U Health_horizontal_cmyk.eps

**Biomedical Informatics**

**Probability and Statistics for Biomedical Data Science (BMI 6106)**

**Spring 2020**

***Time and location***: Mondays 6:10 pm – 8:10 pm and Wednesday 5:10 pm – 6:10 pm (DBMI Room 1016)

Refer to the [University’s online schedule](http://registrar.utah.edu/academic-calendars/index.php) for the meeting time and location for this term. Lecture content will be available online using mostly Jupyter notebooks and notifications will be posted on Canvas, the class will meet for hands-on activities twice a week for a total of three hours.

***Resources***: to access Canvas, the University’s online class resource, navigate to here: <https://utah.instructure.com/> Each student will need access to a laptop for in-class hands-on activities.

***Class numbers and Credit Hours***: Not-for-credit distance-education class: BMI 6106. For-credit class BMI 6106 (3 credit hours).

***Prerequisites***: Undergraduate level introduction to statistics course and BMI Introduction to Programming (Python) or permission of the instructors.

***Faculty***: Edgar Javier Hernandez, PhD, [edgarjavi@gmail.com](mailto:edgarjavi@gmail.com) (University of Utah).

***TA:*** Andrew Miller (Andrew.C.Miller@utah.edu)

***Text Books:*** Introduction to Probability and Statistics Using R (2010 - <https://cran.r-project.org/web/packages/IPSUR/vignettes/IPSUR.pdf> ), Kerns, Jay. and Think Bayes (2012 - http://www.greenteapress.com/thinkbayes/thinkbayes.pdf), Downey, Allen. (both freely available online). Also a suggested text book is Think Stats: Exploratory Data Analysis in Python (2014 - <http://greenteapress.com/thinkstats2/thinkstats2.pdf> ) Downey, Allen.

# Description of the course:

This course offers an introduction to an extensive array of methods for mathematical biomedical data analysis with emphasis on three major topics (probability analysis, statistical inference, and the basic concepts of statistical pattern recognition through machine learning), with a clear emphasis on the biomedical field. We will cover basic probability concepts such as recognizing the importance of the analysis of random events in real life applications using probability axioms and rules. This course will present descriptive and inferential data methods for predictive analysis on samples and populations. This introductory course lays the foundation for more advance classes offered at the Biomedical Informatics Department. As an additional component of this class will be the extensive use of the statistical software R, which is one of the most used statistical packages in many disciplines.

# Learning Objectives:

1. By the end of this course, you will understand basic probability concepts, use probability rules, distinguish between discrete and continuous variables and the methods for their analysis, and solve problems related to random events.
2. By the end of the course, you will be able to describe the main methods for data exploration and analysis. You will use descriptive statistics to understand the nature of your data, and you will use hypothesis testing to predict the behavior of variables from samples and populations.
3. By the end of the course, you will understand the concepts of machine learning and pattern recognition, such as cluster recognition, data classification, dimensionality reduction, and temporal patterns.
4. By the end of the course, you will have experience with of the most widely use statistical software in the sciences, R. You will be able to use many of the statistical and graphical packages available for this software.

The skills we will be focusing on include:

* Identifying and explaining the characteristics of discrete and continuous variables.
* Distinguishing the various methods used to predict random events.
* Comparing inferential statistical methods, existing tools, and terminological resources.
* Proposing adequate statistical methods for specific clinical cases.
* Programming statistical analysis with R

# ADA Statement

The University of Utah seeks to provide equal access to its programs, services and activities for people with disabilities. If you will need accommodations in the class, reasonable prior notice needs to be given to the Center for Disability Services, 162 Olpin Union Building; phone: 581-5020 (V/TDD); email: [info@disability.utah.edu](mailto:info@disability.utah.edu); URL: <http://disability.utah.edu/>. CDS will work with you and the instructor to make arrangements for accommodations. All written information in this course can be made available in alternative format with prior notification to the Center for Disability Services.

# Departmental Teaching Philosophy

*Biomedical Informatics* is defined as “…the interdisciplinary, scientific field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving and decision making, motivated by efforts to improve human health” by the [American Medical Informatics Association](http://www.amia.org). We will work together with open communication in this course to help each other succeed, be collegial, and build skills and knowledge.

