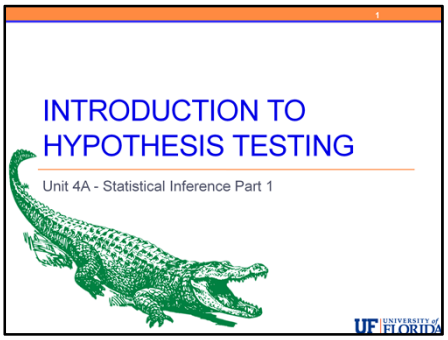
INTRODUCTION TO HYPOTHESIS TESTING

Unit 4A - Statistical Inference Part 1



Now we will begin our discussion of hypothesis testing. This is a complex topic which we will be working with for the rest of the semester.

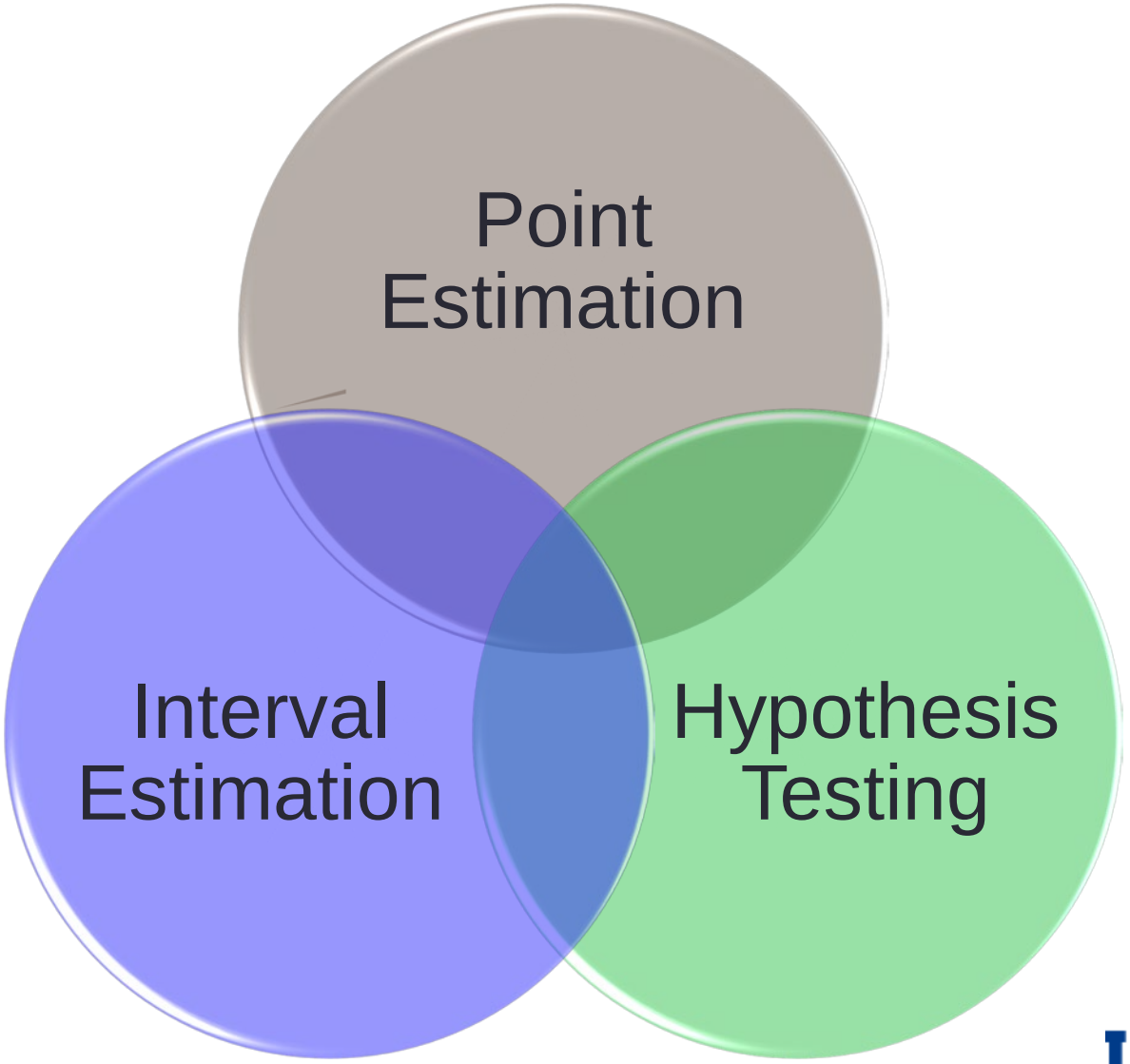
If this is your first statistics course (or even if it is not) you may need to spend a good amount of time on these ideas over the next few weeks to become comfortable with them.

