**Biology/Statistics 214 – Biostatistics**

**Instructor:** Danielle Presgraves, Ph.D.

**(Required) Text:**  Whitlock, M., Schluter, D. (2020). *The Analysis of Biological Data, 3rd Edition.* Macmillan Publishers, NY.

I have the expectation that you will keep up with the reading of this course. This means you will need to read the chapter(s) or recitation prior to watching the video about it or attending the recitation class. You will take a short online quiz by a specific day/time to ensure that you have done so (see schedule). You are also responsible for the 'interleafs' in the textbook.

**Content:** This course will familiarize students with statistical concepts necessary to evaluate the primary literature in the biological sciences. It will improve students’ statistical literacy and sharpen analytical thinking. Topics covered in the course will include: descriptive statistics and graphics, estimation, elementary probability theory, Bayes’ theorem, statistical distributions, hypothesis testing, goodness of fit tests, experimental design, correlation, regressions, analysis of variance and maximum likelihood.

As a hybrid in-person/online format course, you have access to a number of benefits: you may complete aspects of this course of asynchronously — that is, you are free to complete most assignments, and quizzes when you see fit to do so before the completion dates and you still have the opportunity to attend office hours to answer any outstanding questions and review chapter concepts. Scheduled exams will be held during class times and attendance is required on those days (see exam dates on schedule).

**Teaching Philosophy**: Statistics is not just a related series of equations that you can memorize; it involves training yourself to an entirely new way of thinking. Statistics is about evaluating evidence and deciding consistent criteria to apply in order to support decisions. At the end of this class, you should be able to break down complex problems into their component pieces, identify which of these components are important to the hypothesis that you are testing, apply the appropriate statistical test and, of course, to quantify the uncertainty inherent in your conclusion and your data. Embrace diagraming flowcharts – they will be helpful in developing a heuristic about what tests to utilize under what circumstances! Much of the field of applied statistics, and therefore this course, is concerned with attributing variation to either 'noise' or 'signal' and logically justifying the steps you take in analyzing data sets.

Bio/Stt 214 is focused on applying statistical concepts and tests rather than with the derivation of the powerful mathematics that govern why these tests and distributions work in the first place.  During recitations, you will be exposed to additional problems and resources that will help deepen your practical understanding of the complex concepts covered in the syllabus but in a less formal environment and with less formal topics.

**Learning Management System:** We will be using Blackboard.

**Module Quiz:** The quizzes are designed to ensure that you have generally understood the concepts that are covered in the textbook chapters assigned for each module. You will not be able to view or take the quiz until after you have checked off the "reviewed" button in the "reading assignment" section of each module. You can only attempt the quiz twice so you should not do so until after you have finished the assigned reading. You must complete the quiz once you begin it. Only your last attempt at the quiz will be graded.

**Recitations:** Recitations serve two functions: to deepen your understanding of concepts taught during lecture that students have historically struggled with (such as probability) and to provide an opportunity to apply statistical thinking and logic to problems of larger scope than the ones discussed in class. To create the recitations, I have borrowed exercises from a variety of books, including "How to Lie with Statistics" by Darrell Huff, assorted texts from Andrew Gelman, a number of data science blogs and, my current favourite applied course "Calling Bullshit: Data Reasoning in a Digital World" by Professors Bergstrom and West (University of Washington, the syllabus with links can be found here: http://callingbullshit.org/syllabus.html). Each recitation has at least one activity that you will need to complete and attach to the Problem Set for that particular module (so you would attach the answer for recitation 1 to the bottom of your problem set 1).

**How do you do well in a course like this?** The same way you get to Carnegie Hall: practice, practice, practice. I have met very few individuals who were able to assimilate the necessary skills to do well in this course by simply being a sideline observer and only listening to lectures; *the only way to learn a new skill is to exhaustively practice it!* Everyone has the native ability to do well in this course but working through the practice problems at the end of the chapter will ensure that you are prepared for the exams. (see: The Atlantic review article from October, 2013 on the dangerous myth of being “bad at math”: <http://www.theatlantic.com/education/archive/2013/10/the-myth-of-im-bad-at-math/280914/> or more recently: <https://bigthink.com/mind-brain/bad-at-math-myth?rebelltitem=3#rebelltitem3> and one of my favourite blogs has also weighed in: <https://mathbabe.org/2017/08/16/math-still-not-everywhere/> ).

