



**B2CR 101**  
**BIOSTATISTICS AND BIOINFORMATICS 101 FOR CANCER RESEARCH**  
**The University of Florida Health Cancer Center**

*Promoting Rigors and Reproducible Statistical and Bioinformatics Approach for Your Research*

**Fall 2022**

**Director**  
**Ji-Hyun Lee**  
**Professor and Director**  
**Division of Quantitative Sciences & Biostatistics Shared Resources**  
[Jihyun.Lee@ufl.edu](mailto:Jihyun.Lee@ufl.edu)  
**352-273-9079**

**Administrative Assistant:**  
**Ken Thompson**  
[Kenneththompson@ufl.edu](mailto:Kenneththompson@ufl.edu)  
**352-273-8822**

**CaRTECC Administrative Manager**  
**Wendy Malorzo**  
[wmalorzo@ufl.edu](mailto:wmalorzo@ufl.edu)

## Background

Understanding biostatistics and computational biology is paramount for cancer research. BIOSTATISTICS AND BIOINFORMATICS FOR CANCER RESEARCH 101, *B2CR 101*, for cancer researchers at the University of Florida Health Cancer Center (UFHCC) is a **diverse program presented in a 12 lecture** series. The lectures introduce the basic principles of statistics, bioinformatics, and computational biology including AI and are intended for those who are in the **process of learning biostatistical and bioinformatics applications** or for those who **desire a refresher course**. Types of data, descriptive statistics, estimation, hypothesis testing, correlation/regression, survival analysis, sample size calculation/power analysis, and introduction to Bioinformatics and AI will be taught in this course series. Some classes will introduce basic theory and application of computing as a tool. Over the course, **reproducibility and replicability** of scientific research will be emphasized considerably.

<b>Course Goal</b>	The goal of this class is to introduce the basic statistical and computational biology concepts and methods for cancer research and promote rigors and reproducible statistical approach for cancer research.
<b>Course Prerequisites</b>	None
<b>Course Format</b>	The course will consist of both lecture and hands-on instruction. The lecture materials (slides) will be posted before the class.
<b>Registration Policy</b>	There is <i>no fee</i> for this lecture series. However, faculty and students should be registered before September 5 <sup>th</sup> via <a href="#">the website</a> *Note that some students may want to take only selected lectures, and this will be acceptable as long as registered.
<b>Who will take this course?</b>	Clinicians, Fellows, Cancer Researchers, and cancer research focused students. No prerequisites on graduate school level statistical courses are required. Note that only UF Health Science employees are eligible, and exceptional cases should be discussed with the coordinator prior to the registration deadline.
<b>Where</b>	<a href="#">Via Zoom</a>
<b>Time</b>	Sept 7 – Nov 30 (there will not be a lecture on Wed, Nov 23) 12 weeks: Every Wednesday 2:00 pm – 3:00 pm
<b>Credit policy</b>	No credit for this course.
<b>Homework</b>	Every week there will be a homework assignment (5 questions), and this will be considered a test that engages if the student understood the material.
<b>Grade</b>	50% Homework, 50% Attendance → Pass or Fail
<b>Certificate</b>	The class <b>certificate</b> will be given to those who completed and passed the course ( <b>10 attendances of 12 lectures are required as the minimum attendance</b> ).
<b>Course evaluation</b>	At the end of each class, students are recommended to complete <b>an evaluation</b> form for the lecture.

**Faculty Profile for BIOS 101**

<b>Name</b>	<b>Academic Rank</b>	<b>Education</b>	<b>Research Focus</b>
Dr. Ji-Hyun Lee	Professor	DrPH. Biostatistics, University of North Carolina at Chapel Hill, NC	Correlated data analysis, Group randomized trials, Generalized linear mixed models, and Best statistical practices.
Dr. Jason Orr Brant	Research Assistant Professor	PhD. Biomedical Sciences, University of Florida, FL	Role of epigenetics in disease initiation and progression, Bioinformatics
Ms. Zhanna Galochkina	Senior Biostatistician	MS. University of New Mexico	Data analysis using R, SEER and administrative data management
Dr. Shu Wang	Research Assistant Professor	PhD. Biostatistics, University of Pittsburgh, PA	Unsupervised machine learning and its application in medical data, early phase clinical trial design