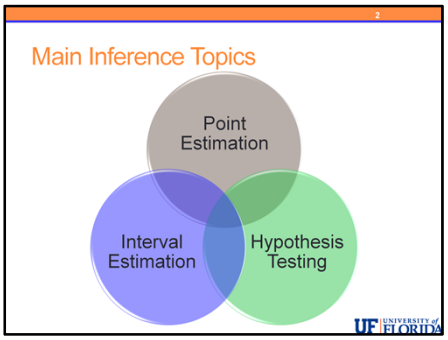
The materials are presented so that you will cover a good bit of information now but we don't expect it to all make sense immediately.

The materials in Unit 4B will offer opportunity to continue developing your understanding of estimation and hypothesis testing in more realistic situations where we will rely on software to perform the calculations.

For now, we will still be working by hand to develop some instincts and understanding about this process – for hypothesis testing we will focus on the case for one population proportion to illustrate by hand and then allow software to help us in all other situations.

Main Inference Topics





We have discussed point estimation, where we estimate an unknown parameter using a **single number** calculated from the sample data and

Interval Estimation where we estimate an unknown parameter using an **interval of values** which

- Quantifies the magnitude of estimation error – with a certain confidence level
- And is likely to contain the true value of that parameter

Now we are moving into much different territory. For hypothesis testing, the idea, logic, and goal are quite different.

In Hypothesis Testing: we begin with a claim about the population, and we check **whether or not the data** obtained from the sample **provide evidence AGAINST this claim.**

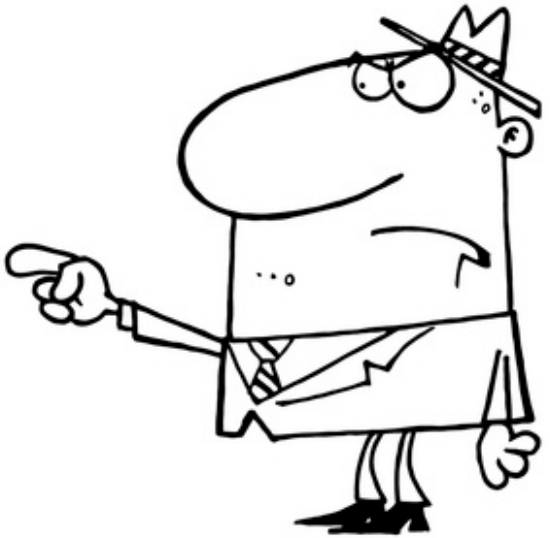
Logic of Hypothesis Testing

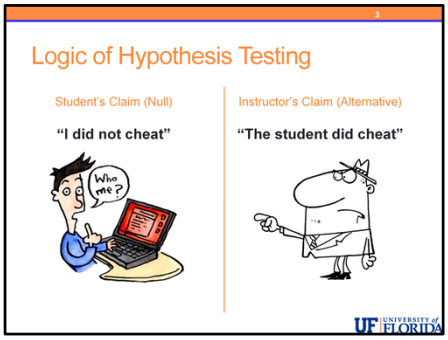
Student's Claim (Null)

"I did not cheat"

Instructor's Claim (Alternative)

"The student did cheat"





For now we want to focus on the concepts and not the details of any particular test.

Consider the situation where a case of suspected cheating on an exam is brought before the disciplinary committee of a university.

We have two opposing claims.

- The **student's claim:** I did not cheat on the exam.
- The **instructor's claim:** The student did cheat on the exam.

Adhering to the principle **"innocent until proven guilty,"** the committee asks the instructor for **evidence** to support his claim.

In this situation, the student's claim is assumed to be true unless sufficient evidence can be brought forth. Note that this will be called the null hypothesis in our statistical definitions.

The opposing claim by the instructor requires evidence to support it in order to reject the student's claim. This will be called the alternative hypothesis.

The instructor explains that the exam had two versions, and shows the committee members that on three separate exam questions, the student used in his solution numbers that were given in the other version of the exam.

