

Modes of Statistical Inference for Causal Effects

Plus an overview of the testing based approach to causal
inference for experiments on networks

CSBS Causal Inference Workshop @ Illinois

Jake Bowers

Political Science & Statistics

<http://jakebowers.org>

Senior Scientist, <http://thepolicylab.brown.edu>

Methods Director, <http://egap.org>

jbowers@illinois.edu

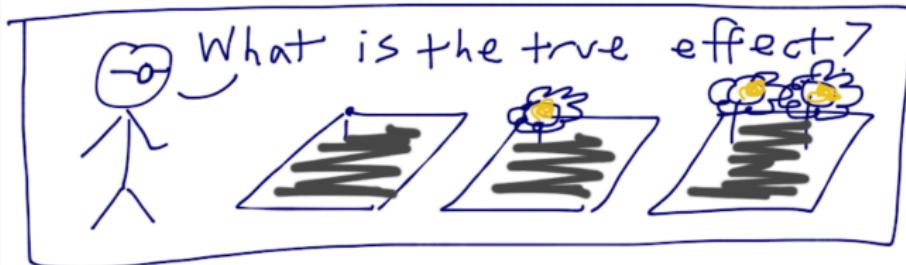
May 28, 2020

An overview of approaches to
statistical inference for causal
quantities

Three General Approaches To Learning About The Unobserved Using Data



Three Approaches To Causal Inference: Potential Outcomes



We don't know.



Imagine we would observe so many bushels of corn, y , if plot i were randomly assigned to new fertilizer, $y_{i,Z_i=1}$ (where $Z_i = 1$ means “assigned to new fertilizer” and $Z_i = 0$ means “assigned status quo fertilizer”) and another amount of corn, $y_{i,Z_i=0}$, if the same plot were assigned the status quo fertilizer condition. These y are are potential or partially observed outcomes.

Three Approaches To Causal Inference: Notation

- Treatment $Z_i = 1$ for treatment and $Z_i = 0$ for control for units i

Three Approaches To Causal Inference: Notation

- Treatment $Z_i = 1$ for treatment and $Z_i = 0$ for control for units i
- In a two arm experiment each unit has at least a pair of potential outcomes $(y_{i,Z_i=1}, y_{i,Z_i=0})$ (also written $(y_{i,1}, y_{i,0})$ to indicate that $y_{1,Z_1=1,Z_2=1} = y_{1,Z_1=1,Z_2=0}$)

Three Approaches To Causal Inference: Notation

- Treatment $Z_i = 1$ for treatment and $Z_i = 0$ for control for units i
- In a two arm experiment each unit has at least a pair of potential outcomes $(y_{i,Z_i=1}, y_{i,Z_i=0})$ (also written $(y_{i,1}, y_{i,0})$ to indicate that $y_{1,Z_1=1,Z_2=1} = y_{1,Z_1=1,Z_2=0}$)
- Causal Effect for unit i is τ_i , $\tau_i = f(y_{i,1}, y_{i,0})$. For example, $\tau_i = y_{i,1} - y_{i,0}$.

Three Approaches To Causal Inference: Notation

- Treatment $Z_i = 1$ for treatment and $Z_i = 0$ for control for units i
- In a two arm experiment each unit has at least a pair of potential outcomes $(y_{i,Z_i=1}, y_{i,Z_i=0})$ (also written $(y_{i,1}, y_{i,0})$) to indicate that $y_{1,Z_1=1,Z_2=1} = y_{1,Z_1=1,Z_2=0}$)
- Causal Effect for unit i is τ_i , $\tau_i = f(y_{i,1}, y_{i,0})$. For example, $\tau_i = y_{i,1} - y_{i,0}$.
- Fundamental Problem of (Counterfactual) Causality We only see one potential outcome $Y_i = Z_i * y_{i,1} + (1 - Z_i)y_{i,0}$ manifest in our observed outcome, Y_i . Treatment reveals one potential outcome to us in a simple randomized experiment.

Design Based Approach 1: Compare Models of Potential Outcomes to Data

1. Make a guess about (or model of) $\tau_i = f(y_{i,1}, y_{i,0})$. For example $H_0 : y_{i,1} = y_{i,0} + \tau_i$ and $\tau_i = 0$ is the sharp null hypothesis of no effects.

I don't know the truth,
but I can assess specific
claims about the truth.



C	Z_i	y_i	$y_{i,1}$	$y_{i,0}$
A	0	16	?	16
B	1	22	?	?
C	0	7	?	?
D	1	14	14	?

