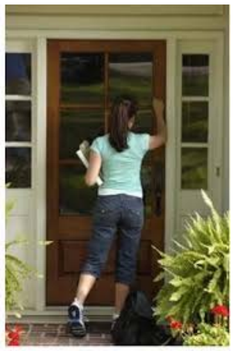
One-sided noncompliance

- Let's start with an application
  - Treatment: Canvasser talks about the upcoming election. Control: Nothing. (N = 1000 in each).
  - Measure: Voter turnout (available in the U.S.)
  - But some people in the treatment are not at home: don't receive the treatment



- One-sided noncompliance: some subjects assigned to treatment don't receive the treatment
  - 250 participants received the treatment, *compliers* (i.e. 75% of the treatment group goes untreated, *never-takers*)
- We could compare everyone anyway 1000 vs 1000, or 250 vs 1000, or 1750 vs 250?. What's the meaning of this?

One-sided noncompliance: Causal effect

- Let's come back to the definition of z_i for treatment assignment (1,0) and d_i for treatment received (1,0)
- Let's define $d_i(z)$ as:
  - $d_i(1) = 0$: subject assigned to treatment is untreated
  - $d_i(0) = 1$: subject assigned to control is treated

  (complete the other two cases)
- For one-sided noncompliance, what is the value of $d_i(0)$ and $d_i(1)$?
- We would say that $\text{ATE}|d_i(1) = 1$ is the average treatment effect among compliers

Intent-to-Treat effects (*ITT*)

- The causal effect of assignment, in this case, is called *ITT*
- We can write the intent-to-treat of z_i on d_i (ITT_D) as:

$$ITT_D = E[d_i(1)] - E[d_i(0)]$$

  - But we are assuming one-sided noncompliance: $E[d_i(0)] = 0$
- The, we can write the intent-to-treat of z_i on Y_i (ITT) as:

$$ITT = E[Y_i(z=1,d(1))] - E[Y_i(z=0,d(0))]$$

- What does it mean?
- When is this important? Do you want to know the average effect of d_i on Y_i, or the average effect of z_i on Y_i?
- What does it happen when compliance = 100%?

Complier average causal effects (CACE)

- If we want to know: ATE = $E[Y_i(d = 1) - Y_i(d = 0)]$
- When there is noncompliance we have: "Complier Average Causal Effect" (CACE), also called local average treatment effect (LATE):

$$\text{CACE} = \underbrace{E[Y_i(d = 1) - Y_i(d = 0)}_{\text{Average Treatment Effect}} | \underbrace{d_i(1) = 1]}_{\text{Among compliers}}$$

$\Leftrightarrow$

ITT and CACE: Application

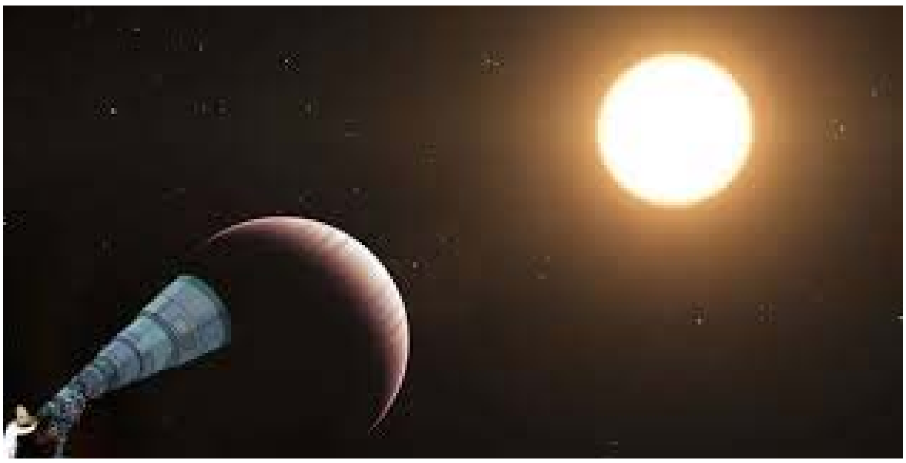
- New Haven voter mobilization experiment (Gerber & Green, 2000)

	Treatment group	Control Group
Turnout rate among those contacted by canvassers	54.43 (395)	
Turnout rate among those not contacted by canvassers	36.48 (1,050)	37.54 (5,645)
Overall turnout rate	41.38 (1,445)	37.54 (5,645)

- $\widehat{ITT}$?, $\widehat{ITT}_D$?, $\widehat{CACE}$? Interpretation (what do you care about?)

- We will see the statistical inference of this type of cases later on in this course

Power analysis: Motivation



Ellis (2010)

Remember this?

- Type I error: Seeing something when there is nothing
- Type II error: Seeing nothing when there is something

Ellis (2010)

Power analysis

- Asymptotic approximation for statistical power using a randomized intervention with N participants in a N/2 binary treatment

$$\beta = \Phi\left(\frac{|\mu_t - \mu_c|\sqrt{N}}{2\sigma} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)$$

β: statistical power; $\Phi(\cdot)$: normal cumulative distribution function (CDF); $\Phi^{-1}(\cdot)$: Inverse of the normal CDF; μ_i: normally distributed outcome with mean μ_i for the treatment and control group $i = \{t,c\}$; σ: pooled standard deviation; α: level of statistical significance

  – You can try this in Stata or Excel (e.g. use N = 500; $\mu_c = 60$; $\mu_t = 65$; $\sigma = 20$; $\alpha = 0.05 \Rightarrow \beta = 0.80$)
- What's the interpretation of this?

Power analysis: Practical issues

- Remember that a certain "power" will tell us the likelihood of making a type II error
- A power analysis should be conducted before a study is conducted → determine sample size
  - This depends on the effect size (from previous study, from experts' judgement). This may difficult for new studies, online and lab studies.
- We may start by answering what's the minimum acceptable sample size?
- What are the ways to increase power?

Power analysis: Sensitivity

- A manufacturing company wants to test a new training program for its workers. They know that in average workers can assemble 78 pieces/day (SD = 4.5), and the new program would be cost-effective if it improves productivity to 80 pieces/day