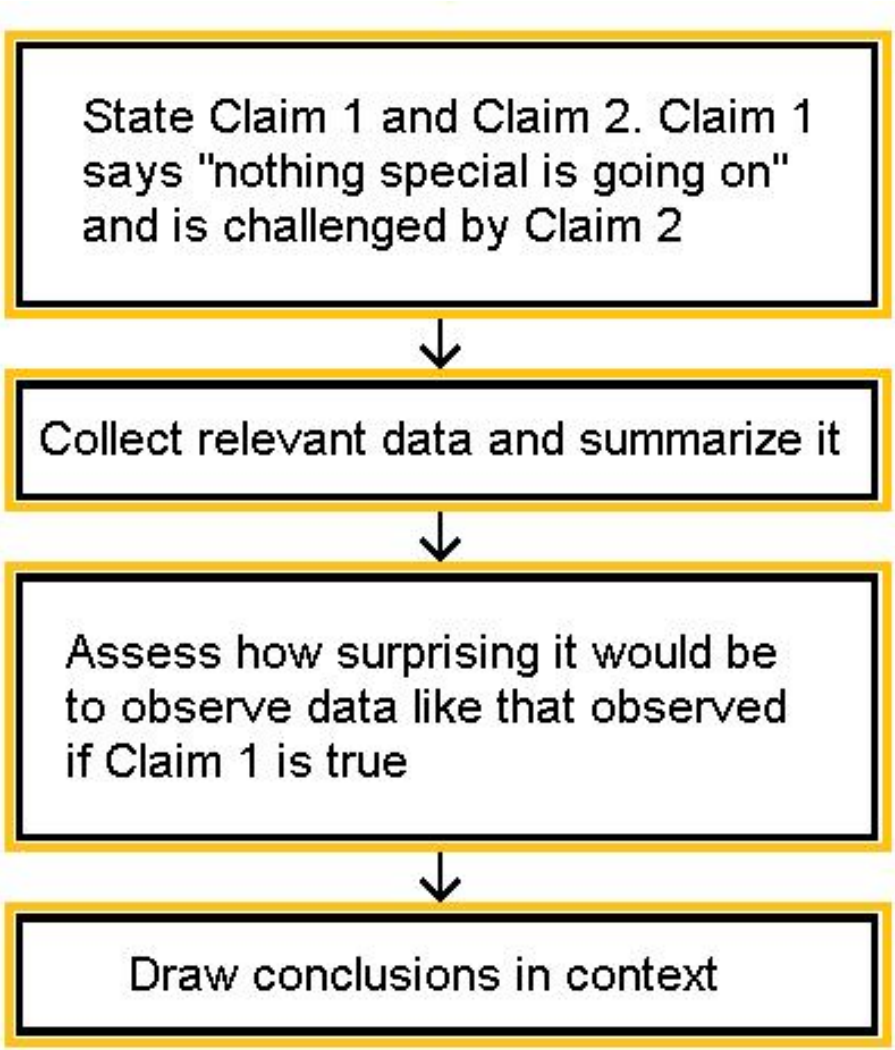
Is this enough evidence?

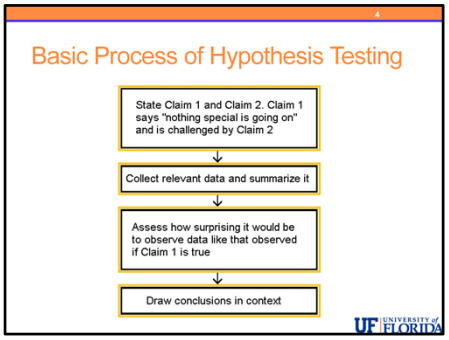
The committee members all agree that **it would be extremely unlikely to get evidence like that if the student's claim of not cheating had been true.**

In other words, the instructor brought forward strong enough evidence to reject the student's claim, and conclude that the student did indeed cheat on the exam.

Although we will be using data to make these decisions and our hypotheses will have very specific forms for each scenario, this example illustrates the logic of hypothesis testing nicely.

Basic Process of Hypothesis Testing





Here is how the process of statistical hypothesis testing works:

- We have **two claims** about what is going on in the population. Let's call them **claim 1 (this will be the null claim or hypothesis)** and **claim 2 (this will be the alternative)**.
    - Similar to our example where the student's claim is challenged by the instructor's claim, in general the null, claim 1 is challenged by the alternative, claim 2.
    - In statistics, these claims are about the value of a population parameter.

