HSCI 3117 - Principles of Biostatistics

Spring 2023 A Session - January and February Term
Instructor: Vimal Rao (vimalrao@gwu.edu)
Office Hours: Tuesdays 2p-3p eastern time, and by appointment

Course Information

HSCI 3117 is an intermediate undergraduate level course focused on developing students' statistical literacy in the health sciences. This course is one of two undergraduate statistics courses in the health sciences, the other being HSCI 2117, which aims to introduce students to quantifying uncertainty amidst variability.

In HSCI 3117, we build on that foundation and introduce statistical models used either to predict, explain, or control for variability. Special attention will be given to interpretation of results from statistical models that are common to biostatistics and epidemiology.

The course primarily engages students by reading and discussing applications of statistics in the health sciences, although some computation is required. While the goal of HSCI 2117 is to have students understand and recognize how statistical study design characteristics support different types of inferences, the goal of HSCI 3117 is to introduce students to statistical methods for making sense of patterns and relationships amidst variability.

Textbook

The required textbooks for the course are:

> Motulsky, H. (2014). *Intuitive Biostatistics: a nonmathematical guide to statistical thinking* (4th ed.). Oxford University Press. ISBN-13: 978-0190643560
>
> Goss-Sampson, M. A. (2020). *Statistical Analysis in JASP 0.14: A Guide for Students*. (November 2020 ed.). DOI: 10.6084/m9.figshare.9980744
>
> Navarro, D.J., Foxcroft, D. R., & Faulkenberry, T. J. (2020). *Learning Statistics with JASP: A tutorial for psychology students and other beginners*. Available at https://learnstatswithjasp.com/

The required textbooks are available in online formats and print format. *Statistical Analysis in JASP* and *Learning Statistics with JASP* are free. *Intuitive Biostatistics* can be purchased for approximately $40. Be sure to purchase the correct edition of *Intuitive Biostatistics*, as there have been substantial changes from previous editions.

Expected Prior Experience

This course has no strictly enforced prerequisites. However, it is highly recommended that you take HSCI 2117 before taking HSCI 3117, especially if you do not have previous experience and familiarity with mathematics and statistics. In the first few weeks of the course, we will briefly review the foundational concepts of reasoning about uncertainty, which is essential for the rest of the course. For a more thorough foundation, complete HSCI 2117, which spends eight weeks focusing on this topic, the statistical principles behind study design characteristics such as random sampling and random assignment, and statistical tools such as the confidence interval and hypothesis tests.

Course Format

This is a fully online remote asynchronous 8-wk course. While this format allows for greater flexibility on the part of the student, it also requires increased effort to maintain open communication. You are expected to complete all assignments in a timely manner. If circumstances prevent you from doing so, you are expected to notify the instructor at least 24 hours *before* the deadline to request an extension. The instructor will hold a weekly office hour and be available via email to answer any questions you may have. You are encouraged to remain in frequent contact with the instructor, especially if you need any additional help with course content.

As this is an 8-wk course, you are expected to spend 14 *quality* hours studying *every* week. While it is understood that many students will be balancing responsibilities from home, work, and school, this course moves quickly and you should keep in constant contact with the instructor so that you do not fall behind with the material.

You are advised not to take any other courses if you also work full-time or part-time while you are taking HSCI 3117.

Communication

Email is the primary source of communication for this course. As such, you are expected to check your university email frequently (i.e., at least once per day). As per University policy, students are responsible for all information sent to them via their university assigned email account.

Course Requirements

Quizzes

Each week you will be assigned readings from the textbook and additional sources. Based on these readings, you will complete a readiness quiz. You will have two attempts to complete the quiz, and are expected to review the questions you got incorrect on your first attempt and review your notes in order to score the maximum number of points on your second attempt. If you score below 75% on your first attempt, contact the instructor for some extra help and guidance. Quizzes should be completed before moving on to the other activities in each week.

Mini-homeworks

You will complete a series of mini-homework assignments (MHW) throughout the course in which you will practice the procedures for conducting statistical analyses and the interpretation of results. Datasets for the MHWs will be from real studies conducted and published in the health sciences. Those studies will be required readings, and linked to the MHWs each week.

Discussion Activities

By far the most important assignment each week is the discussion activity (DA). DAs will focus on reading and discussing aspects of statistics from real studies in the health sciences. Some DAs will explore in-depth details of the readings associated with the MHWs, while other DAs will evaluate readings at a more general conceptual level. You are expected to be an active

participant in all discussions, posting at least 2 different days of the week, responding to all questions posed to you by the instructor or your classmates, and driving the conversation forward by asking questions to your classmates. Further instructions are provided in blackboard.

Project

Towards the end of the course, you will work on a project designed to foster your statistical reasoning and thinking in a problem context important and interesting to you. You will work with the instructor to select a peer-reviewed publication and apply your knowledge to extracting and summarizing the key statistical information. The project will be in two parts, each time focusing on different content that has been covered in the course. Further instructions are provided in blackboard.

Grades

Grades will be calculated based on a weighted average as follows: 20% quizzes, 25% MHWs, 25% DAs, and 30% project. Grading will utilize the plus-minus system.

Software Requirements

Blackboard

This course utilizes the blackboard learning management system. You are expected to familiarize yourself with blackboard and the course page.

JASP

All required data analysis for the MHWs will be conducted using JASP. JASP is a free open-source statistical software. JASP can be downloaded at https://jasp-stats.org/. Further instructions for using JASP are provided in the blackboard course page associated with each assignment.

Course outline

Review
- The uncertainty of inference
- Heuristics and biases
- Types of variables

Sources of variability:
- Inherent variability: Measurement error, Individual differences, Group differences
- Study design variability: Random sampling, Random allocation

Designing studies:
- Populations of interest and inclusion/exclusion criteria
- Sampling bias, random sampling and generalizing sample results
- Confounding variables, random assignment, and causal claims
- Test sensitivity and specificity
- Common study designs in the health sciences

Describing variability:
- Standard deviation and standard error
- Margins of error and interval estimates
- Confidence intervals for means, proportions, survival data, and count data

Hypothesis testing:
- Null models and *p*-values
- Statistical significance and decision errors
- Practical and clinical significance

Comparing two groups:
- Prospective and retrospective designs
- Within-subject comparisons
- Means, prevalence, incidence, and mortality

Simple statistical models:
- Simple linear regression (SLR) and correlation
- Analysis of Variance (ANOVA) and multiple comparisons
- Comparing models (e.g., fit and interpretation)

Statistical tests:
- Simulation based tests (e.g., bootstrap resampling, permutation, re-allocation)

- Distribution based tests (e.g., T, Z, Chi-sq, F)
- Nonparametric tests (e.g., signed-rank, rank-sum, Kruskal-Wallis)

Multiple Linear Regression (MLR):

- Predicting, explaining, and controlling variation
- Transformations, interactions, and collinearity of explanatory variables
- Procedural differences and conceptual similarities of MLR and ANCOVA models

Generalized Linear Regression Models (GLM):

- Adjusted odds ratios and logistic regression models
- Adjusted relative risks and log-binomial regression models
- Survival curves, hazard ratios, and Cox regression models
- Adjusted incidence and prevalence rates and poisson regression models

Special topics:

- Meta-analysis
- Reproducibility
- Reporting guidelines