Survey Experiments in Social Science
Session 1 – Control

Federico Vegetti
University of Turin

University of Newcastle

Goals of the course

Understand the theoretical foundations to causal inference and the logic of experimental research

Learn how to design an experiment, and see some types of design that can be implemented in surveys

Learn some good practices in survey experiments

Learn how to design and analyze a conjoint experiment, and see some examples of how to report the results of the analyses

Explaining phenomena

Usually, our theories are about relationships between concepts

In scientific method, explanations are often **causal**

- *Explanandum*: what gets explained
- *Explanans*: what gives the explanation

Concepts are measured, so we test relationships between variables

The *explanandum* becomes the <u>dependent variable</u>, Y

The *explanans* becomes the <u>independent variable</u>, T

On research questions

Causal questions can be framed in two ways:

- Causes of effects
  - What are the causes of Y?
  - More easily formulated as “why” questions
  - “Why do people vote?”
  - Focus on the dependent variable, and look for all possible explanations
- Effects of causes
  - What are the effects of T?
  - Harder to formulate as “why” questions (though they are)
  - “What is the effect of candidate competence on voters’ preferences?”
  - Focus on a key predictor, and prove its effect on Y
  - Contributes to building a more general “causes of Y” model
  - More common in political science

Running example

Do TV debates help people in their vote choices?

Reasons why this should be the case (theory) and why it is relevant (implications)

Theory: TV debates reduce the uncertainty of the viewers, decreasing the number of undecided voters

Hypothesis: $T \longrightarrow Y$

How do we investigate this?

We can run an observational study

- After the debate (and before the election), we draw a random sample of the population
- We ask them whether they watched the debate on TV
- We ask them for which candidate they intend to vote, including the response option “Still undecided”
- We run a regression model with the variable “Decided/Undecided” as DV and the variable “Watched the debate/Did not watch the debate” as main predictor
- We include statistical controls to rule out spurious correlations
- We find a significant effect
- Can we say that watching the TV debate causes people to make up their mind?

The Data Generating Process

- The DGP is the natural development of a certain phenomenon
- (“natural” from the point of view of the researcher)
- It includes *all* the factors influencing the data that we draw for our experiment, whether we observe them or not
- In an observational study, the DGP is the only process behind the observed outcome
- In our example, watching the debate on TV (or not) is part of the DGP
- However, by looking at the DGP alone we can not really isolate the effect of watching the debate from other **confounders**

Intervention and manipulation

To prove causation, we need to intervene in the DGP and manipulate some aspects of it so that they are no longer naturally occurring

For instance, before the interview, we can ask some people to watch the debate on TV and some other people not to do it

This “cleans” the DGP from all the “natural” factors that may affect the probability that people watch the debate

Experiment as comparison

In an experiment, we compare different states of the world.

In our example:

- One in which the respondents watched the debate
- One in which they did not watch the debate

By comparing Y between the two groups, we can see whether watching the debate affected the outcome

Ideally, we would *run history twice*

Since it’s not possible, we must engage in **control**

Control

We engage in control when we hold elements of the DGP constant between groups

The more elements we hold constant, the better able we are to isolate some specific effects

This also affects the interpretation of the causal effect

- Is watching the debate that makes people less undecided, or it is watching *any* political message on TV?

It is up to us to decide what to hold constant between groups

Experiments, unlike observational studies, allow us to have control over the DGP

Treatment

Treatment is the part of the DGP that we manipulate

It corresponds to the variable T of which effect we want to study

In common terminology:

- "Treatment group" is the group for which manipulation occurs
- "Control group" is the groups for which manipulation does not occur

However, even if we were to compare different treatments, as long as we hold everything else constant, we would still be engaging in "control"

Who goes in which group?

How do we decide which respondents watch the debate and which ones do not?