Remember the old expression: **"The difference between an amateur and a professional is that an amateur practices until they get it right but a professional practices until they can't get it wrong".**

|  |  |  |  |
| --- | --- | --- | --- |
| **Week/**  **Concepts** | **Learning Objectives** | **Means of Assessment**  **(order of activity in module)** | **Instructional Activities and Instructional Tools & Resources** |
| **Module 0: Week -1**  Welcome to the course!  Course expectations, Syllabus, Resources available  **Jan 5-12** | 1. Students will demonstrate awareness of course schedule (due dates) and expectations of participation | 1. Introduction and course expectations quiz | Activities:   1. Fill out the survey in the course overview folder, please! |
| **Module 1: Week 1**  Descriptive statistics: Location, spread, displaying data  **Jan 13-19** | 1. Students will compute and distinguish between descriptors of the location and spread of data such as mean, mode, median and standard deviation, interquartile range. They will explain the effect of bias and non-random sampling on moments. 2. Students will display data with appropriate visual tools 3. Students will judge examples of egregious misuse of statistics – especially with respect to data visualization – by providing examples found in recent newspaper or science articles. | 1. **Pre-Quiz** on assigned reading to test comprehension before accessing module. (LO: 1,2) 2. Short multiple-choice **questions** embedded in lecture videos. The tested information varies from knowledge of covered concepts to implementation and calculations. (LO: 1,2) 3. **Problem set 1 (+ recitation 1 question):** Upload to Blackboard by the due date. (LO: 1,2) | Activities:   1. **Read chapters: 1-3** 2. **Watch lecture videos**: panopto introduction 3. **Read and work through Recitation 1:**  Topics covered:  * How to lie with statistics * How to spot bullshit * Unfair comparisons * Failing to converge: Bias influence on accuracy. * Displaying Data / Describing Data  1. **Office hours on Thursday, January 13th will focus on:**    * Brief Introduction to the course and expectations   **Resources (supplementary):**  There are a number of youtube channels where you can learn basic statistics (other than this course). If you are struggling with concepts early on, here is a good one:   * Crash course in statistics: [**https://www.youtube.com/watch?v=sxQaBpKfDRk**](https://www.youtube.com/watch?v=sxQaBpKfDRk) * There are a lot of definitions in this initial part of the course and a useful way to familiarize yourself with terms is to make flashcards: [**http://flashcardmachine.com**](http://flashcardmachine.com) * The following is a free online stats book that includes simulations:   <http://onlinestatbook.com/2/index.html>  In particular, I encourage you to try out the sampling applet: <http://onlinestatbook.com/2/introduction/sampling_demo.html>  And descriptive statistics visualization (they range from means and median to variance) in section III: Summarizing Distributions. For instance, the median versus mean:  <http://onlinestatbook.com/2/summarizing_distributions/mean_median_sim.html>  and boxplot graphing applet: <http://onlinestatbook.com/2/graphing_distributions/boxplot_demo.html>  This website uses flowcharts to show you, based on your data, what the best way to display it is (and gives you R code to do it):  <https://www.data-to-viz.com/#boxplot>  Fascinating website on **WEB du Bois** data visualization (1890s):  <https://www.smithsonianmag.com/history/first-time-together-and-color-book-displays-web-du-bois-visionary-infographics-180970826/>  **Ida Wells-Barnett**: using data visualizations to display lynchings:   1. <https://www.vox.com/2015/7/16/8979771/ida-b-wells-lynching-data> 2. https://scatter.wordpress.com/2021/03/22/say-their-names-ida-b-wells-and-the-humanizing-of-data/ |
| **Module 2: Week 2**  Quantifying uncertainty  **Jan 20-26** | 1. Students will qualitatively interpret and quantify common descriptions of uncertainty such as confidence intervals and standard error 2. Students will demonstrate when estimation is used in the wider world by summarizing Fermi estimation and constructing their own Fermi estimation problems. | 1. **Pre-Quiz** on assigned reading to test comprehension before accessing module (LO: 1) 2. Short multiple-choice **questions** embedded in lecture videos. The tested information varies from knowledge of covered concepts to their implementation (LO: 1) 3. P**roblem set 2 (+ recitation 2 question):** upload to Blackboard by date given. (LO: 1) | 1. **Read chapter 4** 2. **Watch lecture videos:** Pantopto 3. **Read and work through Recitation 2.** Topics include: Estimation – Fun with Fermi Estimation   **Resources (supplementary):**  Confidence interval simulation:  <https://www.zoology.ubc.ca/~whitlock/Kingfisher/CIMean.htm>  Standard error simulation:  <http://www.zoology.ubc.ca/~whitlock/Kingfisher/SamplingNormal.htm>  From <http://onlinestatbook.com/2/index.html>  Section VII (Sampling distributions) and Section VIII (Estimation) will help you understand Sampling Distributions, Confidence Intervals and Standard Error. |
