Experimental Economics: Inference

Kochubey Center, Radischeva street 4
Pushkin, St.Petersburg

July 15, 2019

Measurement

Sir R.A.Fisher (1890-1962), British statistician, biologist etc, known for F-test, ANOVA etc.

According to Fisher, all inferences from data should be based on

1. manipulation of one or more independent variables
2. use of controls, such as random assignment of subjects to independent variables
3. careful measurement of dependent variable(s)

Experiments are characterized by the presence of 1) and 2); 3) is typical of all data analysis (incl.observational data). Absence of 2) $\Rightarrow$ *quasi-experiment*.

Principles of experimental design

Methodologically, since R.A.Fisher (1926; 1935), treatment experiments are supposed to allow for inference on the treatment effect, which requires

Comparison — results across different treatments should be comparable.

Randomization — assignment of units (subjects) to treatments (conditions) by the experimenter should be *random*. This is crucial inasmuch as we wants to be able to say something about causal effect of treatment on measurable characteristics of the units.

Replication — results should be replicable across samples.

Orthogonality — different interventions should be independent from (orthogonal to) each other, e.g. changes of incentives across groups should be independent of changes in frames, or you should not change both conditions at the same time (or in any other correlated way).

Principles of experimental design II

To improve efficiency of data analysis:

Blocking (noise-reducing) — partition of units into blocks to be subjected to various treatments.

To test the quality of new shoe sole, one may give shoes with different soles (old or new) to different users, but people may have different walking habits. To control for this, let the same person wear one shoe with old (control) and another — with new (experimental) sole, put randomly on left or right shoe each. Each person then acts as 'block' with lower variability than within the population, which reduces noise from individual variances in walking habits.

Factorization (signal-enhancing) — independently vary several dimensions of the experiment to control effectively for the interaction of factors - e.g., $2 \times 2, 3 \times 2$ etc.

If we want to see whether risk attitudes differ depending on lottery scale, as well as depending of pricing conditions (willingness to accept– WTA vs. willingness to pay – WTP), we have to randomly split subjects to WTA and WTP treatment, and offer each of these treatments either low- and high-prize lottery. Controlling for one factor strengthens the effect of the other, if any is found.

Ways to enhance power of the design

Glazed window in Caius college,
commemorating R.A.Fisher and Latin Square

1. *Randomized block design* splits sample into *n* blocks with similar units within it, and assigns separate treatment to each.
2. *Latin square* is a repeated design which subjects the same block to exactly one treatment at each level of every factor (see picture)
3. *Greko-Latin square* is Latin square 'crossed' (multiplied) by its orthogonal (Greek) square to cover more treatments with the same number of observations.

Problem

In practical applications of experiments, especially natural or field, we (usually) want to make *causal* (not casual!) *inferences*:

- Patients receive new drug: does it help to recover more quickly?
- Subjects are given more information: does it help them to make more efficient decisions?
- Local communities have larger representation of women: does it result in better budgetary policy?
- People receive higher incomes: does it make them happier?

Generally, is it true that *treatment* T causes changes in characteristic of interest Y applied to subjects $i = 1, \dots n$?

Formalization: the Rubin model

Introduce the following notation

- units (individuals, communities, firms...) $i = 1, \ldots n$
- treatment $T_i = 1$ for treated and $T_i = 0$ for untreated unit
- potential outcome $Y_i^1 \equiv Y_i(T_i = 1)$ for treated and $Y_i^0 \equiv Y_i(T_i = 0)$ for untreated unit
- causal effect $\tau_i = Y_i^1 - Y_i^0$ of treatment T on unit i

If treatment effect is systematic across units, we want to conclude that treatment causes the effect.
The problem, however, is that in practice we observe each unit in only one state: it is either treated or untreated

$$Y_i = Y_i^0(1 - T_i) + Y_i^1 T_i \tag{1}$$

Solution: randomization

To solve the problem, we have to *randomly* assign units to treatment and control: $(Y_i^1, Y_i^0) \perp T_i$
In words, there are no grounds to believe that units assigned to one of the group are systematically different from units assigned to another.
Admittedly, this is a matter of judgment (!), but if met, we can argue the Average Treatment Effect (ATE) is

$$ATE = \mathrm{E}\tau_i = \mathrm{E}(Y_i^1 | T_i = 1) - \mathrm{E}(Y_i^0 | T_i = 0) \tag{2}$$

Methods, such as factorial designs (Latin or Greek squares) do exist to control for various factors.

