
Applied Data Management

Fall 2013

Syllabus

Thursday, 4:30-7:00 pm

Instructor:

Michael Elliott

Assistant Professor

Salus Center Room 478

3545 Lafayette Ave.

Saint Louis, MO 63104-1314

314-977-4027 Phone

314-977-3234 Fax

elliottmb@slu.edu

Office Hours:

I am available to meet with students by appointment.

Course Overview

Course Description

This class is an advanced seminar intended for graduate students in the health and social sciences who plan to engage in applied research. This course is a survey of important data management topics and techniques. Topics include: data programming and manipulation, data storage and security, data cleaning, relational database theory, and legal and ethical issues of data management. Software tools covered in the course include SPSS, SAS, R, Excel, MS Access, and SQL. A number of guest lecturers will present case studies of real-world data management examples and will discuss data management aspects of each public health core area. A key aspect of the course will be brief weekly discussions by in-class working groups, culminating in a final 20-30 minute long presentation. The only formal prerequisite for this course is BSTC-500 or the equivalent. A familiarity with basic statistics and computer statistical software is required.

Core Competencies:

Section 1: Core Cross-cutting Competencies

Domain 1: Translation and Dissemination: As part of the change process (from discovery to improving public health) translation is the process of adapting information so that it is relevant to users, and dissemination is the widespread communication of that information.

1. Tailor language, style, delivery modalities, and communication channels to specific audiences.
 - Use a variety of communication/dissemination methodologies including media
 - Communicate complex information into written forms that are understandable to target audiences, e.g., policy makers, administrators, etc.

Learning Objectives 3 and 4

4. Use innovative methods (both quantitative and qualitative) to implement and evaluate translation and dissemination activities.
 - Use innovative methods to communicate lessons learned.
 - Articulate and apply methods for evaluating the relative advantages and priority of one discovery/innovation over others (including burden, preventability, cost, readiness to change).

Learning Objectives 1 and 4

Domain 2. Communication and Informatics: The ability to collect, manage and organize information, and to give, solicit, and receive oral, written, graphic and numerical information in both formal and informal settings of professional public health practice.

1. Use the public health information infrastructure to access and apply data for public health.
 - Use the public health information infrastructure to collect and process data
 - Translate data into information for application to public health problems

Learning Objectives 1, 2, 3, and 4

2. Recognize and use appropriate verbal, non-verbal, and visual forms of communication
 - Communicate effectively in writing, orally and graphically with different audiences

- Present accurate demographic, statistical, programmatic and scientific information for different audiences

Learning Objective 4

3. Accurately comprehend and interpret when listening to others.

- Promote the expression of diverse opinions and perspectives
- Demonstrate sensitivity to social, cultural and linguistic diversity in receiving and interpreting information

Learning Objectives 3 and 4

4. Facilitate collective information sharing, discussion and problem solving

- Solicit input from and information sharing among individuals and organizations.
- Define the components of organizational cultural competency and how it influences health, health disparities, research and practice.

Learning Objectives 3 and 4

MPH Core Disciplinary Competencies

Domain 12. Biostatistics: The development and application of statistical reasoning and methods in addressing, analyzing and solving problems in public health; health care; and biomedical, clinical and population-based research.

2. Compute and interpret descriptive and bivariable statistical tests and measures of association using statistical software (e.g., SPSS, Excel, SAS).

- Choose appropriate statistical methods given the measurement scales of the variables of interest and their distributions, as well as the study design and sampling method.
- Apply descriptive techniques commonly used to summarize data.

Learning Objectives 1,2, and 4

4. Follow ethical norms and rules for acquiring, managing, sharing, securing, and analyzing data.

- Know and abide by guidelines and regulations related to use of confidential data and ethical design of research studies.
- Complete the course of instruction offered by the Collaborative Institutional Review Board Training Initiative (CITI).

Learning Objectives 1,2, 3, and 4

Course Description

Course Format: (Lectures, class discussions, guest lectures)

Learning Objectives: Objectives

1. To provide students with an in-depth understanding of important data management tools and techniques.

2. To give students hands-on experience with modern statistical and database programming environments.
3. To show students how data management practices are implemented in several real-world research settings.
4. To expose students to data management aspects of a research project from start to finish.

