

LECTURE 01: INTRODUCTION & THE BASICS OF SAMPLING

- I. What is statistics?
 - a. *Statistics* is the application of mathematics to understanding and collecting of sample data.
 - i. Some of this application is through a controlled experiment. For example, determining if a focus group really likes a new product more than the current one, or if their higher opinion is just due to random chance.
 - ii. But statistics is also useful when you can't run a controlled experiment. Some of the most important questions in the social sciences—from business to economics to psychology—are addressed using statistical techniques we'll explore.
 - b. Consider one of the most important questions in economics: How do turn poor countries into wealthy ones?
 - i. Economists have lots of ideas, but we don't know what will actually work. For example, does giving the country's government a bunch of money help?
 - ii. Ideally the World Bank or IMF would randomly select half of the poor countries and then give that half some large amount of money (adjusted by population). Then we can look at the results.
 - iii. But that's not an option and not just because each country is so different. There are ethical and legal constraints. The struggling countries that didn't get anything would wonder why they are left out. And, by chance, some of that money would go to countries that we know are corrupt. Even if we learn a lot, it would be a short-term disaster.
- II. Why sample?
 - a. Look: whenever we want to figure something out, we want to know what's going on for all instances, not just a few.
 - i. Pepsi doesn't care how popular a new drink for a few people. They want to know how popular it will be for everyone.
 - ii. Scientists don't care that much about how a drug affects a few people. They want to know how it affects everyone.

- iii. Policy makers aren't interested if just a few criminals commit crimes after a rehabilitation program. They want to know how effective that program will be for all criminals.
- b. But checking a whole population is really hard. So we take a sample, or a subset of the population that, ideally, represents the population.
 - i. By *population* we mean all possible subjects of interest. Note this can include subjects which don't exist yet (like the future recipients of a drug treatment).
 - ii. A *parameter* is a characteristic about a population.
 - iii. A *statistic* is a characteristic about a sample.
 - iv. We care about this smaller size not because we're interested in how change affects just the sample but because the sample *represents* a larger population that we do care about.
- c. Sampling has a lot of advantages:
 - i. It's cheaper;
 - ii. It allows greater depth in questioning;
 - iii. It's faster;
 - iv. It's more practical (you have to use a sample for crash testing cars, or you'll smash all your cars and have none left to sell)