European Field Experiments Summer School
Potential Outcomes

Dr. Florian Foos

Department of Political Economy
King's College London

florian.foos@kcl.ac.uk
@florianfoos

June 25, 2018

Some Notation

Some Notation

- Indexing experimental subjects/units: the subscript $_i$ refers to subjects 1 through N.

Some Notation

- Indexing experimental subjects/units: the subscript $_i$ refers to subjects 1 through N.
- Defining treatment: The variable d_i indicates whether the ith subject is treated.

Some Notation

- Indexing experimental subjects/units: the subscript $_i$ refers to subjects 1 through N.
- Defining treatment: The variable d_i indicates whether the ith subject is treated.
- $d_i = 1$

Some Notation

- Indexing experimental subjects/units: the subscript $_i$ refers to subjects 1 through N.
- Defining treatment: The variable d_i indicates whether the ith subject is treated.
- $d_i = 1$ means the ith subject receives the treatment.

Some Notation

- Indexing experimental subjects/units: the subscript $_i$ refers to subjects 1 through N.
- Defining treatment: The variable d_i indicates whether the ith subject is treated.
- $d_i = 1$ means the ith subject receives the treatment.
- $d_i = 0$

Some Notation

- Indexing experimental subjects/units: the subscript $_i$ refers to subjects 1 through N.
- Defining treatment: The variable d_i indicates whether the ith subject is treated.
- $d_i = 1$ means the ith subject receives the treatment.
- $d_i = 0$ means the ith subject does not receive the treatment.

Some Notation

- Indexing experimental subjects/units: the subscript $_i$ refers to subjects 1 through N.
- Defining treatment: The variable d_i indicates whether the ith subject is treated.
- $d_i = 1$ means the ith subject receives the treatment.
- $d_i = 0$ means the ith subject does not receive the treatment.
- It is assumed that d_i is observed for every subject.

Potential outcomes

Potential outcomes

- Potential outcomes: Regardless of which treatment a subject receives, that subject may be said to have a potential response in the event that treatment is or is not received.

Potential outcomes

- Potential outcomes: Regardless of which treatment a subject receives, that subject may be said to have a potential response in the event that treatment is or is not received.
- In general, potential outcomes may be written $Y_i(d)$, where the argument d indexes the treatment.

Potential outcomes

Potential outcomes

- $Y_i(1)$ is the potential outcome if the ith subject was treated.

Potential outcomes

- $Y_i(1)$ is the potential outcome if the ith subject was treated.
- $Y_i(0)$ is the potential outcome if the ith subject was not treated.

Potential outcomes

- $Y_i(1)$ is the potential outcome if the ith subject was treated.
- $Y_i(0)$ is the potential outcome if the ith subject was not treated.
- Potential outcomes are fixed attributes of each subject and represent the outcome that would be observed hypothetically if that subject were treated or untreated.

Conditional potential outcomes

Conditional potential outcomes

- Potential outcomes for a subset of subjects.

Conditional potential outcomes

- Potential outcomes for a subset of subjects.
- $Y_i(d) \mid X = x$ denotes potential outcomes when the condition X = x holds.

Conditional potential outcomes

Conditional potential outcomes

- $Y_i(0) \mid d_i = 1$

Conditional potential outcomes

- $Y_i(0) \mid d_i = 1$: untreated potential outcome for subjects that receive the treatment.

Conditional potential outcomes

- $Y_i(0) \mid d_i = 1$: untreated potential outcome for subjects that receive the treatment.
- $Y_i(0) \mid d_i = 0$

Conditional potential outcomes

- $Y_i(0) \mid d_i = 1$: untreated potential outcome for subjects that receive the treatment.
- $Y_i(0) \mid d_i = 0$: untreated potential outcome for subjects that do not receive the treatment.

Conditional potential outcomes

- $Y_i(0) \mid d_i = 1$: untreated potential outcome for subjects that receive the treatment.
- $Y_i(0) \mid d_i = 0$: untreated potential outcome for subjects that do not receive the treatment.
- $Y_i(1) \mid d_i = 1$

Conditional potential outcomes

- $Y_i(0) \mid d_i = 1$: untreated potential outcome for subjects that receive the treatment.
- $Y_i(0) \mid d_i = 0$: untreated potential outcome for subjects that do not receive the treatment.
- $Y_i(1) \mid d_i = 1$: treated potential outcome for subjects that receive the treatment.

