1

INTRODUCTION

Unit 3A: Probability

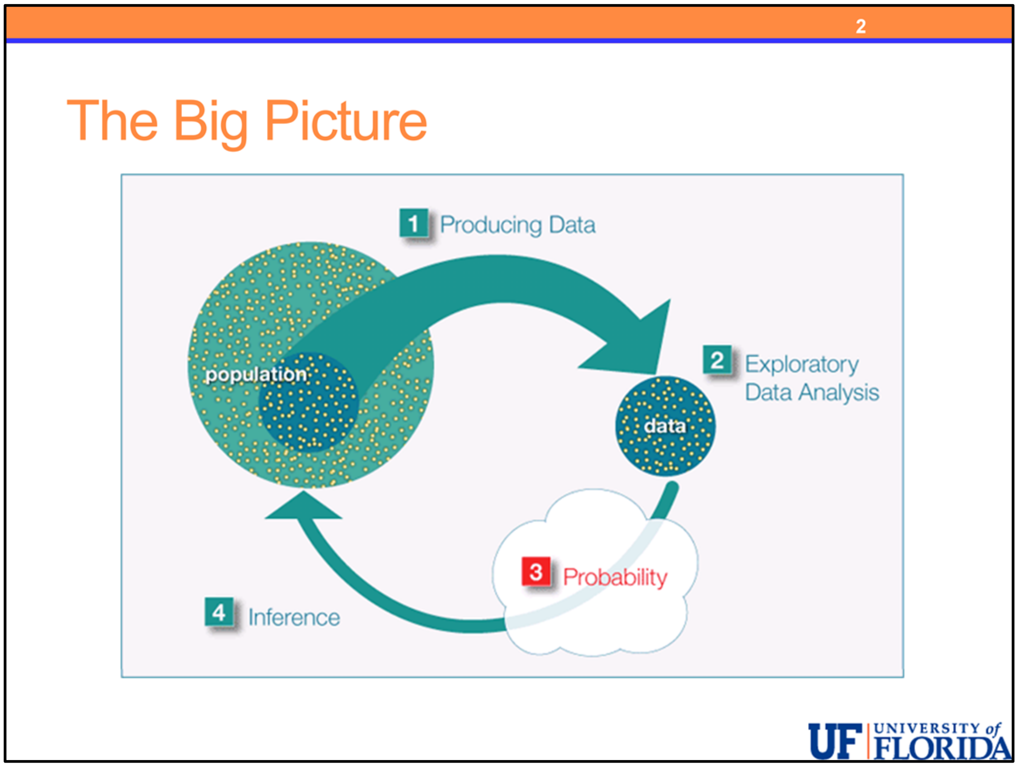
UF UNIVERSITY of FLORIDA

Now we begin our discussion of probability.

As we do, I want to begin by pointing out that probability is an extremely vast subject.

In Unit 3A, we will look only at a few basic probability concepts and rules.

Focus on understanding the examples we present and try not to make your current study of probability more difficult than necessary.



In this course, we are on a conceptual journey toward understanding the process of statistical inference presented in the big picture.

