is the “new way to be smart” (and successful) in commerce, public service... and the law.
Within the legal academy, The Society for Empirical Legal Scholarship (SELS) brings together diverse legal empiricists. SELS publishes a journal (JELS) and hosts a conference (CELS)...

If you are interested in statistical research and practice, SELS, JELS, and CELS have a lot to offer...
Data, Statistics, and Legal Research

The Conference on Empirical Legal Studies (CELS) showcases cutting-edge statistical scholarship, data sets, and statistical strategies.

Learn more:
Bibliography

Recommended readings


3. Acock, Alan C. (2012). *A gentle introduction to Stata* (3rd ed.). College Station, TX: StataCorp. YLS Call #: HA32.A26

Bibliography

Bird images in order of appearance

1. http://digitalcollections.library.yale.edu/138399.jpo?qdl0=creator&qdlv0=Desportes%2c+Alexandre-Francois%2c+1661-1743&qdlt0=contains&qqid=1147033
2. http://digitalcollections.library.yale.edu/138398.jpo?qqid=1147033
3. http://digitalcollections.library.yale.edu/96774.jpo?q=girl+with+bird&qqid=1147051
4. http://digitalcollections.library.yale.edu/2148380.jpo?q=birds&qs=1225&qqid=1102552
5. http://digitalcollections.library.yale.edu/1905214.jpo?q=birds&qqid=1102552
7. http://digitalcollections.library.yale.edu/450853.jpo?q=birds&qqid=1102552
8. http://digitalcollections.library.yale.edu/134029.jpo?q=birds&qqid=1102552
10. http://digitalcollections.library.yale.edu/433929.jpo?q=birds&qqid=1102552
11. http://digitalcollections.library.yale.edu/132095.jpo?q=birds&qs=721&qqid=1102552
14. http://digitalcollections.library.yale.edu/search.dl?q=birds&qqid=1102552&qs=2233
15. http://digitalcollections.library.yale.edu/476608.jpo?q=birds&qs=721&qqid=1102552
17. http://digitalcollections.library.yale.edu/2064410.jpo?q=birds&qs=721&qqid=1102552
18. http://digitalcollections.library.yale.edu/27296.jpo?q=birds&qs=1081&qqid=1102552
19. http://digitalcollections.library.yale.edu/3536.jpo?q=birds&qs=73&qqid=1102552
20. http://digitalcollections.library.yale.edu/439936.jpo?q=birds&qs=433&qqid=1102552