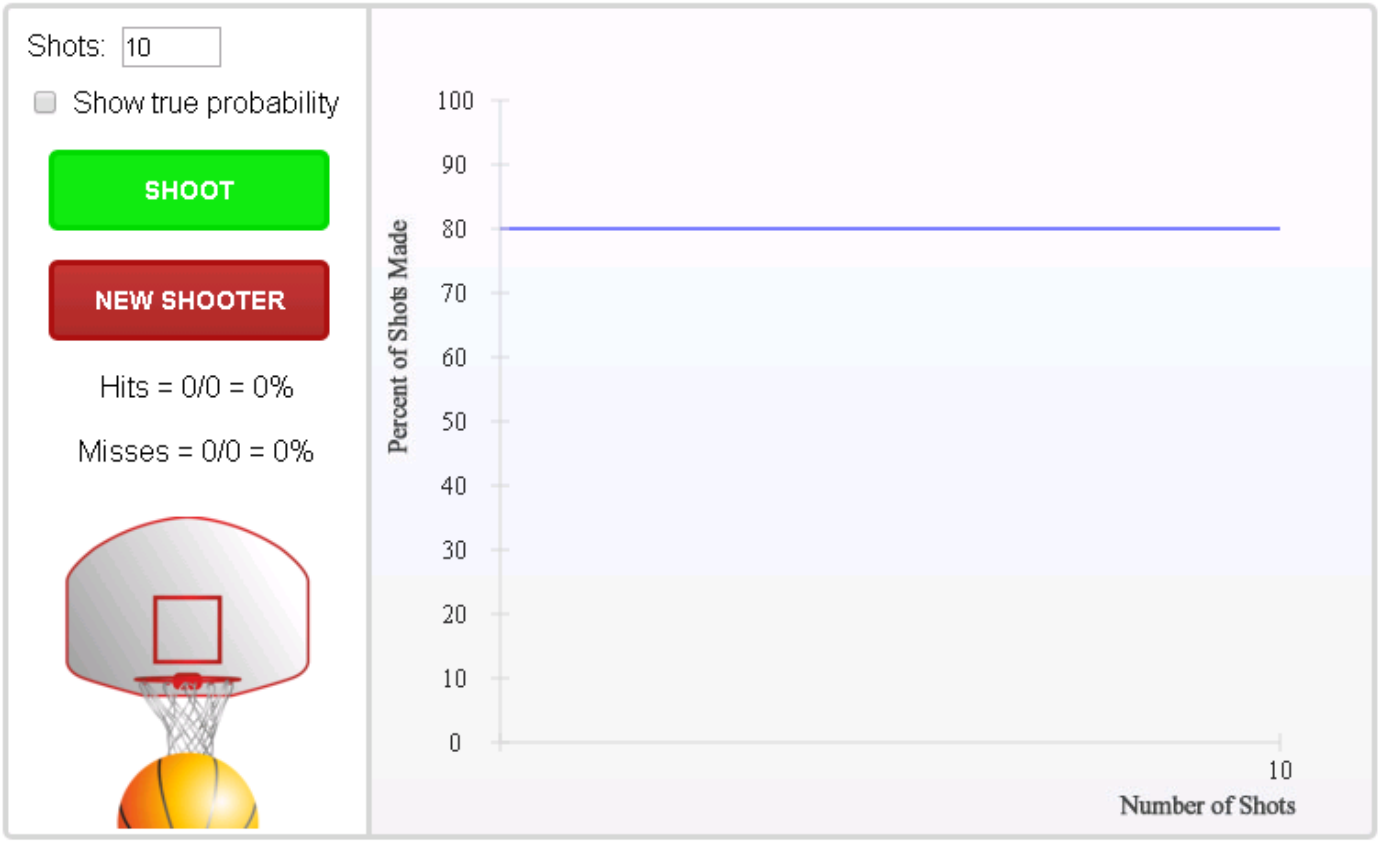
- We choose a sample, collect relevant data and summarize them (this is similar to the instructor collecting evidence from the student's exam). For statistical tests, this step will also involve checking any conditions or assumptions.

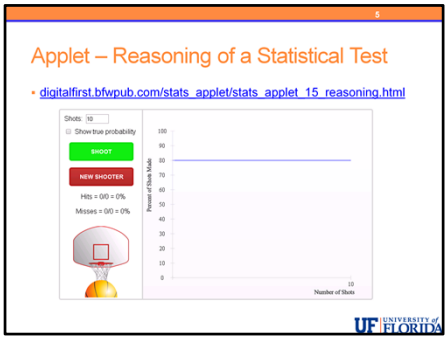
- Then, we figure out how likely it is to observe data like the data we obtained (or more extreme), if claim 1 is true. (Note that the wording "how likely …" implies that this step requires some kind of probability calculation).
    - In the story, the committee members assessed how likely it is to observe evidence such as the instructor provided, had the student's claim of not cheating been true.

- Based on what we found in the previous step, we make our decision:
    - If, after assuming claim 1 is true, we find that it would be **extremely rare** to observe data as strong as ours or stronger in favor of claim 2, then we have strong evidence against claim 1, and we reject it in favor of claim 2. Later we will see this corresponds to a small p-value.
    - If, after assuming claim 1 is true, we find that observing data as strong as ours or stronger in favor of claim 2 is **NOT VERY RARE**, then we do not have enough evidence against claim 1, and therefore we cannot reject it in favor of claim 2. Later we will see this corresponds to a p-value which is not small.
    - In our story, the committee decided that it would be extremely unlikely to find the evidence that the instructor provided had the student's claim of not cheating been true.
    - In other words, the members felt that it is extremely unlikely that it is just a coincidence (random chance) that the student used the numbers from the other version of the exam on three separate problems.
    - The committee members therefore decided to reject the student's claim and concluded that the student had, indeed, cheated on the exam.