Ideally, the two groups should be *identical in all characteristics except for the treatment*

However, this is not always possible due to observable and unobservable confounding factors

Formally, $Y = T + X + U$

Random assignment

With random assignment, respondents are assigned to one of the groups randomly

With every respondent we flip a coin: if we get heads, s/he is told to watch the debate, if we get tails, s/he is told not to do it

The randomness of the process should ensure that, on average, the differences between groups cancel out

The potential confounding factors are still there, but they are not systematically correlated with our manipulation: they become noise

Control and random assignment

- In theory, random assignment is not mandatory in experiments
- We could make groups that are identical in all the aspects that might affect the outcome
- However, this is very difficult, as the phenomena that we observe in social science tend to be very complex (i.e. affected by many factors)
- This is made even more difficult by the fact that many confounders are not measurable
- The random assignment ensures that the groups that we produce are <u>almost identical on average</u>
- We can still statistically control for observables to make our effect more precise

Experimental vs. Statistical Control

- When we run a regression, we say that we “control” for one or more other predictors when we include them in the model along with the predictor(s) that we are interested in
- For instance, we may be interested in the effect of education on income. However, education increases with age, which may have an effect on income by itself (e.g. because of seniority)
- If we include only education in our model, the effect of the variable will also capture the effect of age that is correlated with education
- In other words, ”age” is a confounding factor

Experimental vs. Statistical Control

- By including age in our model as a “control”, we remove from the variation of income all the variation that is attributed to age
- Hence, the effect of education on income will be “independent” from the effect of age
- Implicitly, we are <u>holding age constant</u> when assessing the effect of education on income
- In experiments, “control” means the same thing: holding constant all the effects that might correlate with the effect of the treatment
- In an experiment our ability to control for confounding factors is much higher, because we assign respondents to groups independently from the values of the confounding factors
- However, we can still use statistical control to reduce the noise

Reduction of noise

In experiments we can reduce the noise introduced *in* the treatment (not *by* the treatment) by exerting control over potential confounding factors during the design of the experiment

Different ways to control for confounding factors in experiments

- Including statistical controls (when analyzing the data)
- Observing time and effort spent on a task
- Incentives
- Within-subject design
- Holding observable factors constant across groups

Statistical control in experiments

Recall: $Y = T + X + U$

Where T is the treatment, X are observable and U are unobservable factors that influence Y beyond the treatment

When we analyze experimental data, we can try to *clean up* as much as we can the variance of Y from the effect of X

To do so, we can include control variables together with the treatment variable when we analyze the experimental data

Statistical control in experiments

Although this is done in the analysis stage, we must plan to collect more variables in the design stage

Why is this different from control in observational studies?

- Here we can assume that X and U are uncorrelated with the treatment
- Hence their effect is only noise, it does not affect our causal inference
- Unless the effect of T is <u>conditional</u> on X and U

Tracking the subjects

A typical unobservable factor is how much time and effort subjects put into the task

In some settings, like laboratory experiments or online survey experiments, it is possible to measure the time a respondent spent on a specific task

In some (sophisticated) cases, researchers used eye-tracking technology to follow the gaze of the respondents

However, this is rarely used as a mere control

Incentives

Incentives might increase subjects' motivation to do the assigned tasks properly, reducing the variance in effort they put into it

For instance, we can tell our subjects that they will be paid if they watch the debate

- This implies we are able to determine whether they actually did it or not

Incentives are the rule in experiments in economics, the exception in psychology

Incentives may reduce the external validity by making the experimental setting less realistic

Within vs Between Subject Design

In between-subject design we compare groups of different individuals

- Subjects are only observed in one state of the world
- The counterfactual is another group of subjects who share the same characteristics (on average)

In within-subject design, we observe individuals in different states

- Subjects are observed in multiple states of the world
- The counterfactual of every individual is him/herself in a different state
- For instance, we could run a survey before the debate, ask all respondents to watch the debate, and interview them again after the debate. We can compare their indecision before and after the debate and see whether it has been reduced

Within Subject Design and Inference

What are the implications of within subject design?

Can we say that it solves the causal inference problem?

