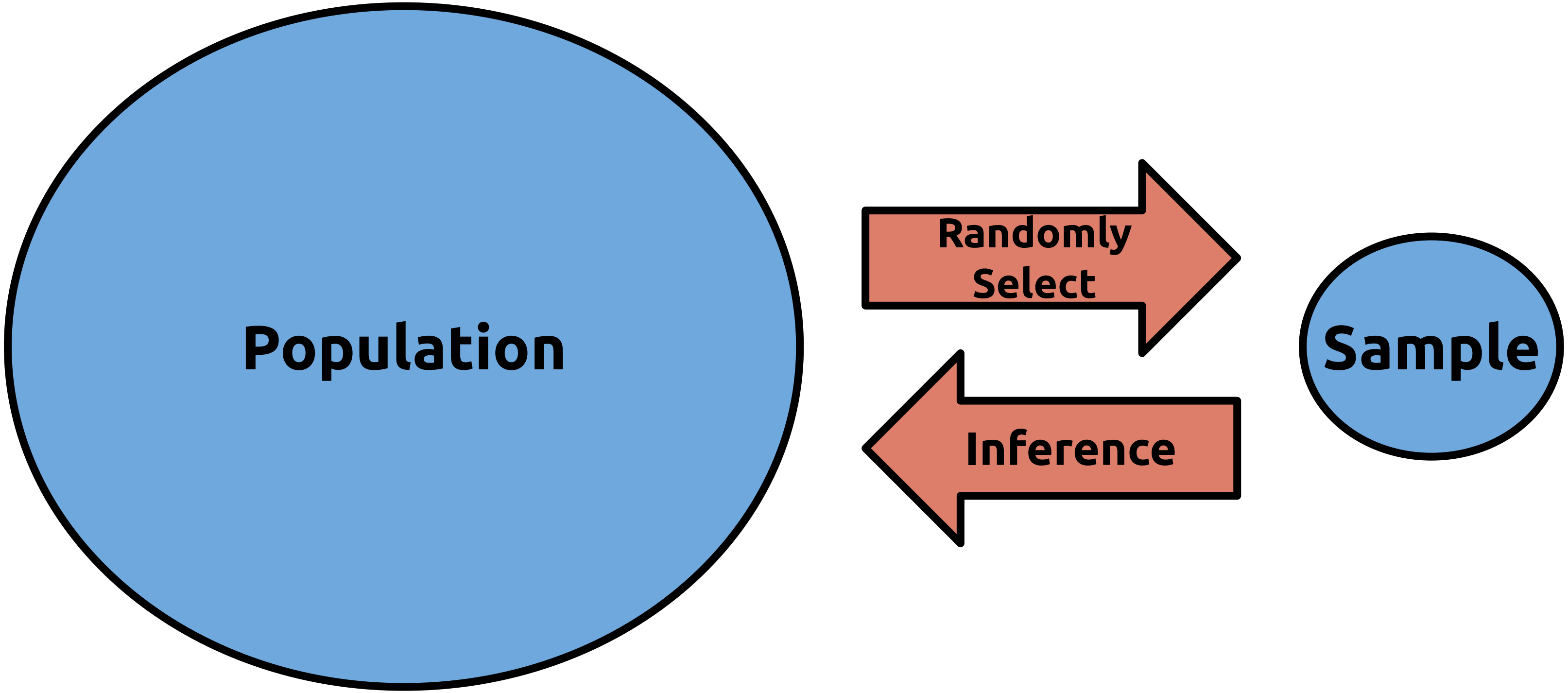
Introduction to Hypothesis Testing

Recall: Populations and Samples



Hypothesis Testing

Goal: Test whether some hypothesis about a population parameter is true, by inspecting only a sample.

Remember that we don't get to know whether we are right or not because we don't get to see the entire population. Instead we are only looking at a subset.

Since we are looking at a subset, we'll have to estimate the parameter of interest using the sample, but since we have a sample, there will be some variance/randomness.

Hypothesis Testing

We will be trying to answer the question “Given a sample and an apparent effect, what is the probability of seeing such an effect by chance?”

What we are testing for is **statistical significance.**

A set of measurements or observations is said to be statistically significant if it is **unlikely to have occurred by chance**.