Design Based Approach 1: Compare Models of Potential Outcomes to Data

1. Make a guess about (or model of) $\tau_i = f(y_{i,1}, y_{i,0})$. For example $H_0 : y_{i,1} = y_{i,0} + \tau_i$ and $\tau_i = 0$ is the sharp null hypothesis of no effects.
2. Measure consistency of the data with this model given the research design and choice of test statistic (summarizing the treatment-to-outcome relationship).

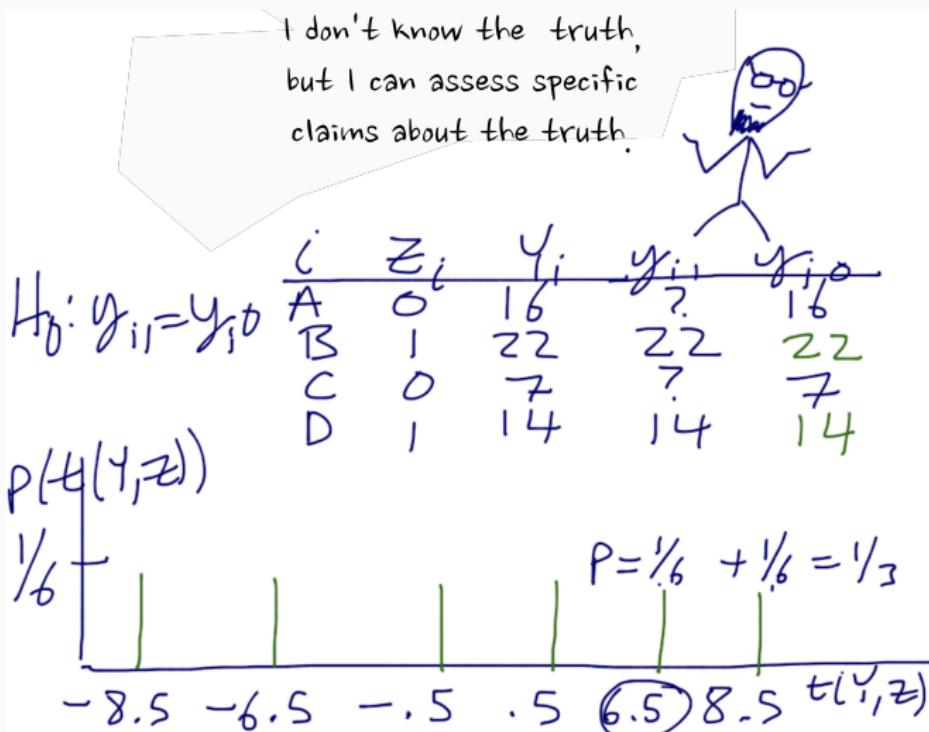
I don't know the truth,
but I can assess specific
claims about the truth.



C	Z_i	y_i	$y_{i,1}$	$y_{i,0}$
A	0	16	?	16
B	1	22	??	?
C	0	7	?	7
D	1	14	14	?

Design Based Approach 1: Compare Models of Potential Outcomes to Data

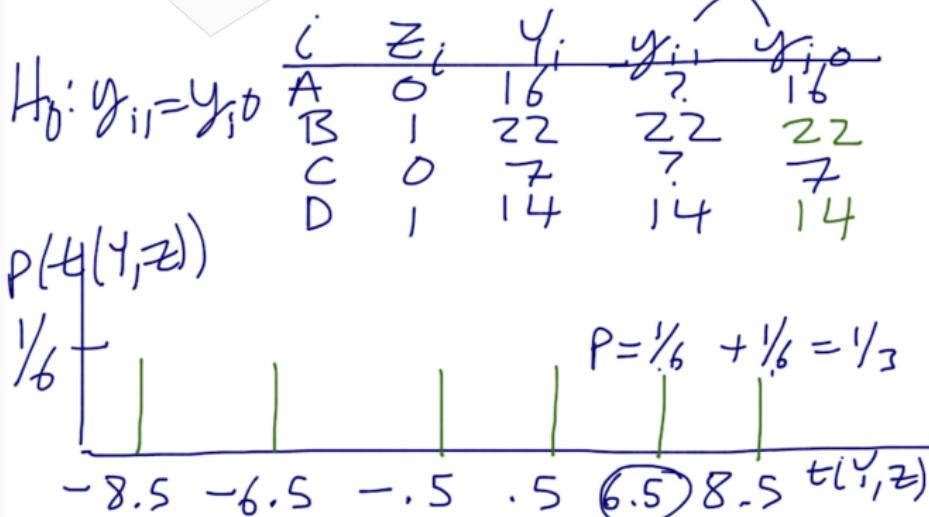
1. Make a guess (or model of) about τ_i .