At the end of the class, students will be able to:

- Understand the importance of data documentation
- Design an effective data codebook
- Design an on-line survey
- Understand the strengths and weaknesses of major statistics packages for data management
- Import, export, and transfer data among various database and statistical platforms
- Fully document SAS and SPSS datasets
- Manage, manipulate, transform, and recode datasets using SAS and SPSS programming
- Find and use national level data
- Design a simple relational database
- Perform simple SQL queries

Course Elements and Requirements:

1. Class participation, including readings as assigned. Participation is very important for a skills-based class such as Applied Data Management. Evidence of participation will include regular attendance, active involvement in class discussions, evidence of having read assignments, and active involvement with small groups. (20%)
2. Completion of weekly assignments. There will be 10 short homework assignments given out throughout the semester. These will focus on applied data management skills, such as data input, data cleaning, data import and export, and data transformations. **Homework will be due the week after it is assigned. Late homework will be docked 5 percentage points per week.** (50%)
3. Final presentations. There will be two final presentations: a group presentation and an individual presentation. The group presentation is the culmination of the group work that has been performed throughout the class. It is worth 10% of the final grade. The individual presentation will be a 15-20 Powerpoint presentation on a data management topic of the student's choosing. This topic can either be a novel data management technique (not discussed in class, such a cell phone data collection), or a novel data management environment/need (such as market research, "heat" maps). It is worth 20% of the final grade.

Required Materials

Delwiche, L. D., & Slaughter, S. J. (2012). *The little SAS book A primer (5th Edition)*. Cary, NC: SAS Institute. (Not available at bookstore. Can be ordered directly from Amazon or through other online bookstores. I would suggest the Kindle version for \$24.99.)

Levesque, R. (2011). *SPSS Programming and Data Management for IBM SPSS 20, A Guide for IBM SPSS Statistics and SAS Users*, available for download at:
http://www.spsstools.net/spss_programming.htm

Watson, R. (2013). *Data Management Databases and Organizations* (6th Ed.). (Available only in Kindle version through Amazon for \$9.99).

You will also need to have access to SAS and SPSS software to do the assigned exercises. We will discuss this during the first class meeting.

In addition to the statistical software, you will need to be able to use a reliable computer.

Grading Determination and Policy

Final grades for this class will be assigned as following: A (≥ 90 to 100 points), B (≥ 80 to 89 points), C (≥ 70 to 79 points) and F (< 70 points).

Feedback on Assignments: Timely feedback on assignments is needed in order to assure that students are aware of their progress. For routine assignments, quizzes, presentations, and exams feedback will be provided within two weeks after the due date. For longer assignments such as term papers and PhD exams, feedback will be provided within three weeks after the due date of the assignment or the completion date of the doctoral exam. In the rare event that these deadlines cannot be met, students should be informed of the delay and the extra time needed in providing feedback.

Attendance: Regular class attendance is an important part of one's graduate education in public health. Students are expected to attend all scheduled class meetings. In rare circumstances (e.g., illness, accident, death in one's family), absences will be excused. However, if a student misses more than seven (7) hours of a three-credit course, they may be asked to withdraw from the course and re-take the course at a later time.

Course Calendar

Session	Date	Topic	Assignment (due following week)
1	August 29	Introduction to data management Causal models Good questions	-create causal model (in class) -find scales (#1) -create additional questions (#1)
2	September 5	Documentation/Codebooks	-create codebook (in class) (#2)
3	September 12	Data Entry basics (Excel, ASCII)	-take survey (in class) -enter data in Excel (#3)
4	September 19	On-line data collection	-create on-line survey (#4) (in class)
5	September 26	DBMS theory	-conduct group data collection (#5) -create data model (in class)
6	October 3	NO CLASS-View recorded session	
7	October 10	<i>Guest Speaker: Angela Recktenwald Using a SQL database</i>	
8	October 17	Data Security/Database development <i>Guest Speaker: Cynthia Hudson Storing Data</i>	-create query and report (in class) (#6)
9	October 24	Importing Data into SAS and SPSS Catch Up and Intro. To R	- import data (#7)
10	October 31	Transferring Data between Programs	-transfer data (#8)
11	November 7	Transforming Data in Programs	-transform data (#9)
12	November 14	Qualitative Data	
13	November 21	Data quality/Cleaning	-clean data (#10)
14	December 5	Data presentation / GIS Group Presentations	-geocode data (in class)
15	December 12	Individual presentations	