Applet – Reasoning of a Statistical Test

- digitalfirst.bfwpub.com/stats_applet/stats_applet_15_reasoning.html





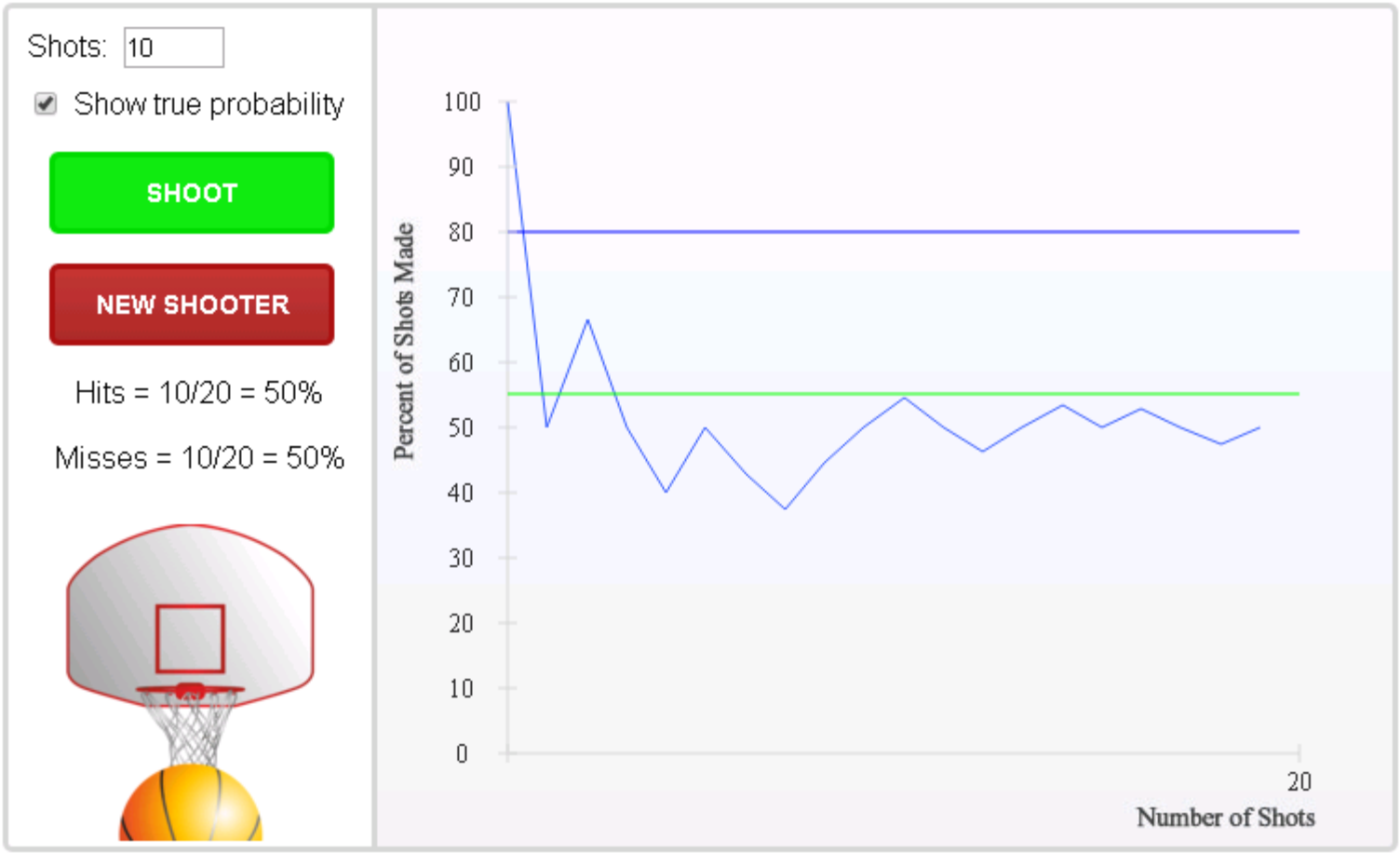
http://digitalfirst.bfwpub.com/stats_applet/stats_applet_15_reasoning.html

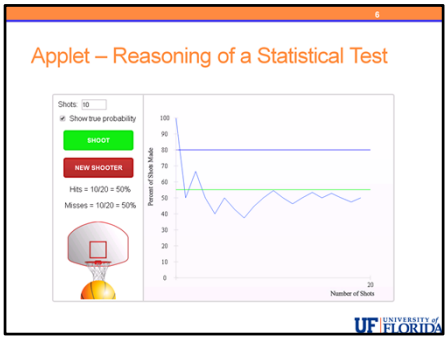
We will use an applet to illustrate how we might make decisions from gathered data.

We are asked to determine if the player's free throw percentage is 80% or not 80%.

To begin we click shoot – by default it will take 10 shots at a time.

