Course Title: Identifying and Preparing Data for Health Service Research and Policy Analysis

Course Number: HSAP 0515J: Data for Hlth Svc Res & Pol

Course Location: Public Health – Piscataway Rm 204 2A

Course Date & Time: Tuesday and Thursday 6:10 pm to 9:00 pm

Course Instructor: Instructor: Rizzie Kumar MA

Office: 683 Hoes Lane West, Rm 308
Phone: (732) 235-5009
E-mail: rizzie.kumar@rutgers.edu

Office Hours: By Appointment Only

Required Course Text:


Additional/Supplemental Readings/Resources:

Optional:

Students may want to purchase a one year license of SAS ($100.00)

Please see link below:

http://computer-repair.rutgers.edu/software/ and select https://software.rutgers.edu/

Any problems, the people who can help would be the New Brunswick Helpdesk at (848)445-HELP(4357)

SAS is also available through the Remote X Application Server:

https://oirt.rutgers.edu/software/remotexserver/

Students may want to purchase a USB flash drive to store data and program files.
Free download (for SAS v9.13):


Course Description:

The purpose of this course is twofold. First it will introduce students to methods for identifying and employing relevant data sources for use in health services research and policy analysis. Second, it will provide a comprehensive introduction to data management.

Students will be introduced to several secondary data sources that frequently are used for health services research and policy analysis, and that are publicly available. They will learn the questions to ask when trying to locate appropriate data for a specific research question. Advantages and limitations of several national data resources will be discussed. Students will gain knowledge of the sample design of each of these surveys and will be introduced to the statistical methods needed to adjust survey estimates for the different sample designs.

The second part of the course will focus on how to manage data for research projects. Students will learn how to take a public use data file and create an analytic file for research purposes. Other topics to be discussed include: locating and downloading data, documenting data files derived from public use data, creating new variables from survey data, assessing data quality, formatting the values of specific kinds of variables, and combining a source data file with other data files. Additional attention will be directed to the appropriate unit of analysis for specific projects, how to handle observations with multiple records, and understanding the data structure (long data file formats versus wide data file formats). The students will also learn techniques used to create tables and graphics for the presentation of research findings.

Selected Department Competencies Addressed:

Each Department identifies competencies for each degree offered. The competencies addressed in this course for the MPH for the Department of Health Systems and Policy includes:

- Apply quantitative research methods in the analysis of health service and policy issues

Please visit the Department webpages on the School of Public Health’s website at http://sph.rutgers.edu/ for additional competencies addressed by this course for other degrees and departments.
Course Objectives:

Upon completion of this course the student will be able to:

- Describe a general process for locating existing data resources that are potentially useful in public health services research and policy analysis, and better understand how to identify and explore these data.

- Indicate the steps required to take an unedited data set and produce an analytic file.

- Describe the basic structure of data files: unit of analysis, unique observations, data file organization and structure (i.e., flat-files, relational and hierarchical data resources, and multiple records per observation).

- Create tables/graphs for presentation of results in a clear and concise manner using appropriate survey-based statistical adjustments.

- Describe the components of a major data set used for health services research and policy analysis: the Medical Expenditure Panel Survey.

- Apply the rudiments of SAS programming language to create analytic data files from secondary data, and to implement and edit SAS programs for use in developing basic statistical information.

Course Requirements and Grading:

1. Grading:
   Homework – 40% (8 homework assignments- each worth 5 points)

   Homework assignments will be based upon the application of essential data programming and data manipulation techniques. These techniques will eventually help students to create a codebook for a specific analytical data file and for use in a final data analysis project. Assignments are due at the start of the class on the day they are due.

   Reading/participation – 5%

   Assigned readings must be completed before the class in which they will be discussed. Readings are critical for useful class discussion and understanding of the material.
Report (on a Medical Expenditure Panel Survey (MEPS) Publication) – 10%

Students will select a MEPS data publication of their choice and make a brief presentation to the class about its content and key findings. Particular attention will focus on the data section of the report. Specific details will be provided during our early class meetings.

Creating a data codebook for an analytical data file – 20%

Each student will create a codebook relating to their final data analysis project. The codebook will include a full list of the variables that they will use in their data analysis project. It will include labels and frequency distributions on all the variables and a written description of how newly created variables were constructed.

Data analysis project – 25%

The final project will be a descriptive analysis of a health services or policy-related topic. The analysis will take the form of a research findings report that introduces the topic and its importance, provides descriptive data estimates (e.g. mean values and/or proportions) of the key outcome of interest by relevant population characteristics. The report will highlight key statistically significant estimates that the author believes are among the important findings of interest from the report. It will include tables and/or graphs to illustrate their findings and will use the MEPS data.

Course Schedule:

Week 1 (May 20, 2014): Introduction to Secondary Data Sources and Analysis.

Week 1 will provide an overview of the course and establish the domains of secondary data analysis: Do available data sets determine your research or does the research question limit the data sets you can use? What are the pros and cons of using secondary data for research? How do we begin the process of identifying the best fit between our research question and available secondary data?

