I. Course Description

The goal of this course is to learn about the methods used to collect publicly available survey data that can be used for policy research so that students can appropriately use these data to answer policy relevant questions. Students will learn about the methods used to collect survey data, how to develop researchable policy questions that can be answered with the survey data, and about the limitations of the survey data for answering policy research questions. In order to analyze policy questions using available survey data, students will also learn about actual survey instruments, survey sample designs, survey data processing, and survey data systems that the major public policy relevant surveys use. The course will also examine specific measurement and analysis issues that are of interest to policy research (e.g., measuring public program enrollment and public program eligibility simulation). By the end of the course each student will understand the methods used to collect survey data, have developed a researchable policy question, carried out the appropriate analysis to answer the question, produced high quality analytical tables, and written up descriptions of the methods used to produce the numbers in the tables in a style that is consistent with professional policy research.
Who should take this course:

This course is targeted to students interested in obtaining skills for developing policy-relevant research questions that can be answered by using existing survey data for empirical evidence. We will use a variety of publicly available survey datasets, including the decennial census, the Current Population Survey, the American Community Survey, the General Social Survey, the Behavioral Risk Factor Surveillance System, and the National Health Interview Survey to illustrate important issues in demographic data collection and analysis. Students may use this course to develop analytic projects that will assist them in meeting Master's or Ph.D. research requirements. The course expands on required statistics and research methods courses within various disciplines (sociology, political science, public policy, and public health among others). The tools learned in basic statistics and research methods are essential but not sufficient to work with the complex realities of using existing survey data resources to conduct policy relevant analyses.

II. Course Prerequisites

Graduate level research methods course, basic graduate level statistics course, or permission of instructor.

III. Course Goals and Objectives

Specific Skills Students Will Acquire

1. Develop a better understanding of survey methodology.
2. Learn to ask research questions that can be answered with available survey data.
3. Understand the strengths and limitations of using survey data to answer policy research questions.
5. Produce publication quality survey methodology descriptions and tables.

IV. Methods of Instruction and Work Expectations

The class will be run with a mixture of lecture, discussion and computer lab work (about 60% lecture, 25% hands on lab work, and about 15% group discussion). All three are essential for students to gain the skills and knowledge necessary to work with publicly available demographic data. Students will learn directly from the instructors, TA and from one another how to solve their analysis problems.

V. Software, Readings, Course Text

Software

INSTRUCTIONS FOR ACCESSING the UCHICAGO VLAB COMPUTERS

This document provides instructions for accessing vlab.uchicago.edu, which provides Stata, R, SPSS and ArcGIS for student use. To access the server, you will need the Chicago VPN software (if working off campus) and a Remote Desktop Connection to the lab.

Chicago VPN

cVPN is the University's Virtual Private Network (VPN) concentrator that provides secure access to University network resources as if you were on campus, no matter where you are in the world. cVPN is necessary to access restricted resources, such as the Virtual Lab, when you are working off campus.

The Cisco AnyConnect VPN software can be downloaded from https://cvpn.uchicago.edu and installed on Windows, Apple, or Linux systems. Once installed, launch Cisco AnyConnect Secure Mobility Client. You will connect to cvpn.uchicago.edu using your CnetID and password to authenticate.

Remote Desktop Connection (RDC)
On Windows systems, RDC is part of the standard install. You should be able to search for the application in the search window. A sample configuration for connecting to the vlab is shown below, note that ADLOCAL\username is ADLOCAL\your-cnetID.

On Apple systems, you will need to download Microsoft Remote Desktop from the Apple Store. After it is installed, launch the Application and create a “New” remote desktop. The standard configuration is shown below.
Once you have configured your session, click start and your session should begin. Subsequently, when you start Microsoft Remote Desktop, you will see your connection displayed on the opening screen and you can simply click on the session name.

If you are using a **linux operating system**, your system should have the rdesktop application installed already. You need to open a terminal and run the following command:

```
rdesktop -g 1024x768 vlab.uchicago.edu
```

When you have connected, you will be prompted for your CnetID and password. For example:

Login: ADLOCAL\fbooker
Password: **************

---

**Note about ado files:**
Currently, students cannot install ado files in the Stata folder(s) or run updates; however, they can install files in the PLUS or PERSONAL folder if they redirect those locations to their U: drive, e.g.,

```
sysdir set PLUS "u:\rado"
```
or
sysdir set PERSONAL "u:\ado"
You need to change the relevant sysdir each time you want to use the ado.