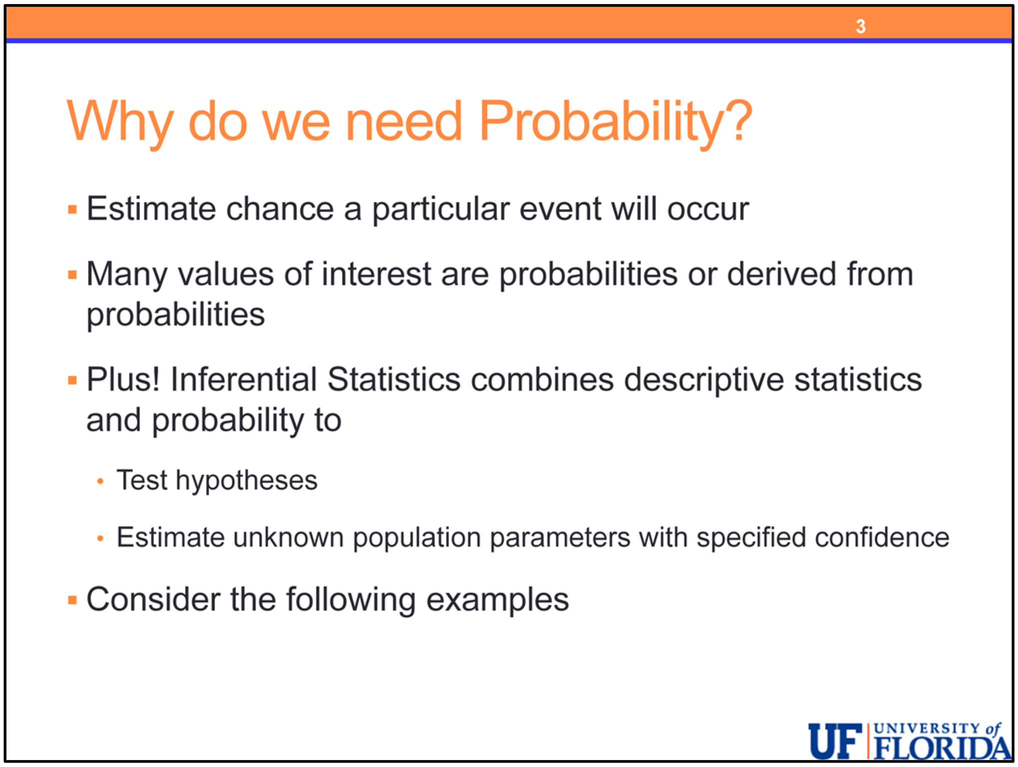
So far, we have discussed producing data and exploratory data analysis (although in the opposite order).

In producing data we discussed methods for obtaining data and issues that need to be considered about the process of sampling and study design.

In exploratory data analysis, we looked at methods that allow us to explore our data descriptively - featuring visual displays and numerical summaries. We learned to explore, describe, and compare distributions leading to methods of investigating relationships between two variables.

Our goal is inference where we will learn to draw conclusions about our population based upon the results from our single sample but before we can do that, we will need learn about probability which is the underlying foundation for statistical inference.

In Unit 3, we break down our study of probability into two parts. In Part A, we will look at basic probability rules. In Part B, we will extend our knowledge of probability to the realm of random variables, probability distributions, and sampling distributions which are all fundamental to the process of statistical inference.



So why do we need probability? In general, probabilities estimate the chance that a particular event will happen.

There are obviously many situations where we want to estimate the probability that something will happen; the chance that something will occur.

- The chance that you'll survive this heart attack
- The chance that you'll die before fifty

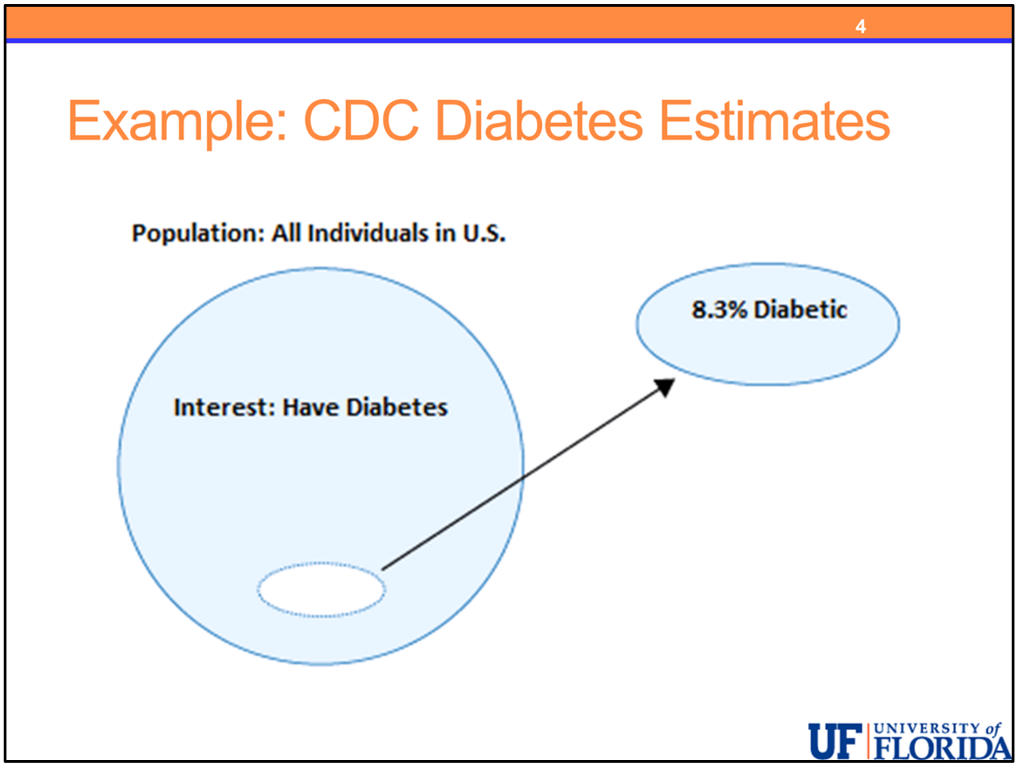
Whatever the chance is that you want to calculate.

