Experimental Methods: Lecture 1

Experimental Design: The Basics

Raymond Duch

April 26, 2023

DPIR University of Oxford

Road Map to Lecture 1

- Why Should We Do Experiments?
- Illustration I – Government Audits and Belief Updating
- Fundamental Components of Experimental Design I
  - Illustration II – Tullock Contest
  - Research Question, Variables and Measures
- Fundamental Components of Experimental Design II
  - Incentives Induced-value Theory
  - Within and Between Subject Design
  - Factorial Design
- Illustration III – Fear and Dissent Decisions in Zimbabwe
- Potential Outcome Causal Inference
- Validity and Data Quality

Why Should We Do Experiments?

- Experiments offer a new source of data for social scientists
- Easier to interpret in terms of cause and effect
- Define strategies of immediate utility (e.g. select vote-getting themes for a presidential campaign)
- Uncover empirical regularities in new domains
- Assess applicability of existing/competing theories and methodologies
- Methodological advances expanded the set of theoretical propositions and phenomena testable empirically (e.g. media experiments)

Illustration I – Government Audits and Belief Updating

- Duch Torres (2022), Government Audits of Municipal Corruption and Belief Updating.
- Do individuals update their beliefs about corruption when informed about audit results for their local governments?
- Does corruption information cause belief updating?
  - Random assignment to audit information
 - Measure beliefs about corruption in local government.

Illustration I – Government Audits and Belief Updating

Table 1: Models of Belief Updating

Model	Specification
Baseline	$\text{Corrupt}_{(t+1)i} = \beta_0 + \beta_1 \text{Corrupt}_{ti} + \beta_3 T_i + \omega_k X_{k,i} + \epsilon_i$
Malfeasance	$\text{Corrupt}_{(t+1)i} = \beta_0 + \beta_1 \text{Corrupt}_{ti} + \beta_2(\text{Log Malfeasance}_i) + \omega_k X_{k,i} + \epsilon_i$
Negative	$\text{Corrupt}_{(t+1)i} = \beta_0 + \beta_1 \text{Corrupt}_{ti} + \beta_2(\text{Negative}_i) +$ $\beta_3 T_i + \beta_4(\text{Negative}_i \times T_i) + \omega_k X_{k,i} + \epsilon_i$
Bayesian	$\text{Corrupt}_{(t+1)i} = \beta_0 + \beta_1 \text{Corrupt}_{ti} + \beta_2(T_i \times (\text{Corrupt}_{ti}^{True} - \text{Corrupt}_{ti})) +$ $\beta_3 T_i + \beta_4(\text{Corrupt}_{ti}^{True} - \text{Corrupt}_{ti}) + \omega_k X_{k,i} + \epsilon_i$
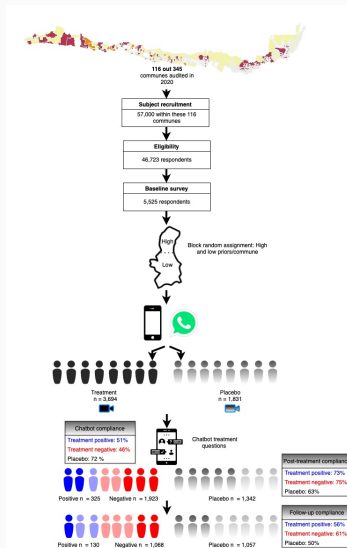
Within Subject Design
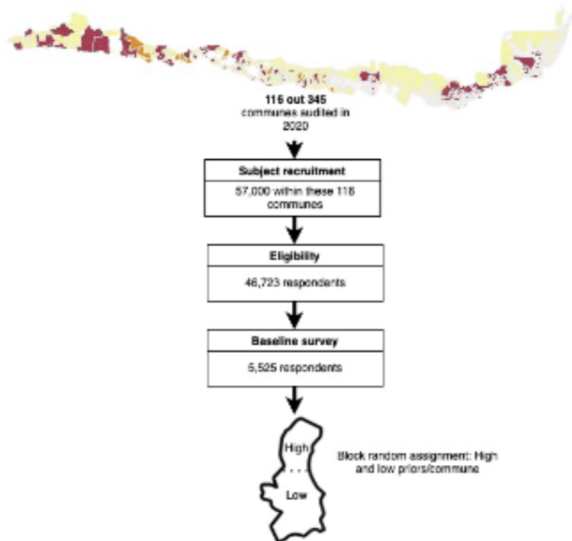
Average Within Subject Change
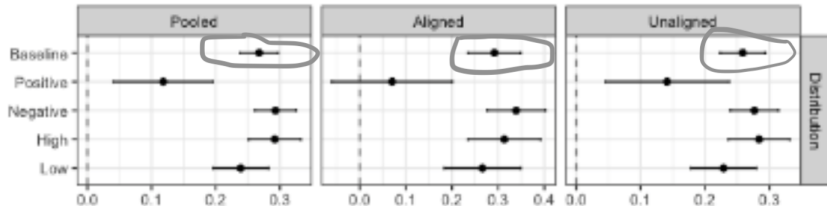
Average Treatment Effect

