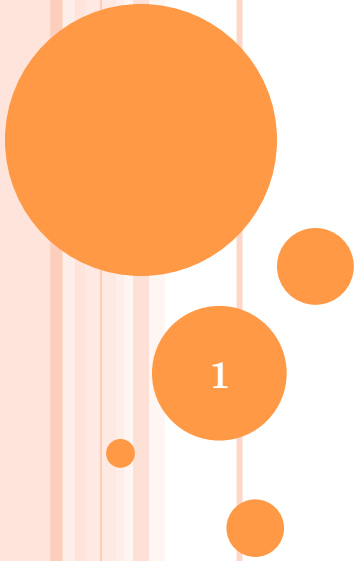
EITM



1

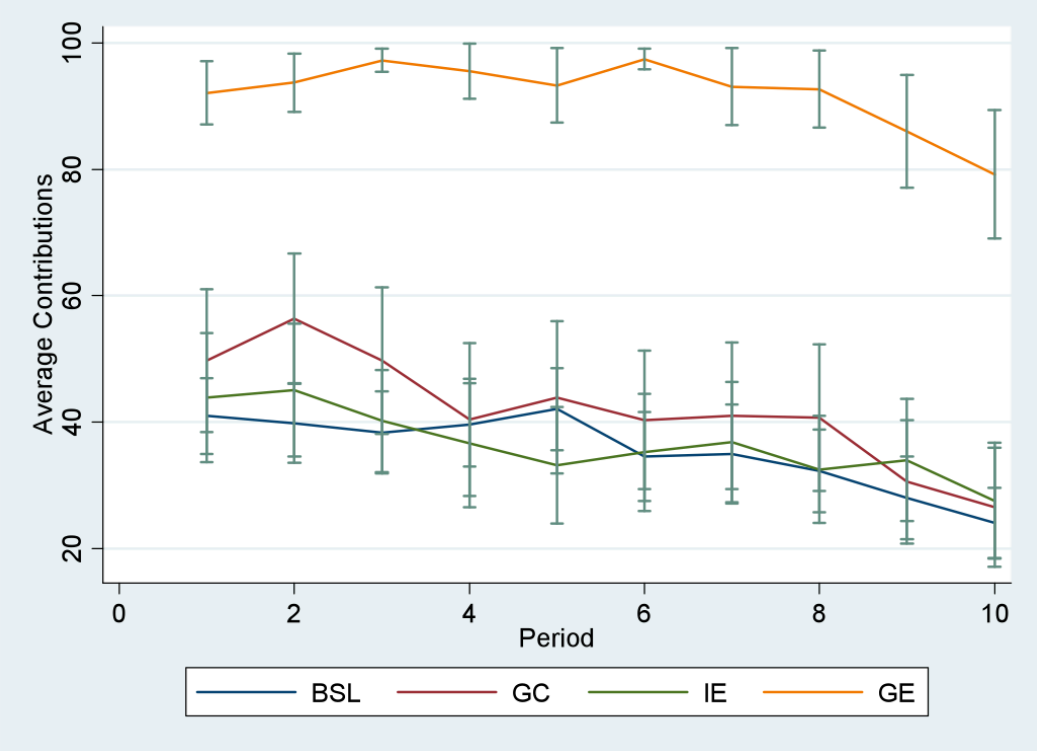
Experimental Design

AND NOW FOR SOMETHING COMPLETELY DIFFERENT

A WELL PLANNED EXPERIMENT ONLY NEEDS

$$\mu_T \neq \mu_C$$

OR



What Is an Experiment?

- Definition: A test of a hypothesis or demonstration of a fact under conditions manipulated by a researcher.
- Key elements:
  - Control, control, control
  - Simplify, simplify, simplify
  - Randomize, randomize, randomize
- Replication
  - Direct
  - Extensions by population or concept

WHY?

- Only Method which can prove causality
- You can create the conditions you wish to investigate your variables of interest without having to wait for it to occur in real world
- You can control and manipulate treatment conditions
- You can create and design measurement tools specifically for your topic of interest

Objectives of experiments

- Testing theories
- Establish empirical regularities as a basis for new theories
- Testing institutions and environments
- Policy advice and wind-tunnel experiments
- The elicitation of preferences
  - Goods, risk, fairness, time

What an experiment doesn't do

- Substitute for thinking
- Generate hypotheses
  - Caveats:
    - Ideas from subject debriefing
    - Ideas from failures/limitations of your own or others' previous work
- Not an all-purpose tool

WHEN?

- Need clear causal information
- Past work has show inconsistent or contradictory results
- Multi-method validation of formal theories
- Investigate underlying phenomenon
- Triangulation

Theory and experiments

- Test a theory or discriminate between theories
  - Formal theory provides the basis for experimental design
  - Test a theory on its own domain:
    - Implement the conditions of the theory (e.g., preference assumptions, technology assumptions, institutional assumptions)
  - (Best to have an alternative hypothesis)
  - Compare the prediction(s) with the experimental outcome

THEORY, CONT'D

- What if the results reject theory?
  - Helpful if you can design an experiment so that either outcome can inform a particular perspective, but not always possible
- Explore the causes of a theory's failure
  - Check each of the assumptions
  - Check population demographics
  - Explore parameter space
  - Find out when the theory fails and when it succeeds
  - Design proper control treatments that allows causal inferences about why the theory fails

THE ELICITATION OF PREFERENCES

- Inform Policy
  - How much should the government spend on avoiding traffic injuries?
  - How much should be spend on the conservation of nature?
- Measuring people's values is hard
  - Are people risk seeking/averse?
  - Who is cooperative?
  - How can you really measure the value of life/risk?
- Requires a theory of individual preferences and knowledge about the strength of particular "motives" (preferences).
- Sometimes you are interested in fixed preferences (or emotional states)

Objections to experiments

- **Objection: *"Experiments are unrealistic."***
  - All models are unrealistic
    - They leave out many aspects of reality.
    - Simplicity is a virtue – focuses on critical aspects of a situation (a causal mechanism or logic of a complex relationship)
  - Experiments are like models
    - They leave out many aspects of reality
    - Focus on critical aspects (cause or precision of estimate)
  - Realism may be important but so is control.
  - Experimental vs mundane realism

