VIRGINIA COMMONWEALTH UNIVERSITY

BIOS 625: CATEGORICAL DATA ANALYSIS AND GENERALIZED LINEAR MODELS

Spring 2018

Tuesday/Thursday: 10:00AM – 11:50AM

One Capitol Square, 5th Floor, Room # 5009

Instructor: Dipankar Bandyopadhyay, Ph.D.
One Capitol Square, 7th Floor, Room # 737
Tel: 804-827-2058
Fax: 804-828-8900
Email: dbandyop@vcu.edu

Teaching Assistants:
None

Prerequisites:

The course catalog states that students should have completed and passed BIOS 514 Mathematical Statistics II, BIOS 554 Analysis of Variance, and BIOS 572 Statistical Analysis of Biomedical Data.

Course Objectives and Expectations:

According to the course catalog, this course will provide "an introduction" to the theory and methods of analysis of categorical data. Topics include exact and asymptotic analysis of contingency tables; measures of association and agreement; theory and applications of

generalized linear models, maximum likelihood estimation and related numerical methods; linear models with different link functions and distributions; model fitting; and diagnostics.

Importantly, we will focus on developing your competency in four distinct areas: theory, application, computation and communication. We will cover the theoretical foundation of categorical methods and generalized linear models in order to prepare students for both their comprehensive examination and their PhD-level dissertation research. The application and practical use of categorical statistical models are covered in such a way that students can adequately perform all aspects of the analysis, including study design and planning, power analysis, data analysis, and interpretation of results. Students will learn how to use existing software (primarily, SAS) to perform traditional analyses, and will also learn computational strategies for more challenging cases, especially when no commercial software program is available. Students will also enhance their communication abilities, focusing on both the written and oral communication skills needed for inferential dissemination and communication with collaborators.

Lecture time will be split between traditional lectures, guided data analyses and demonstrations, class discussions, paper presentations, etc. The eventual goal of this course is to equip students to perform not only original (PhD level) research, but also involve in consultations with non-statisticians at ease.

Textbook:

There are two textbooks for this course:

- *Categorical Data Analysis, Third Edition*. Author: Alan Agresti. Publisher: Wiley.
- *An Introduction to Generalized Linear Models, Third Edition*. Authors: Annette J. Dobson and Adrian G. Barnett. Publisher: CRC Press.

I will also provide general outlines of my notes for which you may follow along and complete during lecture, and I will provide copies of any other assigned readings (see Literature Reviews below). There are several textbooks you may use as reference guides:

- *Generalized Linear Models, Second Edition*. Authors: P. McCullagh and J.A. Nelder Publisher: Chapman and Hall.
- *An Introduction to Categorical Data Analysis, Second Edition*. Author: Alan Agresti Publisher: Wiley.

Computing Requirements:

Computational instruction for this course will be provided primarily using the SAS (Cary, North Carolina, U.S.A.) software. You will need access to SAS version 9.2 or higher, as earlier versions do not offer some of the procedures and capabilities that we will need. The Dept.of Biostatistics has SAS 9.4. Note that this is not a course to teach SAS; hence some preliminary background of SAS is expected. You can make yourself familiar with SAS Documentations from http://support.sas.com/documentation/. I will use both "canned" SAS procedures and some SAS IML procedures (where you need to hard code yourself). This is a learning process, so we will do it together. Towards the end and time permitting, I will also introduce some Bayesian computing for categorical data via the freeware WinBUGS.

You are welcome to use R for computing; however, because of the lack of a teaching assistant, I would prefer homeworks submitted using SAS.

Communication:

All course documentation – syllabus, course notes, guidelines, etc. – will be posted on my course webpage, available at:

http://people.vcu.edu/~dbandyop/BIOS625.18.html

I will not print hard copies of anything. However, note that HWs (and their solutions) will be posted via the Blackboard link for the course, and you are advised to upload your HW solutions in Blackboard. Information related to posting course notes, HWs, etc, will be communicated via the Blackboard. Office hours are on **Thursdays 2PM – 3PM**, during which you are more than welcome to visit my office for any reason related to the course. You can stop by at other times, though I may be unavailable or unable to meet with you, and I reserve the right to deny service during off-hours. You may feel free to email me (dbandyop@vcu.edu) for any topic that you are unable (or unwilling) to bring up in class or during office hours. I will use your VCU administered email address for any communication.

Lecture Attendance and Participation:

Class begins at the time listed above. We will not wait for all enrolled students to be present, and we will not stop and repeat lecture material for late arrivals. Lecture is not solely time to reproduce on a white-board or screen what you can read in a book; rather it is an opportunity for you to actively engage with the material, your instructor, and your classmates. Though accomplished by several means, your willingness to ask questions and share your thoughts will undoubtedly enhance your chances for success in this course. Note that while attendance is not taken, it will be clear as to who attends class and who does not. I reserve the right to conduct impromptu quizzes for which no make-up opportunities will be offered.

