STAT 448   **Advanced Data Analysis**   **Spring 2024**

Course Policies and Syllabus

1 Contact Information

Instructor	Lelys Bravo de Guenni, PhD
Office	Computer Applications Building (CAB), Room 252
Email	lbravo@illinois.edu
Office Hours	Mon, Wed: 1:00-2:00 pm or by Zoom
Lecture Times	Tue and Th, 2:00-3:20 pm
Lecture Location	Everitt Lab 2233
Credit hours	4 (2UG and 2GR)
Course Space	canvas.illinois.edu

Teaching Assistant	BeongJip Kim
Office	616 E. Green St., Champaign, IL, Suite 211-212
Email	bk20@illinois.edu
Office Hours	Tue 10:00 am - 12:00 m by Zoom
Course Assistants	Jiayi Pan (jiayip3) and Nicholas Wong (nwong27)

2 Course Description

The topics in the course include applications of descriptive statistics and visualization, hypothesis tests for population locations and distributional goodness of fit tests, basic introduction to categorical data analysis and independence test for contingency tables, ANOVA models (balanced and unbalanced), simple and multiple linear regression, logistic regression, generalized linear models, Principal Components Analysis, Hierarchical cluster analysis, and Discriminant analysis (LDA and QDA) in a computational setting using SAS.

3 Course Learning Goals

Students are expected to get confident in data analysis in a computational setting, and select the appropriate tools according to the final analysis objectives:

- Apply descriptive Analysis and Visualization of a given Data Set.

- Make inference on the localization parameter of one population and comparisons between two populations.

- Assess the potential linear relationship between quantitative variables.

- Apply distributional goodness of fit tests to assess the normality assumption.

- Make inference for categorical data analysis and do hypothesis testing on the relationship between two or more categorical variables.

- Demonstrate the Analysis of Variance model for the balanced and unbalanced case on specific examples.

- Perform Simple and Multiple linear regression for modeling and prediction.

- Apply logistic regression for Binary data and Binomial responses.

- Use Generalized Linear Models for non-normally distributed responses.

- Apply Principal Component Analysis for dimensionality reduction.

- Demonstrate the use of Hierarchical Cluster Analysis and Discriminant Analysis in a real case example.

4 Course Information

The course is 16 weeks long and consists of 11 modules. You should dedicate approximately 8 - 10 hours per week to working on the course itself, but actual time commitments will vary depending on your input, needs, and personal study habits.

Pre-requisites: Stat 400 or Stat 409, and credit for or concurrent registration in Stat 410; familiarity with basic data analysis (e.g. regression, ANOVA) would be helpful.

Required and Recommended Textbook:

- **Requested**: A Handbook of Statistical Analyses using SAS (3rd edition) by G. Der and B.S. Everitt, CRC Press (2009). Data sets accompanying the text are available at `http://support.sas.com/HandbookofStatisticalAnalyses/`.

- **Recommended**: "SAS Statistics by Example" by Ron Cody, SAS Institute (2011): `https://sasinstitute.redshelf.com/book/1878390`.

- **Recommended**: "SAS and R : data management, statistical analysis, and graphics" (2nd Edition) by K. Kleinman and N.J. Horton, CRC Press, Taylor & Francis Group (2014).

Additional Resources:

- In-product documentation via the Help menu

- SAS$^{\copyright}$ v. 9.4 online documentation.

- SAS e-learning materials from the Webstore

Reading: The material in this course may go quickly. It is expected that you will stay up to date in reading the relevant sections of the text and/or class notes. The tentative schedule is at the end of this document and it will be also available online in the class web page through Canvas. The reading material for each class is listed there. Selected examples from each chapter will be part of the reading material and will be published in Canvas each week.

Required Software:

- SAS© version 9.4, SAS institute

- PDF Reader

- Word processing software compatible with Microsoft Word

- Students should be able to use their phone, tablet or laptop to connect to the course LMS site during class.

SAS Software Access:

All students are expected to gain sufficient knowledge on SAS (version 9.4, SAS Institute) at the end of this class. A brief introduction to SAS using examples will be done during week 1. All software needed for this class is available in labs across campus. The UIUC AnyWare program currently provides remote access to *on-campus* software. ATLAS, CITES, and library labs also have in-person access.

You can use one of the following options to access SAS:

- Run SAS from UIUC AnyWare mentioned above.

- Purchase from the webstore to buy a license expiring on Dec-31-2024.

- Use the SAS's free cloud version: SAS Studio available on SAS OnDemand for Academics.

Our preferred option is SAS OnDemand for Academics. You will receive further instructions on how to sign-up and access this free cloud version. An initial SAS code will be provided for each topic, so you can practice your SAS coding skills by using these examples. **Make sure you save you code each time you have any updates.**

5   Course Assessment and Grading Policy:

- **Homework (45%)**

  Seven graded homework will be assigned during the semester. The dates for the assignments will be announced in Canvas. Assignment approximate dates are also provided in the class schedule at the end of this document.