Required Readings:

Chapter 1 “Secondary Data Sources for Public Health” pp 1-11.


Week 2 will discuss available survey data commonly used to estimate health services utilization and key indicators related to other health services and policy issues. Sources of U.S. health
data including the Inter-university Consortium for Political and Social Research will be described. If time permits, sources of international health-related data such as the Demographic Health Surveys will also be discussed.

Required Readings:


Homework 1 will be assigned.

Week 3 (May 27, 2014): Introduction to SAS statistical software

The Statistical Analysis System (SAS) software will be introduced in week 3. SAS will be the basis for our data file construction and subsequent data analyses. We will learn how to create a SAS file from an Excel data file. We will learn how to use data libraries, working data files, and permanent data files. We will learn about alternative units of analysis and variables in the data file.


Homework 2 will be assigned.

Week 4 (May 29, 2014): Introduction to the Medical Expenditure Panel Survey (MEPS)

Week 4 will provide an in-depth look at the Medical Expenditure Panel Survey. The MEPS is conducted by the Agency for Healthcare Research and Quality (AHRQ) to collect information on health care utilization and costs in the United States as well as characteristics of the civilian, non-institutionalized population. Students will begin to appreciate the specific details about the MEPS survey including its cross-sectional and panel designs. We will begin working with SAS to download the data files and begin creating our research data set. The data set will be used for a small research project that we will do together so that we can begin to understand the principles of data management and documentation. We will create a flow chart that will help us understand how to merge two data files together. We will be downloading the following files:

MEPS HC-138: 2010 Full Year Consolidated Data File
MEPS HC-137: 2010 Medical Conditions File
MEPS HC-135G: 2010 Office Based Medical Provider Visits File

Required Readings:

Download the 2010 Full Year Consolidated Data File pdf documentation (MEPS HC-138: 2010 Medical Conditions File).

Students should become familiar with the Medical Expenditure Panel Survey website:
Students should also take a closer look at the link for downloading data files:

http://meps.ahrq.gov/mepsweb/data_stats/download_data_files.jsp

**Homework 3 will be assigned.**

**Week 5 (June 3, 2014): What kinds of variables can be used in Multiple Regression.**

In week 5 we will take a closer look at the MEPS Full Year Consolidated Data file. We will have a brief review of Multiple Regression (linear regression and logistic regression) and we will look at different variables on the data file. We will also begin to understand the unit of analysis by comparing individuals to families. SAS – creating totals by group and using the first. And last. variables.

**Homework 4 will be assigned.**

**Week 6 (June 5, 2014): MEPS Medical Conditions File**

In week 6 we will take a closer look at the MEPS Medical Conditions File. We will discuss ICD-9 CM codes (International Classification of Diseases, Ninth Revision) used generally in the US, and the Clinical Classification Codes (CCC) used specifically for MEPS data. We will begin to understand what types of inquiries can be made using the condition file. We also will link the condition file to the person level file (MEPS HC-138: 2010 Full Year Consolidated Data File).

**Required Readings:**

Download the 2010 Medical Conditions File pdf documentation (MEPS HC-137: 2010 Medical Conditions File).

Students are required to become familiar with the Priority Conditions Enumeration (PE) section and the Condition Enumeration (CE) section of the 2010 Survey Questionnaire.

http://meps.ahrq.gov/mepsweb/survey_comp/survey_questionnaires.jsp

**Additional information:**

**Link to information on the Medical Conditions File:**

http://meps.ahrq.gov/mepsweb/data_stats/MEPS_topics.jsp?topicid=32Z-1

Merging data files via data step (not proc sql):

http://statistics.ats.ucla.edu/stat/sas/modules/merge.htm

Using PROC SORT and the BY statement (removing duplicates):

http://statistics.ats.ucla.edu/stat/sas/modules/sort.htm
Homework 5 will be assigned.

Week 7 (June 10, 2014): MEPS Office Based Medical Provider Visits data file

In week 7 we will take a closer look at the MEPS utilization and expenditure files. We will also discuss the MEPS prescription medicine files. We will pay close attention to one of the utilization files: the Office-Based Medical Provider Visit file. We will begin to understand what types of inquiries can be made using this file, and we will link the office-based visit file to the person-level file (MEPS HC-138: 2010 Full Year Consolidated Data File).

Required Readings:


Homework 6 will be assigned.

Students will present a journal article that uses the Medical Expenditure Panel Survey (MEPS).

Week 8 (June 12, 2014): Introduction to Complex Surveys: How to Adjusted for the Non-Random Sampling of Survey Data

Survey data are typically not obtained from data collection designs based on simple random samples. For reasons we will describe in class, survey data may be obtained through disproportionate sampling of specific groups and through clustered sampling designs. Using the MEPS data set as an example, we will learn to adjust for the nature of data collection in order to produce unbiased estimates and reliable standard errors for statistical testing.

Required Readings:

NESUG 17 “Variance Estimation with Complex Surveys: Some SAS-SUDAAN Comparisons” By Xiuhua Chen and Paul Gorrell

Homework 7 will be assigned.