Design Based Approach 1: Compare Models of Potential Outcomes to Data

1. Make a guess (or model of) about τ_i .
2. Measure consistency of data with this model given the design and test statistic.

I don't know the truth,
but I can assess specific
claims about the truth.



Design Based Approach 1: Compare Models of Potential Outcomes to Data

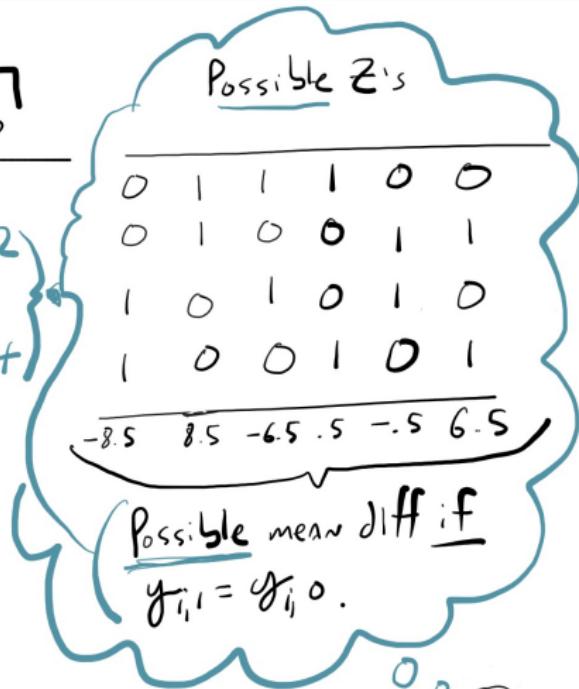
i	fully observed		part observed	
	Z_i	Y_i	$y_{i,z_i=1}$	$y_{i,z_i=0}$
units	A	0	16	?
	B	1	22	22
	C	0	7	?
	D	1	14	14
		6.5		

mean diff

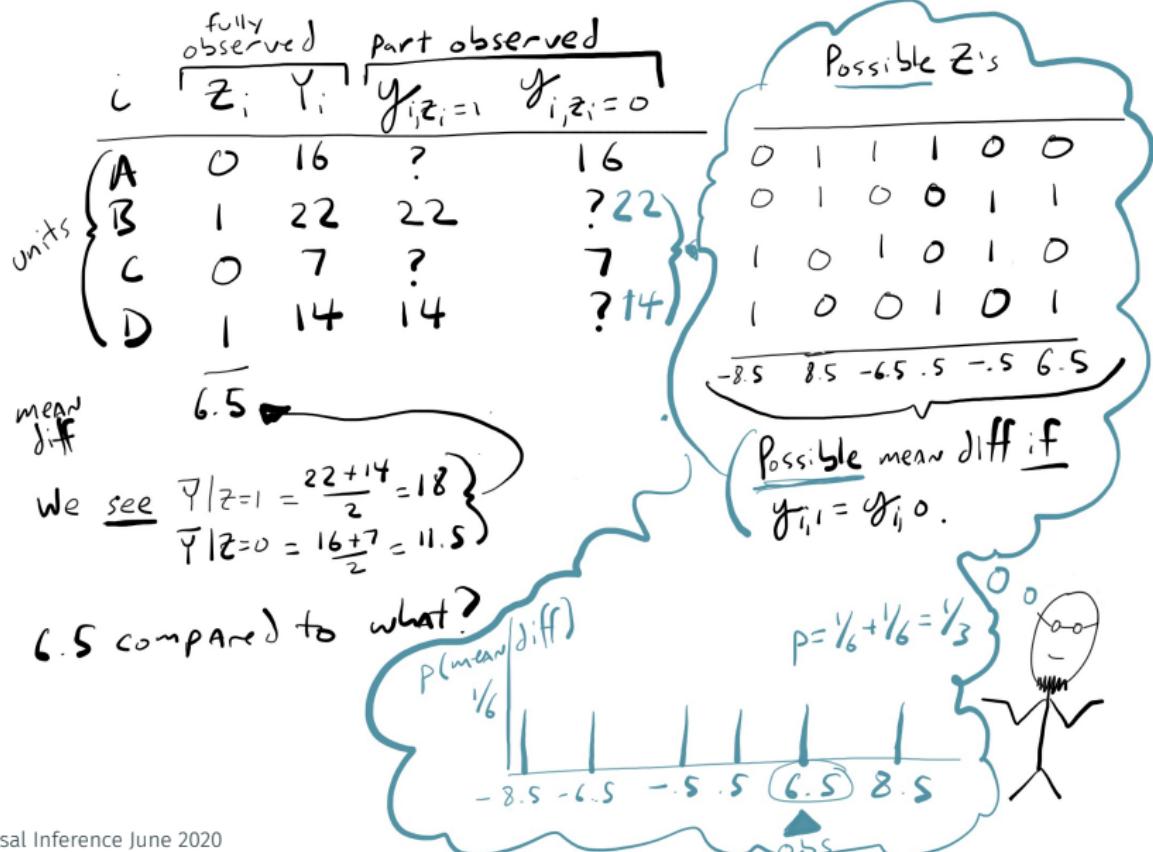
We see $\bar{Y}|z=1 = \frac{22+14}{2} = 18$