* When finished, don't forget to logoff. Click the Start button and select Log off

Commented [MD1]: Need to verify the above material is up to date with Fay and make sure we have Apple computers working better now (or if there are any tricks make sure to get them from Fay). Also can you check on the best way to do the "ado" files based on last year’s experience?
**Course Readings:** Most of the required readings are available on the Internet or will be handed out.

**Course Reference Materials Available On-Line (alphabetical by author)**


**Course Reference Materials Available From Instructor**


Supplemental Readings


<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Class Topics</th>
<th>Due at the Start of Class</th>
<th>Readings and Course Materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/10/2019</td>
<td>(1) Syllabus (2) Review list of public data sources (3) Fitting research questions to existing data (4) basic background on US federal survey data MD (FL)</td>
<td>Review Links to Survey Data handout</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1/17/2019</td>
<td>Introduction to surveys, survey quality and survey error Complex survey sample designs, sampling error, variance estimation and survey data analysis MD</td>
<td>Group List. One person per group submit the list; If not in a group one will be assigned by Instructor</td>
<td>Federal Committee Statistical Methodology (FCSM) (2001): Chapter 1 and FCSM Chapter 3</td>
</tr>
<tr>
<td>2</td>
<td>Lab 1</td>
<td>STATA basics, accessing data, creating usable data files, data management; Show IPUMS site and structure of the files</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1/24/2019</td>
<td>Unit level non-response, coverage error and response rates FL</td>
<td>Lab 1; 1st group paper due (1a)</td>
<td>FCSM (2001) Chapters 4 and 5 and Groves (2006) paper</td>
</tr>
<tr>
<td>3</td>
<td>Lab 2</td>
<td>Additional data management and basic processing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lab</td>
<td>Date</td>
<td>Description</td>
<td>Lab</td>
<td>References</td>
</tr>
<tr>
<td>-----</td>
<td>------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>4</td>
<td>1/31/2019</td>
<td>Complex Sample design and determining a necessary sample size for a survey</td>
<td>Lab 2</td>
<td>Davern et al 2006, Boudreaux et al (2012)</td>
</tr>
<tr>
<td>5</td>
<td>2/7/2019</td>
<td>Survey estimates basics in STATA</td>
<td>Lab 3; Out of Class Assignment Due (Sample Size)</td>
<td>FCSM (2001) Chapter 6</td>
</tr>
<tr>
<td>5</td>
<td>Lab 4</td>
<td>Tabling STATA output</td>
<td>Lab 3</td>
<td>FCSM (2001) Chapter 6</td>
</tr>
<tr>
<td>6</td>
<td>2/14/2019</td>
<td>Measurement Error in Surveys (cont’d)</td>
<td>Lab 4; 2nd group paper due (1b)</td>
<td>Idler and Benyamini, 1997; Miltumalo, Seppo, et al., 1997; Klerman et al., 2009; see SHADAC 2017</td>
</tr>
<tr>
<td>6</td>
<td>Lab 5</td>
<td>Survey means and t-tests in STATA</td>
<td>Lab 5</td>
<td>FCSM (2001) Chapters 6 and 7</td>
</tr>
<tr>
<td>7</td>
<td>2/21/2019</td>
<td>Processing Error and Total Survey Error</td>
<td>Lab 5</td>
<td>FCSM (2001) Chapters 6 and 7</td>
</tr>
<tr>
<td>7</td>
<td>Lab 6</td>
<td>STATA survey regressions procedures</td>
<td>Lab 6</td>
<td>Glied, Remler, and Zivin, 2002 (and commentaries).</td>
</tr>
<tr>
<td>8</td>
<td>Lab 7</td>
<td>STATA Replicate weight variance estimation</td>
<td>Lab 7</td>
<td>Couper 2013, Lazar and Radford 2017, Johnson and Smith 2017</td>
</tr>
<tr>
<td>9</td>
<td>3/7/2019</td>
<td>The future of surveys/Surveys in the context of other data</td>
<td>Lab 7; 3rd group paper due (1c)</td>
<td>Couper 2013, Lazar and Radford 2017, Johnson and Smith 2017</td>
</tr>
<tr>
<td>9</td>
<td>Lab 8</td>
<td>STATA Hotdeck Imputation</td>
<td>Lab 8</td>
<td>Couper 2013, Lazar and Radford 2017, Johnson and Smith 2017</td>
</tr>
<tr>
<td>10</td>
<td>3/14/2019</td>
<td>Group Presentations and Discussion of National Academy Report on Combining Data Sources MD (FL)</td>
<td>Lab 8; One copy of Group presentation which counts as Lab 9;</td>
<td>National Academies of Sciences, Engineering, and Medicine, 2017</td>
</tr>
</tbody>
</table>
VII. Evaluation and Grading

