

## 2.1: Why (Bio)Statistics?

### What are the basic elements of biostatistics?

What skills are you going to get from all of this? You will get different opinions on the elements of an essential first course in statistics or biostatistics. Certainly the basics are a foundation in probability and a breadth of classical elementary statistical procedures, which will include [descriptive statistics](#), [analysis of variance](#) and [linear regression](#), and an introduction to [multivariate analysis](#). In preparation for your course in epidemiology you will also be introduced to [risk analysis](#) and [survival analysis](#). However, the primary return for your time, I hope, will be a deeper appreciation for how to think about problems in biology from experimental design and data analysis perspectives. Practical skills you will learn include how to process and clean data for analysis, data visualization, and a foundation in parametric and nonparametric statistical methods.

### Why do we require you to take (Bio)statistics as part of your major?

At [Chaminade University](#) we require all biology students to take biostatistics, and we do so with an emphasis on use of data analysis skill development. This requirement aligns our program to national expectations of biology undergraduate education (e.g., AAAS, NAS, NIH, NSF). As stated in *Bio210: Transforming Undergraduate Education for Future Research Biologists*,

*“Biology majors should be adept at using computers to acquire and process data, carry out statistical characterization of the data and perform statistical tests, and graphically display data in a variety of representations (p. 15).”*

Learning biostatistics from a course like BI-311 — which relies heavily on use of the R programming language and data sets — helps the biology student develop these skills.

In the next pages I will outline a [history of statistics \(Chapter 2.3\)](#), but here I wish to make the point that biostatistics is now considered to be a core skill set for biologists. Biostatistics as a discipline came into its own in the 1930's, but extensive reliance on statistics in research really dates to more recent times because of the ubiquity of personal computers (Salsburg 2002). Modern biological and biomedical research requires computational and quantitative methods to collect, process, analyze, and interpret large data sets. And yet, even a casual survey of required courses in the year 2014 for entry into graduate programs in biology will reveal that biostatistics is not expected of candidates; so what gives?

The first point is that programs list only minimum requirements. The second point is that many programs (genomics, ecology, etc..) will expect the graduate student to take a year or more of statistics. The need is so crucial that at Harvard Medical School, all biology graduate students are expected to take a crash-course in computing and statistics to work with data (Stefan et al 2015).

Moreover, while graduate programs are not listing statistics as a requirement, many biology undergraduate curricula now require a course in biostatistics to reflect the increasingly data-driven state of modern biology — where the jobs are!

I'll make you a bet — or at least, I'll make this part of your required homework (see BI311 Workbook)! Even a causal search of a research journal article in a biology discipline of your choosing will prove that there is no doing biology research today without an understanding of statistics.

### But, you may be thinking, I'm pre-med and plan to apply to medical school ...

Even a cursory look at the literature will result in finding many authors strongly calling for this kind of preparation for a successful career in medicine (e.g., Brieger and Hardin 2012). It's obvious, but needs stating — you're applying to medical school to become a doctor — you'll spend the majority of your adult life as a doctor. Statistical thinking is crucial to answering the daily question: “My patient tested positive for biomarker X, what's the chance that the patient has disease Y?” If you answer is, the patient has the disease, then you definitely need this course! Hint: there are four possible outcomes of a test, see [Chapter 7.3 – Conditional Probability and Evidence-Based Medicine](#).

Need more convincing? Take a look at the targets of questions intended to evaluate Skill 4 of the Scientific Inquiry and Reasoning Skills standard of the revised MCAT<sup>2015</sup> Exam (p. 107, What's on the MCAT<sup>2015</sup> exam?).

- Using, analyzing, and interpreting data in figures, graphs, and tables.
- Evaluating whether representations make sense for particular scientific observations and data.
- Using measures of central tendency (mean, median, and mode) and measures of dispersion (range, inter-quartile range, and standard deviation) to describe data.

- Reasoning about random and systematic error.
- Reasoning about statistical significance and uncertainty (e.g., interpreting statistical significance levels, interpreting a confidence interval).
- Using data to explain relationships between variables or make predictions.
- Using data to answer research questions and draw conclusions.
- Identifying conclusions that are supported by research results.
- Determining the implications of results for real-world situations.

I won't trouble you now with further justifications.

### In what disciplines are biostatisticians employed?

One way to begin this discussion is to think about where statisticians work. The job market includes:

#### Health Science

- Drug design, causes of diseases (many "causes" of cancers).
- Health Professional (nurses, physical therapists).
- Type of care and recovery period (importance of a person's mood on health).
- Exercise regime and recovery from injury.
- Nutrition: vitamins and health, diet and health.

#### Ecology & Evolution

- Causes of changes in population sizes (conservation biology).
- Effects of pollution on organisms and ecosystems.
- Evolution of traits in populations over time.
- Global environmental changes and changes in population sizes or species diversity.

#### Genetics & Molecular Biology

- Identifying genes that influence traits (e.g., breast cancer, cystic fibrosis).
- Nature vs. nurture (heredity and environment effects on phenotypes).
- Multiple sequence alignment in comparative genomics.

#### Agriculture

- Fertilizer effects on plant growth and productivity.
- Compare farming and harvesting methods (e.g., organic vs conventional farming).
- Compare plant hybrids for differences in productivity.

Here's a web site that keeps track of statistics jobs: [Jobs in Biostatistics](#). I would go on to add that experience and competence in statistics would also translate to employment in non-biology fields, e.g., [business analytics](#).

### Conclusions

Moving forward, we have much to do — you will be exposed to many specific examples of statistical tests, how to calculate estimators, and how to make inferences from experiments. An important goal of this course is for you to be introduced and develop your ability to design experiments. Why should you, as biologists and future health care providers, learn biostatistics?

1. Develop statistical reasoning skills. Most, if not all graduate students will need to take several courses in statistics.
  - Statements about research findings, new and better products, sociological and political issues often depend in large part on some form of statistical analysis.
  - By learning a little about experimental design, sampling, and statistical testing, you will be much closer to being able to participate fully in these debates.
2. Most, if not all graduate students will need to take several courses in statistics.
3. Most, if not all jobs in biology require some training in statistics.

So, there's really no doing biology without at least some knowledge of statistics. You're getting a head start!

## Questions

1. Compare the table of contents for *Mike's Biostatistics Book* and our *BI 311 Workbook* against the key terms listed from the MCAT<sup>2020</sup> Skill 4 expectations. Which chapters do you think cover the key terms in the MCAT expectations?
2. Find and copy definitions for **data processing** and **data cleaning** from
  1. one peer-reviewed, primary source\* (e.g., search Google Scholar).
  2. one peer-reviewed, secondary source (e.g., search Google Scholar).
  3. Wikipedia. From these three sources you collected, write your own definitions for data processing and data cleaning.(\* Not sure what is meant by "sources in science?" Search the phrase in Google 😊)
3. In what field or discipline do you see yourself studying or working by the year 2030? What are the data and analytical skills needed for this field? Cite your source (blogs are fine for this).

---

This page titled [2.1: Why \(Bio\)Statistics?](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michael R Dohm](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.