Hypothesis Testing

Example: I have a coin which I believe to be fair, meaning that it is has the same chance of landing on heads as it does of landing on tails.

How can I test this?

One option is to flip it some number of times, say 100 times, and observe what happens.

Population of interest: All possible tosses of this particular coin

Sample: The 100 coin tosses that I record.

Hypothesis Testing

Before I can proceed, I need to decide what my default position is. Since I have reason to think otherwise, I am going to assume that I do in fact have a fair coin.

I will only change my mind if my test reveals very compelling evidence - evidence so compelling that I would feel silly not to change my mind.

I will not change my mind about the coin being fair unless my test reveals something that is very unlikely to happen just due to chance.

In hypothesis testing, this is known as the **null hypothesis**, or H_0

If I see compelling evidence to change my mind, I will instead adopt the **alternative hypothesis,** H_A

Hypothesis Testing

Scenario 1: I flip the coin 100 times. It lands on heads 47 times and tails 53 times.

Is this a compelling reason to change from my default position that the coin is fair?

Not particularly. Flipping a coin is a random process. There will be variability in the proportion of times that it lands on heads, and a fair coin landing on heads 47 out of 100 times is not that unusual.

Here, I do not have enough evidence to reject the null hypothesis.

Note: that I haven't *proven* the null hypothesis; I've just not

Hypothesis Testing

Scenario 1: I flip the coin 100 times. It lands on heads 47 times and tails 53 times.

Here, I do not have enough evidence to reject the null hypothesis.

Note: that I haven't *proven* the null hypothesis; I've just not rejected it.

Hypothesis Testing

Scenario 2: I flip the coin 100 times. It lands on heads 38 times and tails 82 times.

Is this a compelling reason to change from my default position that the coin is fair?

While it is *possible* for a fair coin to land on heads only 38 times out of 100, how *probable* is it?

We could compute the probability (using a binomial distribution), but I think it's safe to assume that it is improbable enough that I would be silly to not conclude that the coin is not fair.

Hypothesis Testing

Scenario 2: I flip the coin 100 times. It lands on heads 38 times and tails 82 times.

In this scenario, I'm much more skeptical that the coin is fair.

I will reject the null hypothesis, in favor of the alternative hypothesis that the coin is <u>not</u> fair.

Again, I have not proven anything. We cannot know with absolute certainty that the coin is not fair, but our evidence does not support the null hypothesis being true.

Hypothesis Testing - Types of Errors

What could have gone wrong in the above example?

In scenario 2, we rejected the null hypothesis in favor of the alternative hypothesis.

If the coin really was fair, and we just had a particularly unlikely run of coin flips, then we would have committed what is called a **Type I Error**. That is, we incorrectly rejected the null hypothesis.

The coin was fair, but we concluded that it was not.

Hypothesis Testing - Types of Errors

What could have gone wrong in the above example?

In scenario 1, we did not reject the null hypothesis. We would have been wrong if in reality the coin was not fair.

This is an example of a **Type II Error**. That is, failing to reject the null hypothesis when in reality it is false.

Hypothesis Testing - Types of Errors

		Reality	
		H_0 is True	H_0 is False
Our Decision	Do not Reject H_0	Correct Decision	False Negative / Type II Error
	Reject H_0	False Positive / Type I Error	Correct Decision

Hypothesis Testing

When doing hypothesis testing, we choose the null hypothesis H_0 so that it serves as the “default decision”.

This means that in the absence of compelling evidence, we can feel good about falling back on the null hypothesis.

The null hypothesis gives a skeptical view of what we are trying to show.

We can think of a hypothesis test as being like a trial. Our default decision is to find the defendant *not* guilty unless the prosecution can present compelling enough evidence to change our minds.

Hypothesis Testing - Types of Errors

		Reality	
		H_0 is True: Not Guilty	**H_0 is False: Guilty**
Our Decision	**Do not Reject H_0: Not Guilty**	Correct Decision	False Negative / Type II Error
	Reject H_0: Guilty	False Positive / Type I Error	Correct Decision

Hypothesis Testing

The way that hypothesis testing is done, the goal is to avoid a Type I error.

That is, we don’t want to conclude that there is some effect when there is none, just like we wouldn’t want to incorrectly find a person who is not guilty to be guilty.

The data must show us compelling enough evidence to reject the null hypothesis.