Conditional potential outcomes

- $Y_i(0) \mid d_i = 1$: untreated potential outcome for subjects that receive the treatment.
- $Y_i(0) \mid d_i = 0$: untreated potential outcome for subjects that do not receive the treatment.
- $Y_i(1) \mid d_i = 1$: treated potential outcome for subjects that receive the treatment.
- $Y_i(1) \mid d_i = 0$

Conditional potential outcomes

- $Y_i(0) \mid d_i = 1$: untreated potential outcome for subjects that receive the treatment.
- $Y_i(0) \mid d_i = 0$: untreated potential outcome for subjects that do not receive the treatment.
- $Y_i(1) \mid d_i = 1$: treated potential outcome for subjects that receive the treatment.
- $Y_i(1) \mid d_i = 0$: treated potential outcome for subjects that do not receive the treatment.

Treatments as random variables

Treatments as random variables

- We distinguish between d_i the treatment that a given subject receives and D_i, the treatment that could be administered hypothetically.

Treatments as random variables

- We distinguish between d_i the treatment that a given subject receives and D_i, the treatment that could be administered hypothetically.
- D_i is a random variable (the ith subject might be treated in one hypothetical study and not in another).

Treatments as random variables

- We distinguish between d_i the treatment that a given subject receives and D_i, the treatment that could be administered hypothetically.
- D_i is a random variable (the ith subject might be treated in one hypothetical study and not in another).
- $Y_i(0) \mid D_i = 1$

Treatments as random variables

- We distinguish between d_i the treatment that a given subject receives and D_i, the treatment that could be administered hypothetically.
- D_i is a random variable (the ith subject might be treated in one hypothetical study and not in another).
- $Y_i(0) \mid D_i = 1$: untreated potential outcome for subjects that would receive the treatment under a hypothetical random assignment.

Treatments as random variables

- We distinguish between d_i the treatment that a given subject receives and D_i, the treatment that could be administered hypothetically.
- D_i is a random variable (the ith subject might be treated in one hypothetical study and not in another).
- $Y_i(0) \mid D_i = 1$: untreated potential outcome for subjects that would receive the treatment under a hypothetical random assignment.
- We use D_i when talking about the statistical properties of treatments.

Example

- What's the effect of private tutoring on exam scores (ranging from 1 to 6)?

Full schedule of potential outcomes

	$Y_i(0)$	$Y_i(1)$
	Test score if	
subject i	not tutored	tutored
1	3	4.5
2	5	5
3	5	4.5
4	4.5	5
5	4	5.5
6	6	6

Definition of a subject-level treatment effect

- The individual level treatment effect τ_i for a given subject i is defined as:

Definition of a subject-level treatment effect

- The individual level treatment effect τ_i for a given subject i is defined as: $\tau_i = Y_i(1) - Y_i(0)$.

Full schedule of potential outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	3	4.5	
2	5	5	
3	5	4.5	
4	4.5	5	
5	4	5.5	
6	6	6	

Full schedule of potential outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	3	4.5	1.5
2	5	5	0
3	5	4.5	-0.5
4	4.5	5	0.5
5	4	5.5	1.5
6	6	6	0

Definition of Average Treatment Effect

$\mu_{(Y(1))} - \mu_{(Y(0))}$,

where

$\mu_{(Y(1))}$ is the average value of $Y_i(1)$ for all subjects and
$\mu_{(Y(0))}$ is the average value of $Y_i(0)$ for all subjects.

Full schedule of potential outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	3	4.5	1.5
2	5	5	0
3	5	4.5	-0.5
4	4.5	5	0.5
5	4	5.5	1.5
6	6	6	0
Average			

Full schedule of potential outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	3	4.5	1.5
2	5	5	0
3	5	4.5	-0.5
4	4.5	5	0.5
5	4	5.5	1.5
6	6	6	0
Average	4.58	5.08	0.5

Potential and observed outcomes

Potential and observed outcomes

- The $Y_i(1)s$ are observed for subjects who are treated, and the $Y_i(0)s$ are observed for subjects who are not treated. For any given subject, we observe either $Y_i(1)$ or $Y_i(0)$, never both at the same time.

Potential and observed outcomes

- The $Y_i(1)s$ are observed for subjects who are treated, and the $Y_i(0)s$ are observed for subjects who are not treated. For any given subject, we observe either $Y_i(1)$ or $Y_i(0)$, never both at the same time.
- The fact that we observe only one potential outcome is called the "fundamental problem of causal inference" (Holland 1986).

Potential and observed outcomes

- The $Y_i(1)s$ are observed for subjects who are treated, and the $Y_i(0)s$ are observed for subjects who are not treated. For any given subject, we observe either $Y_i(1)$ or $Y_i(0)$, never both at the same time.
- The fact that we observe only one potential outcome is called the "fundamental problem of causal inference" (Holland 1986).
- A subject's treatment effect is unobserved.