| **Module 3:**  **Weeks 3, 4**  Probability: Frequentist and Bayesian (and the differences between them)  **Jan 27 – Feb 9** | 1. Students will re-construct from first principles the laws of basic probability including conditional and Bayesian logic 2. Students will compute probabilities using the addition and multiplication laws | 1. **Pre-Quiz** on assigned reading to test comprehension before accessing module (LO: 1,2) 2. Short multiple-choice **questions** embedded in lecture videos. The tested information varies from knowledge of covered concepts to implementation (LO: 1,2) 3. P**roblem set 3 (+ Recitation 3 question)**: upload to Blackboard by date given. (LO: 1,2) | 1. **Read chapter 5** 2. **Read supplementary materials, if you want more background on:** probability concepts including Bayesian applications 3. **Watch lecture videos:** Panopto      1. **Read and work through Recitation 3** topics:  * Introduction to probability: Gambling * Intuitions in probability: The birthday and Monty Hall problems   **Resources (supplementary):**  <http://onlinestatbook.com/2/index.html> Section V (Probability) will take you through conditional probability, the Gamblers' Fallacy, the Birthday Problem, the Monty Hall problem, Bayes' Theorem and the Binomal Distribution |
| **Exam 1 – Module 1-3 (Tuesday, Feb 15th)** | | | |
| **Module 4:**  **Weeks 5, 6, 7**  Intro to Hypothesis Testing principles and some simple applications: Binomial test and Chi squared (contingency and goodness of fit)  **Feb 10- Mar 2** | 1. Students will define and describe the relationship between type I errors, type II errors, power and sample size. Students will calculate an example of inflated alpha given simultaneous multiple tests (ie. GWAS) 2. Students will constructbinomial, normal and Poisson distributions. 3. Students will clearly demonstrate the four components of hypothesis testing:  * Students will justify use of hypothesis testing and will construct the appropriate null hypothesis * Students will select the appropriate statistical test and will calculate the test value * Student will either calculate a p-value (if using Binomial Distribution) or interpret the critical value from the appropriate table. * Students will justify their conclusion and will offer support in the form of calculating a confidence interval | 1. **2 Pre-Quizzes** on assigned reading to test comprehension before accessing module (LO: 1, 2, 3) 2. Short multiple-choice **questions** embedded in lecture videos. The tested information varies from knowledge of covered concepts to implementation (LO: 1, 2, 3) 3. **Problem set 4 (+ Recitation 4 question):** upload to Blackboard by date given. (LO: 1, 2, 3) | 1. **Read chapters** **6, 7, 8, 9** 2. **Watch lecture videos:** Pantopto 3. **Read and work through Recitation 4**. Topics include,  * Hypothesis testing * Refuting bullshit (the importance of deploying null models) * Fitting models * Read paper: https://www.nature.com/articles/d41586-019-00857-9   **Resources (supplementary):**  <http://www.intuitor.com/statistics/T1T2Errors.html> This will help you understand Type I (alpha) and Type II (Beta) errors and their relationship trade-off. You can also investigate what happens as you increase **n**.  Contingency Analysis: <http://www.zoology.ubc.ca/~whitlock/Kingfisher/ContingencyAnalysis.htm>  <http://onlinestatbook.com/2/index.html> Section XI addresses Power, and Section XIV will give you experience with Chi squared tests. |
| **Module 5:**  **Week 8**  Normal Distribution  **Mar 3 – Mar 16**  (Spring break interrupts this module) | 1. Students will be able to interpret and compute z-scores and will assess any assumptions that z-scores include | 1. **Pre-Quiz** on assigned reading to test comprehension before accessing module (LO: 1) 2. Short multiple-choice **questions** embedded in lecture videos. The tested information varies from knowledge of covered concepts to implementation (LO: 1) 3. **Problem set 5 (+ Recitation 5 question):** upload to Blackboard by date given. (LO: 1) | 1. **Read Chapter 10** 2. **Watch lecture videos:** panopto 3. **Read and work through Recitation 5**, topic: the normal distribution (especially with respect to continuous traits)   **Resources (supplementary):**  Explore the Central limit theorem: <http://www.zoology.ubc.ca/~whitlock/Kingfisher/CLT.htm>  <http://onlinestatbook.com/2/index.html> Section VI Normal distributions |
| **Exam 2: Module 4, 5 (Tuesday, March 22th)** | | | |