Objections continued

- **Objection: "*Experiments are artificial.*"**
  - Biased subject pool (students)
  - Low stakes
  - Small number of participants
  - Inexperienced subjects
  - Anonymity
- All can be tested in the lab
- Such testing has never overthrown an important result

Objections continued

- **Objection: "*Experiments say nothing about the real world.*"**
  - External validity
  - Generalization
- The experiment, if properly designed, is "real" for the subjects
- What is the aim of the experiment?
  - Internal validity – ensuring that the causal inference is correct
  - Minimizing general claims

Limits of experiments

- **Control is never perfect**
  - Weather, Laboratory environment
  - No real control about all other motives (no dominance)
  - Self-selection: who takes part in the experiment?
- **Randomization is difficult**
- **Experiments (like models) are never general, just examples**
- **Lab experiments compared to field experiments**
  - Difference in control
  - Difference in randomization
  - Problems with ITT

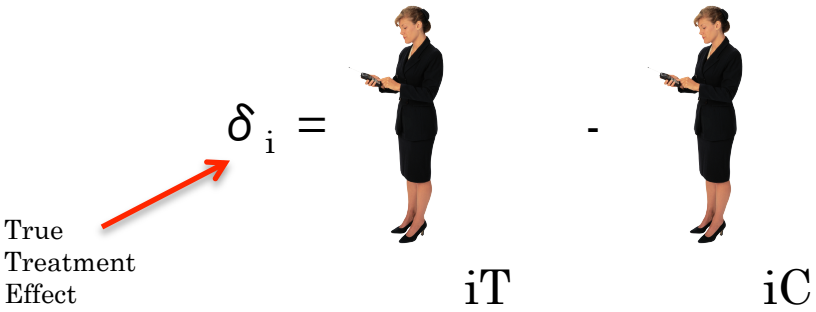
What Experiments Do Well

- Test for causal claims
- Inform theory
- Allow for replication
- Develop measures (problem of reliability and validity)
- Explore parameters of interest
- Control for effect of alternative hypotheses
- Develop counterfactuals

Causal Considerations

RUEBEN CAUSAL MODEL (RCM)

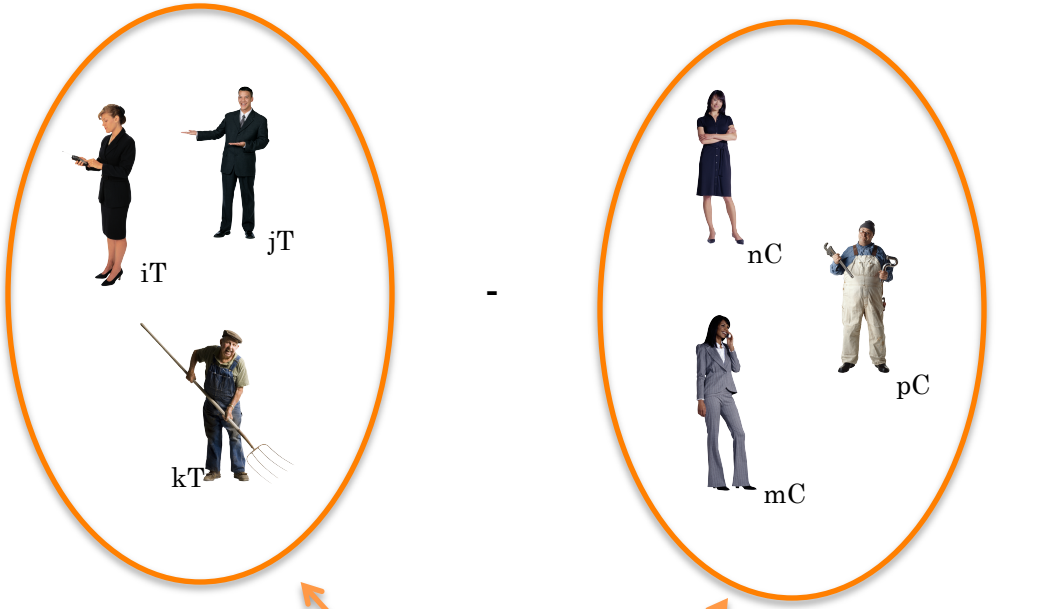
- The dilemma:



- The same "i" can't be in two states at the same time!

