

1. Introduction

In these notes, we'll try to follow the book fairly closely throughout the semester. You may occasionally find some things in the notes and not in the book and vice versa. First, let's set the stage for the semester by giving a broad overview of statistics. What is statistics? Here's a "textbook" definition:

Definition: Statistics. Broadly speaking, statistics is the art and science of collecting, describing, analyzing, and interpreting data. (Gee, thanks!)

The field of statistics may be categorized into 3 main areas:

- 1) **Data Collection.** Chapter 22 of Principles of Biostatistics (POB) gives a brief introduction into this. While this is a very important aspect of statistics, we will not spend much time on it. We will typically use existing data sets and concentrate on the next 2 areas.
- 2) **Data Analysis.** The use of graphical and numerical summaries to describe and explore data.
- 3) **Inference.** Often, we are not only interested in the objects (e.g., patients) from which we've obtained our data, but would like to use our data to draw conclusions (i.e., make inferences) about a larger collection of objects.
 - a) E.g., Study the response of a group of patients to a new drug to help gain information on the effectiveness of the drug for all potential users of the drug.

Roughly, chapters 2-5 of POB are concerned with data analysis. Chapters 6-8 will give us some tools to help us use our data to make inference about a larger collection of objects

than the group of objects from which our data was collected. The remaining chapters (except Chapter 22) are mostly devoted to making inference, but also include elements of previous chapters.