	Baseline
Intercept	1.122***
	(0.153)
Prior	0.421***
	(0.024)
Treat	0.347***
	(0.028)
Negative	
Log malfeasance	
Corrup diff	
Treat x Negative	
Treat x Log malfeasance	
Treat x Corrup diff	
Covariates	Yes
Num.Obs.	3439
R2	0.203
R2 Adj.	0.196

Fundamental Components of Experimental Design I

- Experimental design as method of planning and conducting experiments to test hypotheses in a controlled and systematic way.
- It aims to isolate the effects of specific variables and to determine cause-and-effect relationships between them.
- What are the some fundamental components of an experimental design based on the experiment that has been illustrated?
- Let's review some of those components looking at another experiment.

Illustration II – Tullock Contest

Theoretical Model

- Agents exert efforts in order to win a prize.
- In a one-stage contest model, the winning probability for a risk-neutral agent i is given by

$$p_i(e_i, e_{-1}) = \frac{e_i}{\sum_{j=1}^{N} e_j}$$

where N is the number of identical agents and e is an agent's effort level.

Illustration II – Tullock Contest

- Agent i's expected utility is therefore

$$E(\pi_i) = p_i(e_i, e_{-1})V - e_i$$

  where V is the prize value.

- Differentiating with respect to e_i and accounting for the symmetric Nash equilibrium gives Tullock's classic solution for the optimal effort level

$$e^* = \frac{(N-1)}{N^2}V$$

Illustration II – Tullock Contest

- R.M. Sheremeta (2010), Experimental comparison of multi-stage and one-stage contests, Games and Economic Behavior, 68:731-747
- Does the model outlined above predict actual behaviour?

Research Question

- Main guide of the experimental design
- Needs to be clear, specific, and well-defined
- Grounded in the existing theory and answerable through empirical investigation
- A few examples (good or bad?):
  - How does exposure to violent media affect aggressive behaviour in children?
  - How can we achieve world peace?
  - Do investment in a repeated public good game decrease over time?
  - Does the Tullock's model of contests predict actual behaviour?

Illustration II – Tullock Contest

- Relevant behaviour is difficult to measure in the real world.
- When possible, measures can be noisy and control over features and parameters of the decision setting is limited.
- Do experiments offer a method to answer questions on the predictive validity of Tullock's model? What kind of experiments?

Illustration II – Tullock Contest

- Sheremeta (2010) tests Tullock's model in a laboratory experiment.
- 84 undergraduate students
- 7 sessions of 12 students per session
- Instructions were given in written format and read aloud at the onset of each session.
- The experimental tasks started only after all participants had passed a comprehension quiz.
- Comprehension quiz and the experimental tasks were administered on computers.
- Experiment programmed using zTree (Fischbacher, 2007).