$$\bar{Y}|z=0 = \frac{16+7}{2} = 11.5$$

6.5 compared to what?



Design Based Approach 1: Compare Models of Potential Outcomes to Data



Design Based Approach 1: Compare Models of Potential Outcomes to Data

Testing Models of No-Effects.

Here is some fake data from a tiny experiment with weird outcomes.

```
Z   y0    y1      Y zF rY
1 0    16    16    16  0  2
2 1    22    24    24  1  3
3 0     7    10     7  0  1
4 1 3990 4000 4000  1  4
## A mean difference test statistic
tz_mean_diff <- function(z, y) {
  mean(y[z == 1]) - mean(y[z == 0])
}
## A mean difference of ranks test statistic
tz_mean_rank_diff <- function(z, y) {
  ry <- rank(y)
  mean(ry[z == 1]) - mean(ry[z == 0])
}
## Function to repeat the experimental randomization
newexp <- function(z) {
  sample(z)
```

Design Based Approach 1: Compare Models of Potential Outcomes to Data

Testing Models of No-Effects.

```
rand_dist_md <- with(smdat, replicate(1000, tz_mean_diff(z = newexp(Z), y = Y)))
rand_dist_rank_md <- with(smdat, replicate(1000, tz_mean_rank_diff(z = newexp(Z), y = Y)))
obs_md <- with(smdat, tz_mean_diff(z = Z, y = Y))
obs_rank_md <- with(smdat, tz_mean_rank_diff(z = Z, y = Y))
c(observed_mean_diff = obs_md, observed_mean_rank_diff = obs_rank_md)

  observed_mean_diff observed_mean_rank_diff
              2000                      2
table(rand_dist_md) / 1000 ## Probability Distributions Under the Null of No Effects

rand_dist_md
-2000.5 -1992.5 -1983.5  1983.5  1992.5  2000.5
  0.188   0.163   0.161   0.172   0.155   0.161
table(rand_dist_rank_md) / 1000

rand_dist_rank_md
 -2     -1      0      1      2
0.172  0.197  0.318  0.161  0.152
p_md <- mean(rand_dist_md >= obs_md) ## P-Values
p_rank_md <- mean(rand_dist_rank_md >= obs_rank_md)
c(mean_diff_p = p_md, mean_rank_diff_p = p_rank_md)

  mean_diff_p mean_rank_diff_p
              0.161                 0.152
```

Design Based Approach 1: Compare Models of Potential Outcomes to Data

Testing Models of Effects.

To learn about whether the data are consistent with $\tau_i = 100$ for all i notice how treatment assignment reveals part of the unobserved outcomes:

$$Y_i = Z_i * y_{i,1} + (1 - Z_i) * y_{i,0}$$

and if $H_0 : \tau_i = 100$ or $H_0 : y_{i,1} = y_{i,0} + 100$ then:

$$Y_i = Z_i(y_{i,0} + 100) + (1 - Z_i)y_{i,0} \quad (1)$$

$$= Z_i y_{i,0} + Z_i 100 + y_{i,0} - Z_i y_{i,0} \quad (2)$$

$$= Z_i 100 + y_{i,0} \quad (3)$$

$$y_{i,0} = Y_i - Z_i 100 \quad (4)$$

Design Based Approach 1: Compare Models of Potential Outcomes to Data

Testing Models of Effects.