RCM – Best Correction

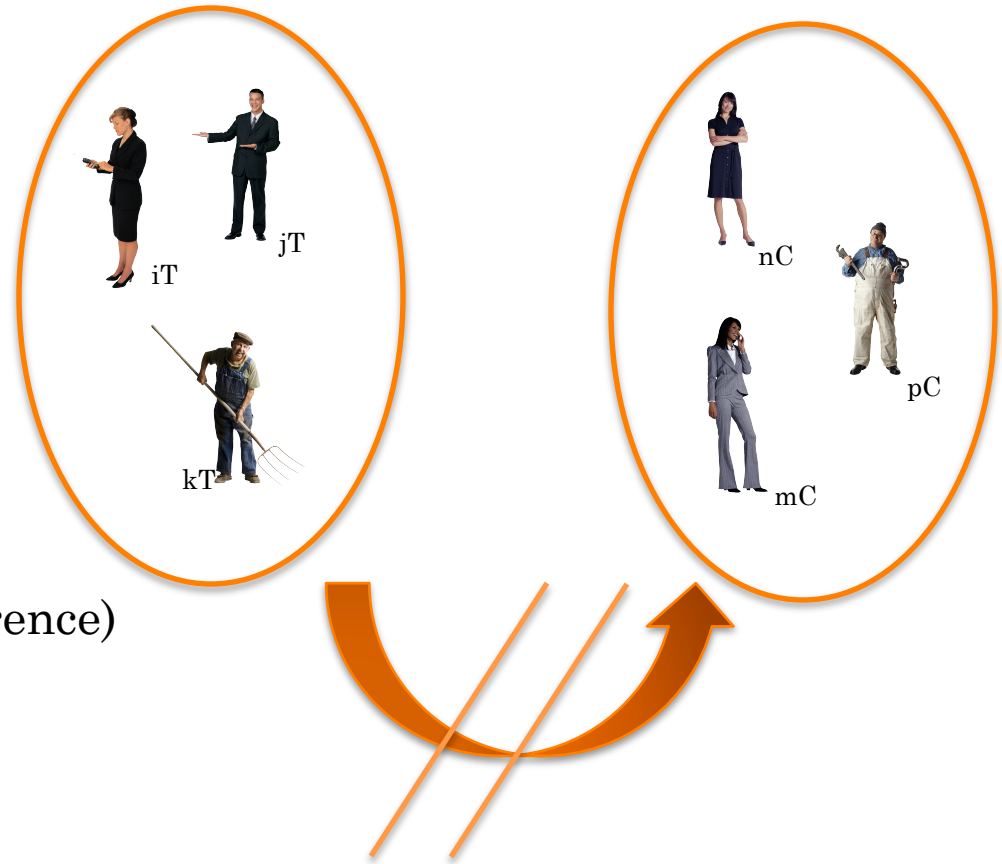
ATE =



Analog: $y = \alpha + \beta_1 X_1 + \varepsilon$

SUTVA – Stable Unit Treatment Value Assumption(s)

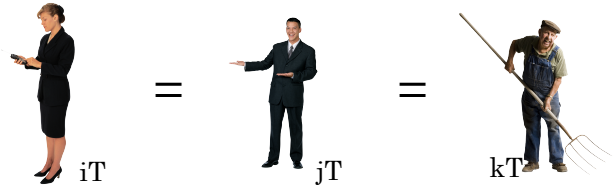
Assumption 1: Treatment ONLY affects the treated.



No spillover
(non-interference)

SUTVA – Assumption 2

Assumption 2: Average treatment effect is homogeneous across individuals



Everyone responds the same to the same dosage/
The treatment for everyone is the same

SUTVA – ASSUMPTION 3

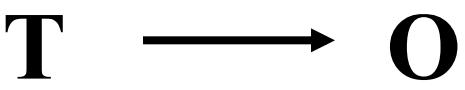
Assumption 3: Treatment is invariant to manner delivered

SUTVA – Assumption 4

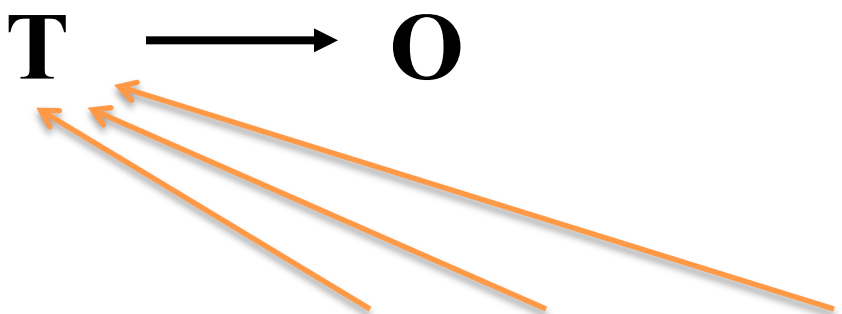
The treatment precedes the action by subject – no simultaneity

Design Considerations

Common Designs: One-Shot Design



Common Designs: One-Shot Design



Inference: none

Statistics: descriptive or kitchen sink

$$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m + \varepsilon$$

SUTVA violations: Everyone is treated (maybe)?

Pre/Post-Test Design

$$O_1 \longrightarrow T \longrightarrow O_2$$

Pre/Post-Test Design

$$O_1 \longrightarrow T \longrightarrow O_2$$

Inference: Something might have caused a difference

Statistics: $O_1 \neq O_2$

SUTVA violations:

Everyone is treated (maybe)?

All possible states of the world are not observed.

Static Group Comparison

Group A $\quad T \longrightarrow O_{1A}$

Group B $\quad O_{1B}$

Static Group Comparison

Group A $\quad T \longrightarrow O_{1A}$

Group B $\quad O_{1B}$
(Control group?
Nope, not Randomized.)

Inference: Something might have caused a difference

Statistics: $O_{1A} \neq O_{1B}$ $\quad y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m + \varepsilon$

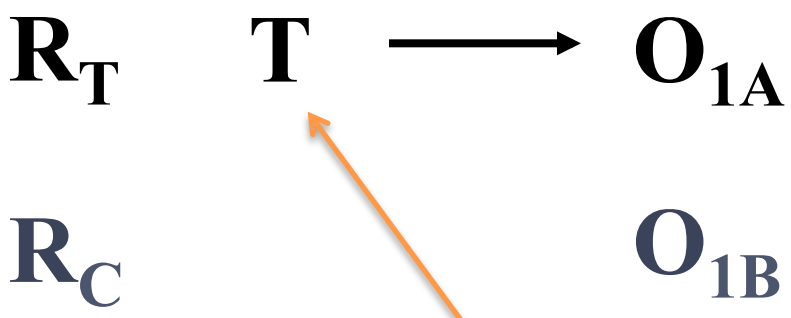
SUTVA violations:
Not clear that treatment only affects the treated
Average treatment effect is not homogeneous across individuals