# Course Logistics

Jupyter Notebooks will be used for the didactic (lecturing) portion of this course. The advantage of using this platform is the practical convergence between lecture and programming practices within the software. Canvas, the University of Utah online teaching resource, will be used for tests and communication. Students will be given instructions about using Canvas and Jupyter after they enroll. ***Please*** limit all communication to the faculty to the Canvas email/conversations system (except emergencies). This greatly simplifies spreading the word when students spot an error or encounter a problem that affects the entire class. For technical assistance with Canvas, please contact the University’s Teaching and Learning Technologies (TLT) group: phone: 801-581-6112; email [classhelp@utah.edu](mailto:classhelp@utah.edu); or browse their URL at <http://tlt.utah.edu/>.

# Office Hours

In our experience students rarely avail themselves of online office hours, for this reason the instructor will be available after the in-class session to address any problems, concerns, or questions you might have during the course.

# Grading and course evaluation

Module Homeworks (50%): At the end of each practical module (Mondays) there will be a set of exercises to reinforce each topic. This homework will be due at the beginning of the next module.

Final project (25%): During the course we will work on developing a final project that will aim in addressing a question(s) or method(s) learned through the semester. For this exercise, the student will collect its own data, explore it, and analyze it using more than one method seen in class. There are no topic requirements, but an effective R script should be submitted with the analysis of the data. This will require R programming.

Exams (25%): We will have three exams presented during the semester. The grading criteria include conciseness, clarity, and proper use of statistical techniques.

***Participation:*** While there are no participation points awarded, weekly postings on the discussion board and contributing to the in-class sessions are strongly encouraged. Active participation can boost your final grade if your final total score is borderline between two grades. Discussion topics will be posted, but additional topics for discussion are welcome. When developing the system for the final project, students are encouraged to post questions and issues on the board for general discussion.

# Course Topics

|  |
| --- |
|  |
| **Probability theory** |
| Introduction to R: installation, basic functions, reading data, data classes, installing packages |
| Probability Review: Notation, conditional, marginal and joint probabilities, Bayes’ rule, chain/product rule, Markov chains, Odds Ratios and Risk |
| Discrete and continuous distributions, Probability density functions, mass functions, kernel density estimation, Maximum likelihood estimation (MLE) |
| Introduction to Bayesian inference, naive bayes, Belief Networks, Resampling methods, Confidence intervals |
| **Estimation** |
| Histograms and other visualization tools for hypothesis testing (qq-plots plots, box plots, etc.) |
| Hypothesis testing: parametric and non-parametric methods, power analysis (t-test, ANOVA,Kruskal-Wallis), Post-Hoc Analysis |
| Linear regression and correlation, logistic regression. L1 and L2 Regularization Methods, introduction to cross-validation. |
| Discrete random variables estimation, contingency tables, chi square, Poisson distribution, exact test |
| **Advanced topics** |
| Clustering Analysis (k-means, hierarchical clustering, PCA), factor analysis, permutation and randomization: bootstrap |
| Survival Analysis, Time series, autocorrelation, ARIMA |
| Information Theory: Entropy, Information Gain, KL Diverge, and Mutual Information |
| Case studies |
| Case studies |

# Course Policies

***Accommodation Policy:*** Some of the readings, lectures, films, or presentations in this course may include material that may conflict with the core beliefs of some students. Please review the syllabus carefully to see if the course is one that you are committed to taking. If you have a concern, please discuss it with the instructors at your earliest convenience. For more information, please consult the University of Utah’s Accommodations Policy, which appears at: [www.admin.utah.edu/facdev/accommodations-policy.pdf](http://www.admin.utah.edu/facdev/accommodations-policy.pdf)

***Attendance policy:*** The University expects regular attendance at all in-class meetings. Instructors must communicate any particular attendance requirements of the course to students in writing on or before the first class meeting. Students are responsible for acquainting themselves with and satisfying the entire range of academic objectives and requirements as defined by the instructor. (PPM, Policy 6-100III-O)

***Faculty Responsibilities: (from the University of Utah Policies and Procedures Manual)***

1. Faculty members are expected to meet their regularly scheduled classes. Failure to meet scheduled classes without prior notice to students is excusable only for reasons beyond the control of faculty members. Alteration of schedules, cancellation or rescheduling of classes may be done only for valid reasons and after adequate notice to students.

2. Faculty members shall engage in reasonable and substantial preparation for the teaching of courses assigned to them, consistent with their scope and nature and appropriate to the educational objectives sought to be achieved.