To test a model of causal effects we adjust the observed outcomes to be consistent with our hypothesis about unobserved outcomes and then repeat the experiment:

```
tz_mean_diff_effects <- function(z, y, tauvec) {  
    adjy <- y - z * tauvec  
    radjy <- rank(adjy)  
    mean(radjy[z == 1]) - mean(radjy[z == 0])  
}  
rand_dist_md_tau_cae <- with(smdat, replicate(1000, tz_mean_diff_effects(z = newexp(Z), y = Y, tauvec = c(100, 100, 100, 100))))  
obs_md_tau_cae <- with(smdat, tz_mean_diff_effects(z = Z, y = Y, tauvec = c(100, 100, 100, 100)))  
mean(rand_dist_md_tau_cae >= obs_md_tau_cae)  
[1] 0.505
```

Design Based Approach 1: Compare Models of Potential Outcomes to Data

Testing Models of Effects.

Now let's test $H_0 : \tau = \{0, 2, 3, 10\}$

```
rand_dist_md_taux <- with(smdat, replicate(1000, tz_mean_diff_effects(
  z = newexp(Z), y = Y,
  tauvec = c(0, 2, 3, 10)
)))
obs_md_taux <- with(smdat, tz_mean_diff_effects(z = Z, y = Y, tauvec = c(0, 2, 3, 10)))
mean(rand_dist_md_taux >= obs_md_taux)
[1] 0.178
```

Design Based Approach 2: Estimate Averages of Potential Outcomes

1. Notice that the observed Y_i are a sample from the (small, finite) population of unobserved potential outcomes $(y_{i,1}, y_{i,0})$.



I don't know the truth, but I can provide a good guess of the average causal effect.

i	Z_i	Y_i	$y_{i,1}$	$y_{i,0}$
A	0	16	?	16
B	1	22	22	?
C	0	7	?	7
D	1	14	14	?

$$\widehat{ATE} = \bar{Y}_i | Z_i=1 - \bar{Y}_i | Z_i=0 = \frac{\bar{Y}_{i,1}}{\bar{Y}_{i,0}}$$

Design Based Approach 2: Estimate Averages of Potential Outcomes

1. Notice that the observed Y_i are a sample from the (small, finite) population of unobserved potential outcomes ($y_{i,1}, y_{i,0}$).
2. Decide to focus on the average, $\bar{\tau}$, because sample averages, $\hat{\tau}$ are unbiased and consistent estimators of population averages.



I don't know the truth, but I can provide a good guess of the average causal effect.

i	Z_i	Y_i	$y_{i,1}$	$y_{i,0}$
A	0	16	?	16
B	1	22	22	?
C	0	7	?	7
D	1	14	14	?
			$\bar{Y}_{i,1}$	$\bar{Y}_{i,0}$

$\widehat{ATE} = \bar{Y}_i | Z_i=1 - \bar{Y}_i | Z_i=0$

Design Based Approach 2: Estimate Averages of Potential Outcomes

1. Notice that the observed Y_i are a sample from the (small, finite) population of unobserved potential outcomes ($y_{i,1}, y_{i,0}$).
2. Decide to focus on the average, $\bar{\tau}$, because sample averages, $\hat{\tau}$ are unbiased and consistent estimators of population averages.
3. Estimate $\bar{\tau}$ with the observed difference in means as $\hat{\tau}$.



I don't know the truth, but I can provide a good guess of the average causal effect.

i	Z_i	Y_i	$y_{i,1}$	$y_{i,0}$
A	0	16	?	16
B	1	22	22	?
C	0	7	?	7
D	1	14	14	?
			$\bar{y}_{i,1}$	$\bar{y}_{i,0}$

$$\widehat{ATE} = \bar{Y}_i | Z_i=1 - \bar{Y}_i | Z_i=0$$

Design Based Approach 2: Estimate Averages of Potential Outcomes



I don't know the truth, but I can provide a good guess of the average causal effect.

i	z_i	y_i	y_{i1}	y_{i0}
A	0	16	?	16
B	1	22	22	?
C	0	7	?	7
D	1	14	14	?

$$\begin{aligned}\hat{ATE} &= \bar{Y}_i | z_i = 1 - \bar{Y}_i | z_i = 0 \\ &= \frac{22+14}{2} - \frac{16+7}{2} = 6.5\end{aligned}$$

Design Based Approach 2: Estimate Averages of Potential Outcomes

Here using Neyman's standard errors (same as HC2 SEs) and Central Limit Theorem based *p*-values and 95% confidence intervals:

```
est1 <- difference_in_means(Y ~ Z, data = smdat)
est1

Design: Standard
   Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF
Z     2000      1988    1.006    0.498    -23259    27260    1
```