Potential and observed outcomes

- The connection between the observed outcome Y_i and the underlying potential outcomes is given by the "switching equation":

$$Y_i = d_i Y_i(1) + (1 - d_i) Y_i(0)$$

Independence assumption

Treatment status is statistically independent of potential outcomes and background attributes (X):

$$D_i \perp\!\!\!\perp Y_i(0), Y_i(1), X$$

Independence assumption

Treatment status is statistically independent of potential outcomes and background attributes (X):

$$D_i \perp\!\!\!\perp Y_i(0), Y_i(1), X$$

If a subject is randomly assigned to treatment, knowing whether a subject is treated provides no information about the subject's potential outcomes, or background attributes.

Excludability assumption

Excludability assumption

- When we only define two potential outcomes, $Y_i(1)$ and $Y_i(0)$, based on whether the treatment is administered, we assume that the only relevant causal agent is receipt of the treatment.

Excludability assumption

- When we only define two potential outcomes, $Y_i(1)$ and $Y_i(0)$, based on whether the treatment is administered, we assume that the only relevant causal agent is receipt of the treatment.
- The exclusion restriction breaks down if treatment assignment z_i sets in motion causes of Y_i other than the treatment d_i.

Non-interference assumption

Non-interference assumption

- The value of the potential outcomes for subject i depend only on whether the subject itself is treated (whether d equals 1 or 0).

Non-interference assumption

- The value of the potential outcomes for subject i depend only on whether the subject itself is treated (whether d equals 1 or 0).
- Each subject is unaffected by the treatments and assignments of other units.

Expectations

The expectation of a discrete random variable is defined as

$E[X] = \sum xPr[X = x]$,

Expectations

The expectation of a discrete random variable is defined as

$E[X] = \sum xPr[X = x]$,

where $Pr[X = x]$ denotes the probability that X takes on the value x, and where the summation is taken over all possible values of x.

Definition of Average Treatment Effect

- Under random assignment:

$$E[Y_i(1) \mid D_i = 1] = E[Y_i(1) \mid D_i = 0]$$

Definition of Average Treatment Effect

- Under random assignment:

$$E[Y_i(1) \mid D_i = 1] = E[Y_i(1) \mid D_i = 0] = E[Y_i(1)]$$

Definition of Average Treatment Effect

- Under random assignment:

$$E[Y_i(1) \mid D_i = 1] = E[Y_i(1) \mid D_i = 0] = E[Y_i(1)]$$

$$E[Y_i(0) \mid D_i = 1] = E[Y_i(0) \mid D_i = 0]$$

Definition of Average Treatment Effect

- Under random assignment:

$$E[Y_i(1) \mid D_i = 1] = E[Y_i(1) \mid D_i = 0] = E[Y_i(1)]$$
$$E[Y_i(0) \mid D_i = 1] = E[Y_i(0) \mid D_i = 0] = E[Y_i(0)]$$

Definition of Average Treatment Effect

- Under random assignment:

$$E[Y_i(1) \mid D_i = 1] = E[Y_i(1) \mid D_i = 0] = E[Y_i(1)]$$

$$E[Y_i(0) \mid D_i = 1] = E[Y_i(0) \mid D_i = 0] = E[Y_i(0)]$$

$$E[Y_i(1) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0]$$

Definition of Average Treatment Effect

- Under random assignment:

$$E[Y_i(1) \mid D_i = 1] = E[Y_i(1) \mid D_i = 0] = E[Y_i(1)]$$

$$E[Y_i(0) \mid D_i = 1] = E[Y_i(0) \mid D_i = 0] = E[Y_i(0)]$$

$$E[Y_i(1) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0]$$

$$= E[Y_i(1)] - E[Y_i(0)] = E[\tau_i] = ATE.$$

Estimator and estimand

Estimator and estimand

- Without prior knowledge, an estimate from just one experiment is only a best guess about the true value of the ATE. Any one ATE might be a little too high or a little too low.

Estimator and estimand

- Without prior knowledge, an estimate from just one experiment is only a best guess about the true value of the ATE. Any one ATE might be a little too high or a little too low.
- Our data set is just one of many possible data sets that could have been created via random assignment. If we would redo the exact same random assignment procedure, different units would be allocated to treatment and control groups!