3. Faculty members must maintain regular office hours during which they are available for consultation with students or otherwise assure their accessibility to students.

4. Faculty members must, at the beginning of a course, give reasonable notice to students of the general content of the course, what will be required of the students, and the criteria upon which their performance will be evaluated. Evaluations must be performed promptly, conscientiously, without prejudice or favoritism, and consistently with the criteria stated at the beginning of the course. The criteria for evaluating student performance must relate to the legitimate academic purposes of the course. Grade appeals submitted by students are not considered charges of misconduct under this code. [For the appeals procedure, see the student code, PPM 8-10.2, Article III, section 3.04.]

5. Faculty members must not misuse the classroom by preempting substantial portions of class time for the presentation of their own views on topics unrelated to the subject matter of the course. Where faculty members find it pedagogically useful to advocate a position on controversial matters, they must exercise care to assure that opportunities exist for students to consider other views. Faculty members must not reward agreement or penalize disagreement with their views on controversial topics.

6. Faculty members must not use their position, authority, or relationship with students to obtain uncompensated labor for their own personal or pecuniary gain. They may not ask students to perform services unrelated to legitimate academic requirements of a course unless the student is adequately compensated for such services. Faculty members must not solicit gifts or favors from students. They must not accept gifts or favors where they have reason to believe that such gift or favor is motivated by a desire to secure some academic advantage.

7. Faculty members must not plagiarize the work of a student. Where a faculty member and a student work together, appropriate credit must be given to the student. Faculty members may not limit or curtail the right of a student to publish or otherwise communicate the result of the student's own scholarly activities.

8. Faculty members must not reveal matters related in explicit confidence by a student, except as required by law or university policy. Personal matters relating to a student must not be revealed by faculty members except to persons entitled to such information by law or university policies. Faculty members may, however, report their assessment of a student's academic performance and ability to persons making legitimate inquiry provided such disclosure is in accordance with the Family Educational Rights and Privacy Act ("FERPA").

***Student Responsibilities (From University of Utah student handbook)***

Students are expected to follow the Code of Student Rights and Responsibilities (“Student Code”) as delineated in the University of Utah Policies and Procedures Manual (http://www.admin.utah.edu/ppmanual/8/8-10.html).

1. Students are responsible for satisfying the entire range of academic objectives, requirements and prerequisites as defined by the instructor.

2. The University expects regular attendance at all class meetings. If you are absent from class to participate in officially sanctioned University activities, religious obligations, or with instructor's approval, you will be permitted to make up assignments and examinations.

3. In order to ensure that the highest standards of academic conduct are promoted and supported at the University, students must adhere to generally accepted standards of academic honesty, including but not limited to refraining from cheating, plagiarizing, research misconduct, misrepresenting one's work, and/or inappropriate conduct.

***Wellness statement***  
Personal concerns such as stress, anxiety, relationship difficulties, depression, cross-cultural differences, etc., can interfere with a student’s ability to succeed and thrive at the University of Utah. For helpful resources, contact the Center for Student Wellness: [www.wellness.utah.edu](file:///C:\Users\jennifer\AppData\Local\Microsoft\Windows\Temporary%20Internet%20Files\Content.Outlook\P7CHDFTQ\www.wellness.utah.edu) or 801-581-7776.

***Veterans Center***

The University of Utah has a Veterans Support Center on campus. They are located in Room 418 in the Oplin Union Building. Browse here for more information: <https://veteranscenter.utah.edu/>; phone 801-587-7722; email: [vetcenter@utah.edu](mailto:vetcenter@utah.edu).

***LGBT Resource Center***

Please let the instructor know if there is anything she can do to make the classroom environment more welcoming and respectful. Additionally, please recognize that the University of Utah has an LGBT Resource Center in campus. They are located in Room 409 in the Oplin Union Building. Hours: Monday – Friday, 8:00 am – 5:00 pm. Visit their website to find more about the support they can offer, a list of events through the center, and links to additional resources: [http:/lgbt.utah.edu/](http://lgbt.utah.edu/)

***Learners of English as an Additional Second Language***

If you are an English language learner, please be aware of several resources on campus that will support you with your language development and writing. These resources include: the Department of Linguistics EAS Program (http://linguistics.utah.edu/eas-program/index.php), the Graduate Writing Center, and the English Language Institute. Browse here for more information: <https://continue.utah.edu/eli>.