BIOSTAT 702: Module 0

Introduction to Applied Biostatistics

Dr. Marissa Ashner

Department of Biostatistics and Bioinformatics

Fall 2025

Module Goals

- Motivate the class by talking about what applied biostatistics is
- Understand the lifecycle of a clinical research study
- Become familiar with the VAI framework for statistical analyses

Resources for this Module

Websites

- Biostatistics Series Module 1: Basics of Biostatistics

What is Statistics?

What is Statistics?

- “a set of tools that enable us to learn from data”

What is Statistical Thinking?

- “a systematic way of thinking about how we describe the world and use data to make decisions and predictions, all in the context of the inherent uncertainty that exists in the real world”

What is Statistical Thinking?

Let's break this down further...

- A way to DESCRIBE the world
- Use DATA to make DECISIONS and PREDICTIONS
- Capture how UNCERTAIN we are about these descriptions, decisions, and predictions

What is Statistical Thinking?

Let's break this down further...

- A way to DESCRIBE the world
- Use DATA to make DECISIONS and PREDICTIONS
- Capture how UNCERTAIN we are about these descriptions, decisions, and predictions

Removes BIASES from human judgement

What is BIOstatistics?

What is BIOstatistics?

- development and application of statistical methods to research in health-related fields
- biostatisticians collaborate with scientists, and do not simply crunch numbers
- “central to all of science, because science needs that gathering of evidence and the evaluation of evidence to make a judgment”

How the First Semester Courses Fit Together

- **BIOSTAT701:** focuses on the theory behind the analytic methods we will be learning here
- **BIOSTAT703:** focuses on the clinical/scientific research question, study aims, and study design behind the contextual examples we will be using here
- **BIOSTAT721:** focuses on learning the statistical software we will use here to implement our analyses

Note: It may be the case that we cover something in this course before you learn it in another. I am happy to go over any details that may be helpful in further understanding the concept in this course

Life Cycle of a Clinical Research Study

- Observation / pattern
- Research Question
- Hypothesis
- Study design / data collection
- Statistical analyses
- Results / conclusions

Life Cycle of a Clinical Research Study

- **Observation / pattern**
  - Example: A primary care physician notices that many of her hypertensive patients also tend to have elevated creatinine levels, a sign of worsening kidney function.
- Research Question
- Hypothesis
- Study design / data collection
- Statistical analyses
- Results / conclusions

Life Cycle of a Clinical Research Study

- Observation / pattern
  - Example: A primary care physician notices that many of her hypertensive patients also tend to have elevated creatinine levels, a sign of worsening kidney function.
- **Research Question**
  - Example: Is there a link between hypertension and kidney?
- Hypothesis
- Study design / data collection
- Statistical analyses
- Results / conclusions

Life Cycle of a Clinical Research Study

- Observation / pattern
  - Example: A primary care physician notices that many of her hypertensive patients also tend to have elevated creatinine levels, a sign of worsening kidney function.
- Research Question
  - Example: Is there a link between hypertension and kidney?
- **Hypothesis**
  - Example: There is a link between systolic blood pressure (SBP) and estimated glomerular filtration rate (eGFR)
  - Must be *measurable*, *clinically meaningful*, and of *greatest importance* for the research question at hand
- Study design / data collection
- Statistical analyses
- Results / conclusions

Life Cycle of a Clinical Research Study

- Observation / pattern
  - Example: A primary care physician notices that many of her hypertensive patients also tend to have elevated creatinine levels, a sign of worsening kidney function.
- Research Question
  - Example: Is there a link between hypertension and kidney?
- Hypothesis
  - Example: There is a link between systolic blood pressure (SBP) and estimated glomerular filtration rate (eGFR)
- **Study design / data collection**
  - What to think about: control over the study environment, timing of the study, type of sample, timing of measurements, etc.
- Statistical analyses
- Results / conclusions

Life Cycle of a Clinical Research Study

- Observation / pattern
  - Example: A primary care physician notices that many of her hypertensive patients also tend to have elevated creatinine levels, a sign of worsening kidney function.
- Research Question
  - Example: Is there a link between hypertension and kidney?
- Hypothesis
  - Example: There is a link between systolic blood pressure (SBP) and estimated glomerular filtration rate (eGFR)
- Study design / data collection
  - What to think about: control over the study environment, timing of the study, type of sample, timing of measurements, etc.
- **Statistical analyses**
  - A statistical analysis plan (SAP) should be in place before analyses begin
  - **V**isualize and **A**nalyze
- Results / conclusions

Life Cycle of a Clinical Research Study

- Observation / pattern
  - Example: A primary care physician notices that many of her hypertensive patients also tend to have elevated creatinine levels, a sign of worsening kidney function.
- Research Question
  - Example: Is there a link between hypertension and kidney?
- Hypothesis
  - Example: There is a link between systolic blood pressure (SBP) and estimated glomerular filtration rate (eGFR)
- Study design / data collection
  - What to think about: control over the study environment, timing of the study, type of sample, timing of measurements, etc.
- Statistical analyses
  - A statistical analysis plan (SAP) should be in place before analyses begin
  - **V**isualize and **A**nalyze
- **Results / conclusions**
  - **I**nterpret

Which steps in the life cycle are biostatistcians (at least partially) responsible for?

- ALL OF THEM!!!

Our Primary Focus in this Course Sequence

- Statistical analyses
  - A statistical analysis plan (SAP) should be in place before analyses begin
  - **V**isualize and **A**nalyze
- **Results / conclusions**
  - **I**nterpret

Visualize, Analyze, Interpret (VAI)

Visualize

- The first step in any data analysis is to get to know your data!
- Typically referred to as *exploratory data analyses*
- Examples of things to look at:
  - Are your values clinically reasonable?
  - What do the values of each variable mean and how were they coded?
  - Are there missing values and how are they coded?
  - Plot the univariate distributions of your variables to investigate the shape and spread of the data
  - Plot 2-way visualizations to explore how your variables are related to each other
  - Perform any data cleaning, wrangling, and variable derivations needed to create a *clean* analytic dataset

Analyze

- Once you have a clean, final analytic dataset, then the statistical analyses as outlined in the SAP can be performed
- This is typically the first step taken in many stats classes, as you are often given the cleanest form of a dataset
- Which analyses are outlined in the SAP depends on many factors
  - study design
  - sampling
  - variable definitions
  - theoretical assumptions
  - *reality of the data once observed*
  - *exploratory data analyses*
- Analyses should map back to hypotheses being tested
- Sensitivity and post-hoc analyses

Interpret

- Your job doesn't end when you output the code results from your analyses
- We must interpret the results and map them back to the context at hand to draw conclusions related to the research questions / hypotheses, and in a way someone with less statistical knowledge can understand
- Interpret the results in terms of *statistical* and *clinical* significance
- Discuss limitations of the study / analyses