In the Health Sciences, many values we wish to estimate are probabilities or are directly derived from probabilities.

Sensitivity and specificity are probabilities
Relative risk is derived from two probabilities
Odds ratios are actually derived from four probabilities.

It's just a matter of what it is that you're looking to do.

And also a main reason we're covering probability in this course is that inferential statistics requires us to understand some aspects of probability since it uses probability to help us test hypotheses and estimate unknown population parameters.



The CDC estimates that 8.3% of individuals in the US have diabetes.

How does such a number come about?

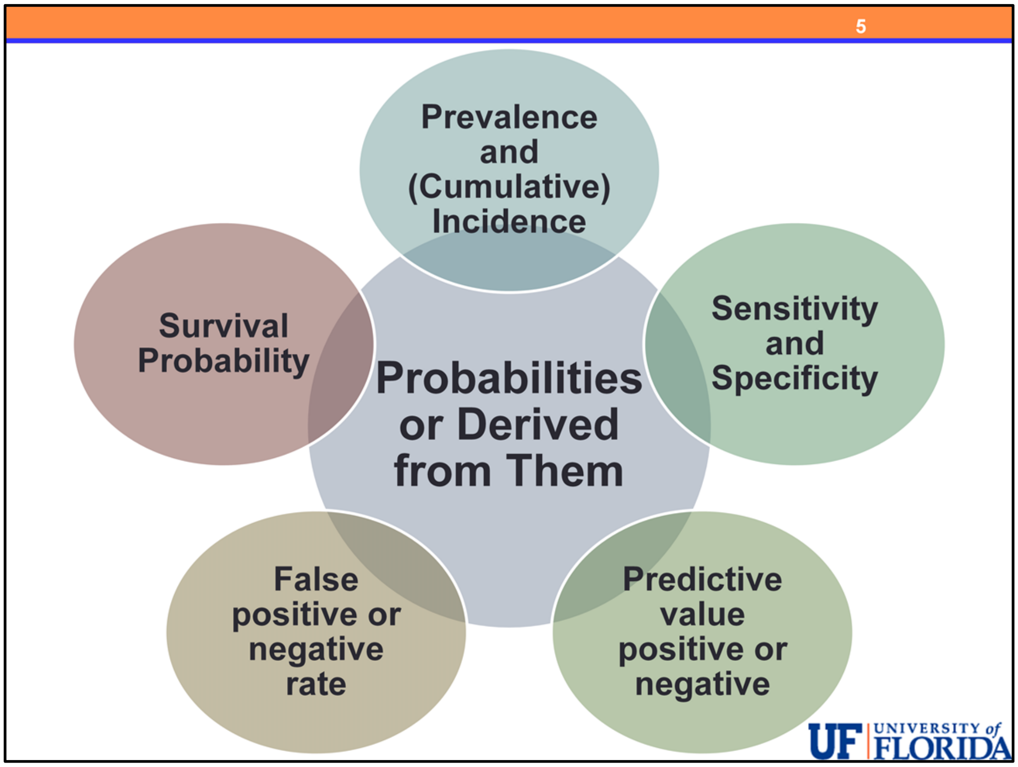
The CDC conducts surveys of the US population. These surveys are often more complex than the simple random sampling plans we will focus on this semester, however, the ideas are the same.