Estimation

In practice, the finite sample unbiased estimator of the treatment effect is

$$ATE = \hat{\tau} = \frac{\sum_{i=1}^{n^1} Y_i T_i}{n^1} - \frac{\sum_{i=1}^{n^0} Y_i (1 - T_i)}{n^0} \quad (3)$$

and the variance of that estimator is

$$Var(\hat{\tau}) = \frac{\sigma^2_{Y_i^1}}{n^1} + \frac{\sigma^2_{Y_i^0}}{n^0} \quad (4)$$

Key conditions

For this trick to work, we need to ensure that potential outcome depends solely on the unit itself, which requires

Excludability : the only relevant causal agent affecting the treated units is receipt of the treatment, or no other variable $Z = \{z, z', \ldots\}$ affecting Y_i can cause the effect: $Y_i^1(z) = Y_i^1(z')$ and $Y_i^0(z) = Y_i^0(z'), \forall z, z' \ldots$.

SUTVA – Stable Unit Treatment Value Assumption, or non-interference: treatment of one unit is not affected by treatment status of any other unit, $Y_i(T_i) = Y_i(\mathbf{T})$ for any vector of treatments $\mathbf{T}$ and all treated units.

Meeting these conditions in practice can be problematic.

Sampling

Participants of an experiment can be recruited in the following way:

Convenience : Those whom you can find to participate, e.g. students

Sampling

Participants of an experiment can be recruited in the following way:

Convenience: Those whom you can find to participate, e.g. students

Snowball: Invite your friends, and ask them to invite their friends etc. (cheap way to ‘randomize’)

Sampling

Participants of an experiment can be recruited in the following way:

- **Convenience**: Those whom you can find to participate, e.g. students
- **Snowball**: Invite your friends, and ask them to invite their friends etc. (cheap way to ‘randomize’)
- **Quota**: Select participants according to some criteria, e.g. no more than 20% of students

Sampling

Participants of an experiment can be recruited in the following way:

Convenience: Those whom you can find to participate, e.g. students

Snowball: Invite your friends, and ask them to invite their friends etc. (cheap way to ‘randomize’)

Quota: Select participants according to some criteria, e.g. no more than 20% of students

Representative: Proportion of participants representing a particular group is the same as in general population

Sampling

Participants of an experiment can be recruited in the following way:

- **Convenience**: Those whom you can find to participate, e.g. students
- **Snowball**: Invite your friends, and ask them to invite their friends etc. (cheap way to ‘randomize’)
- **Quota**: Select participants according to some criteria, e.g. no more than 20% of students
- **Representative**: Proportion of participants representing a particular group is the same as in general population
- **Population**: Every member of the population in question participates (e.g. census)

Types of experiment

	Control	Incentives	Sampling	Realism
Lab	High	High (?)	Convenience	Low
Lab in the field	Medium	High (?)	Case/Snowball	Medium
Survey	Low	Low	Quota/Random	Medium
Field	Medium/Low	Medium	Quota	High
Natural	Medium/Low	Medium(?)	Population	High
			Self-selection	

Experimental validity

Internal validity Claimed inference is valid within the target population of subjects (e.g. convenience sample of HSE students)

Experimental validity

Internal validity Claimed inference is valid within the target population of subjects (e.g. convenience sample of HSE students)

External validity Claimed inference is valid beyond the target population (e.g. any random sample of students)

Experimental validity

Internal validity Claimed inference is valid within the target population of subjects (e.g. convenience sample of HSE students)

External validity Claimed inference is valid beyond the target population (e.g. any random sample of students)

Ecological validity Claimed inference is valid in the real environment the experiment attempts to model/represent.

Project topics

Groups

Finance	Artur, Daniil, Egor
Mechanism design	Aida, Dmitry Pra, Vitaliia, Igor, Marina
Matching	Alla, Dmitry Pok, Daniel, Elizaveta, Georgy
Cooperation	Anna Sed, Maxim, Alexey, Anna Sok
Honesty	Evgeniia, Elena, Alexandra
Inequality Morality	Ekaterina, Roman, Alexander, Dmitry Kis, Xenya