Estimator and estimand

- Without prior knowledge, an estimate from just one experiment is only a best guess about the true value of the ATE. Any one ATE might be a little too high or a little too low.
- Our data set is just one of many possible data sets that could have been created via random assignment. If we would redo the exact same random assignment procedure, different units would be allocated to treatment and control groups!
- The average estimated ATE across all possible random assignments is equal to the true ATE.

Estimator and estimand

- Without prior knowledge, an estimate from just one experiment is only a best guess about the true value of the ATE. Any one ATE might be a little too high or a little too low.
- Our data set is just one of many possible data sets that could have been created via random assignment. If we would redo the exact same random assignment procedure, different units would be allocated to treatment and control groups!
- The average estimated ATE across all possible random assignments is equal to the true ATE.
- On average we recover the true ATE. Our *estimator* is unbiased.

Observed Outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ tutored	τ_i Treatment effect
1	?	4.5	?
2	5	?	?
3	?	4.5	?
4	4.5	?	?
5	4	?	?
6	?	6	?

Observed Outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	?	4.5	?
2	5	?	?
3	?	4.5	?
4	4.5	?	?
5	4	?	?
6	?	6	?
Average	4.5	5	

Observed Outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	?	4.5	?
2	5	?	?
3	?	4.5	?
4	4.5	?	?
5	4	?	?
6	?	6	?
Average	4.5	5	0.5

Observed Outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	?	4.5	?
2	5	?	?
3	?	4.5	?
4	4.5	?	?
5	?	5.5	?
6	6	?	?

Observed Outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	?	4.5	?
2	5	?	?
3	?	4.5	?
4	4.5	?	?
5	?	5.5	?
6	6	?	?
Average	5.17	4.83	

Observed Outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	?	4.5	?
2	5	?	?
3	?	4.5	?
4	4.5	?	?
5	?	5.5	?
6	6	?	?
Average	5.17	4.83	-0.32

Sampling distribution of the ATE

- In our example, how many different ways are there of assigning 3 subjects to the treatment group?

Sampling distribution of the ATE

- In our example, how many different ways are there of assigning 3 subjects to the treatment group?

$$\frac{N!}{m!(N-m)!} =$$

Sampling distribution of the ATE

- In our example, how many different ways are there of assigning 3 subjects to the treatment group?

$$\frac{N!}{m!(N-m)!} = \frac{6!}{3!3!} =$$

Sampling distribution of the ATE

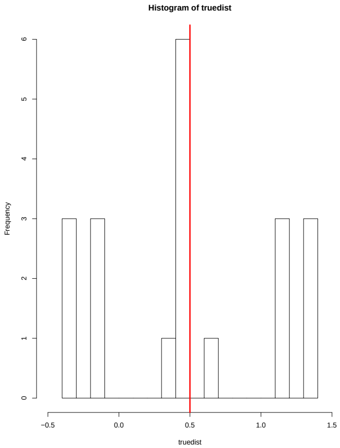
- In our example, how many different ways are there of assigning 3 subjects to the treatment group?

$$\frac{N!}{m!(N-m)!} = \frac{6!}{3!3!} = \frac{720}{36} = 20$$

ATEs

	Estimated ATEs	Frequency
1	-0.33	3
2	-0.17	3
3	0.33	1
4	0.50	6
5	0.67	1
6	1.17	3
7	1.33	3
	0.5	20

Sampling Distribution of estimated ATEs



Standard error

Standard error

- Sampling distribution:

Standard error

- Sampling distribution: The collection of estimates that could have been generated by every possible random assignment.

Standard error

- Sampling distribution: The collection of estimates that could have been generated by every possible random assignment.
- Sampling distribution of the ATE:

Standard error

- Sampling distribution: The collection of estimates that could have been generated by every possible random assignment.
- Sampling distribution of the ATE: The collection of ATE estimates that could have been generated by every possible random assignment.

Standard error

- Sampling distribution: The collection of estimates that could have been generated by every possible random assignment.
- Sampling distribution of the ATE: The collection of ATE estimates that could have been generated by every possible random assignment.
- The standard error is the standard deviation of the sampling distribution.

Standard error

- Sampling distribution: The collection of estimates that could have been generated by every possible random assignment.
- Sampling distribution of the ATE: The collection of ATE estimates that could have been generated by every possible random assignment.
- The standard error is the standard deviation of the sampling distribution.
- How to:

Standard error

- Sampling distribution: The collection of estimates that could have been generated by every possible random assignment.
- Sampling distribution of the ATE: The collection of ATE estimates that could have been generated by every possible random assignment.
- The standard error is the standard deviation of the sampling distribution.
- How to: Calculate the squared deviation of each estimate from the average estimate, divide by the number of possible randomizations, and take the square root of the result.