Randomized Group Comparison

$$R_T \quad T \longrightarrow O_{1A}$$

$$R_C \qquad\quad O_{1B}$$

Randomized Group Comparison



Inference: Very likely the treatment caused the difference

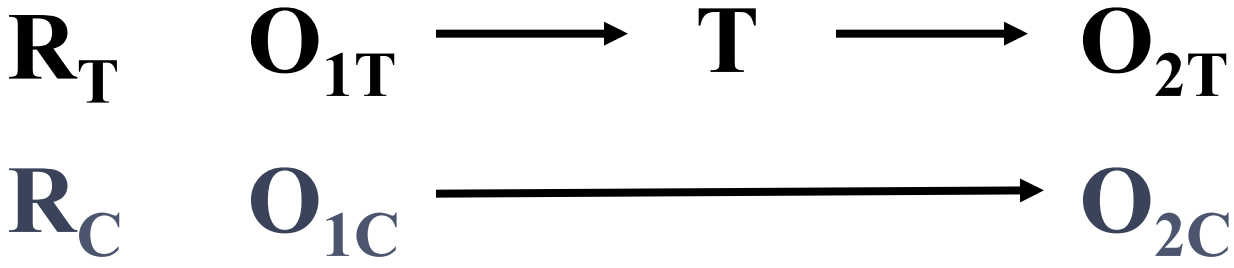
Statistics: $O_{1A} \neq O_{1B}$

$$y = \alpha + \beta_1 X_1 + \varepsilon$$

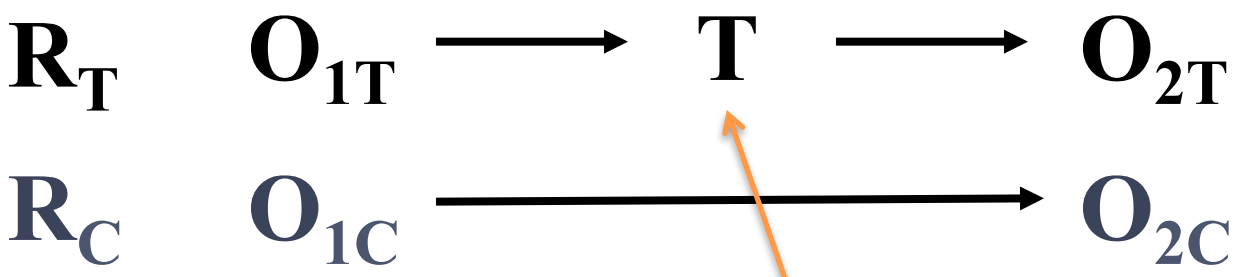
SUTVA violations:

Something might have happened between T and O

PRE/POST CONTROL



PRE/POST CONTROL



Inference: Treatment probably caused a difference

Statistics: $O_{1T} = O_{1C}$; $O_{2T} \neq O_{2C}$

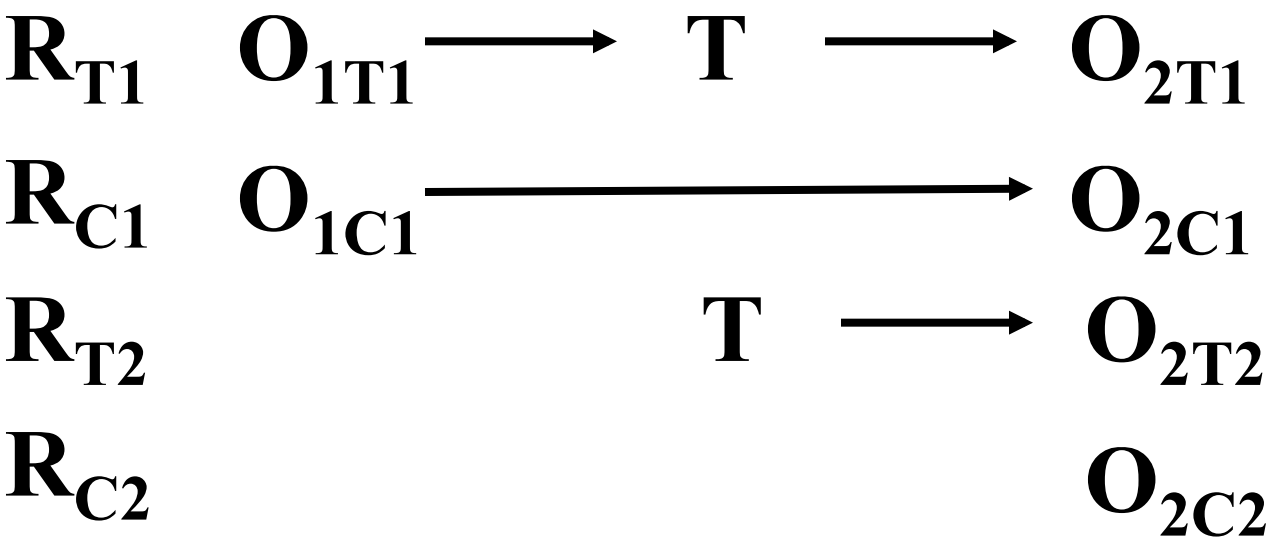
$$y = \alpha + \beta_1 X_1 + \varepsilon$$

where $y = O_2 - O_1$

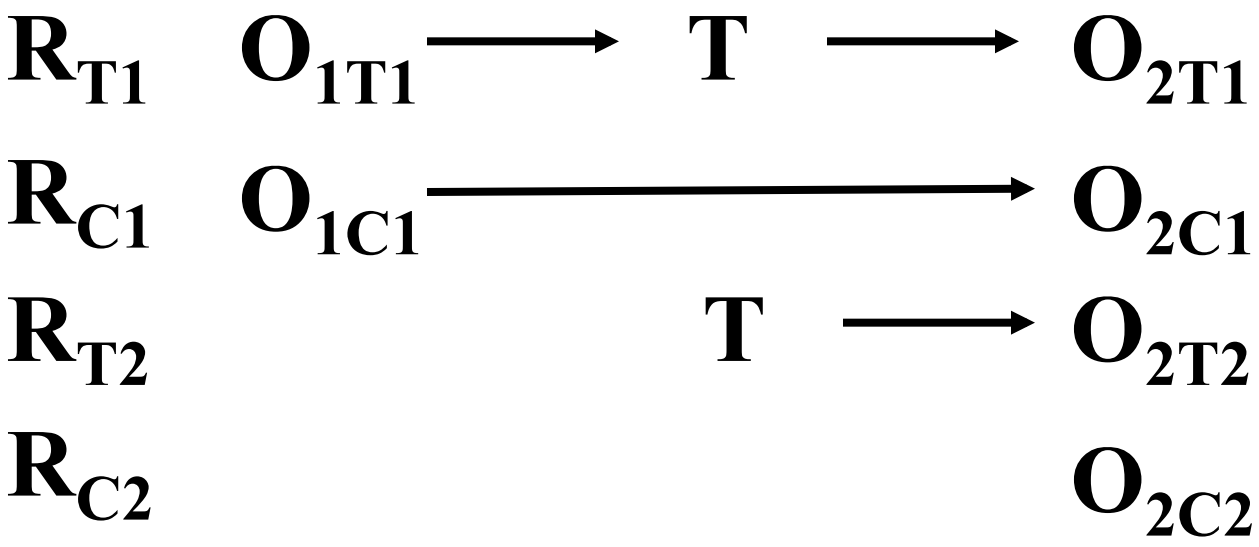
SUTVA violations:

- Treatment ONLY affect the treated?
- Treatment homogeneous across individuals?
- Treatment invariant to delivery method?