Model Based Approach 1: Predict Distributions of Potential Outcomes

I dew nut knew thee truth,
 but, given prors, I cane
 predikte itf
 probabeeleetee.



i	Z_i	Y_i	y_{i1}	y_{i0}	Z_i	\hat{Z}_i
A	0	16	16	16	0	0
B	1	22	22	22	1	1
C	0	7	7	7	0	0
D	1	14	14	14	1	1

$$\text{prob}(\hat{Z}_i | Y_i, Z_i, y_{i1}, y_{i0}) \propto \text{Prob}(Y_i | Z_i, \hat{Z}_i, y_{i1}, y_{i0}) \cdot \text{prob}(\hat{Z}_i) \dots$$

$$Y_i \sim N(Z_i \hat{Z}_i, \sigma_i)$$

$$Z_i \sim \text{Bernoulli}(\pi_i)$$

$$\hat{Z}_i \sim \text{Beta}(a, b) \dots$$

Model Based Approach 1: Predict Distributions of Potential Outcomes

adapted from https://mc-stan.org/users/documentation/case-studies/model-based_causal_inference_for_RCT.html

- Given a model of Y_i :

$$\Pr(Y_i^{obs} | Z, \theta) \sim \text{Normal}(Z_i \cdot \mu_1 + (1 - Z_i) \cdot \mu_0, Z_i \sigma_1^2 + (1 - Z_i) \cdot \sigma_0^2) \quad (5)$$

where $\mu_0 = \alpha$ and $\mu_1 = \alpha + \tau$.

- And a model of the pair $\{y_{i,0}, y_{i,1}\} \equiv \{Y_i(0), Y_i(1)\}$ but random not fixed as before (and so written as upper-case):

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \mid \theta \sim \text{Normal} \left(\begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho \sigma_0 \sigma_1 \\ \rho \sigma_0 \sigma_1 & \sigma_1^2 \end{pmatrix} \right) \quad (6)$$

- And a model of Z_i is known because of randomization so we can write:

$$\Pr(Z | Y(0), Y(1)) = \Pr(Z)$$

- And given priors on $\theta = \{\alpha, \tau, \sigma_c, \sigma_t\}$ (here make them all independent $\text{Normal}(0, 5)$).

We can generate the posterior distribution of α, τ, σ_c , and σ_t and thus can impute $\{Y_i(0), Y_i(1)\}$ to generate a distribution for T_i .

Model Based Approach 1: Predict Distributions of Potential Outcomes

```
## rho is correlation between the potential outcomes
stan_data <- list(N = 4, y = smdat$Y, w = smdat$Z, rho = 0)
# Compile and run the stan model
fit_simdat <- stan(file = "rctbayes.stan", data = stan_data, iter = 5000, warmup = 2500, chains = 4, cont
res <- as.matrix(fit_simdat)

## Summary of the 2000 Predicted Treatment effects for units 1 and 4
t(apply(res[, c("tau_unit[1]", "tau_unit[4]")], 2, summary))

parameters      Min. 1st Qu.   Median     Mean 3rd Qu.    Max.
tau_unit[1] -501   -100.1   -3.863   -3.229   93.43   535.7
tau_unit[4] 3945   3986.7  3990.675 3991.025 3994.99 4030.3
## Probability that effect on unit 1 is greater than 0
mean(res[, "tau_unit[1]"] > 0)

[1] 0.4885
## Overall mean of the effects:
mean_tau <- rowMeans(res[, c("tau_unit[1]", "tau_unit[2]", "tau_unit[3]", "tau_unit[4]")])
summary(mean_tau)

Min. 1st Qu.   Median     Mean 3rd Qu.    Max.
825     968    1002     1002    1036    1197
```

Summary: Modes of Statistical Inference for Causal Effects

We can infer about unobserved counterfactuals by:

1. assessing claims or models or hypotheses about relationships between unobserved potential outcomes (Fisher's testing approach via Rosenbaum)

Summary: Modes of Statistical Inference for Causal Effects

We can infer about unobserved counterfactuals by:

1. assessing claims or models or hypotheses about relationships between unobserved potential outcomes (Fisher's testing approach via Rosenbaum)
2. estimating averages (or other summaries) of unobserved potential outcomes (Neyman's estimation approach)

Summary: Modes of Statistical Inference for Causal Effects

We can infer about unobserved counterfactuals by:

1. assessing claims or models or hypotheses about relationships between unobserved potential outcomes (Fisher's testing approach via Rosenbaum)
2. estimating averages (or other summaries) of unobserved potential outcomes (Neyman's estimation approach)
3. predicting individual level outcomes based on probability models of outcomes, interventions, etc. (Bayes's predictive approach via Rubin)

Summary: Modes of Statistical Inference for Causal Effects

Statistical inferences — formalized reasoning about “what if” statements (“What if I had randomly assigned other plots to treatment?”) — and their properties (ex.bias, error rates, precision) arise from:

1. Repeating the design and using the hypothesis and test statistics to generate a reference distribution that describes the variation in the hypothetical world. Compare the observed to the hypothesized to measure consistency between hypothesis, or model, and observed outcomes (Fisher and Rosenbaum's randomization-based inference for individual causal effects).

