

# Advanced Biostatistics with R – Design & Data Analysis with Case Studies

## 2016 - Syllabus

### Instructor:

Yu Shyr, Ph.D.  
Harold L. Moses Chair in Cancer Research  
Director, Vanderbilt Center for Quantitative Sciences  
Director, Vanderbilt Technologies for Advanced Genomics Analysis and  
Research Design (VANGARD)  
Professor of Biostatistics, Biomedical Informatics, Cancer Biology, and  
Health Policy  
Vanderbilt University Medical Center,  
Nashville, Tennessee, USA

**E-mail:** [yu.shyr@vanderbilt.edu](mailto:yu.shyr@vanderbilt.edu)

### Co-Instructor:

Qi Liu, Ph.D.  
Assistant Professor  
Department of Biomedical Informatics  
Vanderbilt University Medical Center  
Nashville, Tennessee, USA

**E-mail:** [qi.liu@vanderbilt.edu](mailto:qi.liu@vanderbilt.edu)

### Teaching Assistant:

Sheau-Chiann Chen, Ph.D.  
Postdoctoral Research Fellow  
Vanderbilt University Medical Center  
Nashville, Tennessee, USA

**E-mail:** [sheau-chiann.chen@vanderbilt.edu](mailto:sheau-chiann.chen@vanderbilt.edu)

## I. Design and Conduct of Biomedical Research

The theoretical and practical challenges to be considered in advanced biostatistics including designing and conducting a **Receiver Operating Characteristic (ROC) curve experiment**, **meta-analysis for multiple studies**, **longitudinal data experiment design**, **lasso/group lasso/elastic net based high-dimensional Omics research design**, **bootstrap resampling**, **zero-truncated and zero-inflated model**, and Markov Chain Monte Carlo (**MCMC**) Hidden Markov Models (**HMM**) Parameter Estimates will be presented.

## II. Analysis of Biomedical Research

Topics to be discussed in the advanced biomedical experiments include the specification of a **primary objective**, **biomarker based prediction model guidelines**, the role of **internal validation**, false discovery rate (**FDR/local fdr**) and the means for their implementation, the assessment of model performance, the choice of design strategy and design strengthening features, and the considerations involved in **common mistakes** in advanced biomedical research. Methods of analysis appropriate to various study objectives, class comparison and class prediction will be presented. The statistical approach will be based on **empirical use of methodologies rather than formal algebraic knowledge**, the emphasis on understanding what the procedures do and applications to data analysis. Methods of novel data analysis and various visualization tools will be discussed.

## III. R

The R software will be introduced, including coding for graphics, hypothesis testing, regression, lasso/group lasso/elastic net method, and MCMC.

## IV Topics:

Date	Topic
July 6 (Wed) Morning	FDA case study ROC course experimental design and data analysis
July 6 Afternoon	MCMC in R
July 7 (Thu) Morning	Meta-analysis Regression analysis: GEE, Mixed effect model, lasso/group lasso/elastic net model
July 7 Afternoon	Regression analysis in R (I)
July 8 (Fri) Morning	Resampling & bootstrap Missing data analysis Robust regression Zero truncated and zero inflated model
July 8 Afternoon	Regression analysis in R (2)
July 9 (Sat) Morning	Statistical methods for multiple comparisons FDR Local fdr Biomarker based prediction model guidelines
July 9 Afternoon	Regression analysis in R (3)