Variables and Operationalization

- Two crucial components of experimental design.
- Variables definition
  - Which variables should be manipulated experimentally?
  - Which variables might confound the experimental results and needs to be controlled?
  - What is the outcome (dependent variable) that we want to measure?
- Operationalization
  - Defining how a phenomenon or a concept is measured in the context of the experiment.

Back to the Contest Experiment

- Contest
    - participants are randomly placed into groups.
    - Each participant receives an endowment of 120 ECU to bid for the prize.
    - Each participant bids for a reward of 120 ECU.
- A contest is repeated for 30 periods within groups of 4 participants.
- Monetary parameters are expressed in Experimental Currency Units (ECU), which will be converted into dollars at a known exchange rate when the experiment is over.

Illustration II – Tullock Contest

Illustration II – Tullock Contest
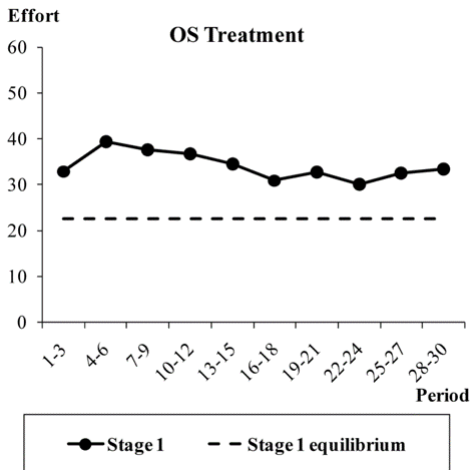
$$e^* = 22.5$$

Illustration II – Tullock Contest

- Efforts are distributed on the entire strategy space
- Inconsistent with a pure Nash equilibrium

Illustration II – Tullock Contest

- The aggregate dissipation (total efforts / prize value) exceeds the models' equilibrium predictions.
- Which factors could explain this pattern of over-dissipation?

Induced-value Theory
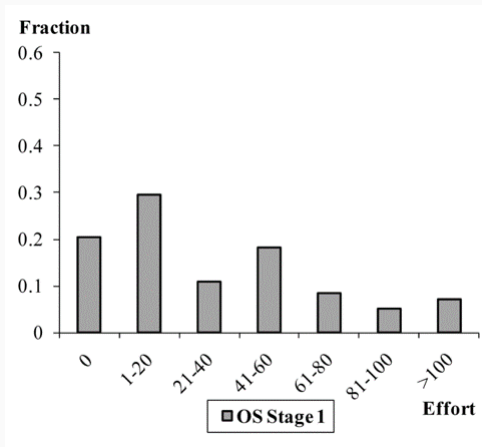
Proper use of reward allows the experimenter to *induce* pre-specified characteristics in experimental participants.

- Monotonicity - more reward is better non-satiation.
- Salience - reward is contingent to decisions and design rules (the latter being understood by the participants).
- Dominance - reward is the main influence on the participant's utility.
- What type of empirical research strategy does not comply with induced-value theory?
- What practical implications for incentivised experiments?

Back to the Contest Experiment

Design extension

- Do agents derive non-monetary utility from winning *itself*?

- Introduce within-subjects variation of the reward's value

- In one part, participants bid for a reward of 120 ECU

- In the following part, repeated for 1 period, participants bid for a reward of 0 ECU

- Bids are costly in both parts

Illustration II – Tullock Contest

- The within-subjects design allows to compare bidding behaviour with/out prize for each participant

Table 4.2
Effort in a contest with no prize.

Effort in a contest with no prize	Percent of subjects	Average effort in contests with prize
0	57.7%	31.3
0–10	17.3%	33.4
10–20	2.6%	39.9
20–30	10.3%	45.1
30–40	1.3%	50.6
40–50	2.6%	73.2
50–60	2.6%	74.3
>60	5.8%	54.2

Illustration II – Tullock Contest

- Utility of winning itself predicts over-dissipation in the one-stage contest (see column 3)

Determinants of effort in contests with prize.