Summary: Modes of Statistical Inference for Causal Effects

Statistical inferences — formalized reasoning about “what if” statements (“What if I had randomly assigned other plots to treatment?”) — and their properties (ex.bias, error rates, precision) arise from:

1. Repeating the design and using the hypothesis and test statistics to generate a reference distribution that describes the variation in the hypothetical world. Compare the observed to the hypothesized to measure consistency between hypothesis, or model, and observed outcomes (Fisher and Rosenbaum's randomization-based inference for individual causal effects).
2. Repeating the design and the estimation such that standard errors, p -values, and confidence intervals reflect design-based variability. Probability distributions (like the Normal or t-distribution) arise from Limit Theorems in large samples. (Neyman's randomization-based inference for average causal effects).

Summary: Modes of Statistical Inference for Causal Effects

Statistical inferences — formalized reasoning about “what if” statements (“What if I had randomly assigned other plots to treatment?”) — and their properties (ex.bias, error rates, precision) arise from:

1. Repeating the design and using the hypothesis and test statistics to generate a reference distribution that describes the variation in the hypothetical world. Compare the observed to the hypothesized to measure consistency between hypothesis, or model, and observed outcomes (Fisher and Rosenbaum's randomization-based inference for individual causal effects).
2. Repeating the design and the estimation such that standard errors, p -values, and confidence intervals reflect design-based variability. Probability distributions (like the Normal or t-distribution) arise from Limit Theorems in large samples. (Neyman's randomization-based inference for average causal effects).
3. Repeatedly drawing from the probability distributions that generate the observed data (that represent the design) — the likelihood and the priors — to describe a posterior distribution for unit-level causal effects. Calculate posterior distributions for aggregated causal effects (like averages of individual level effects). (Bayes and Rubin's predictive model-based causal inference).

Summary: Applications of the Model-Based Prediction Approach

Examples of use of the model-based prediction approach:

- Estimating causal effects when we need to model processes of missing outcomes, missing treatment indicators, or complex non-compliance with treatment Barnard et al. 2003

Summary: Applications of the Model-Based Prediction Approach

Examples of use of the model-based prediction approach:

- Estimating causal effects when we need to model processes of missing outcomes, missing treatment indicators, or complex non-compliance with treatment Barnard et al. 2003
- Searching for heterogeneity (subgroup differences) in how units react to treatment (ex. Hahn, Murray, and Carvalho 2020 but see also literature on BART, Bayesian Machine Learning as applied to causal inference questions).

Summary: Applications of the Testing Approach

Examples of use of the testing approach:

- Assessing evidence of pareto optimal effects or no aberrant effect (i.e. no unit was made worse off by the treatment) (Caughey, Dafoe, and Miratrix 2016; P. Rosenbaum and Silber 2008).

Summary: Applications of the Testing Approach

Examples of use of the testing approach:

- Assessing evidence of pareto optimal effects or no aberrant effect (i.e. no unit was made worse off by the treatment) (Caughey, Dafoe, and Miratrix 2016; P. Rosenbaum and Silber 2008).
- Assessing evidence that the treatment group was made better than the control group (but being agnostic about the precise nature of the difference) (ex. $p > .2$ with difference of means but $p < .001$ with difference of ranks in Office of Evaluation Sciences study of General Services Administration Auctions)

Summary: Applications of the Testing Approach

Examples of use of the testing approach:

- Assessing evidence of pareto optimal effects or no aberrant effect (i.e. no unit was made worse off by the treatment) (Caughey, Dafoe, and Miratrix 2016; P. Rosenbaum and Silber 2008).
- Assessing evidence that the treatment group was made better than the control group (but being agnostic about the precise nature of the difference) (ex. $p > .2$ with difference of means but $p < .001$ with difference of ranks in Office of Evaluation Sciences study of General Services Administration Auctions)
- Focusing on detection rather than on estimation (for example to identify promising sites for future research in experiments with many blocks or strata) (Bowers and Chen 2020 working paper).

