BIOM 611: **Statistical Methods for the Design and Analysis of Experiments**
Spring 2017

Description: This introductory course provides an overview of fundamental concepts in biostatistics as they relate to experimental design and analysis. As well, we introduce students to the related concepts of premise and reproducibility. The course has three units:

- Unit 1 introduces the scientific method as a series of six steps (define the research question, design the experiment & collect data, explore the data, carry out inference, formulate conclusions, look forward and back). In the first unit, we explore the strength of evidence supporting a particular research hypothesis, consider the generalizability of our conclusion and estimate the magnitude of the effect. We work largely with individual samples and consider single sample tests of proportion, and means. For hypothesis testing, emphasis is placed first on simulation, and then connecting ideas based on simulation to theory-based inference.
- Unit 2 specifically introduces methods for comparing means and proportions between groups, and revisit the analysis of paired data.
- Unit 3 concludes by considering differences in means and proportions between multiple groups and associations between quantitative variables. Analysis of variance (ANOVA), correlation and linear regression are explored.

Both parametric and non-parametric (rank-based) approaches to inference will be discussed. Statistical methods will be implemented using the software package R in Rstudio. Students will have an opportunity to code directly in R or to use the graphical user interface, Rcmdr to facilitate their work in R. We will explore and graph data, conduct simple tests, create confidence intervals, develop sample size estimates for a simple two-group comparisons, carry out ANOVA and linear regression. R markdown will be introduced. Students who successfully complete the course will be able to develop reproducible code in R.

Note: This course is **not recommended** for students with a moderate to strong quantitative background. Those of you with these backgrounds **should** take a statistics course with more rigorous mathematical underpinnings and/or with exposure to more advanced statistical methods. If you fall in this category, please discuss with your advisor

Active Learning: We will poll in the lectures, to encourage participation and to help your instructor assess understanding. We use Poll Everywhere (PE) to encourage student participation. A web-enabled device is needed. Polling is not graded and you do not need to register with PE.

Labs: The lab will require a laptop with access to the internet. R is currently not available through mobile devices.

Textbook: Tintle et al. 2015 Introduction to Statistical Investigations (ISI, Highly Recommended first half of course).

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

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

Instructors & Activities:

Course Director: Mary Putt, PhD, ScD (621 Blockley), Ph (215) 573-7020

Teaching Assistants: *Christy Hulling, BS, hcristy@mail.med.upenn.edu*

Teaching Assistants: *Carrie Caswell, MS*

Ruohui ('Roo-h-wee') Chen, BS

Jordan Dworkin, BS
Lingjiao ('Ling-zhi-ow') Zhang, MS

Office Hours: After lab & TBD: See CANVAS Announcements

Activity	Instructor	Time	Location
Lectures	*Putt*	*Tu 11.00-12.30*	*Austrian Auditorium*
Labs		*Tu-Fr: 15.30-17.00*	*252 BRB II/III*
Coursework Review		*Mn: 15.30-17.00*	*Austrian Auditorium (OPTIONAL)*

Week	Unit	Date	Topic
1		17 January	*Lecture:* Course Organization, Scientific Premise: The Six-step Method *Lab:* Installing R, Rstudio, Rcmdr and RcmdrPlugin.EZR, Graphics Devices, Simple explorations
2	1	24 January	*Lecture:* Ch. 1.1 & 1.2 Statistics, Simulations, Strength of Evidence, Significance how strong is the evidence. *Lab:* Subset data, save dataset, create categorical variable, R markdown, ggplot2, explore data
3	1	31 January	*Lecture:* Ch. 1.3 & 1.4 What impacts strength of the evidence One sample test of proportion, Standardized Statistics; *Lab:* Data manipulations, one sample tests of proportion, Exact Binomial Test, Large Sample Tests
4	1	7 February	*Lecture* Ch. 2 Quantitative Variables One-sample tests, Generalization to paired data *Lab:* One-sample T-test, Wilcoxon Signed Rank Test, Sign-test, Assessing Normality
5	1	14 February	*Lecture* Ch. 2 Formal Hypothesis Testing, Sample Size Determination *Lab:* Sample size determination, Class Experiment: Mystery Data
6	1	21 Feburary	*Lecture: Ch. 3* Estimation & Confidence Intervals (Single Sample) Lab: Estimation, Single sample proportions (large sample and exact), T-intervals
		27 February	*Optional Review session: Student questions*
7	2	28 February	*Lecture:* Ch. 4. Causation: Experimental Design, Confounding Lab: (short) Sourcing an R program. How to randomize, Class Experiment, Mystery Data **In-class Timed Quiz**
8	2	7 March	*Lecture:* Ch. 5 Comparing two groups and paired data (Proportions) Lab: Tests of proportions for two groups; Pearson's Chi-square test, Fisher's Exact test, McNemar's test
NOTE: WE HAVE CLASS & LAB DURING UNDERGRADUATE 'SPRING BREAK'			
9		Week of 14 March	Faculty and TAs at Conference **No Class No Labs: Assignment Due**