  Homework due date will be indicated in the homework file. Normally it is one week after the assignment date.

Late assignments are not allowed. The lowest score will be dropped. This will account for any missing assignment you might have due to unforeseen circumstances.

Homework presentation should be neat and submitted online through Canvas. You must show all your work for full credit. You must submit two files: a file with your .sas code and a second file with your homework question answers in .pdf format.

If you feel it would help, you are encouraged to work together on homework, but you have to present assignments individually using your own words. The aim of the homework is to learn the material and practice for the exams and final project. Copying homework solutions from another student, from past Department solutions, or from online solutions is cheating and plagiarism, and is a violation of Academic Integrity.

- **Quizzes (20%):**

Quizzes will be scheduled during the weeks with no homework assignments. Quizzes can be a great tool for self-assessment.**All quizzes will be available in Canvas and they are open book. You are only allowed to use the class notes and slides, as well as examples given in class. Using a different source of information will be considered as a violation of academic integrity**.

Quizzes questions might be presented in different formats: multiple selection, multiple choice, blank completion or short answers. The quizzes have generally 20 min duration and are worth 10 points. On the week a quiz is scheduled, the quiz will be held normally on Thursdays at the beginning of the standard class time.

Quizzes should be taken in class online using your preferred device. The lowest score will be dropped. This will account for any missing Quiz you might have due to unforeseen circumstances.

- **Final Project (30%):**

Instructions and data for the final project will be published in Canvas in a separate section. You are expected to work in teams of **up to three people maximum** for the final project. Project team members can be self-assigned during a specific open period. **All final project teams should be in place by week 6**. Beyond that date, I will assign the remaining students to the incomplete teams or will create new teams.

A final project progress report should be submitted by week 10. This will allow me to check your progress and make any corrections if necessary. **The final report should be submitted by Monday May 6th at midnight (11:59 pm, CST)**. Instructions for preparing the final project report will be published in Canvas.

- **Discussion Board (DB) Participation (5%):**

A Discussion Board topic will be available in Canvas every week. You are expected to participate in the discussion board at least every other week. You can start a new discussion, or reply to an existent discussion already started by your instructor or any of your class mates. You are expected to have participated in at least 8 of the 16 weeks discussion boards to

have the full DB participation points. Participation more than once in any particular week is counted as participation for one week.

- **Grading scale:**

A+	97-100%
A	93-97%
A-	90-93%
B+	87-90%
B	84-87%
B-	80-84%
C+	77-80%
C	74-77%
C-	70-74%
D+	67-70%
D	64-67%
D-	60-64%
F	60-0%

6 General Statements

Academic Integrity Statement

The University of Illinois at Urbana-Champaign Student Code should also be considered as a part of this syllabus. Students should pay particular attention to Article 1, Part 4: Academic Integrity. Read the Code at the following URL: http://studentcode.illinois.edu/ Academic dishonesty may result in a failing grade. Every student is expected to review and abide by the Academic Integrity Policy: http://studentcode.illinois.edu/. Ignorance is not an excuse for any academic dishonesty. It is your responsibility to read this policy to avoid any misunderstanding. Do not hesitate to ask the instructor(s) if you are ever in doubt about what constitutes plagiarism, cheating, or any other breach of academic integrity.

Copying homework solutions from another student, from past Department solutions, or from online solutions is cheating and plagiarism, and is a violation of Academic Integrity.

Disability Accommodations

To obtain disability-related academic adjustments and/or auxiliary aids, students with disabilities must contact the course instructor and the Disability Resources and Educational Services (DRES) as soon as possible. To contact DRES, you may visit 1207 S. Oak St., Champaign, call 333-4603, e-mail disability@illinois.edu or go to the DRES website. If you are concerned you have a disability-related condition that is impacting your academic progress, there are academic screening appointments available on campus that can help diagnosis a previously undiagnosed disability by

visiting the DRES website and selecting Sign-Up for an Academic Screening at the bottom of the page.

Family Educational Rights and ELEMENTS Privacy Act (FERPA) Statement

Any student who has suppressed their directory information pursuant to Family Educational Rights and Privacy Act (FERPA) should self-identify to the instructor to ensure protection of the privacy of their attendance in this course. See https://registrar.illinois.edu/academic-records/ferpa/ for more information on FERPA.

Safety

We have been asked by public safety to share the following information in case of weather or security emergencies.
Emergency Response Recommendations:
`https://police.illinois.edu/emergency-preparedness/run-hide-fight/`
Video on Emergency Response: `https://youtu.be/8j0_8PCWASE`

Sexual Misconduct Policy and Reporting

The University of Illinois is committed to combating sexual misconduct. Faculty and staff members are required to report any instances of sexual misconduct to the Universitys Title IX and Disability Office. In turn, an individual with the Title IX and Disability Office will provide information about rights and options, including accommodations, support services, the campus disciplinary process, and law enforcement options.