Solomon Four-Group



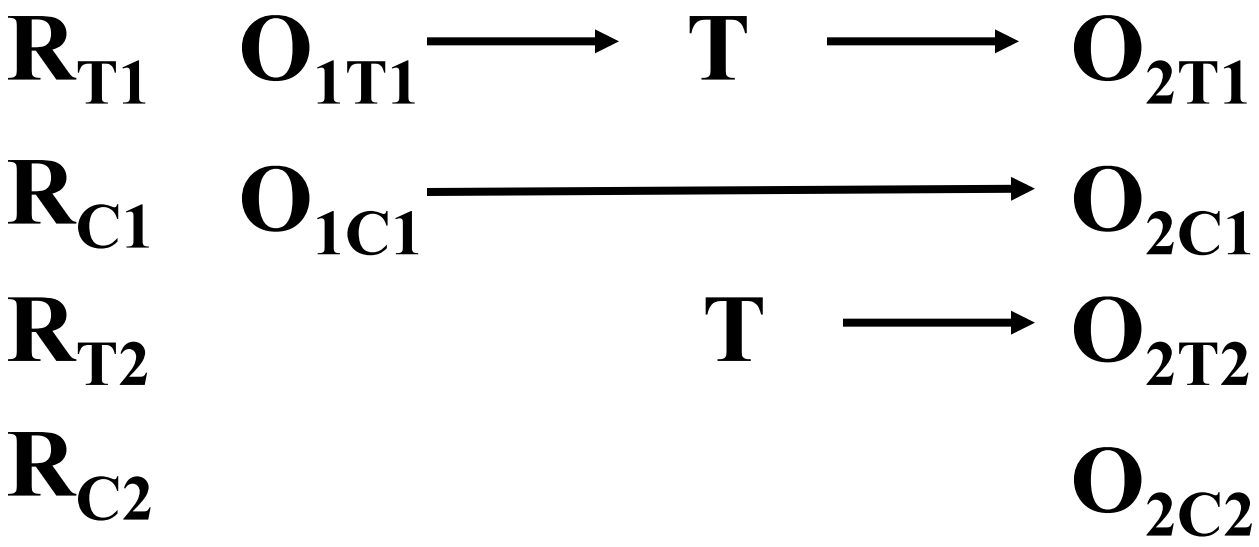
SOLOMON FOUR-GROUP



Inference: Treatment very likely caused a difference

Statistics: $O_{1T1} = O_{1C1} = O_{2C1} = O_{2C2}$;
$O_{2T1} = O_{2T2} \neq O_{2C1} = O_{2C2}$

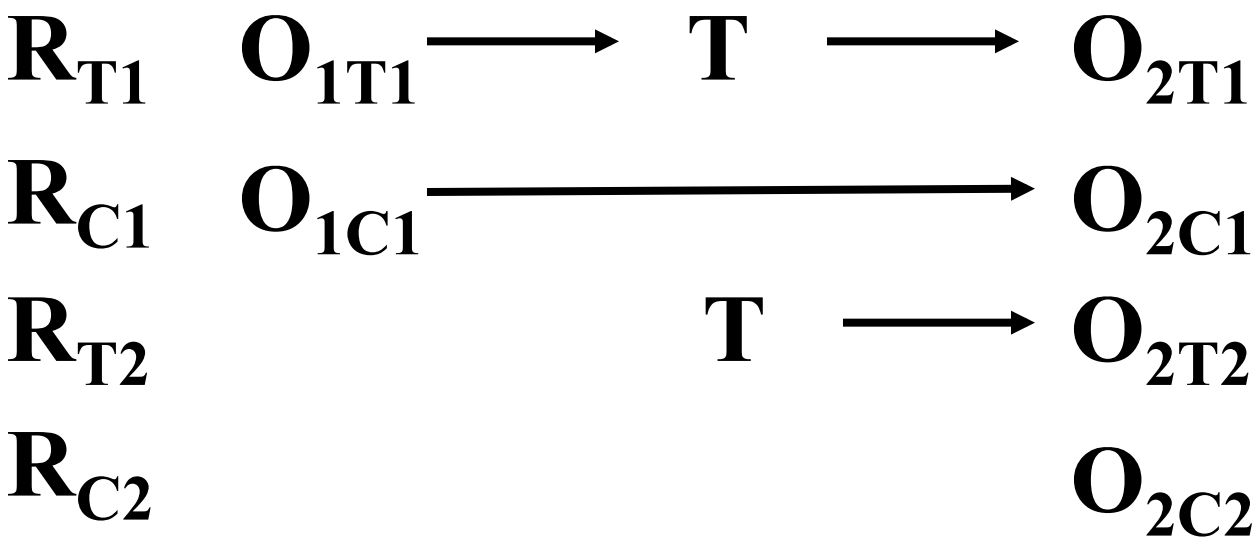
Solomon Four-Group



SUTVA violations:

Treatment ONLY affect the treated?

SOLOMON FOUR-GROUP



NOTE:

This is easy to accomplish in the Lab. It is a nightmare in the field.

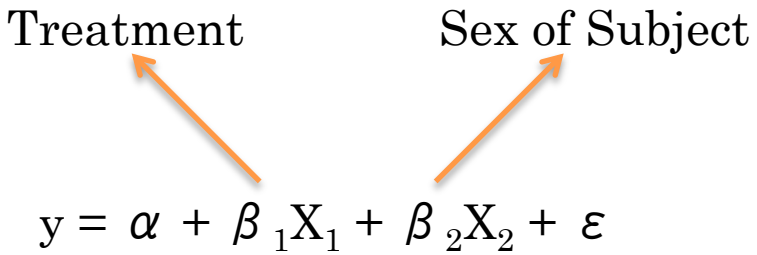
Efficiency in Design: Counter-balance/Within Subject

- Counter-balanced designs
  - $O_1\ T_A\ O_2\ T_B\ O_3$ and $O_1\ T_B\ O_2\ T_A\ O_3$
    - Builds within subject design
    - Decreases the number of trials
    - Accounts for treatment ordering effect
- Cross-over designs
  - $O_1\ T_A\ O_2\ T_B\ O_3\ T_A\ O_4$
    - Accounts for treatment plus learning

Efficiency In Design: Blocking (and Matching)

- Randomized Blocking Designs
  - Suppose a 2(H,L)x2(B,S) within subject design with a 4 game ordering effect (HB, HS, LB,LS).
    - Would require 16 cells under complete factorial design
    - Blocking allows 4 cells: (1) HB,HS,LB,LS (2) HS,LB,LS,HB (3) LB,LS,HB,HS (4) LS,HB,HS,LB
  - Assumptions
    - Blocks must be homogeneous
    - Blocks must be randomly assigned
- Blocking on "nuisance" variables
  - Sex is not randomly assigned

Note: Similar to Imai et al. APSR 2011 "parallel encouragement design"



Practical Design Considerations

WHAT MUST BE DESIGNED?

