**Methods in Biostatistics I and II**

**140.651 and 140.652**

**Course Web Page**: CoursePlus/Course List/Methods in Biostatistics I

**Important: all lectures, labs, and office hours are online via zoom meetings. Links below. Please do not share zoom links with anybody to avoid zoom bombing/spamming. In the unlikely event of zoom bombing the lecture or lab will be immediately interrupted and suspended for live teaching. A video of the lecture will still be posted online, and a new zoom link will be sent to everybody in the class. Plan to arrive early (3-5 minutes), as all zoom meeting have a waiting room and classes will start on time.**

**Instructor**:

Ciprian Crainiceanu, email: [ccraini1@jhu.edu](mailto:ccraini1@jhu.edu), office: E3636

Office hours: Email to set zoom appointment

**Teaching Assistants:**

Marta Karas ([mkaras2@jhmi.edu](mailto:mkaras2@jhmi.edu))

Nathan Dyjack ([ndyjack1@jhmi.edu](mailto:ndyjack1@jhmi.edu))

Jingning Zhang ([jzhan218@jhu.edu](mailto:jzhan218@jhu.edu))

**Lectures**: 10:30 – 11:50am Tuesday and Thursday

Online zoom link (do not share this information): <https://us02web.zoom.us/j/86595689888?pwd=WDcranRpOTEyKytKY0Fra291TU1iUT09>

**Labs**:

Tuesday 1:30 – 2:20pm

Online zoom link (do not share this information)

<https://JHUBlueJays.zoom.us/j/92343078671?pwd=NlVBeWJwTGJjRDZlT3lsbk8wd0sxZz09>

Wednesday 9:30 – 10:20am

Online zoom link (do not share this information)

<https://JHUBlueJays.zoom.us/j/98279730728?pwd=eElqV05jNHVHWXo5ZnlrNnArZExOdz09>

**TA office hour**:

Tuesday 5:00-6:00PM (starting September 8th, 2020)

Online zoom link (do not share this information)

<https://JHUBlueJays.zoom.us/j/93397501747?pwd=djN1citDSlRSREtyZjh3eEZacGdYZz09>

**Required Textbook**: Methods in Biostatistics with R, Crainiceanu, Caffo, Muschelli. This is an online book only. Here you have a coupon for getting the book free of charge. Please do not share this coupon with anybody and do not share the information without the express consent of the course instructor. Follow the link and chose $0.00 when you set the price

<http://leanpub.com/biostatmethods/c/MolbZYdb5wj6>

**Statistical Software and Programming Language** (suggested) R, freely available at [www.cran.rproject.org](http://www.cran.rproject.org) and RStudio, freely available at [www.rstudio.com](http://www.rstudio.com)

**Evaluation:** Students will be evaluated by ~ four take home problem sets, one midterm and one final online exam *each quarter*. Both the midterm and final exams will be timed and instructions for taking online exams will be provided. Grading will be 20% problem sets, 40% midterm and 40% final exam each term. Problem sets will be graded for completeness and a random sample of problems will be graded for correctness (the sample is the same for all students). Solving all problems and reporting the results will help improve performance on the exams. Take home assignments must be handed in on time. Deadlines will be indicated online and on the assignments. Students must complete assignments individually. Students with special needs should contact the course instructor as soon as possible to make the necessary learning and evaluation accommodations.

**Exam schedule** (all are online exams, open book, open laptop):

**Midterm 651:** September 24, 2020

**Final 651:** October 22, 2020

**Midterm 652:** November 24, 2020

**Final 652:** December 22, 2016

**If you have personal- or faith-related reasons that do not allow you to take the exam on these dates, please let the instructor know as soon as possible and alternative accommodations will be provided.**

**Prerequisites:** Calculus, linear algebra and a moderate level of mathematical literacy are prerequisites for this class. Simply having the prerequisites for this class does not necessarily mean that it is the correct class for you. For example, a student with a PhD in theoretical mathematics who would like a broad overview of biostatistics and immediately applicable techniques would be better off in the 620 series.