Specification	(1)	(2)	(3)	(4)
Dependent variable, total effort	OS+TS	OS+TS	OS	TS
Period-trend	−0.27***	−0.27***	−0.11**	−0.56***
(inverse of a period trend, $1/t$)	(0.05)	(0.05)	(0.05)	(0.10)
Non-monetary	0.28***	0.26***	0.22***	0.41***
(effort in a contest with no prize)	(0.08)	(0.08)	(0.08)	(0.13)
Quiz		−2.93*	−1.97	−5.48*
(number of correct quiz answers)		(1.72)	(1.64)	(3.05)
OS dummy	−6.24***	−6.24***		
(1 if OS treatment)	(1.12)	(1.12)		
Observations	3960	3960	2520	1440

Note. Robust standard errors in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%. Random effect models account for individual characteristics of subjects. In each regression we control for session, period, and treatment effects.

Within-subject designs

- Costs
    - Demand effects
    - Confounders from multiple treatments (assumptions stated above)
    - Complexity and validity: willingness to pay in Charness et al 2012
- Benefits
    - Internal validity unrelated to assignment mechanism
    - Statistical power
    - Proximity to theory
- Example: Bellemare and Shearer 2009

Between-subject designs

- Each individual exposed to 1 treatment
- ATE as difference in means of *different* groups of individuals
- Costs
    - Natural anchor with respect to (economic) decision-making?
    - Statistical power
- Example: Gneezy and List (2006)
- Overall: confounders $\succ$ power

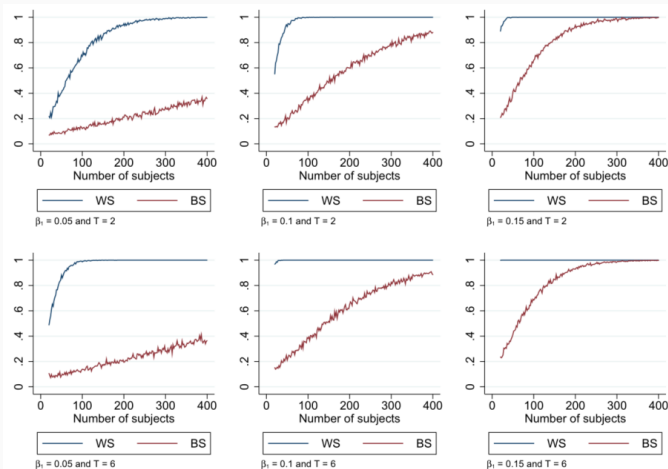**Fig. 2** Simulated statistical power of BS and WS designs with $T = 2$ and $T = 6$ for the high-noise scenario. Simulations based on values $\sigma_\mu^2 = 0.09$ and $\sigma_e^2 = 0.02$. Results for the BS design are computed by allocating the same number of subjects to control and treatment conditions for all periods. Results for the WS design are computed by assigning all subjects to the same number of control and treatment periods

Factorial Design and Multiple treatment arms

- Define a factorial experiment as an experiment involving factors 1 and 2, with factor 1 conditions being A and B, and factor 2 conditions being C and D and E
- Then, allocate subjects at random to every possible combination of experimental conditions
- $\{AC, AD, AE, BC, BD, BE\}$

Resume Field Experiments

- Designed to investigate discrimination of demographic and socio-economic characteristics in the labor market.
- Ficticious resumes sent out to employers in response to job adverts.
- Resumes randomly varies relevant characteristics.
- Callbacks for interviews as dependent variable.
- Bertrand and Mullaithan(2004) classic study
  - Resumes randomly assigned Black- or White-sounding names.
  - Each resume assigned a random selection of other attributes (e.g. gender, education level, experience).
  - Were resumes with White-sounding names called back at the same rate as resumes in the other group?

Factorial Design and Multiple treatment arms

From Rosen 2010

	Colin		Jose	
	Good grammar	Bad grammar	Good grammar	bad grammar
% Received reply	52	29	37	34
(N)	(100)	(100)	(100)	(100)

This design requires us to be especially careful with defining the causal estimand – what quantity are we interested in in this application?