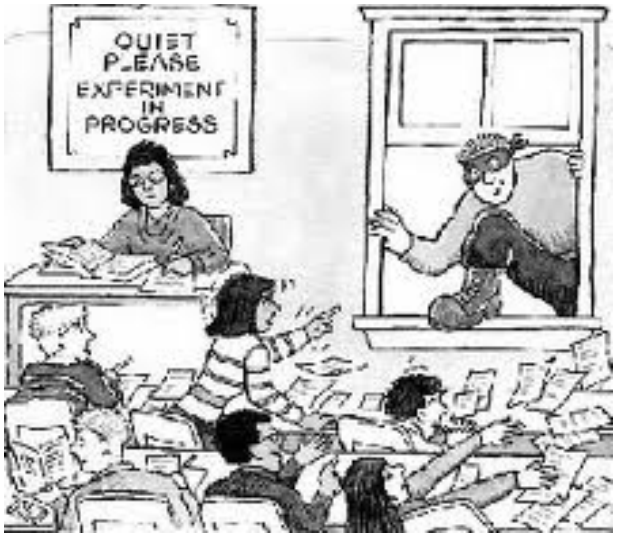
- "Laboratory experimental design involves designing a microeconomic system"
  - Vernon Smith, AER, December, 1982
- Environment:
  - Agents (Number, type, motivation)
  - Commodities -- what do decisions get made over?
  - Endowments -- what do the decision-makers have at the outset?
  - Mechanism by which learning can occur (search opportunities, practice)
- Institution:
  - Decisions available to subjects
    - Rules about choices
    - Rules about communication
  - Connection between decisions and payoffs

Fatal Errors in Design

- Inadequate or inappropriate incentive
- Nonstandardized instructions
- Uncontrolled effects of psychological biases
- Insufficient statistical power
- Failure to provide a calibrated baseline
- Change in more than one factor at a time (confounds)
- Subjects bring themselves to the experiment: You do yourself a serious injustice not to recognize what this means for your study

Additional Sources of Error

- Demand Characteristics
- Experimenter Bias: Protocol
- Expectancy Effects: Communication
- Incentives



INCENTIVES: INDUCED VALUE THEORY
SMITH (AER 1976; AER 1982)

- In many experiments the experimenter wants to **control** subjects' preferences. How can this be achieved?
- Subjects' homegrown preferences must be "neutralized" and the experimenter "induces" new preferences. Subjects' actions should be driven by the induced preferences but may not be.
- Reward Medium: Money
- Assumption: People care about money and some other motives.
  - Note 1: money may function as the "price" of other motives
  - Note 2: sometimes you are interested in "homegrown preferences." But be willing to adjust for heterogeneous treatment effects (Imai et al. APSR 2011). Example: partisan preferences.

Incentives Continued
Minimal Conditions for Control

- **Monotonicity/nonsatiation:** Subjects must prefer more of the reward medium to less and not become satiated.
- **Salience:** The reward depends on a subject's actions (note: show up fee is not salient).
- **Dominance:** Changes in a subject's utility from the experiment come predominantly from the reward medium and the influence of the other motives is negligible (this assumption is the most critical).
- If these conditions are satisfied, the experimenter has **control** of the subjects' preferences, i.e., there is an incentive to perform actions that are paid.

Discussion of incentive effects

- In experiments in which incentives have an effect, the difference between no and low incentives is often bigger than the difference between low and high incentives.
- Higher incentives often lead to a reduction of the variance of decisions (Smith&Walker, IntJGameTheory 1993).
- Treatment effects are often at least as high as incentive effects.
- Payment of subjects necessary for getting published (in econ)
  - 100% of experimental papers since 1970 have monetary incentives
  - 26% of papers in Journal of Behavioral Decision Making do

Incentives Continued: Qualifications

- Subjective costs (controlled by a commission or by raising stakes)
- Utility of winning or earning points
  - In some environments this can look like risk aversion (overbidding in common value auctions)
- Payoffs to others may matter
  - Envy, egalitarianism
- Desire to please the experimenter
- Potential solutions
  - Make the change in monetary payoffs sufficiently large
  - Avoid public information about payoffs
  - Do not give hints about the purpose of the experiment
  - Use a neutral language in the instructions
  - Use a well trained assistant to run the experiment

NON-MATERIAL INCENTIVES

- Economists think money is the only incentive and you have to pay subjects to get published in those journals.
- As you know, they are wrong.
- Other incentives that work:
  - Class credit (vast majority of psychology experiments)
  - Food and drink (some behavioral econ; this has the added effect of making subjects happy which you or may not want)
  - Status elevation
  - Sex, drugs and music (often causes IRB issues☺)

Uncontrolled Psychological biases

- Loss aversion
  - Avoid losses or zero payoff options
- Status quo bias
  - Avoid accidentally anchoring subjects
  - Experimenter demand: experimenter can accidentally set the status quo by signaling expected behavior
- Endowment effect
  - Willingness to accept v. willingness to pay

Uncontrollable Psychological Bias (con't)

- Emotion
  - Subjects show up in different moods which can affect decision making, some predictable, others not
- Emotion can be manipulated for experimental purposes (some easier than others)
  - Film
  - Music
  - Food
  - Writing
- Always induce happy before they leave
- Always debrief