Week 9 (June 17, 2014): Creating An Analytic File: Step 1 -- Which Variables do we need?

In week 9 we will start to conceptualize a simple research question. We will think about how we can create different socio-demographic groups based on the available data. This will help us to identify key independent variables to be used in our descriptive project.

Students will identify a primary dependent variable of interest. We will learn how to keep only the variables we need and how this can sometimes be an ongoing process.


Homework 8 will be assigned.
Week 10 (June 19, 2014): Creating an Analytic File: Step 2 -- Organizing the project and extracting and recoding variables from the MEPS data.

In week 10 we will actually begin constructing our analytic file for our small research project. We will illustrate the difference between continuous and categorical variables and describe how to use them properly in analyses (e.g., dummy variable construction). We will discuss how each variable is coded and how the missing data are represented in a variable. We will learn how to recode data and run consistency checks of your original and recoded variables. We will also discuss how to handle missing data.

Required Reading:

Required Readings: Cody, Ron, “SAS Statistics by Example”, Chapter 4 and 5.

Creating and recoding variables:

http://statistics.ats.ucla.edu/stat/sas/modules/vars.htm

Using SAS functions for making/recoding variables:

http://statistics.ats.ucla.edu/stat/sas/modules/funct.htm

Subsetting variables and observations:

http://statistics.ats.ucla.edu/stat/sas/modules/subset.htm

Week 11 (June 24, 2014): Creating an Analytic File: Step 2 continued -- Organizing the project and extracting and recoding the variables from MEPS data.

In week 11 our data project will be conceptualized. We will describe the objectives of the project, the data we are using, and any specific hypotheses that may be related to project objectives. We will continue recoding variables and finalize the variables needed for our research project.

Required Readings:


Missing values in SAS:

http://statistics.ats.ucla.edu/stat/sas/modules/missing.htm

Labeling data, variables, and values:

http://statistics.ats.ucla.edu/stat/sas/modules/labels.htm
Week 12 (June 26, 2014): Creating an Analytic File: Step 3 -- Creating a personal codebook based on your data analysis project.

In week 12 we will begin work on the codebook. The codebook will include a full list of the variables that you use in your data analysis project. It will include labels and frequency distributions on all the variables (raw and newly created) with a written description of how the newly created variables were constructed. The importance of codebook creation and its relevance to data management/documentation will be emphasized.

Required Readings:

Link to proc_codebook macro

http://www.cpc.unc.edu/research/tools/data_analysis/proc_codebook

Link: To an example a codebook from the HRS survey.

http://hrsonline.isr.umich.edu/modules/meta/rand/randhrs1/randhrs1.pdf

Week 13 (July 1, 2014): Creating an Analytic File Step 4: Running the analysis.

In Week 13 we will learn how to examine the data and use inferential statistics to examine our hypothesis. We will learn how to do cross-tabulations/Chi square with proc freq and how to generate t-tests to understand the differences in means. We will create these descriptive statistics on our key outcome variable (dependent variable) by relevant population characteristics (independent variables).

Required Reading:


Week 14 (July 3, 2014): It's all about the Presentation.

In Week 14 we will learn how to present the findings from our data analysis project. We will learn how to present visuals: tables, graphs and charts. Students will learn that it is important for the visuals to be a numeric representation of the findings in their data analysis project. We will emphasize the importance of the specific parts of the visuals including the title and footnotes.

Required Readings:


Optional Readings:

Link:

http://www.policy.rutgers.edu/faculty/miller/tasks07.pdf

Another link:


**Week 15 (July 8, 2014): Student Presentations and summing up key points.**

During this time students will provide a ten minute presentation of their research projects, discussing the importance of the topic, the methods used to assemble the data, how the estimated were obtained, and the key findings from the study. The class will conclude with a brief summary of the key points emphasized during the semester.

**School of Public Health Honor Code:** The School of Public Health Honor Code is found in the student bulletin (sph.rutgers.edu/academics/catalog/index.html). Each student bears a fundamental responsibility for maintaining academic integrity and intellectual honesty in his or her graduate work. For example, all students are expected to observe the generally accepted principles of scholarly work, to submit their own rather than another’s work, to refrain from falsifying data, and to refrain from receiving and/or giving aid on examinations or other assigned work requiring independent effort. In submitting written material, the writer takes full responsibility for the work as a whole and implies that, except as properly noted by use of quotation marks, footnotes, etc., both the ideas and the works used are his or her own. In addition to maintaining personal academic integrity, each student is expected to contribute to the academic integrity of the school community by not facilitating inappropriate use of her/his own work by others and by reporting acts of academic dishonesty by others to an appropriate school authority. It should be clearly understood that plagiarism, cheating, or other forms of academic dishonesty will not be tolerated and can lead to sanctions up to and including separation from the Rutgers School of Public Health.

**Policy Concerning Use of Recording Devices and Other Electronic Communications Systems:**

When personally owned communication/recording devices are used by students to record lectures and/or classroom lessons, such use must be authorized by the faculty member or instructor who must give either oral or written permission prior to the start of the semester and identify restrictions, if any, on the use of mobile communications or recording devices.