**Important Questions**

**How should you ask questions about lectures and labs?** Some questions will be answered during lectures; please remember that prerequisites are not optional! Focus on those questions during lectures that you think are of common interest to most students.You can also ask questions during my TAs labs and office hours. If your questions cannot be answered by the TAs then you can ask me during my office hours. In zoom I recommend to use the chat feature for asking questions, though you can also use the raise hand feature.

**Is lab attendance required, is class attendance required?** No. However, you should go to a lab and the lectures. Occasionally, lecture material will be presented in the labs. You should make sure that you get the notes for whatever class/labs that you miss. Be considerate of the instructor and TA's time; do not skip lots of classes and labs and try to make it up during office hours.

**Do I need to go to both labs each week?** No. We run two labs to help accommodate your schedule. Pick one to go to.

**Can I hand in the homework late?** No. At least, not without a penalty. Homeworks should be turned in at the end of class on the due date.

**Must I use R?** No. Use whatever you would like. You will be required to do a small amount of simulation and programming, for which software like R is well suited. The R programming language is the language of choice for research statisticians. However, you can use whatever language works best for you. SAS and STATA are also widely used. Programs such as SPSS, Minitab and Excel are insufficient for the needs of this class. Our TAs are R gurus, so they can offer help. No guarantees that they will know other software.

**There are many ways to learn R**. Here it is a particularly neat one, Swirl developed at JHU Biostatistics. In R or RStudio type

install.packages("swirl")

library(swirl)

swirl()

**Useful R reference card.**

<http://cran.r-project.org/doc/contrib/Short-refcard.pdf>

**What should I use to edit programs?** You should use an editor that has a fixed width font syntax highlighting and automatic indenting. Don't use Microsoft Word or Notepad to edit programs. Get latex or use Rmarkdown.

**When are TA's office hours?** They are listed in this document and they are online only via zoom. Please try your best to use this time. In special circumstances you can email your TAs to explore other times when they may be available.

**Can I miss lots of lectures?** Yes. Miss as many lectures as you would like. However, do not use TA and instructor office hours as a substitute for attending the lectures. Lectures and labs will be recorded and posted online.

**I have an event with dates that conflict with the final exam. Can I take the exam early?** No. I probably won't have the exam written to be taken early.

Course outline

1. Probability

A) Set theory

B) Axioms

C) Probability

D) Random variables

E) Bayes Theorem

F) Probability distributions

G) Expectations, variances

2. Inferences for one sample

A) Confidence intervals for a normal mean

B) Central limit theorem

C) The t distributions, t intervals

D) Hypothesis testing, significance levels, p-values, power

E) Sample size consideration

F) Inference for proportions

G) Maximum likelihood

3. Summarizing and Describing Data

A) Summary statistics

B) Empirical quantiles

C) QQ plots

D) Histograms, density estimates, stem and leaf plots

E) Boxplots

F) Percentile bootstrap confidence intervals

4. The two sample problem

A) Comparing 2 means (paired versus independent)

B) Comparing two proportions

C) Comparing dependent proportions

D) Sample size

5. Contingency tables

A) 2x2 tables

B) RxC tables

C) Three way contingency tables

D) Kappa

6. Statistical methods in epidemiology

A) Prospective, case control and cross sectional studies

B) Odds ratio relative risk

C) Simpson's paradox

D) CMH test

E) Methods for matched/paired observations

5. Introduction to nonparametric methods (possibly skipped)

A) Sign test

B) Signed rank test

C) Rank sum test

D) Permutation testing

E) Randomization testing and causal effects

8. Analysis of count data

A) The Poisson distribution

B) Inference based on count data and applications

9. One way ANOVA (possibly skipped)

A) F test

B) Decomposition of sums of squares

10. Large data sets (possibly skipped)

A) Multiplicity in significance testing

B) FDR, FWER, Bonferroni

C) Borrowing information, Bayes

11. Issues and Controversies (possibly skipped)