Statistical Considerations

Insufficient Statistical Power

- You must have enough data to do a statistical test
- Plan ahead – decide what test you want to do and run the experiment that will let you do it
  - Comparative Statics?
  - Panel Design
    - (problems of independence)
- Avoid too many treatments
  - Complete Factorial Designs
  - (# factors)*(#factors)*(#factors)
- Calculate your power test

Insufficient Statistical Power: Power Tests, I

- Need three elements:
  - Significance criterion – specify the trade off between Type I and Type II errors (both α and β). (Even a Bayesian has to worry about low power for updating beliefs)
  - Magnitude of the effect
    - ATE or LATE: $(MEAN_T - MEAN_C)$
    - Standardized Effect Size (with common variance) $(MEAN_T - MEAN_C)/\sigma$
    - Maximize the expected difference in effects!
    - Pretest Data can inform you about means and variance
  - Sample size
    - Obviously related to the sample error – as sample size goes up, sampling error goes down
    - Measurement precision helps here as well – decrease variance

Insufficient Statistical Power: Power Tests, II

- Many tools available
  - in r: power.t.test(n, delta, sd, sig.level, power, type, alternative) – omit n and it will be calculated.
  - in STATA: sampsi mean1 mean2, sd1(value) sd2(value)

Insufficient Statistical Power: Power Tests, II

- Examples – Precision in the SD
  - **Between** Ss

```
. sampsi 5.1 4.3, sd1(3.3) sd2(3.0)

Estimated sample size for two-sample comparison of means

Test Ho: m1 = m2, where m1 is the mean in population 1
                  and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.9000
            m1 =      5.1
            m2 =      4.3
           sd1 =      3.3
           sd2 =        3
         n2/n1 =     1.00

Estimated required sample sizes:

            n1 =      327
            n2 =      327
```

```
. sampsi 5.1 4.3, sd1(1.3) sd2(1.0)

Estimated sample size for two-sample comparison of means

Test Ho: m1 = m2, where m1 is the mean in population 1
                  and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.9000
            m1 =      5.1
            m2 =      4.3
           sd1 =      1.3
           sd2 =        1
         n2/n1 =     1.00

Estimated required sample sizes:

            n1 =       45
            n2 =       45
```

Insufficient Statistical Power: Power Tests, II

- Examples – Maximize differences in means
  - **Between** Ss

```
. sampsi 5.1 4.3, sd1(3.3) sd2(3.0)

Estimated sample size for two-sample comparison of means

Test Ho: m1 = m2, where m1 is the mean in population 1
                  and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.9000
            m1 =      5.1
            m2 =      4.3
           sd1 =      3.3
           sd2 =        3
         n2/n1 =     1.00

Estimated required sample sizes:

            n1 =      327
            n2 =      327
```

```
. sampsi 5.1 3.3, sd1(3.3) sd2(3.0)

Estimated sample size for two-sample comparison of means

Test Ho: m1 = m2, where m1 is the mean in population 1
                  and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.9000
            m1 =      5.1
            m2 =      3.3
           sd1 =      3.3
           sd2 =        3
         n2/n1 =     1.00

Estimated required sample sizes:

            n1 =       65
            n2 =       65
```

Insufficient Statistical Power: Power Tests, III

- **Within** Subjects

```
. sampsi 4.3 5.1, sd1(4.3) onesample

Estimated sample size for one-sample comparison of mean
  to hypothesized value

Test Ho: m =    4.3, where m is the mean in the population

Assumptions:

        alpha =   0.0500  (two-sided)
        power =   0.9000
 alternative m =      5.1
           sd =      4.3

Estimated required sample size:

            n =      304
```

```
. sampsi 4.3 5.1, sd1(1.3) onesample

Estimated sample size for one-sample comparison of mean
  to hypothesized value

Test Ho: m =    4.3, where m is the mean in the population

Assumptions:

        alpha =   0.0500  (two-sided)
        power =   0.9000
 alternative m =      5.1
           sd =      1.3

Estimated required sample size:

            n =       28
```

Small Sample Problem

- Small samples create several problems
  - Distributions are difficult to calculate (normal? etc.)
  - Outlier can have dramatic impact
- But experimental design gets you around many statistical problems

Control	TRT
10	14
12	16
13	17
15	19
17	21
34	23
21	25
24	29
29	31

Classic Parametric Tests

- t-tests
  - Null Hypothesis: (Control – Treatment)=0 (2-tail)
  - Pluses
    - Useful in 2 factor designs
  - Minuses
    - Small sample problems (distributions, outliers, power)
- Example

Classic Parametric Tests

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
1	9	19.44444	2.734101	8.202303	13.1396	25.74929
2	9	21.66667	1.950783	5.85235	17.16815	26.16518
combined	18	20.55556	1.651346	7.006066	17.07152	24.03959
diff		-2.222222	3.3587		-9.342348	4.897903

diff = mean(1) - mean(2) t = -0.6616

Ho: diff = 0 degrees of freedom = 16

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0

Pr(T < t) = 0.2588 Pr(|T| > |t|) = 0.5176 Pr(T > t) = 0.7412

Within Group Parametric

- Paired t-test
  - Null Hypotheses: $(O_1-O_2)=0$
  - Pluses
    - Each pair of observations is on the individual
    - Differences presumably due to treatment
  - Minuses
    - Testing threats to internal validity

Within Group Parametric

Paired t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
control	9	19.44444	2.734101	8.202303	13.1396	25.74929
trt	9	21.66667	1.950783	5.85235	17.16815	26.16518
diff	9	-2.222222	1.673136	5.019407	-6.08048	1.636035

mean(diff) = mean(**control - trt**) t = **-1.3282**

Ho: mean(diff) = 0 degrees of freedom = **8**

Ha: mean(diff) < 0 — Pr(T < t) = **0.1104**

Ha: mean(diff) != 0 — Pr(|T| > |t|) = **0.2208**

Ha: mean(diff) > 0 — Pr(T > t) = **0.8896**

Losing the Outlier

(Still no power)