Factorial Design and Multiple treatment arms

Quiz: Why would these two models estimate the same quantities from the Rosen 2010 experiment?

$\{CG, JG, CB, JB\}$ are indicator variables for each of the 4 treatment groups

$J_i = 1$ if Jose Ramirez; $G_i = 1$ if good grammar

$$Y_i = b_1 CG + b_2 JG + b_3 CB + b_4 JB + u_i$$
$$Y_i = a + bJ_i + cG_i + d(J_i G_i) + u_i$$

What quantity in the table do each of the coefficients represent?

Illustration III – Fear and Dissent Decisions

- Lauren E. Young (2019) The Psychology of State Repression: Fear and Dissent Decisions in Zimbabwe. *APSR* 113(1):140–155.
- What is the effect of the emotion of fear on citizen dissent in autocracies?
- How would we demonstrate that the effect of fear on dissent is unconfounded by other variables?
  - Characteristics that induce emotions
  - New information about a threat

Lab-in-the-field Experiments

- Occupy a continuous spectrum between
  - Pure lab experiments, conducted in highly controlled settings, using often abstract tasks, usually with a convenience sample of students.
  - Field experiments, conducted in the real-world with participants not aware of being observed.
- Move beyond the lab in important ways
  - Target specific populations, e.g. from small-scale societies, with particular religious views or jobs (CEOs).
  - Complement field experiments with standardised measures of relevant characteristics (e.g. risk aversion, time preferences, in-group bias).

Illustration III – Fear and Dissent Decisions

- Lab-in-the-field experiment in Harare, Masvingo and Manicaland provinces in Zimbabwe
- Random assignment to treatment or control: Affective emotional memory task (AEMT)
- Ethical implications

Illustration III – Fear and Dissent Decisions

- Sample: 647 participants from six communities in Zimbabwe where the NGO Voice for Democracy (VfD) has a network of mobilizer and informants; and affected by state-sponsored violence
    - Treatment 1: Enumerator asks to describe a situation of fear *around politics and elections* [political fear]
    - Treatment 2: Enumerator asks to describe a situation of *general fears other than politics or elections* [general fear]
    - Control: Enumerator asks to describe a situation that makes them relaxed
- Outcome: Propensity to dissent: hypothetical (via index) and behavioral (via selection of political wristband)

Illustration III – Fear and Dissent Decisions

- Young (2019)

TABLE 3. The Fear Treatments Reduce Dissent

| | Hypothetical | | Behavioral | |
	General fear (1)	Political fear (2)	General fear (3)	Political fear (4)
ATE[1]	−0.545	−0.773	−0.104	−0.189
SE[2]	(0.077)	(0.080)	(0.050)	(0.053)
RI p-value[3]	<0.001	<0.001	0.035	<0.001
N	484	486	329	326
Sample	All		Wristband[4]	

[1]The first row presents the estimated average treatment effects (ATEs) of the general and political fear treatments on the hypothetical measure of propensity to dissent in columns 1 and 2, and the behavioral measure in columns 3 and 4. ATEs are calculated based on assignment to treatment and weighted by inverse propensity scores by block.
[2]Robust standard errors (SEs) from linear regression analysis.
[3]The p-value is based on a two-tailed test using randomization inference.
[4]The estimate of the treatment effect on the wristband measure comes from the subset of the sample respondents who were offered a choice between two real wristbands. Results are similar for the full sample.

