BIOM 611: **Statistical Methods for the Design and Analysis of Experiments**
Spring 2020

Description: This **introductory** course provides a foundation for the fundamental concepts in biostatistics as they relate to experimental design and analysis. We focus on defining research questions, carefully choosing appropriate analytic tools and interpreting the results of the analyses, including limitations of the analysis. The course has three units:

- Unit 1 introduces statistical concepts and their application to study designs involving a single sample. We begin with the broad topic of 'reproducibility' and the role of statistics in reproducibility. The unit primarily focuses on two key inferential methods: hypothesis testing and estimation. Parametric and non-parametric (rank-based) approaches to inference are discussed. Validity assumptions and approaches to choosing different methods are considered. Errors in hypothesis testing and sample size determination are covered.

- Unit 2 generalizes the methods from Unit 1 to study designs with two samples.

- Unit 3 concludes by considering differences in means and proportions between multiple groups and associations between quantitative variables. Analysis of variance (ANOVA) including two-way ANOVA is considered in detail. The issue of inferential reproducibility in the context of multiple comparisons is used to motivate approaches to adjust for multiple comparisons. Correlation and regression models are introduced.

Statistical methods will be implemented using the freely available software package Rcmdr. Rcmdr is a menu driven front-end to R. Initially we solely use the menu-driven options; as the course progresses students will learn how to download and use various R packages, and to create simple R code. Some R code is provided, but **at this point, we do not anticipate parallel instruction using R or RStudio.**

Student proficiency in R coding is not a goal of BIOM 611; R is simply a tool to implement statistical methods.

BIOM 611 is **NOT RECOMMENDED** for students with a moderate to strong quantitative/computational background.

Guided Questions. Each week involves a several-page introduction to material covered in lecture along with a set of questions specifically related to the reading. A Canvas quiz is provided. This material is optional; it may help you get a feel for big picture topics in the lecture.

Participation in Lecture: Lecture attendance is encouraged.

Labs: Each week students work together and hand in annotated code from the lab exercises. Lab group assignments will be made by your TA(s). The lab requires a laptop with access to the internet. This year we will initially work exclusively with Rcmdr. As the course progresses, there will be a few opportunities to work directly with R code. Lab reports are due **Fridays at 4:59 PM**. Labs are penalized 2% for each hour late. (By request, an extension will be made for Passover/Good Friday.)

Course Notes: Course notes and/or supplemental reading will be included on the Canvas website each week.

R Materials: We provide detailed commands in lab materials. The Fox text book described below is useful but not necessary. The Karp website provides a nice introduction to Rcmdr.

Fox, J 2016. Using the R Commander: A point and click interface for R. Chapman & Hall. (CRC Press)
(Optional Resource: we will provide detailed commands for implementing methods)

https://cran.r-project.org/doc/contrib/Karp-Rcommander-intro.pdf

Website: *CANVAS through* https://upenn.instructure.com/

All of the materials for student use are linked through modules. Files are not organized for student use.

Instructors & Activities:

Course Director: Mary Putt, PhD, ScD (621 Blockley), Ph (215) 573-7020
mputt@pennmedicine.upenn.edu

Administrative Assistant: *Joyce Jones woodwarj@pennmedicine.med.upenn.edu*

Teaching Assistants:

Andrew Chen (andrewac@pennmedicine.upenn.edu
Rebecca Deek (Rebecca.Deek@pennmedicine.upenn.edu)
Francesca Mandel (Francesca.Mandel@pennmedicine.upenn.edu)
Jill Schnall (jschnall@pennmedicine.upenn.edu)

Office Hours:

Mary Putt (by appointment)

Instructors: TBD

See CANVAS Announcements

Activity	Instructor	Time	Location
Lectures	*Putt*	*Tue 11:00-12:30*	*Austrian Auditorium*
Midterm Exam		*3 Mar 11:00 -1:00*	*Smilow Auditorium*
Final Exam		*5 May 11:00-1:00*	*Austrian Auditorium*
Labs		*Tuesday: 3:00-4:30*	*252 BRB II/III*
		Wednesday 3:30-5:00	*252 BRB II/III*
Office Hours & Review		*Mon: 3:30-5:00*	*BRB II/III 1201*

Week	Unit	Date & Instructor	Topic
1		21 January **Mary Putt**	*Lecture & Lab:* **Course Organization**. Installing R, and Rcmdr.
2	1	28 January **Rui Xiao**	*Lecture:* **Probability Distributions & The Exact Binomial** *Lab*. Probability distributions in Rcmdr, exact binomial test.
3	1	4 Feb	*Lecture:* **Tests based on Sample Means: Choosing between Tests**. Single sample tests of proportion. *Lab:* Working with proportions
4	1	11 February **Warren Bilker**	*Lecture* **Standardized Test Statistics: Single-sample continuous data.** *Lab:* One-sample T-test, Wilcoxon Signed Rank Test, Sign-test, Assessing normality; graphical approaches and hypothesis tests. Choosing between tests; assessing validity assumptions
5	1	18 February	*Lecture* **Errors in Hypothesis Testing** Statistical power and sample size determination.