In some cases, yes

E.g. multiple choice procedure:

- Subjects are asked to make choices in different hypothetical conditions
- The manipulation here affects the hypothetical conditions that are shown to them

Within Subject Design and Inference

In many cases observing people in multiple conditions is just not possible

- Imagine an experiment where we want to see whether people's policy preferences are affected by partisanship
- We want to see to what extent people "follow their party"
- Between design:
  - We ask subjects what is their position on a given issue. In Group 1 we give information about the position of parties. In Group 2 we give no additional information. We expect subjects in Group 1 to take positions that are similar to those of their own party (observed previously)
- Within design:
  - We ask subjects what is their position on a given issue. Then we give information about the position of the parties. Then we ask them what is their position again. We expect subjects to change their position in a way that is more similar to their own party (observed previously) after we have given them the information
- What is the problem here?

Problems of Within Subject Design

Within subject designs may produce some biases

The multiple treatments may not be independent from each other

- People like to appear consistent. They do not want to look like their opinions change easily

Subjects may just follow a pattern

- The pattern may also be the “idea” that subjects have about what the experimenter wants (“demand effect”)

Advantages of Within Subject Design

Within subject designs may help controlling for unobservable confounding factors

By using the same subjects, we hold constant factors that are not easy to observe (e.g. cognitive ability, motivation)

Moreover, within subject designs allow us to have more statistical power, as we compare the full sample with itself (hence we have double as many observations as the between subject design)

Combining Within and Between

- We can also combine between and within
- For instance, we can ask respondents about their vote choice before the debate, then randomly assign half of them to watch the debate and half to watch something else, then ask again
- By comparing the difference between t_1 and t_2 in both groups, we can assess what is the treatment effect more precisely
- This allows us to control for unobservable confounding factors improving the quality of our inference
- In statistics, when this is applied to quasi-experimental data, this is called "difference-in-difference" design

Control by design

- Example: we want to make sure that education does not affect our treatment of watching the TV debate
- We divide subjects in 4 levels of education
  - 1: Up to middle school
  - 2: High school
  - 3: 3-years degree (e.g. B.A.)
  - 4: 5-years degree and higher (e.g. M.A., PhD)
- We assign 10 subjects for each group
- Then, within each group, we assign the 5 subjects picked randomly to each of the 2 conditions (watching the debate; not watching the debate)

Randomized Blocks Design

	Watch the debate YES	Watch the debate NO	Total
Up to middle school	5	5	10
High school	5	5	10
3-years degree	5	5	10
5-years degree+	5	5	10
Total	20	20	40

Randomized Blocks Design

- In RBD the goal is to provide balanced samples between conditions
- To do so, we create blocks so that the variation within blocks is much smaller than the variation in the entire sample
- Then, within each block, we randomly assign subjects to conditions
- The blocks need to be all the same size, and the size needs to be a multiple of the number of conditions that we have
- This reduces the variability between experimental conditions and effectively eliminates the confounding effect of the factor that we use to create blocks (e.g. education)

Generalizing the Randomized Blocks Design

The same design can be generalized to any number of variables

For instance, we may want to group subjects into gender

Note that, the more blocks we have, the larger the sample needs to be

Generalizing the Randomized Blocks Design

	Watch the debate YES		Watch the debate NO		
	F	M	F	M	Total
Up to middle school	5	5	5	5	20
High school	5	5	5	5	20
3-years degree	5	5	5	5	20
5-years degree+	5	5	5	5	20
Total	20	20	20	20	80

Generalizing the Randomized Blocks Design

- This design follows the same logic of **stratification** when constructing survey samples
- It is an effective way to build balanced samples to reduce the error in the treatment
- This should clean the treatment effect from the effect of confounders
- However, like with stratification, the representativeness of the sample may suffer (and so will the external validity)
- Moreover, this allows us to control only for observable confounders (whereas in social science we deal with a ton of unobservable characteristics)
- This is why in social science the “simple” random assignment is by far the most common technique

To sum up

Ways in which we exert control in experiments:

1. Manipulation of the DGP
2. Random assignment to the experimental conditions
3. Reducing the noise