Illustration III – Fear and Dissent Decisions

- Young (2019)

FIGURE 2. The Fear Treatments Cause Substantively Large Increases in the Proportion of Respondents Who Are Very Likely or Sure to Dissent During an Election Period
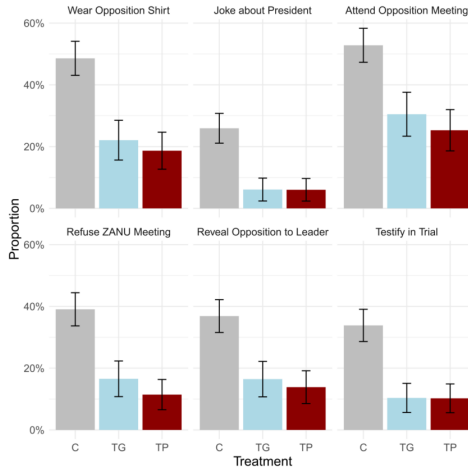
Defining Treatment

- The variable d_i indicates whether the ith subject is treated
- In the typical case of binary treatments, $d_i = 1$ 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

- Y_i: the potential outcome for subject i
- $Y_i(d_i)$: the outcome for subject i, written as a function of the treatment i received; it is generally the case that we observe only one of the potential outcomes for each i
- For the binary-valued treatment, there are two "potential outcomes":
    - $Y_i(1)$, the potential outcome for i conditional on i being treated
    - $Y_i(0)$, the potential outcome for i conditional on i not being treated

Potential Outcome Schedule

- "Hypothetical"
- Comprehensive list of potential outcomes for all subjects
- Rows of this schedule are indexed by i, and the columns are indexed by d
- Potential outcomes for the fifth subject may be found in adjacent columns of the fifth row

Treatments as random variables

- Note on notation: We distinguish between d_i, the treatment that a given subject actually 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).

- $Y_i(0)|D_i = 1$: 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.

Chattopadhyay & Duflo 2004

- Randomized policy experiment in India
- 1990s, one-third of village council heads reserved for women
- women.csv contains subset of data from West Bengal
- Gram Panchayat (GP) = level of government
- Analysis?
    - Was randomization implemented properly?
    - Conjecture: more drinking facilities under women
    - Conjecture: no effect on irrigation

Potential Outcomes Local Budget

	Budget share if village head is male	Budget share if village head is female	Treatment Effect
Village 1	10	15	5
Village 2	15	15	0
Village 3	20	30	10
Village 4	20	15	-5
Village 5	10	20	10
Village 6	15	15	0
Village 7	15	30	15
Average	15	20	5

Potential Outcome Subgroup

- Sometimes useful to refer to potential outcomes for a subset of the subjects
- Expressions of the form $Y_i(d)|X = x$ denote potential outcomes when the condition $X = x$ holds
- For example, $Y_i(0)|d_i = 1$ refers to the untreated potential outcome for a subject who actually receives the treatment

Conditional potential outcomes

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

Individual Level Causal Effect

- For subject i, the effect of the treatment is conventionally defined as the difference between outcomes across the two potential outcomes:

$$\delta_i = Y_i(1) - Y_i(0)$$

- Alternatively:

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

- Often referred to as the Rubin causal model; perhaps more appropriately, the Neyman-Holland-Rubin causal model

- **The Fundamental Problem of Causal Inference** only one of the two potential outcomes is realized, so that δ_i is typically non-operational

Realized Potential Outcomes

- Use lower-case letters for realization of the potential quantities (again, typically only one of the two potential outcomes is realized)

    1. $y_i(1)$, the outcome observed for i conditional on $d_i = 1$ (i is treated)
    2. $y_i(0)$, the outcome observed for i conditional on $d_i = 0$ (i is not treated)

The Fundamental Problem of Causal Inference

Table 2: Table 2.1, p35 Morgan and Winship, *Counterfactuals and Causal Inference*

Group	$Y_i(1)$	$Y_i(0)$
Treatment ($D_i = 1$)	**Observable**	Counterfactual
Treatment ($D_i = 0$)	Counterfactual	**Observable**

Observed Outcomes

- The connection between the observed outcome and the underlying potential outcome is given by the equation
$Y_i = d_i Y_i(1) + (1 - d_i) Y_i(0))$

- This equation indicates that the $Y_i(1)$ are observed for subjects who are treated, and the $Y_i(0)$ are observed for subjects who are not treated