A list of the designated University employees who, as counselors, confidential advisers, and medical professionals, do not have this reporting responsibility and can maintain confidentiality, can be found here:
`wecare.illinois.edu/resources/students/#confidential`.
Other information about resources and reporting is available here:
`wecare.illinois.edu`.

7 General information, guidelines and tips for success

- To do well in this course you should keep up with lectures, readings and assignments. If you feel you are falling behind please contact the TA and/or the Instructor as soon as possible.

- You are expected to attend lectures and attend office hours meetings if you have any questions.

- The class material posted in Canvas is arranged in modules by week. The main topics and activities to be covered on a week will be announced at the beginning of the week. **You are expected to review this web site at least 4 days per week**, and devote at least 8-10 hours/week to prepare all the weekly material, including reading assignments and graded assignments.

- Office hours will be held in person or by Zoom. Office hours' Zoom link can be accessed from Canvas. Office hours can be used to ask questions about the class materials and/or homework questions, or any related grading procedure. Office hours are the main direct

interaction between the professor and the students. You should try to use office hours as much as possible.

- You are encouraged to ask questions about the homework or any other class material using the **Discussion Board** in Canvas. If you have private concerns or questions for me or the course assistants, please write to us directly using the Inbox tool in Canvas. We will be monitoring the space often, and we can provide a response **within the next 24 hours of your post on weekdays**. Rather than emailing questions to the TAs or to me, I encourage you to post your questions on the Discussion Board. If possible, students are encouraged to answer questions posted by other students to the Discussion Forum, rather than waiting for an instructor's response.

- When using the Discussion Board, please be careful with the use of your language. Be respectful to others and avoid offensive language at all times. This a great communication channel in different directions: learners to learners, learners to instructors and instructors to learners. Maintaining a constructive attitude in our communications will be a gain for all.

- Assignments submitted online will be reviewed and graded by the course instructor/TA **within two weeks after submission**. If your instructor/graders are unable to meet this time-line, students will be notified.

- The instructor will respond to email messages within 24 hours of receiving them **during weekdays** unless the instructor notifies you ahead of time of an inability to do so. When sending email, include a subject line that identifies the course number and nature of your question.

- Please note that the Discussion Board is not a substitute for office hours. You are encouraged to use the office hours for a more in-depth discussions related to the class material you want to review, or specific questions about the computational platform use.

- Make-up assignments/quizzes will generally not be accepted unless there is a legitimate reason (e.g., a serious medical emergency). Under such a situation, please contact the instructor as soon as possible. The acceptance of a late assignment or make-up exam will be determined based on each individual case. Proof of the event needs to be provided.

- This syllabus is a guide and every attempt is made to provide an accurate overview of the course and its requirements. However, certain circumstances may make it necessary for me to modify the syllabus during the semester for your benefit and the changes may depend, in part, on course progress and our needs. I will announce any change to the syllabus as early as possible so that you can adjust your schedule. The department/ school will also be notified of any change.

Tentative Schedule and Reading assignments*

Week	Book Chapters and Topics	Assignments
Week 1	Syllabus. Introduction to SAS Portions of Chapter 1	Register/Find SAS access
Week 2	Chapter 2 Data description and simple inference	**Quiz 1 (01/25)** **Homework 1 is assigned (01/23)**
Week 3	Chapter 2 Data description and simple inference	**Homework 1 is Due (01/30)**
Week 4	Chapter 3 Simple Inference for Categorical Data	**Quiz 2 (02/06)** **Homework 2 is assigned (02/06)**
Week 5	Chapters 4 and 5 Analysis of Variance I and II	**Homework 2 is Due (02/13)**
Week 6	Chapters 4 and 5 Analysis of Variance I and II	**Quiz 3 (02/20)** **Homework 3 is assigned (02/20)** **Final project is assigned (02/22)**
Week 7	Chapters 6 and 7 Simple and Multiple Linear Regression	**Homework 3 is Due (02/27)**
Week 8	Chapters 6 and 7 Simple and Multiple Linear Regression	**Quiz 4 (03/05)** **Homework 4 is assigned (03/05)**
Week 9	Spring Break	NA
Week 10	Chapter 8 Logistic Regression	**Homework 4 is Due (03/19)**
Week 11	Chapter 9 Generalized Linear Models	**Quiz 5 (03/26)** **Homework 5 is assigned (03/26)** **Final Project Progress Report is Due (03/28)**
Week 12	Chapter 9 Generalized Linear Models	**Homework 5 is Due (04/02)**
Week 13	Chapter 16 Principal Components Analysis	**Quiz 6 (04/09)** **Homework 6 is assigned (04/09)**
Week 14	Chapter 17 Cluster Analysis	**Homework 6 is Due (04/16)**
Week 15	Chapter 18 Discriminant Analysis	**Quiz 7 (04/23)** **Homework 7 is assigned (04/23)**
Week 16	Chapter 14 Longitudinal data analysis III	**Homework 7 is Due (04/30)**
Week 16	Reading Day (May 2nd)	**Final Project (Due on May 6th at 11:59pm)**

* Please note that this schedule might change if this improves the learning process.