Example

	Estimated ATEs	Frequency
1	-0.33	3
2	-0.17	3
3	0.33	1
4	0.50	6
5	0.67	1
6	1.17	3
7	1.33	3
	0.5	20

True standard error

True standard error

Sum of squared deviations =

True standard error

Sum of squared deviations =

$(-0.33-0.5)^2+(-0.33-0.5)^2+(-0.33-0.5)^2+(-0.17-0.5)^2+(-0.17-0.5)^2+(-0.17-0.5)^2+(0.33-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.67-0.5)^2+(1.17-0.5)^2+(1.17-0.5)^2+(1.17-0.5)^2+(1.33-0.5)^2+(1.33-0.5)^2+(1.33-0.5)^2$

=

True standard error

Sum of squared deviations $=$
$(-0.33-0.5)^2+(-0.33-0.5)^2+(-0.33-0.5)^2+(-0.17-0.5)^2+(-0.17-0.5)^2+(-0.17-0.5)^2+(0.33-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.67-0.5)^2+(1.17-0.5)^2+(1.17-0.5)^2+(1.17-0.5)^2+(1.33-0.5)^2+(1.33-0.5)^2+(1.33-0.5)^2$
$= 6.8846$

True standard error

Sum of squared deviations =
$(-0.33 - 0.5)^2 + (-0.33 - 0.5)^2 + (-0.33 - 0.5)^2 + (-0.17 - 0.5)^2 + (-0.17 - 0.5)^2 + (-0.17 - 0.5)^2 + (0.33 - 0.5)^2 + (0.5 - 0.5)^2 + (0.5 - 0.5)^2 + (0.5 - 0.5)^2 + (0.5 - 0.5)^2 + (0.5 - 0.5)^2 + (0.5 - 0.5)^2 + (0.67 - 0.5)^2 + (1.17 - 0.5)^2 + (1.17 - 0.5)^2 + (1.17 - 0.5)^2 + (1.33 - 0.5)^2 + (1.33 - 0.5)^2 + (1.33 - 0.5)^2$
$= 6.8846$

Square root of average squared deviation =

True standard error

Sum of squared deviations =
$(-0.33-0.5)^2+(-0.33-0.5)^2+(-0.33-0.5)^2+(-0.17-0.5)^2+(-0.17-0.5)^2+(-0.17-0.5)^2+(0.33-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.67-0.5)^2+(1.17-0.5)^2+(1.17-0.5)^2+(1.17-0.5)^2+(1.33-0.5)^2+(1.33-0.5)^2+(1.33-0.5)^2$
$= 6.8846$

Square root of average squared deviation $= \sqrt{\frac{1}{20} * 6.8846} =$

True standard error

Sum of squared deviations =
$(-0.33-0.5)^2+(-0.33-0.5)^2+(-0.33-0.5)^2+(-0.17-0.5)^2+(-0.17-0.5)^2+(-0.17-0.5)^2+(0.33-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.67-0.5)^2+(1.17-0.5)^2+(1.17-0.5)^2+(1.17-0.5)^2+(1.33-0.5)^2+(1.33-0.5)^2+(1.33-0.5)^2$
$= 6.8846$

Square root of average squared deviation $= \sqrt{\frac{1}{20} * 6.8846} = 0.587$

The true standard error using potential outcomes notation

$$SE(\widehat{ATE}) =$$

$$\sqrt{\frac{1}{N-1}\{\frac{mVar(Y_i(0))}{N-m} + \frac{(N-m)Var(Y_i(1))}{m} + 2Cov(Y_i(0), Y_i(1))\}}$$

where

$$Var(Y_i(1)) = \frac{1}{N}\sum_1^N (Y_i(1) - \frac{\sum_1^N Y_i(1)}{N})^2$$

$$Cov(Y_i(0), Y_i(1)) =$$

$$\frac{1}{N}\sum_1^N (Y_i(0) - \frac{\sum_1^N Y_i(0)}{N})\ (Y_i(1) - \frac{\sum_1^N Y_i(1)}{N})$$

Estimating the standard error

$$\widehat{SE} = \sqrt{\frac{\widehat{Var}(Y_i(0))}{N-m} + \frac{\widehat{Var}(Y_i(1))}{m}}$$

We don't know the covariance between $Y_i(0)$ and $Y_i(1)$, so we (conservatively) assume that the treatment effect is the same for all subjects (correlation between $Y_i(0)$ and $Y_i(1) = 1$).

Time for questions.