- For any given subject, we observe either $Y_i(1)$ or $Y_i(0)$, not both

Observed Outcomes Local Budget

	Budget share if village head is male	Budget share if village head is female
Village 1	?	15
Village 2	15	?
Village 3	20	?
Village 4	20	?
Village 5	10	?
Village 6	15	?
Village 7	?	30

Average Treatment Effect

- Average Treatment Effect:

$$E(\delta) = E[Y(1)] - E[Y(0)] = E[Y(1) - Y(0)]$$

- where the expectation is over a population, and so no subscript i

- This is operational, in that we can compute sample estimates of $E[Y(1)]$ and $E[Y(0)]$: e.g., the sample averages:

$$\hat{y}(1) = \frac{1}{n_1} \sum_{i:d_i=1} y_i(1) \text{ and } \hat{y}(0) = \frac{1}{n_0} \sum_{i:d_i=0} y_i(0)$$

- where n_1 and n_0 are the number of subjects in groups $d(1)$ and $d(0)$ respectively

Randomization Generates Unbiased Estimates of Average Treatment Effect

- Rubin (1974) calls this:

$$\hat{\delta} = \hat{y}(1) - \hat{y}(0)$$
$$= \hat{y}_d$$

- Under certain circumstances, this is an unbiased estimate of the population average treatment effect δ

- Why? How?

- Nice, informal treatment in "Two Formal Benefits of Randomization"

Properties of Random Assignment

- Under equal probability random assignment, the conditional ATE among the treated is the same as the conditional ATE among the control group, which is therefore the same as the ATE
- The expected $Y_i(0)$ in the treatment group is the same as the expected $Y_i(0)$ in the control group
- When random assignment is not used (i.e., observational research), the unbiasedness of the difference-in-means estimator becomes a matter of conjecture

Potential Outcomes: Core Assumptions

- We assume that each subject has two potential outcomes $Y_i(1)$ if treated and $Y_i(0)$ if not treated

- Each potential outcome depends **solely** on whether the subject **itself** receives the treatment

- Potential outcomes respond only to the treatment and not to some other feature of the experiment - such as assignment or measurement

The Beauty of Randomization: Independence

- Treatment status is statistically independent of potential outcomes and background attributes **X**

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

- 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.

Validity and Data Quality

Internal validity

- Answer the question: do the experimental data allow for correct causal inferences?

- Depends fundamentally on experimental design and data analysis.

External validity

- Answer the question: do the experimental results generalise to the real world?

- Depends fundamentally on experimental design and data analysis.

Exclusion restriction

- Let $Y_i(z, d)$ be the potential outcome when $z_i = z$ and $d_i = d$ for $z \in (0, 1)$ and for $d \in (0, 1)$

- For example, if $z_i = 1$ and $d_i = 1$, the subject is assigned to the treatment group and receives the treatment

- Or $z_i = 1$ and $d_i = 0$ - subject is assigned treatment but does not receive treatment

- The exclusion restriction is that $Y_i(1, d) = Y_i(0, d)$ - subjects only respond to input from d_i

- The excludability assumption cannot be verified empirically because we never observe both and for the same subject

Classic Drug Experiment Example

- Treatment group receives a new drug
- Control group receives nothing
- Experiment confounds this treatment with receipt of a pill
- If patients respond to the pill rather than the pill's ingredients, excludability is violated
- Jeopardizes unbiasedness of the difference-in-means estimator

Non-interference

- Permits us to ignore the potential outcomes that would arise if subject i were affected by the treatment of other subjects

- Formally, we reduce the schedule of potential outcomes $Y_i(\mathbf{d})$, where $\mathbf{d}$ describes all of the treatments administered to all subjects, to a much simpler schedule $Y_i(d)$, where d refers to the treatment administered to subject i.