Applet – Reasoning of a Statistical Test

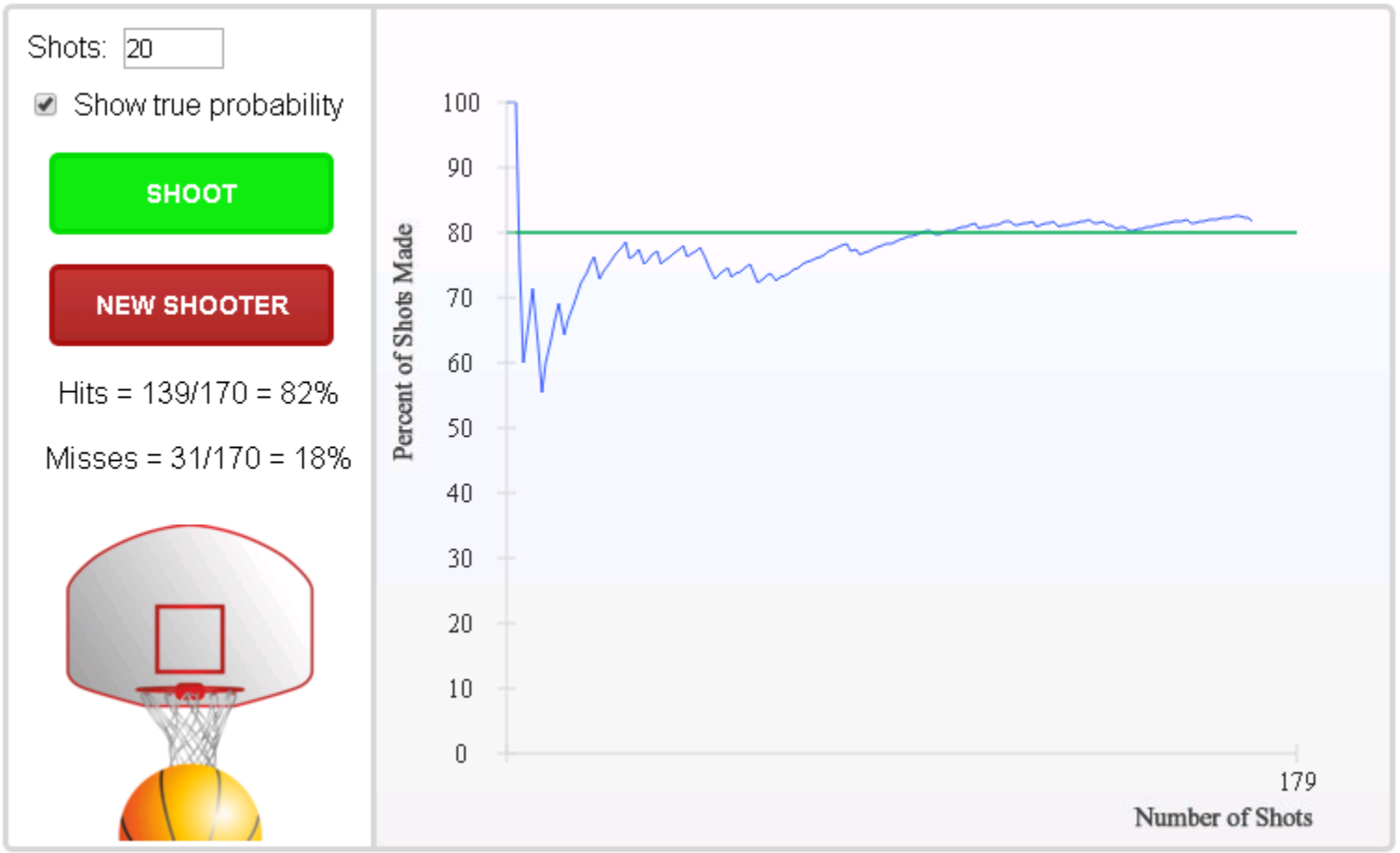


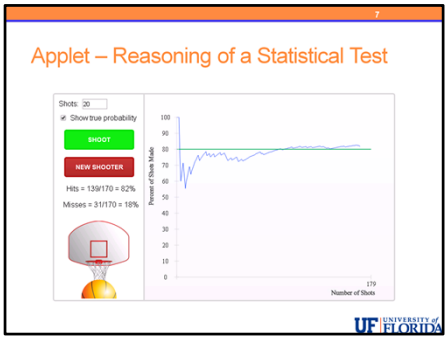


In this example, after around 20 shots, I was willing to conclude that this player's free throw percentage is not 80%.

I clicked the box to show the true probability which seems to be about 55%.

