Experimental Design

Econ 176, Fall 2014

Some Terminology

Subject: A single human being we are doing observations on.

Session: A single meeting at which observations are made on a group of subjects.

Treatment: A set of conditions subjects are being observed under (often a single economic system).

Experiment: A collection of treatments meant to answer a question.

Types of Inference

We usually have two different types of questions that we might ask in a economics experiment.

Measurement: Measure behavior or preferences in some type of environment.

Example: How close is price to a competitive equilibrium in a double auction?

Comparative Statics

Example: How does a change in the number of traders affect the efficiency of a double auction market?

Although all experiments involve measurement, most of them are comparative statics inquiries.

Factors and Levels

When we do comparative statics experiments, we are usually adjusting the level of a factor.

Factor: A variable hypothesized to affect the outcome of an experiment.

- The “cooperation” payoff in a prisoner’s dilemma.
- The number of traders in a double auction.
- Whether a game is repeated or not.
- The auction institution used to sell a good.

Level: A setting for a factor.

- High versus Low.
- 2, 3, 4, 5, 6.
- Yes or No.
- First price auction, Dutch auction, second price auction, English auction

Our Battle With "Noise"

In experimental economics, we often want to measure how the level of a factor *x* affects an outcome *y*.

Ex: What happens to an equilibrium price when we shift out supply.

When we measure an effect we have to deal with to types of causes of behavior..

- **The direct effect of the level of factor x on y.**

  Ex: How supply would affect offers if everything else stayed the same.

- **Factors that have nothing to do with the change in supply.**

  Ex: Subjects on a particular day didn't understand instructions.

Our Battle With "Noise"

What if we invited one set of subjects and had them in a double auction with low supply…

…and then do the same thing with another set of subjects and high supply.

Suppose prices went up when we shifted supply down. What's the problem?

Since we only have one session of subjects, the incidental factors (like bad instructions or unusual personality of a few subjects) might have taken over and ruined our inference.

We sometimes refer to the influence of these incidental factors "noise".

Treatments

A treatment is just a set of observations in which all of the factors are set to one level.

Comparative statics experiments often involve comparing outcomes across treatments.

> How do changes in the treatment variable (level of factors) affect outcomes?

Control treatment:

> We often think about a treatment as a deviation from a normal state of the world.
>
> > Example: Run a prisoner's dilemma with only male students.
>
> In such a case, we might call a normal treatment a control treatment, and the other treatment simply "the treatment."
>
> > Example: The control would be a normal mix of males and females and the treatment would have males only.

Confounding Factors

In our battle against noise, part of what we need to do is to "control" confounding factors.

These are facts about particular subjects or the conditions under which an experiment is run that could systematically affect outcomes.

- Particularly intelligent (or unintelligent) subjects.
- Particularly social subjects.
- A particularly uncommunicative experimenter.
- An uncomfortable room.
- A group of subjects know one another.

Confounding Factors

We have to be careful about two things with confounding factors:

We don't want to confuse confounding factors with our treatments.

- What if a group of super geniuses all happened to go to participate in one treatment but not another?
- What if a particularly clear experimenter ran all of the experiments in one treatment but not the ones in another?
- What if a group of friends all participated in one treatment rather than another (in an experiment studying cooperation)?

Then we might accidentally confuse the effects of the confounding factors for the effects of the treatment!

Experimental Design: Our Arsenal

We need weapons to fight the adverse effects of confounding factors on our observation of experimental results.

This is what experimental design is all about.

- Randomization.
- Blocking.
- Replication.

We exercise control over the assignment of subject to treatments and the degree to which confounding factors systematically influence results.

Weapon 1: Randomization

If we randomly assign subjects to treatments, we can keep individual subjects from *systematically* determining treatments.

Random recruiting: We tend to randomly choose who to invite to experiments.

Random orders: We often do experiments in random order.

Random locations: If we are going to run the experiment at multiple Universities, we often randomly decide what treatments to run.

The idea is that if we randomize, confounding factors still exist but affect one treatment as much as another.

Weapon 2: Blocking

Sometimes we can do even better than randomizing.

We try to identify a "block" that is likely to exhibit the same confounding factor (or likely not to exhibit that factor).

We then run each of our treatments on that block.

Example: We suspect that each individual session is likely to exhibit unique confounding factors

We could then have the session experience one treatment and then the other.

Doing this allows us to later use some statistical techniques that reduce the level of noise.

Have to be careful to control for order effects.

Completely Randomized vs. Randomized Complete Block

We have two strategies so far:

Completely randomize: Randomly assign subjects and everything else (including the order in which you run experiments, where you run them and who runs them if more than one) to treatments.

Randomized Complete Block: Find subsets of your subjects that are homogonous. Make sure all treatments are run on each of these homogonous blocks.

Within and Between

A common form of blocking is to exploit what is known as a "within design."

Within Design: Treatments are run on the same subject or same group of subjects.

Advantage: Less noise, need less data, fewer sessions.

Disdvantage: You might get a lot of learning or other contamination that affects observations on one of the treatments.

Between Design: No subject or group of subjects experiences the same treatment twice.

Advantage: Total separation between treatments.