- Implies that so long as a subject's treatment status remains constant, that subject's outcome is unaffected by the particular way in which treatments happened to be assigned to other subjects

Non-interference violated

- Police patrols displace crime from treated to untreated areas
- Non-interference violated if your estimand is following:
  - Average potential outcome when a block is treated minus average potential outcome when no blocks are treated
- If police patrols displace crime from treated to untreated areas, observed outcomes in control will not be potential outcomes when no treatment administered anywhere
- Estimated ATE will tend to exaggerate the true ATE

Core assumptions violated?

- Public Health: Providing an infectious disease vaccine to some individuals may decrease the probability that nearby individuals become ill
- Politics: Election monitoring at some polling stations may displace fraud to neighboring polling stations
- Economics: Lowering the cost of production for one firm may change the market price faced by other firms
- Marketing: Advertisements displayed to one person may increase product recognition among her work colleagues

Estimation

Difference-in-means is an unbiased estimator of ATE

$$E\left[\frac{\sum_1^m Y_i}{m} - \frac{\sum_{m+1}^N Y_i}{N-m}\right] = E\left[\frac{\sum_1^m Y_i}{m}\right] - E\left[\frac{\sum_{m+1}^N Y_i}{N-m}\right]$$
$$= E[Y_i(1)] - E[Y_i(0)]$$
$$= E[\tau_i] = ATE$$

Estimation

Difference-in-means estimator implemented via OLS

$$
\begin{aligned}
Y_i &= Y_i(0)(1 - d_i) + Y_i(1)d_i \\
&= Y_i(0) + (Y_i(1) - Y_i(0))d_i \\
&= \mu_{Y(0)} + [\mu_{Y(1)} - \mu_{Y(0)}]d_i + Y_i(0) - \mu_{Y(0)} \\
&\quad + [(Y_i(1) - \mu_{Y(1)}) - (Y_i(0) - \mu_{Y(0)})]d_i \\
&= \alpha + \beta d_i + \epsilon_i,
\end{aligned}
$$

$\alpha = \mu_{Y(0)}$ (average of untreated potential outcomes for all N),
$Y_i(0) =$ untreated potential outcome
$\beta = \mu_{Y(1)} - \mu_{Y(0)}$
and ϵ_i comprises idiosyncratic variation in untreated responses plus idiosyncratic variation in treatment effects.

Chattopadhyay & Duflo 2004

- Randomized policy experiment in India
- 1990s, one-third of village council heads reserved for women
- women.csv contains subset of data from West Bengal
- Gram Panchayat (GP) = level of government
- Analysis?
  - Was randomization implemented properly?
  - Conjecture: more drinking facilities under women
  - Conjecture: no effect on irrigation

Name	Description
GP	An identifier for the Gram Panchayat (GP)
village	identifier for each village
reserved	binary variable indicating whether the GP was reserved for women leaders or not
female	binary variable indicating whether the GP had a female leader or not
irrigation	variable measuring the number of new or repaired irrigation facilities in the village since the reserve policy started
water	drinking-water facilities in the village since the reserve policy started

Table 4.6: *The Variable Names and Descriptions of the Women as Policy Makers Data.*

```
women <- read.csv("women.csv")

## proportion of female politicians
## in reserved GP vs. unreserved GP
mean(women$female[women$reserved] == 1)
[1] 1

mean(women$female[women$reserved == 0])
[1] 0.07476636
```

```
## drinking-water facilities
mean(women$water[women$reserved == 1]) -
    mean(women$water[women$reserved == 0])

## [1] 9.25223

## irrigation facilities
mean(women4irrigation[women$reserved == 1]) -
    mean(women$irrigation[women$reserved == 0])

## [1] -0.3693319
```

Lecture 1 Assignment

Analyses of Bertrand and Mullaithan (2004)

- Exercise 1 - Load data and generate a summary statistics table (mean and sd) of resume characteristics.
- Exercise 2 - Check for balance of mean resume characteristics between experimental conditions.
- Exercise 3 - Why do we care about balance characteristics? Would imbalance be a threat to internal or external validity?
- Exercise 4 - Test the null hypothesis that the callback rates are equal across racial groups.
- Exercise 5 - Test whether returns to resume quality differ between racial groups.