PHA 6805

Applied Data Analysis, Interpretation and Reporting of Findings in Pharmacy

Instructors:
Joshua Brown, PharmD, PhD
Assistant Professor, Department of Pharmaceutical Outcomes & Policy
University of Florida College of Pharmacy
E-mail: joshua.brown@ufl.edu
Office hours:

Jingchuan "Serena" Guo, MD, PhD
Assistant Professor, Department of Pharmaceutical Outcomes & Policy
University of Florida College of Pharmacy
E-mail: guoj1@cop.ufl.edu
Office hours:

Hui Shao, MD, PhD
Assistant Professor, Department of Pharmaceutical Outcomes & Policy
University of Florida College of Pharmacy
E-mail: hui.shao@ufl.edu
Office hours:

Class Meetings: Lecture: Monday, 3-4pm, HPNP 2309
Lab: Wednesday, 9am-12pm, HPNP 2309

Course Prerequisites: STA 6166 and 6167, or permission of the instructor.

Recommended Reference Textbooks:

Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. 2007; New York: Cambridge University Press.

Gillingham M. SAS Programming with Medicare Administrative Data. 2014. SAS® Press.

Delwiche and Slaughter. The Little SAS® Book: A Primer, Fifth Edition. SAS® Press.

Recommended Reference Materials: SAS – IDRE Stats – UCLA website (includes many tutorials and code for statistical analyses. Simple Google of, "logistic regression SAS ucla" will pull up many great links, for example.)

Course Description:
This class has been redesigned and the instructors welcome your feedback and suggestions for continuous improvement. Your patience is appreciated!

The purpose of this course is to develop the skills students need to ask good research questions, plan and design studies, present a clear and reproducible protocol, develop

analytical datasets from raw data, test hypotheses that answer the research questions, and report, interpret, and discuss the findings. The course consists of two components: 1) lectures, readings, and activities focused on the conceptual and practical aspects of data analysis; and 2) hands-on laboratory and supplemental lectures regarding types of data, analysis and management of data, and reporting of methods and results.

A major assumption of this course is that students already have taken statistics or biostatistics courses and have introductory SAS programming skills. The primary focus of this course is **NOT** to re-teach statistics or programming, but to ensure that the student can correctly apply logic to programming challenges and apply appropriate statistical techniques that match the research question.

Course Structure:
The course will consist of two different sections: a) lectures over data analysis and concepts and b) experiential lab and project workshop sessions. Schedules of each are below. A project will encompass the entire class and be evidence for concepts learned throughout the course. A final protocol, report, and presentation will be required and described by the instructor. Two exams will assess your conceptual knowledge and class presentations and peer grading will evaluate your ability to work and communicate in a team environment.

Research dataset(s)
Each student is required to have a dataset (selected in consultation with their academic advisor and the course instructors) to work with. The instructors will work with students who do not have appropriate data sets available. In general, administrative claims data, including both MarketScan and Medicare are NOT encouraged or supported for this course. Examples of available and “easy to use” databases are Nationwide Readmissions Database, NHANES, NAMCS, BRFSS, MEPS, and so on. We have access to such databases on our research servers and many can be freely accessed and downloaded to your personal computer without relying on ResVault or other online resources.

For in-class programming, we will use Medicare SynPUF data (https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/DESample01) where appropriate for exposure to claims data: Other class demonstrations will typically use the HCUP NRD database and you will be exposed to other databases throughout the course.

Student Evaluation:

- Student presentation of assignments and project progress reports
- Peer evaluations of code, presentations, and overall class participation
- Full Protocol Development
- Final Report including SAS code
- Two exams

Final grades will be assessed on the basis of the following course requirements:
Class Presentations: 10 points
Peer review activities: 10 points
Protocol: 20 points
Final Report: 30 points
Exams: 30 points (15 points each)

Students who audit or retake the course are expected to follow all class regulations and complete all work mentioned above.

Scores from each of the assignments will be combined to calculate the final grade. Letter grades will be assigned according to the following scheme. Grades will be given only with whole numbers; thus, no discussion about "rounding up" or extra credit will be entertained.

Points earned	100-93	92-90	89-87	86-83	82-80	79-77	76-73	72-70	69-67	66-63	60-62	<60%
Letter grade	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	E

Attendance and Participation: Students are fully responsible for any materials missed during any absence. Your attendance is mandatory AND your participation is necessary. In-person attendance is expected and required until further notice. Zoom connection is provided for those with concerns regarding in-person attendance or those who are self-isolating. All lectures will be recorded.

Academic Dishonesty: Familiarize yourself with the University's policy regarding academic dishonesty. See the Statements regarding the Student Conduct Code on page 11 of the 1999-2000 Graduate Catalog. This policy will be strictly enforced. The University's conduct regulations are available on the Internet at https://www.dso.ufl.edu/sccr.

*Class schedules and topics are subject to change based on course needs
**All deliverables are due at 11:59 pm on the day listed unless noted otherwise

Lecture Session, Monday 3-4pm

Date		**Topic**	**Instructor**	**Deliverable****
W1	August 23	Class introduction, SAS Studio Sign-up	Brown, Guo, Shao	
W2	August 30	Data generating processes	Brown	
W3	September 6 (holiday)	No class		
W4	September 13	Student project concept presentations (5 mins)	Brown, Guo, Shao	
W5	September 20	Measuring outcomes and exposures		
W6	September 27	Exam	Brown	
W7	October 4	Reporting scientific data, best practices, journal formatting	Guo	
W8	October 11	Linear regression	Guo	
W9	October 18	Logistic regression	Guo	
W10	October 25	Survival analysis	Guo	
W11	November 1	Incorporating time, time series, and repeated measures	Guo	
W12	November 8	Generalized linear models, part 1	Shao	
W13	November 15	Generalized linear models, part 2	Shao	
W14	November 22	Missing data	Shao	
W15	November 29	Power analysis, sensitivity analysis	Shao	
W16	December 6	Exam	Shao, Guo	

Lab: Wednesday, 9am – 12pm

Date		**Topic**	**Instructor**	**Deliverable****
W1	August 25	Protocol development, reproducible research, research reporting guidelines	Brown	Project idea submitted to faculty
W2	September 1	Data management & efficiency Introduction to SQL and SAS macros	Henriksen/ Brown	
W3	September 8	Administrative claims data Public survey and HCUP data	Brown	
W4	September 15	Exploratory data analysis and describing data	Brown	
W5	September 22	Curating and manipulating data Defining study variables	Brown	
W6	September 29	Data visualization and R	Maguire	Protocol due
W7	October 6	Creating Table 1, etc.	Guo	
W8	October 13	Linear regression	Guo	
W9	October 20	Logistic regression	Guo	
W10	October 27	Survival analysis	Guo	
W11	November 3	Incorporating time, time series, and repeated measures	Guo	
W12	November 10	Generalized linear models, part 1	Shao	Draft report due
W13	November 17	Generalized linear models, part 2	Shao	
W14	November 24 (holiday)	No class		Peer grading due (Programming and draft final report)
W15	December 1	Missing data, power analysis, sensitivity analysis	Shao	
W16	December 8	Student presentations		
W17	December 13 (TBD)	Final report due		