| **Module 6:**  **Weeks 9-14**  More hypothesis tests: t tests, ANOVA, correlation, Regression, General Linear Models  **March 16 – April 13** | 1. Students will assess and contrast common statistical tests and apply them to novel problems appropriately by incorporating an evaluation of assumptions and experimental design concerns. 2. Students will be able to compute common statistical tests including t-tests, ANOVA, correlation, regression, general linear models including ANCOVA. | 1. **2 Pre-Quizzes** on assigned reading to test comprehension before accessing (LO: 1, 2) 2. Short multiple-choice **questions** embedded in lecture videos. The tested information varies from knowledge of covered concepts to implementation (LO: 1, 2) 3. **Problem set 6 (+ Recitation 6 question):** upload to Blackboard by date given. (LO: 1, 2) | 1. **Read Chapters 11, 12, 15, 16, 17, 18** 2. **Watch lecture videos:** panopto 3. **Work though recitation 6**, topic:  * Discriminating between the 'signal' and 'the noise' in statistical tests: * T-tests * ANOVA * Correlation * Regression * General Linear Models * Correlation ≠ causation   **Resources (supplementary):**  <https://shiney.zoology.ubc.ca/whitlock/Guessing_correlation/>  <http://onlinestatbook.com/2/index.html> Section X allows you to stimulate the t-distribution and Section XIII allows you to play with ANOVA. |
| **Exam 3: Module 6 (Tuesday, April 19th)** | | | |
| **Module 7:**  **Week 15**  *What happens when assumptions of tests are not met?*  Introduction to Non-parametric statistics,  Computationally intensive methods and  Maximum likelihood  **Apr 14– Apr 20** | 1. Students will conduct statistical tests when the assumptions of the test are violated will compare the results of standard parametric test with their analogous non-parametric test 2. Students will be able to identifythe nonparametric alternative to each parametric test 3. Students will evaluate the effects of using an inappropriate test (or a test when its assumptions are violated) on alpha and beta 4. Students will identify when it is appropriate to use computational methods such as Bootstrap and will apply them to problems. 5. Students will calculate Maximum Likelihood Estimates (MLE) and explain when they are used and contrast them to Bayes' theorem. | 1. **Pre-Quiz** on assigned reading to test comprehension before accessing module (LO: 1, 2,3,4,5) 2. Short multiple-choice **questions** embedded in lecture videos. The tested information varies from knowledge of covered concepts to implementation (LO: 1, 2,3,4,5) 3. **Problem set 7 (+ recitation 7 questions):** Upload to Blackboard by date given. (LO: 1, 2,3,4,5) | **1. Read Chapters 13, 19, 20, 21**  **2. Watch lecture videos:** panopto  **3. Recitation 7**, topics   * Examining assumptions * Big Data, Big problems? * How to refute bullshit (how to be vigilant)   **Resources (supplementary):**  How robust is the t-test? <https://shiney.zoology.ubc.ca/whitlock/RobustnessOfT/> |
| **Module 8:**  **Week 16**  Experimental Design  **Apr 21-Apr 27** | 1. Students will be able to identify specific and common errors in poor experimental design and describe qualitative effects of poor design 2. Students will analyze the results of poor experimental design on common statistical tests ability to discriminate against the null hypothesis (power) 3. Students will be able to calculate the sample size necessary to obtain sufficient power for the study. 4. Students will be able to calculate effect size and be able to summarize the challenges and benefits of meta-analysis | * 1. **Pre-Quiz** on assigned reading to test comprehension before accessing module (LO: 1, 2,3,4)   2. Short multiple-choice **questions** embedded in lecture videos. The tested information varies from knowledge of covered concepts to implementation(LO: 1, 2,3,4,5)   3. **Concept map –** draw out a concept map that connects all of the tests discussed in this course. Youmay want to use Coggle.com   4. **Problem set 8 (+ recitation 8 question):** upload to Blackboard by date given. (LO: 1, 2,3,4,5) | 1. **Read Chapter 14 and interleaf s 10 (publication bias) and 11 (meta-analysis)** 2. **Watch lecture videos:** panopto 3. **Recitation 8**, topics:  * Experimental design and (unintentional) bias * Publication bias and Dunning-Kruger effect.   **Supplementary material:**  Students may want to use mind-mapping tool Coggle.com to draw out how these tests are linked and **which test to use under different circumstances (ie. variable type, question asked).** For an example, look here: <http://jobloving.com/infographics/data-visualization/data-visualization-infographics-flow-chart-what-stat-test-to-use-resource-materials-painless-guide-to-statist/attachment/flow-chart-what-stat-test-to-use-resource-materials-painless-guide-to-statist/> |
| **FINAL (Cumulative) EXAM, Scheduled by the Registrar** | | | |