Medical/Family Emergency Absences:

While students are expected attend every lecture, I understand that *on occasion* there may be other priorities. This understanding is reserved for students who responsibly inform the instructor *before* – or while – such instances occur. Regardless of the reason for an absence, the responsibility for obtaining missed lecture material and submitting/receiving assignments is entirely yours. While family and medical emergencies are sure to arise, you still need permission to make-up late assignments. Specifically, you should (a) contact and inform your instructor of your situation as soon as possible, and (b) contact Student Services to provide documented evidence for your absence. All work (homework, presentations and exams) missed for any reason can be made up only at your instructor's discretion.

Homework, Examinations and Grading:

Homework:

There will be about 6 homework assignments throughout the semester. You can choose to write the more statistically rigorous derivations by hand, or in LaTex. Solutions to applied problems (i.e. data analyses) should be typed, and submitted WITH THE COMPUTING CODES involved. They should be **submitted via Blackboard**.

Literature Reviews:

On several occasions throughout the semester, students will be assigned readings to supplement the lecture material and enforce certain concepts. They will be assigned on Thursdays, giving you enough time to go through those over the weekend. We will discuss some of those in class.

Examinations:

There will be 2 mid-terms and a Final examination during this semester. We can decide on the format [take home versus in-class] at a later time. These examinations will consist of questions on all material covered up to that point in the semester.

Plagiarism:

Collaborative *assistance* is allowed for some homework assignments, with the understanding that individuals are responsible for their own results, coding, output and write-ups. Due to the relative importance of developing certain skill sets, collaboration will not be allowed for some homework and in-class assignments. To be clear, at no point is it acceptable for you to write-up an assignment with another student, and submit it as your own work.

Grading:

No late homework assignments are accepted unless the student has made prior arrangements with the instructor. Provided all homework and in-class assignments have been submitted, I will replace the lowest homework score with the second lowest. A make-up exam is rare, so it is imperative that you perform well in the 2 midterms.

Distribution

Assignment	% of Final Grade
Homework + Other Assignments	**40%**
Midterm 1	**20%**
Midterm 2	**20%**
Exam	**20%**

Grades

% of Total Points	Letter Grade
[90%,100%+)	**A**
[80%, 90%)	**B**
[70%, 80%)	**C**
[0, 70%)	**F**

Withdrawal and Incomplete Policy:

The last day to withdraw from the course with a grade of "W" is **Friday, March 23rd**. Except in cases of medical or family emergencies that are documented with Student Services, I abhor and do not grant "incompletes."

Disability Services Policy:

Statement on Americans with Disabilities Act:

Section 504 of the Rehabilitation Act of 1973 and the Americans with Disabilities Act of 1990 require Virginia Commonwealth University to provide an 'academic adjustment' and/or a 'reasonable accommodation' to any qualified individual with a physical or mental disability

who self-identifies as having such. Students should contact the Disability Support Services office on the Monroe Park Campus (828-2253) or on the MCV Campus (828-9782) for appropriate academic adjustments or accommodations. Students are also required to notify their instructor of appropriate needs during the first week of class.

VCU Statement on Safety:

You are encouraged to sign up to receive VCU text messaging alerts (http://alert.vcu.edu/signup/index.php). Keep your information updated. You may visit (http://alert.vcu.edu/) for additional emergency information. The emergency phone number for the VCU Police is (828-1234). Please report suspicious activities and objects.

Honor Code and Student Conduct:

VCU Honor Code:

Virginia Commonwealth University recognizes that honesty, truth and integrity are values central to its mission as an institution of higher education. The Honor System is built on the idea that honor is a student's most cherished attribute. A foundation of honor is essential to a community devoted to learning. Within this community, respect and harmony must coexist. The Honor System is the policy of VCU that defines the highest standards of conduct in academic affairs. The Honor System must be upheld and enforced by each member of the Virginia Commonwealth University community. The fundamental attributes of our community are honor and integrity. We are privileged to operate with this Honor System. For more on the VCU honor system, see

https://medschool.vcu.edu/graduate/student_res/honor_system/

Student Conduct:

Your participation and behavior in class is expected to be courteous and considerate. As registered VCU graduate students, you have tacitly agreed to abide by the University statement on "Student Code of Conduct," which is found at the following website:

https://conduct.students.vcu.edu/student-code-of-conduct/

Tentative Schedule [subject to change]:

- Week 1: Introduction, Distributions and Inference for Categorical Data
- Week 2: Statistical Inference for Binomial parameters
- Week 3: Statistical Inference for Multinomial parameters, Describing Contingency Tables

- Week 4: Generalized Odds ratio and partial Tables, Contingency Table Probability Structures [CTPS]
- Week 5: More on CTPS, Testing Independence
- Week 6: Fisher's Exact Test and Partitioning Chi-squares, Ordinal Associations
- Week 7: Midterm 1 [Feb 27th, tentative], Introduction to Generalized Linear Models [GLM]
- Week 8: GLM for binary data, GLM for Count data
- Week 9: GLM estimation, and Logistic regression
- Week 10: Logistic regression continued
- Week 11: Conditional logistic regression, multinomial logit models
- Week 12: Midterm 2 [April 10th, tentative], Introduction to log-linear models
- Week 13: Log-linear models for contingency tables
- Week 14: Matched pairs models, Models for clustered categorical data (marginal and transition models)
- Week 15: Clustered categorical data: Random effects models, other mixture models. Some overview of Bayesian models for categorical data, with applications.