**Fall 2022 BB 101  
Syllabus**

#	Date	Contents	Goals
1	Wed Sep 7	Introduction	<ul style="list-style-type: none"> <li>• Understand data types</li> <li>• Descriptive statistics/graphical summaries</li> <li>• When to use and how to calculate and interpret different measures of mean and variance</li> <li>• Distributions (Normal, Binomial, Poisson)</li> </ul>
2	Wed Sep 14	Estimation/ Hypothesis Testing and Statistical Inferences	<ul style="list-style-type: none"> <li>• Understand how to lay out scientific questions with hypotheses</li> <li>• Understand basic statistical inference: effect size, p-value, 95% CI, type I error, &amp; statistical power</li> </ul>
3	Wed Sep 21	Common Statistical Tests I	<ul style="list-style-type: none"> <li>• Association/Group Comparisons</li> <li>• Basic assumptions required for common statistical tests including the t-, paired t-tests, Chi-Square, Fisher's exact test, and the corresponding non-parametric tests</li> </ul>
4	Wed Sep 28	Common Statistical Tests II	<ul style="list-style-type: none"> <li>• Association/Modeling</li> <li>• Correlation</li> <li>• Linear Regression</li> <li>• ANOVA/ ANCOVA</li> </ul>
5	Wed Oct 5	Introduction to Clinical Trials	<ul style="list-style-type: none"> <li>• Concepts of Phase I, II, III, and IV</li> <li>• Translational study→ Confirmation study</li> <li>• New Trend for Phase I/II, including Bayesian approach</li> </ul>
6	Wed Oct 12	Survival Analysis	<ul style="list-style-type: none"> <li>• Kaplan-Meier Curve with Log Rank Test</li> <li>• Cox Proportional Hazard Regression Model</li> </ul>
7	Wed Oct 19	Sample Size Calculation I	<ul style="list-style-type: none"> <li>• Understand the role of sample size, effect size, variability, and power in study design</li> </ul>
8	Wed Oct 26	Sample Size Calculation II	<ul style="list-style-type: none"> <li>• For Survival data</li> <li>• For Cell Culture Experiments</li> </ul>
9	Wed Nov 2	Introduction to Bioinformatics/Computational Biology	<ul style="list-style-type: none"> <li>• Introduction to bioinformatics and computational biology concepts and tools.</li> <li>• Overview of RNA-Seq; from libraries to analyzed data</li> </ul>
10	Wed Nov 9	Introduction to R	<ul style="list-style-type: none"> <li>• Introduction to basic R commands and data object types</li> <li>• Reading/Manipulating/Saving data</li> <li>• Graphical illustration</li> </ul>
11	Wed Nov 16	The Cancer Genome Atlas (TCGA)	<ul style="list-style-type: none"> <li>• Access and utilize TCGA genomics and clinical data via the web</li> <li>• Perform simple cohort comparisons</li> <li>• Access gene mutation data and statistics for various cancers</li> </ul>
12	Wed Nov 30	Analyzing TCGA data using R	<ul style="list-style-type: none"> <li>• Access and analyze TCGA data using R via Bioconductor R packages.</li> <li>• Perform analysis of RNA-Seq and clinical data from TCGA and apply statistical tests.</li> </ul>

### Lecture Evaluation Form

This evaluation is anonymous.  
DO NOT PUT YOUR NAME ON THIS SHEET.

**COURSE:** BB 101 at UFHCC, Fall 2022

**Course Sequence #:**

**Please Score**

**(Lowest) 1 – 2 – 3 – 4 – 5 (Highest) for each box below**

Lecture Content	Organization & Presentation	Overall Rating

#### COMMENTS:

- 

- 

- 

---

Question for you: How long has been taken to touch your cell-phone since the class started?

Answer: \_\_\_\_\_ min.

- Please note that the data collected from this question will be used to illustrate the **Survival Data Analysis Method (Lecture 6)**.