**Schedule (and Checklist) Spring, 2022:**

* Note that different problem sets have different point values as some are much longer than others, start problem sets early since late problem sets won't be accepted.
* Some of the activities (such as answering embedded clicker questions) are not required, but might be useful to you. To try to save space, I have not included reminders of my Tues/Thurs Zoom Office hours, but please remember that I will host them every Tuesday and Thursday – depending on Covid, they will take place in Goergen 109 or via Zoom.

|  |  |
| --- | --- |
| Week | Important Activities in each module |
| 0 | * Friday, Jan 7: Module 1 opens * Please take survey |
| 1 | * Sunday, Jan 16:  1. **pre-quiz due before 12:30PM** 🡨 usually this will be due on a Thursday 2. To stay 'on schedule', complete watching course module 1 videos by this date |
| 2 | * Wednesday, Jan 19: **Module 2 opens** * Thursday, Jan 20:  1. **Pre-quiz due before 12:30PM**  * Friday, Jan 21 by 11:59PM:  1. **Problem Set 1 (Chapters 1-3), including recitation 1 activities**  * Sunday, Jan 23 by 11:59PM:  1. To stay 'on schedule', complete watching course module 2 videos |
| 3 | * Wednesday Jan 26: Module 3 opens * Thursday, Jan 27:  1. **Pre-quiz due before 12:30PM**  * Friday, Jan 29 by 11:59PM:  1. **Problem Set 2 (Chapter 4), including recitation 2 activities** |
| 4 | Module 3 continues   * Wednesday, Feb 2:  1. In order to stay 'on schedule', complete watching course module videos and answer embedded clicker questions by this date |
| 5 | * Wednesday, Feb 9 : Module 4 opens * Friday, Feb 11 by 11:59PM:  1. **Problem Set 3 (Chapter 5), including recitation 3 activities**  * Tuesday, Feb 15: **EXAM 1 (modules 1-3: chapters 1, 2, 3, 4, and 5)** |
| 6 | Module 4 continues   * Thursday, Feb 17:  1. Pre-quiz due before 12:30PM **– split it into 2 pre quizzes: Only on chapters 6 and 7** |
| 7 | Module 4 continues   * Thursday, Feb 24 :  1. **Pre-quiz(2) due before lecture (before 12:30PM) – Only on Chapters 8 and 9**  * Friday, Feb 25 by 11:59 PM:  1. In order to stay 'on schedule, finish watching course module 4 videos |
| 8 | * Wednesday, Mar 2: Module 5 opens * Thursday, Mar 3:  1. Pre-quiz due before 12:30PM  * Friday, Mar 4 by 11:59PM:  1. **Problem Set 4 (Chapter 6,7,8,9), including recitation 4 activities** |
| 9 | **Spring Break at UR** |
| 10 | * Tuesday, Mar 15: OH resume * Wednesday, Mar 16: Module 6 opens * Friday, Mar 18:  1. **PS 5 is due by 11:59PM, including recitation 5 activities**  * Tuesday, Mar 22: **EXAM 2 (module 4 and 5 : chapters 6, 7, 8, 9, 10)** |
| 11 | Module 6 continues   * Thursday, Mar 24:  1. **Pre-quiz 6 (part i) due before lecture (before 12:30PM):** Chapters 11 and 12 |
| 12 | Module 6 continues   * Thursday, Mar 31 : Pre-quiz 6 (part ii) due before 12:30PM: Chapters 15,16,17,18 |
| 13 | Module 6 continues   * Friday, April 8 by 11:59PM:  1. In order to be "on track", finish watching course module videos and answer embedded clicker questions |
| 14 | Module 6 continues   * Wednesday, April 13: Module 7 opens * Thursday, April 14th:   **1. Pre-quiz 7 due by 11:59PM**   * Friday, April 15 by 11:59PM:   1. **Problem Set 6** (Chapters 11, 12, 15,16,17,18), **including recitation 6 activities**   * Tuesday April 19th by 11:59 PM: * **EXAM 3 (Module 6: Chapters 11, 12, 15, 16, 17,18)** |
| 15 | * Wednesday, April 20: Module 8 opens * Thursday, April 21:   **1. Pre-quiz 8 due by 11:59PM**   * Friday, April 22 by 11:59PM:  1. **Problem Set 7 (Chapters 13, 19, 20), including recitation 7 activities** |
| 16 | * Tuesday, April 26:   1. **Concept map due** * Wednesday, April 27:   1. **Problem Set 8 (Chapter 14 and 21), including recitation 8 activities** |
|  | **CUMULATIVE (Chapters 1-21) FINAL EXAM: Wednesday, May 4th at 19:15** |