10	2	21 March	*Lecture:* Ch. 6 Comparing two groups (Quantitative Data) Introduction to bootstrap confidence intervals Lab: Means, T-tests, Wilcoxon Tests, Bootstrap Confidence Intervals (Ch 7, paired data)
11		27 March	*Optional Review Session: Student questions*
12	2	28 March	*Lecture*: Sample Size Calculations: Two groups, Student Questions **Take-home Midterm handed out** (replaces coursework) Lab: Abbreviated, sample size calculations for two groups
13	2 & 3	4 April	*Lecture* Ch.8 & Ch. 9 Two categorical variables, Multiple Testing, One-way ANOVA (continuous outcome, single categorical predictor) Lab: Analysis of contingency tables, (Maybe One-way ANOVA)
14	3	11 April	*Lecture*: Ch. 9. One-way ANOVA continued. Kruskal-Wallis; Adjustment for multiple comparisons. Two-way ANOVA (Additive Model, Introduction to Interaction Model) Lab: One-way ANOVA and Kruskal-Wallis, P-value adjustment methods, Tukey HSD confidence intervals, Two-way ANOVA for balanced designs
15	3	18 April	*Lecture* Ch. 10 Two-way ANOVA Interaction Model Continued. Correlation & Linear Regression Lab: Two-way ANOVA, Interaction Model, Unbalanced designs, Correlation, Simple Linear Regression
	All	24 April	*Optional Review Session: Student Questions*
16	3	25 April	*Lecture* Ch. 11 Multi-predictor Linear Regression, Additive model Lab: Linear regression, No assignment (brief coverage on exam)
			Provide solution to assignment: Reading days and office hours
EXAMINATION PERIOD		2 May	**Take-home Final** Due 9 May

Assessment:

Component	Contribution to Grade (%)
Coursework	
Weekly assignments (drop worst 2 grades)	*35*
Examinations	
Quiz	*10*
Midterm	*20*
Final	*35*

An absolute grade of at least 90 guarantees an A- or better; a grade of at least 82 guarantees a B or better.

Coursework: Coursework is intended to reinforce lecture and lab material and provide experience applying statistical methods and interpreting results of the analysis. *Coursework must be submitted using CANVAS.* Late homework receives a grade of 0, unless you have a doctor's note or a verified personal or family emergency.

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

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 (gold standard--nothing beats in-person help). (2) discussion boards through CANVAS. Due to the class size, I will not answer individual emails regarding coursework; I will answer questions posted to the discussion board on CANVAS. Please check with your TA regarding individual policies regarding email and discussion boards. In 2016, multiple students had similar questions and using the shared discussions boards helped us efficiently answer questions and clarify issues.

A 'typical' student might expect to devote around **4-6 hours per week for coursework outside of lab and lecture**. At times, and particularly **around the examinations you should double that estimate**. If you are routinely spending more time LET US KNOW. Either the assignments are too complex/long or you are getting stuck and we can help with efficiency.

Discussion Board Schedule: I will check discussion boards and respond once per day, every day except Tuesday and during our mid-semester break. Students are encouraged to respond to each others' questions.

Optional Review Sessions: Please see schedule for review sessions.

Quiz: The first quiz will be timed and in-class. It will take about ½ hour. You may not consult with other students during the quiz. You may use any materials you wish. I expect better grades for students who understand fundamental concepts with limited need to refer to external materials.

Midterm and Final Examinations: This course will be successful if students are able to carry out and interpret basic data analyses relevant to biomedical research. Thus examinations will be take-home and will involve the analysis and interpretation of data from experiments, and the design of experiments. You may use any materials that you want for these examinations, but you must work alone. For one question of each quiz/exam you can use a 'lifeline'. You will have the option of naming one other student. The grade that student receives on that question of the exam will be the grade you receive on that question of the exam. You do not have to choose the lifeline.

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**. All students enrolled at Penn 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. A student with a disability should contact Dr. Putt to discuss accommodations needed for success in this class.