		Justine Schults	*Lab:* Sample size determination. Type I and Type II error exercises.
6	1	25 February **Alisa Stephens**	*Lecture:* **Estimation & Confidence Intervals** (Homework not graded) Lab: Single sample designs: different types of confidence intervals.
		2 March	*Optional Review session: Student questions 501 BRB 3:30-5:00*
7		**3 March MARY Putt**	**Midterm Exam** In-class Smilow **11:00** AM-1:00 PM
Midterm Break: no lecture or lab			
8	2	17 March **Haochang Shao**	*Lecture:* **Two-sample proportions**. Metrics (risk difference, risk ratio, odds ratio). Z-test. Pearson's Chi-square test. Fisher's exact test. Paired data: McNemar's test. Two-sample T-tests. Lab: Two-sample tests (Proportions and T-tests)
9	2	24 March **Yun Li**	*Lecture:* **Two-sample continuous data** Wilcoxon Rank Sum test. Paired dataPower and Sample size. Results Reproducibility. Introduction to contingency tables. Lab: Wilcoxon Tests, Two-sample Bootstrap Confidence Intervals. Choosing between tests. Sample size calculations.
10	2	31 March **Justine Schults**	*Lecture*: **Contingency tables** continued. Issues with multiple comparison: Inferential reproducibility. Family-wise error rate. Bonferroni method. Introduction to one-way ANOVA. Lab: Contingency tables. Ordered predictor variable: Cochran Armitage test for trend. Pairwise hypothesis testing and confidence intervals. Maintaining family-wise error rate.
11	**2 & 3**	**7 April Jing Huang**	*Lecture* **One-way ANOVA** (continuous outcome, single categorical predictor), Approaches to maintaining family-wise error rates (Tukey HSD, Holm-Bonferonni), Kruskal Wallis. Two-way ANOVA (Additive model) Lab: One-way ANOVA, Kruskal Wallis. Pairwise tests and corrections for multiple comparisons
12	3	14 April **Doug Schaubel**	*Lecture* **Linear Regression/Correlation**. Lab: Linear Regression continued. Assessment of fit.
13	3	21 April **Doug Schaubel**	*Lecture* Linear Regression/**Correlation** Lab: No lab
14	3	28 April **Hongzhe Li**	Topic: To Be Determine
			Reading days and office hours
	All	4 May	*Optional Review session: Student questions 501 BRB 3:30-5:00*
EXAMINATION PERIOD		5 May	**Final Exam (In class)** Austrian 11:00 AM-1:00 PM

Assessment:

Component	Contribution to Grade (%)	Due[2]
Lab Reports[1]	*10*	*4:59 PM Fridays*
Homework	*30*	*11:59 PM Tuesdays*
Examinations (In-class requires Lock-down browser)		
Midterm Exam (In-class requires Lock-down browser)	*25*	
Final (Cumulative)	*35*	
[1] Drop worst 2 grades, not eligible for credit for a lab report if don't attend lab; [2]Any changes to due dates will be made on Canvas Announcements.		

A final absolute grade of at least 90 guarantees an A- or better; a final grade of at least 80 guarantees a B- or better; a final grade of at least 70 guarantees a C- or better.

Weekly Homework: Homework is set up in the CANVAS quiz format and must be submitted through CANVAS. Late submissions are penalized 2% per hour late.

Due times for homework show on CANVAS and will generally be **11:59 PM on Tuesdays for the week following lecture/lab.**

Check the announcements before starting your coursework; any comments, hints, or corrections will be posted there. Announcements are organized by date.

Assignments are learning experiences not mini-tests. Don't get bogged down and overly frustrated. I encourage students to seek help through either: (1) in-person office hours with your instructor or TA (2) homework hints posted through CANVAS Announcements.

A 'typical' student should expect to devote around **6-12 hours per week for coursework outside of lab and lecture**.

Email: Faculty notice email more reliably when students use the CANVAS email system.

Exams: The midterm evaluates understanding of statistical concepts and simple applications of these concepts. The **final examination is cumulative** and involve both concepts and data analysis. Both exams are in-class. Both require a laptop and a lock-down browser. (See CANVAS site)

Grade Changes: Any **grade changes must be requested in person** from the instructor for that week. No grade changes are made based on email.

Academic Integrity: Unless specifically indicated in writing, students may work together but must submit individually constructed responses to questions. **Doing otherwise constitutes a violation of the code of academic integrity**. Students must work independently on all exams. All students enrolled in BIOM 611 are responsible for understanding and following the Penn code of academic integrity. Please see *provost.upenn.edu/policies/pennbook/2013/02/13/code-of-academic-integrity*

Students with Disabilities: The University of Pennsylvania is committed to providing equal educational opportunities to all students, including students with disabilities. Penn does not discriminate against students with disabilities and provides reasonable accommodation to a student's known disability in order to afford that student an equal opportunity to participate in University-sponsored programs.

Excused Work; Unexpected Circumstances
Please contact the Course Director if health problems or life-circumstances impact your progress in the course. Accommodations can be made.