Applet – Reasoning of a Statistical Test



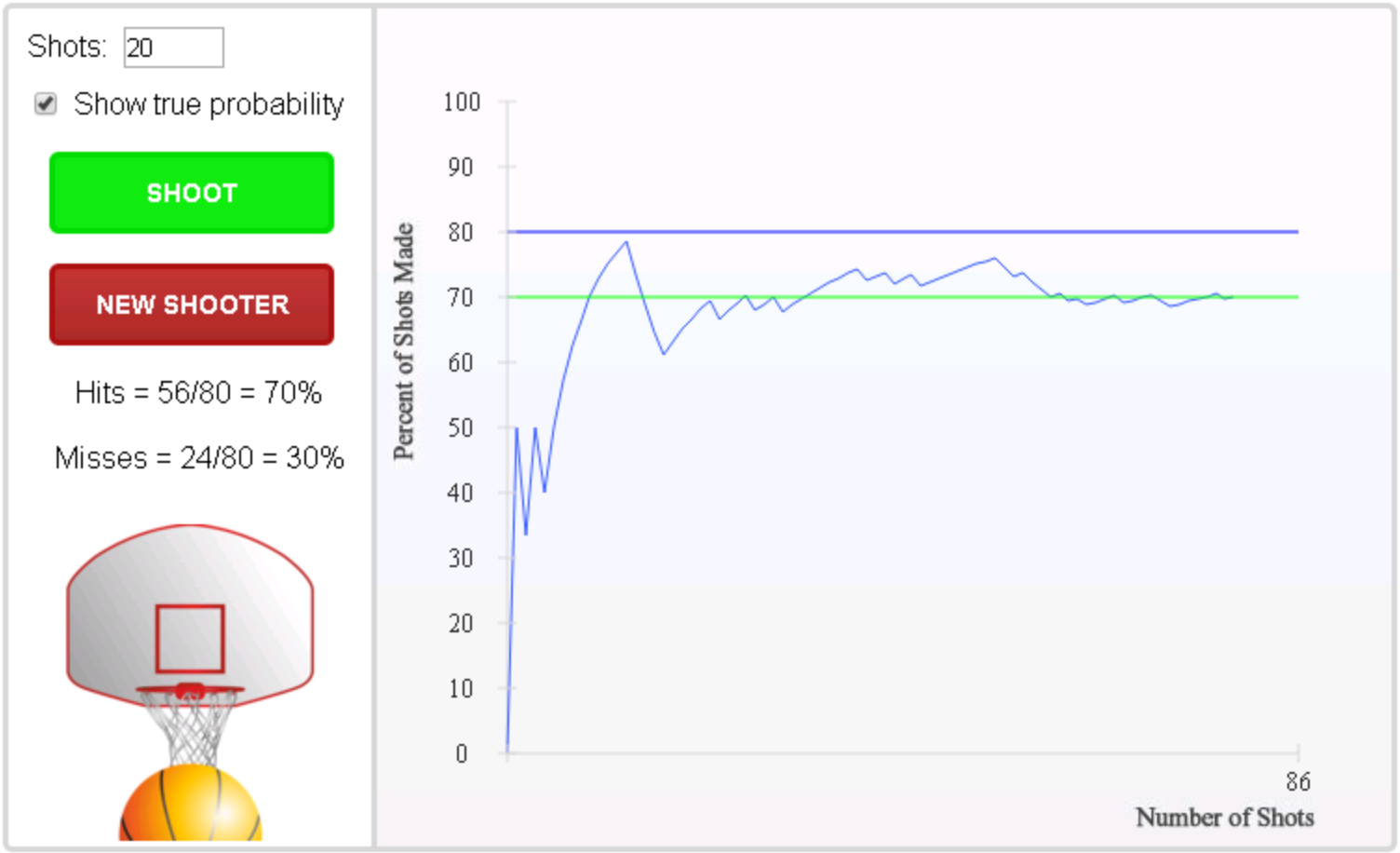


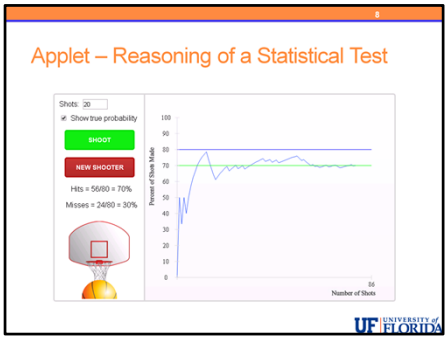
You should find that it will take much more shots to feel comfortable concluding that it absolutely IS 80% and maybe you will never find it easy to make this conclusion without checking the box to see the true value.

In statistics we will never be able to prove this claim of equality.

Here I took 170 shots but thought that maybe it was a little less than 80 early on and revealed the line to myself too soon!

Applet – Reasoning of a Statistical Test





You should also find that it will take more shots to feel comfortable concluding that it absolutely different from 80% the closer the true value is to this claim of 80%.

