A MOTIVATING EXAMPLE

A motivating example

Recall that we emphasized that we should always start with a research question: In this lecture, we will consider a motivating example and discuss potential challenges in experimental design.

Comparing contact lenses for controlling myopia progression

Myopia has become a serious problem among young kids. Now suppose there are two types (A and B) of contact lenses that can potentially slow down the progression of myopia. You want to know which type is more effective. How would you design an experiment to compare the two types of contact lenses?

Can you write down the basic idea of your design? (1 minute)

A basic setup: need more details

I believe most of you will come up with a basic setup similar to the following: we can let two groups of kids try the two types of the contact lenses and then see which group has better results.

However, “the devil is in the details”.

On the next few slides, we will ask many questions to clarify more details. Through these questions, we try to understand the challenges and contemplate key elements and principles of experimental design.

How do we compare the results?

First, in order to compare the two types of contact lenses, there has to be something we can measure.

So one basic element of experimental design has to be **measurements of experimental outcomes**.

In this example, we can measure the change in the degrees of myopia before and after the trial.

Often times, we will measure more than one quantity and we can also consider measuring one or more quantities at multiple time points.

Why do we want to try each treatment on a group of people?

Why use two groups of kids? Can we just ask one kid to try type A lenses and another kid to try type B lenses?

- Our gut feeling tells us this is not good idea: if we try the lenses on two other kids, they may respond differently.

In other words, we believe there is variation among individuals: the individuals vary in their degrees of the myopia; more importantly, they vary in how they respond to the treatments (the two types of lenses).

To assess the variation, we need **replicates**: **replicating the treatments** on different individuals (called **experimental units**).

Understanding the **variation among individuals** is a key task of statistics. In the context of experimental design, we particularly need to understand the variation of outcomes among individuals receiving treatments.

But it is impossible to ask all kids to try the lenses!

That is correct: there is no way we can ask all myopia kids to try these two types of lenses.

But it is still important that we have a **target population** in mind when we design an experiment: the target population is the set of all individuals we want to apply the experimental conclusions to.

- In this case, the target population can be all myopia kids in a region that are potential users of the two types of lenses.

In an experiment, the best we can do is to select a **representative sample** (a subset) from all kids in the target population. The best way to achieve "representativeness" is through **random sampling**.

When analyzing the experimental outcomes, we will try to draw conclusions about (i.e., make statistical inference) the entire population based on what we see in a sample.

But that means each time I will get a different sample?

That is also correct. It is usually not practical to ask all individuals in the entire target population to participate in the experiment: in our example, that means to ask all myopia kids in a target region to try on the lenses.

Since we have to rely on a randomly selected sample (a sample is a subset of individuals from the target population), there will be **uncertainty associated with random sampling** in our experimental outcomes : if someone else were to repeat my experiment, he or she would most likely select a random sample that is different from the sample I selected—and thus he or she would see a different set of results.[1]

It is our task to help the people understand uncertainty (associated with random sampling) and its implications, and provide guidance for people to make decisions in the presence of uncertainty. We have to learn to **communicate** uncertainty.

Again, how do we compare the results—from two groups of individuals?

Now we have two groups of kids trying on the two types of contact lenses. Different kids will have different outcomes—even when they are wearing the same type of lenses.

What do we mean, then, that "one group has better results" than the other?

We have to be able summarize results on a group/set/collection of the individuals. For this, we will need to

1. understand the sample and population **distributions** of a measurement/quantity (e.g., change in the myopia degree).
2. learn how to compare two distributions and interpret the results.

What if I assigned all the "good kids" to one group–by chance?

This is a great question! What if, just by chance, the two groups of kids—besides wearing different types of lenses—also differ in some other ways? By the end of the experiment, how do I know whether the differences I see is due to the lenses or due to other factors?

For example, suppose we know that myopia tends to progress faster among younger kids. If in our experiment, we assigned more younger kids to try type-A lenses, it may lead to worse results for type-A lenses.

Factors—other than the measured treatment factors—can be associated with the treatment factors and at the same time potentially impact the experimental outcomes. Such factors are called **confounding factors**.

I will give you 1 minute to think about this question and suggest a solution.

Blocking

If we know age will potentially impact the experimental outcomes. One possible solution is to **block** by age groups: we will sort the candidates by age groups, and within each age group, assign the two types of lenses to two sub-groups and compare results.

Blocking is a good idea and will work for known confounding factors. For example, for health studies, age, gender, ethnic groups are "the usual suspects" of confounding, and we often block on those factors when designing an experiment.

But what about potential confounding factors that are hidden, unmeasured, or unknown to us ahead of time?

Randomization

When Fisher invented the experimental design almost 100 years ago, that is exactly the question he worried about.

Before Fisher, in crop field experiments, generations of field workers tried extremely hard to eliminate all possible differences between different fields before they start experiments comparing different treatments (crop varieties, fertilizers, etc.) on the fields.

Fisher pointed out, ingeniously, it is not necessary that all individual units receiving treatments are as similar as possible at the beginning of an experiment, by **randomly** assign individual units to the different treatment groups, we can obtain valid comparisons of the treatments even if there is variation.

Randomization, blocking, replication

Fisher also provided answers to questions we asked earlier in his pioneering work in experimental design. We often summarized Fisher's basic principles of experimental design as **randomization, blocking and replication.**

In this class, we will study these principles as we explore different commonly used designs.

More questions?

As you can see, as we ask more questions, we start to see it is not a simple task to compare two treatments in a fair (unbiased) and effective (if there is a true difference, we want to detect it) way.

In this class, we will learn the basic principles and commonly used methods for designing and analyzing experiments. Along the way, we will give more detailed answers to the questions we have raised.

At this point, you are encouraged to think about more potential issues that will affect the fairness and/or effectiveness of an experiment.