Our goal here is inference — to learn and draw conclusions about the prevalence of diabetes in the entire U.S. population based upon the individuals in this sample.

Can we conclude that 8.3% of the population is diabetic? Another sample could give a very different result. So we are uncertain.

We will learn to use probability to describe the likelihood that our sample is within a desired level of precision. For example, probability can answer the question, “How likely is it that our sample estimate is no more than 1% from the true percentage of the U.S. population with diabetes?”

The answer to this question (which we find using probability) is obviously going to have an important impact on the confidence we can attach to any inferences we make about our population using this sample.



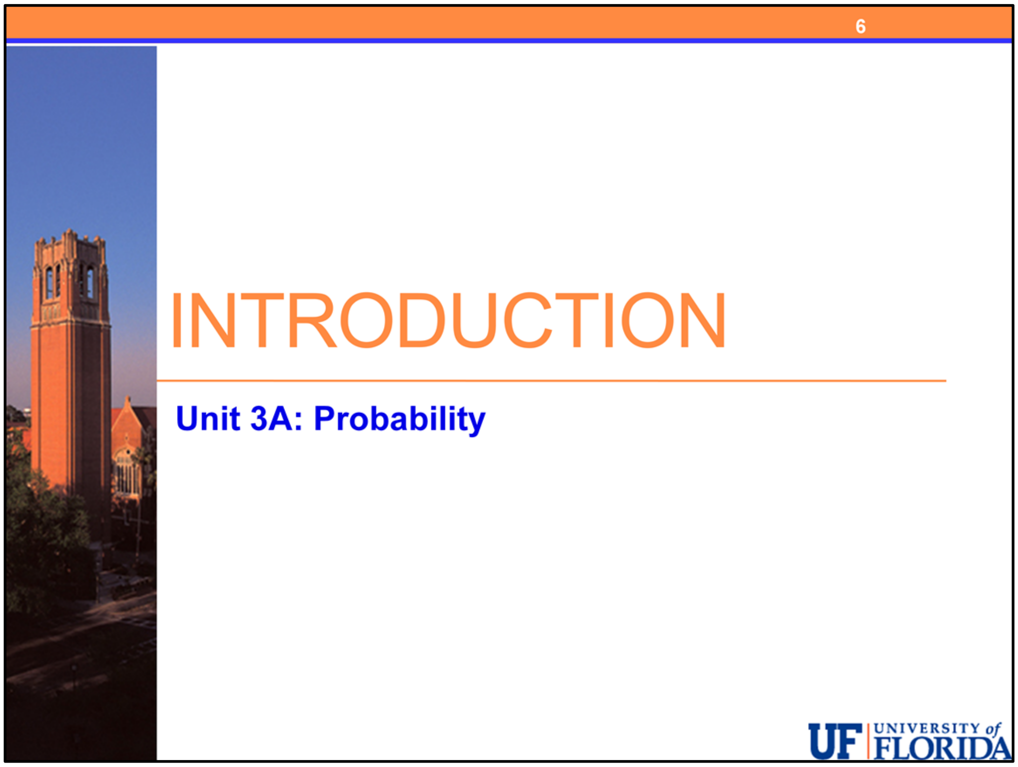
A few other probabilities to get you thinking about the applications of probability in the Health Sciences are:

- Prevalence which we've already mentioned, is defined as the probability of currently having the disease or condition regardless of the duration one has had the disease, so that's just how many people have the condition in the whole population right now at this moment.

- And cumulative incidence is the probability that a person with no prior disease or condition will develop that disease or condition during a certain time period. And we don't usually say cumulative incidence. It's almost always just shortened to incidence. So, that's why that's in parentheses. In the CDC's fact sheet, you can see a statement that is an example of incidence on page 6: "Women who have had gestational diabetes have a 35% to 60% chance of developing diabetes in the next 10–20 years."

We have already briefly mentioned that Sensitivity and Specificity of a diagnostic test are probabilities. As are the predictive value positive and negative as well as the false positive rate and false negative rate. These are measures which are commonly covered in epidemiology courses but we may discuss them on occasion.

Another common example is a survival probability which is the probability that an individual survives beyond a certain time. For example, the 5 year survival probability following treatment for breast cancer.



Next we will begin our formal study of basic probability.