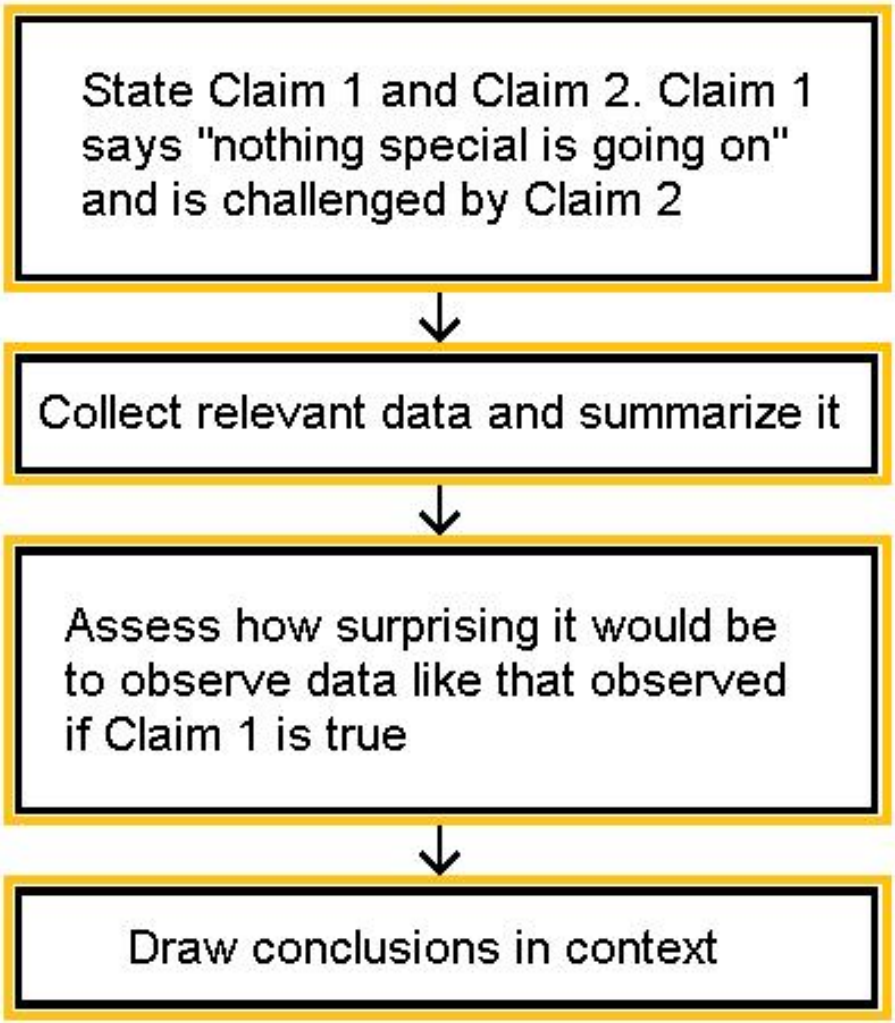
Here I took 80 shots when the true value was 70%.

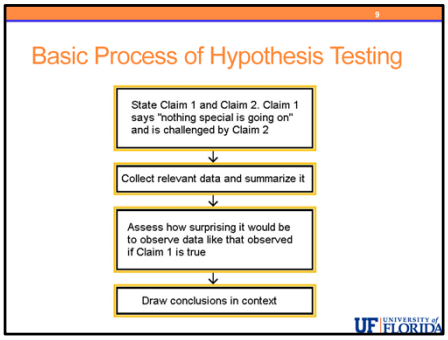
This applet help's illustrate the kind of evidence data can provide.

We will learn to calculate a probability on which we will base our conclusions – here we are just relying on our subjective view of the data and likely if you repeat this applet yourself a few times, you may not always make the correct choice.

We will soon learn that statistical hypothesis testing also has the potential to make the wrong decision too.

Basic Process of Hypothesis Testing





There are more examples in the materials. Review them carefully and pay attention to the wording in the conclusions.

Note that there are two types of conclusions:

- "The data provide enough evidence to reject claim 1 and thus we accept claim 2";

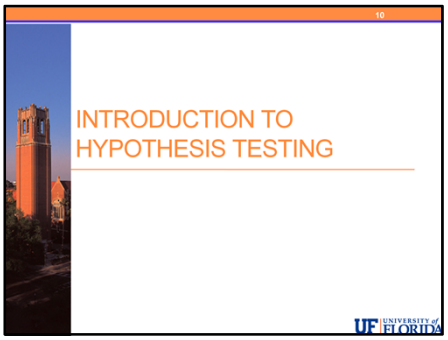
OR

- "The data do not provide enough evidence to reject claim 1."

In particular, note that in the second type of conclusion **we did not say:** "**I accept claim 1**," but only "**I don't have enough evidence to reject claim 1**."

We will come back to this issue later, but this is a good place to begin to make you aware of this subtle difference.

INTRODUCTION TO HYPOTHESIS TESTING



That gives an overview of the ideas and logic of hypothesis testing.

Next we will cover the steps in more detail.