Control	TRT
10	14
12	16
13	17
15	19
17	21
18	23
21	25
24	29
29	31

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
1	9	17.66667	2.041241	6.123724	12.95956	22.37378
2	9	21.66667	1.950783	5.85235	17.16815	26.16518
combined	18	19.66667	1.452966	6.164414	16.60118	22.73216
diff		-4	2.823512		-9.985579	1.985579

diff = mean(1) - mean(2) t = -1.4167

Ho: diff = 0 degrees of freedom = 16

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0

Pr(T < t) = 0.0879 Pr(|T| > |t|) = 0.1758 Pr(T > t) = 0.9121

Losing the Outlier

(Within Subject)

Control	TRT
10	14
12	16
13	17
15	19
17	21
18	23
21	25
24	29
29	31

Paired t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
control	9	17.66667	2.041241	6.123724	12.95956	22.37378
trt	9	21.66667	1.950783	5.85235	17.16815	26.16518
diff	9	-4	.2886751	.8660254	-4.665686	-3.334314

mean(diff) = mean(control - trt) t = -13.8564

Ho: mean(diff) = 0 degrees of freedom = 8

Ha: mean(diff) < 0 Pr(T < t) = 0.0000

Ha: mean(diff) != 0 Pr(|T| > |t|) = 0.0000

Ha: mean(diff) > 0 Pr(T > t) = 1.0000

Non-Parametric

- Go through some basic non-parametric tests. When to use each? What they tell you?

Non-Parametric Tests, I

- Mann-Whitney rank sum (between subject)
  - Given by *U* which is the sum of the ranks from the different samples.
  - Take the Tortoise & Hare example. There are 6 of each all running the race at the same time. The finish order is the following: THHHHHTTTTTH
  - Calculate U_T for the Tortoises. Highest ranking tortoise beat 6 Hare's. Remaining tortoises only beat one Hare. Hence: 6+1+1+1+1+1=11
  - Calculate U_H for the Hares. Highest ranking Hare beat 5 tortoises. Next highest did as well, and so on. Hence: 5+5+5+5+5+0
  - Use smallest U and consult table for p-values

Non-Parametric Tests, I

Sensitive to sample size.

Relies on medians.

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

      group |      obs       rank sum    expected
------------+---------------------------------------
          1 |        9           75.5        85.5
          2 |        9           95.5        85.5
------------+---------------------------------------
   combined |       18            171         171

unadjusted variance        128.25
adjustment for ties         -0.40
                           ------
adjusted variance          127.85

Ho: control(group==1) = control(group==2)
                z =  -0.884
    Prob > |z| =   0.3765
```

Non-Parametric Tests, II

- Wilcoxon signed rank (within subject)
  - Does the same as Mann-Whitney
  - Compares between pairs of observations

Non-Parametric Tests, II

Wilcoxon signed-rank test

sign	obs	sum ranks	expected
positive	1	9	22.5
negative	8	36	22.5
zero	0	0	0
all	9	45	45

unadjusted variance	71.25
adjustment for ties	-4.38
adjustment for zeros	0.00
adjusted variance	66.88

Ho: control = trt

z = -1.651

Prob > |z| = 0.0988

Distributional Tests

- Kolmogorov-Smirnov
  - Good for small data sets that
    - Have the same mean, but skewed distributions OR
    - Have different means, but distributions are not normal
  - Works for two-sample tests
  - Non-parametric (similar to Epps-Singleton)
- Kruskal-Wallis k-sample
  - This allows an k-sample test among distributions

CALIBRATION

- Keep in mind that you are producing a data set
- Include a "baseline" in the experimental design
- Set parameters so you can be sure to tell if hypothesis is supported
- Ideally, you need a competing hypothesis that is "far away" in the design space.
- Try to factor in "noise" in behavior – variability in the performance of the subjects. Lots of noise means results are hard to determine.
- Develop criteria for rejection

Conclusion of Stats

- Note what I have been silent about:
  - Covariate adjustments
    - Fishing
    - Limiting inferential power
  - Inference to Population
    - Mostly concerned with causal effects
    - Danger of heterogeneous treatment effects
  - Interactions
    - If you expect them, design them into the experiment
    - Full factorial design can help.

Nuts and Bolts

First Steps (Practical Advice)

- Begin with Theory. Translate theory to lab.
- Begin with phenomenon. Design experiments to dissect
- Begin with something you want to measure. Design experiment to measure it.

General Principles

- Experimental Design = Selection and Arrangement of Conditions
- Minimum of two conditions required, one of which is control
- Manipulate a minimum number of variables relative to your cases (Simplify)
- Randomize
- Plan for Replication from outset

Experimental Designs

- Between Subject
- Within Subject (A-B-A): Counterbalancing
- Matching (Within)
- Field Experiments
- Natural Experiments

"There's a flaw in your experimental design. All the mice are scorpios."

Experimental Measures

- Self-reports (Surveys)
- Behavioral Measures
- Biological Measures
  - Physiological
  - Hormonal and Genetic
  - MRIs, EEGs, etc
- Observational Measures (Video)

Subject Selection

- Specialized Groups:
  - Elderly
  - Professionals
  - Medical cases
  - Poor
  - Residents of hurricane-vulnerable areas
  - Public officials
- Population Samples
  - Pluses: External validity, Heterogeneity
  - Minuses: Costly, decreased control, heterogeneity

Experimental Ethics

- Informed Consent
- Avoid Harm
- Deception
- Incentives
- Debriefing

An abandoned ham sandwich? Or Psychology Department experiment? There was no way Fred could tell for sure.

Basic Ethical Principles

- DO. NOT. FABRICATE. DATA. EVER.
- DO. NOT. BREAK. THE. LAW. EVER.
- Do get IRB approval for all studies involving human or animal subjects.
- Be transparent about procedures
- Mount data for replication on completed studies
- Respect confidentiality of subjects
- Useful strategies:
  - Always be present when your experiments are being run. Don't run them, but be there to observe and address issues as they come up. Plus you learn useful stuff.
  - Always be your own first subject so you know what the experience is like from the inside out.
  - Always check on data obtained from surveys or on line platforms to verify authenticity

General Remark

- Whether the conditions implemented in the laboratory are also present in reality will probably always be subject to some uncertainty.
- Therefore, laboratory experiments are no substitute
  - for the analysis of field happenstance data
  - for the conduct and the analysis of field experiments
  - and for survey data.
  - And field experiments are no substitute for lab based ones either
- We support use of a combination of all these empirical methods.