Disdvantage: Need more data plus you generate more noise.

Weapon 3: Replication

Perhaps the most important way to limit the influence of uncontrolled factors is to collect a number of observations.

In general the average of a number of observations will be less noisy (less due to idiosyncracies) than a single observation .

Imagine a subject (for personal reasons) chose to play “Hare” in the one shot Stag Hunt game.

> If we only looked at that piece of data we would be mislead.
>
> If we looked at 100 subjects playing Stag Hunt, we’d see that subjects have a tendency to play Stag.

Replication tends to “wash out” idiosyncracies due to noise that can ruin inference.

Experimental economists typically run many replications in any treatment.

Traditional Design

Traditionally, if we wanted to figure out how 2 factors changed behavior we would vary each one independently.

Example: Want to look at how the number of subjects, the information given affect the double auction's ability to reach CE.

- Create a baseline -- 8 subjects and full information
- Study factor 1 (# subjects) by running a second treatment with 16 subjects (keeping full information).
- Study factor 2 information) by running a second treatment with 16 no information (keeping 8 subjects).

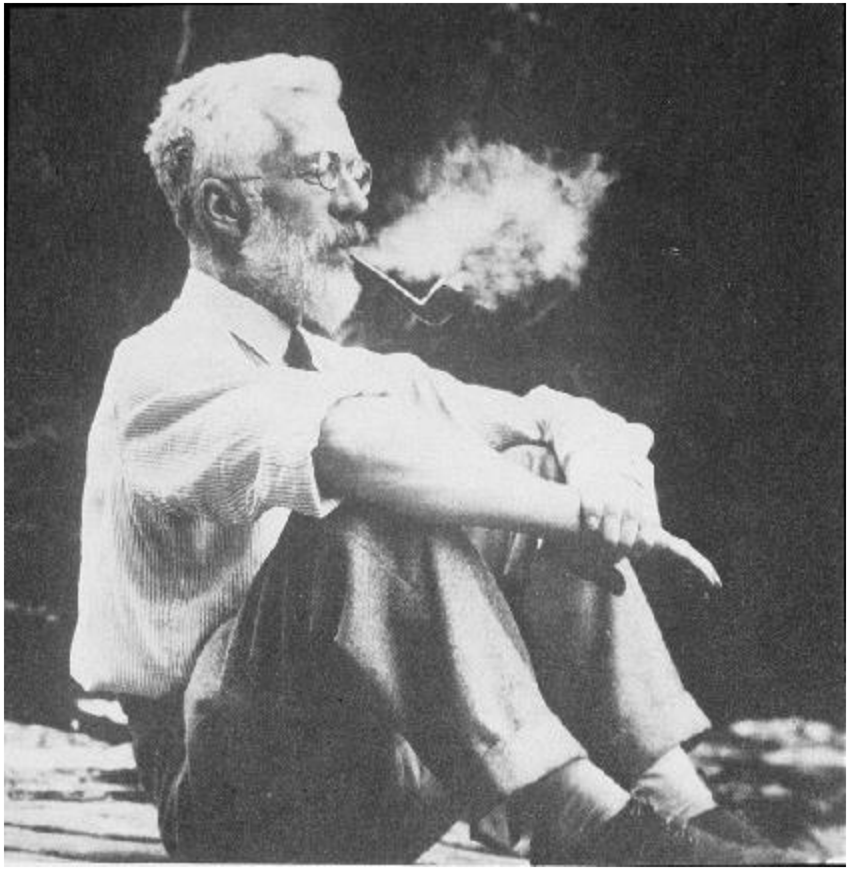
What do we miss by doing this?

Factorial Designs

The problem with this is that we miss the interaction effects.

What if the addition of information only matters when there are a lot of subjects?

The founder of modern statistics, RA Fisher noticed this and developed the idea of a *factorial design.*



Factorial Designs

The idea is to have some observations of every level of every factor being observed along with each level of each other factor.

Meaning, with N factors and 2 levels observed on each, a total of 2^N treatments.

This is a 2 x 2 factorial design meaning 2 factors at 2 levels.

Information

		Informed	Not Informed
# Traders	8		
	16		

Fractional Factorial

When we have more than 2 factors, a full factorial design will give us measurement of high order interactions.

> These show the interactions of interactions
>
> Generally these higher order interactions have little effect (and are hard to interpret).

When you start studying more and more factors in an experiment, the number of treatments you have to run starts to explode (at the rate of 2^N).

To study 6 variables, we'd need to run 64 treatments!

For these two reasons, we have fractional factorial designs.

Fractional Factorial

A fractional factorial design is basically a way of running fewer sessions.

A typical fractional factorial is a half factorial design.

- It captures interactions between each pair of factors, but no interactions between interactions.

- You have to assume that these third order interactions have no effect.

The good news: You can learn a lot with half the number of experiments.

The bad news: You assume the three factor effects don't matter.

Summary

Experimental design is meant to

Avoid confounds

Limit noise.

Allow for accurate measurement.

Clear understanding of the effects of factors using treatments.

By randomizing, replicating and blocking we can do all of these things.

Careful design makes experimental statistics much easier
-- as we'll see next week!