Course Projects

Course projects are to be completed in a group of 3 (one or two groups may have only 2 or as many as 4 members). Students will complete a series of short papers and data tables that present empirical data informing a specific policy-relevant research question using data from IPUMS family of data sets (www.ipums.org) -- which include the Current Population Survey, Census Data (both US and many other countries), the National Health Interview Survey -- the Behavioral Risk Factor Surveillance System (BRFSS) or the General Social Survey. Other datasets can potentially be used but will require instructor permission before use. These papers will build off one another to form the core of an empirical research paper or report. Each paper will be organized and formatted in accordance with the criteria specified in the relevant assignment. Revised versions of all preceding sections are to be turned in with each new section. Upon instructor review of the final assignment, the groups may wish to revise the entire paper and turn it in for further consideration. The goal of this project is to have a fully assembled empirical research paper or report that uses actual analysis of survey data.

Submitting assignments: Please submit assignments in hard-copy in class or on-line using the course Canvas site. Instructions on submitting assignments via Chalk are available here: http://answers.uchicago.edu/page.php?id=20069.

Students will be evaluated as follows:

1. **3 short papers** representing an empirical research study’s components. Each paper should have 1-inch margins, use 11-point Arial or 12-point Times font, and be double-spaced.
   a. Background Paper (3-5 pages) define policy issue, research question, and data set to be used (12%).
   b. Revised assignment (a) and a description of the data set, description of the key variables to be used, and a descriptive statistics table of key variables (12%).
   c. Revised Assignments (a), and (b), a description of the analytic research framework employed, and a table with key analytic statistics and a description of what the findings tell us about the research question (30%).
   d. A revised “final assignment” paper can be resubmitted by March 15th at 1:00 pm if the group wishes to improve upon the initial assessment. Instructors can award up to half of the points deducted on the first submission (for example on submission one if the group received 80/100 points; by rewriting the assignment and turning it back in the group could get up to 10 points more or 90/100 total for the final paper).

2. There will be **9 lab assignments** and the highest eight scores will count toward your final grade (the lowest score will be dropped so only 8 labs count in your final grade) (5% each) and **1 out-of-class assignment** (6%). These assignments will make up a total of 46% of the final grade. Under some circumstances, in-class assignments can be made up if instructors are notified in advance about an absence.
Late work will be docked 10% of the total for each day it is late. Assignments are due by the beginning of class on the date due. Exceptions may be granted by the instructors on a case-by-case basis. The student must contact the instructors in advance of the deadline, and the instructors must agree that the student’s circumstances warrant a new deadline for the student. If you think an excuse may be tenuous, it likely is. Make every effort to turn assignments in on time.

Grading

A/F letter grade will be determined by total effort as follows:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Percentage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>95-100%</td>
<td>(4.0) Represents achievement that is outstanding relative to the level necessary to meet course requirements</td>
</tr>
<tr>
<td>A-</td>
<td>90-94%</td>
<td></td>
</tr>
<tr>
<td>B+</td>
<td>87-89%</td>
<td>(3.0) Represents achievement that is significantly above the level necessary to meet course requirements</td>
</tr>
<tr>
<td>B</td>
<td>83-86%</td>
<td></td>
</tr>
<tr>
<td>B-</td>
<td>80-82%</td>
<td></td>
</tr>
<tr>
<td>C+</td>
<td>77-79%</td>
<td>(2.0) Represents achievement that meets the minimum course requirements</td>
</tr>
<tr>
<td>C</td>
<td>73-76%</td>
<td></td>
</tr>
<tr>
<td>C-</td>
<td>70-72%</td>
<td></td>
</tr>
<tr>
<td>D+</td>
<td>67-69%</td>
<td>(1.0) Achievement below minimum course expectations but sufficient to be awarded credit</td>
</tr>
<tr>
<td>D</td>
<td>63-66%</td>
<td></td>
</tr>
<tr>
<td>D-</td>
<td>60-62%</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>below 60%</td>
<td>Represents failure (no credit) and signifies that the work was either (1) completed at a level of achievement that is not worthy of credit or (2) was not completed and there was no agreement between the instructor and the student that the student would be awarded an I.</td>
</tr>
</tbody>
</table>