Summary: Applications of the Testing Approach

Examples of use of the testing approach:

- Assessing evidence of pareto optimal effects or no aberrant effect (i.e. no unit was made worse off by the treatment) (Caughey, Dafoe, and Miratrix 2016; P. Rosenbaum and Silber 2008).
- Assessing evidence that the treatment group was made better than the control group (but being agnostic about the precise nature of the difference) (ex. $p > .2$ with difference of means but $p < .001$ with difference of ranks in Office of Evaluation Sciences study of General Services Administration Auctions)
- Focusing on detection rather than on estimation (for example to identify promising sites for future research in experiments with many blocks or strata) (Bowers and Chen 2020 working paper).
- Assessing hypotheses of no effects in small samples, with rare outcomes, cluster randomization, or other designs where reference distributions may not be Normal (see ex, (Gerber and Green 2012) or).

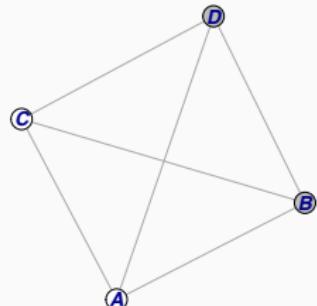
Summary: Applications of the Testing Approach

Examples of use of the testing approach:

- Assessing evidence of pareto optimal effects or no aberrant effect (i.e. no unit was made worse off by the treatment) (Caughey, Dafoe, and Miratrix 2016; P. Rosenbaum and Silber 2008).
- Assessing evidence that the treatment group was made better than the control group (but being agnostic about the precise nature of the difference) (ex. $p > .2$ with difference of means but $p < .001$ with difference of ranks in Office of Evaluation Sciences study of General Services Administration Auctions)
- Focusing on detection rather than on estimation (for example to identify promising sites for future research in experiments with many blocks or strata) (Bowers and Chen 2020 working paper).
- Assessing hypotheses of no effects in small samples, with rare outcomes, cluster randomization, or other designs where reference distributions may not be Normal (see ex, (Gerber and Green 2012) or).
- Assessing structural models of causal effects (for example models of treatment effect propagation across networks) (Bowers, Desmarais, et al. 2018; Bowers, M. Fredrickson, and Aronow 2016; Bowers, M. M. Fredrickson, and Panagopoulos 2013).

Statistical Inference about Causal Models on Networks

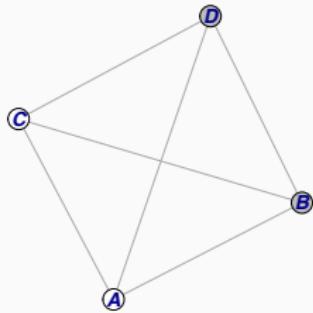
Statistical inference with interference?



i	Z_i	Y_i	$y_{i,1100}$	$y_{i,0101}$	$y_{i,1001}$	$y_{i,0110}$	$y_{i,1010}$	$y_{i,0011}$
A	0	16	?	16	?	?	?	?
B	1	22	?	22	?	?	?	?
C	0	7	?	7	?	?	?	?
D	1	14	?	14	?	?	?	?

On estimation see (Sobel, Aronow, Eckles, Samii, Hudgens, Ogburn, VanderWeele, Toulis, Kao, Coppock, Sicar, Raudenbush, Hong, ...). The key question for that work: What is the function of potential outcomes that we can estimate using observed data?

Statistical inference with interference?



i	Z_i	Y_i	$y_{i,1100}$	$y_{i,0101}$	$y_{i,1001}$	$y_{i,0110}$	$y_{i,1010}$	$y_{i,0011}$	$y_{i,0000} \equiv y_{i,0}$
A	0	16	?	16	?	?	?	?	16
B	1	22	?	22	?	?	?	?	22
C	0	7	?	7	?	?	?	?	7
D	1	14	?	14	?	?	?	?	14

The sharp null of no effects is a model of no interference:

$$H_0 : y_{i,1100} = y_{i,0101} = y_{i,1001} = y_{i,0110} = y_{i,1010} = y_{i,0011} = y_{i,0000},$$

$$y_{i,0} = \mathcal{H}(y_{i,z}, \mathbf{0}) = y_{i,z}, p = 0.33.$$

Introducing the uniformity trial $\equiv y_{i,0000}$ (P. R. Rosenbaum, Ross, and Silber 2007).

Imagine an experiment on a network:

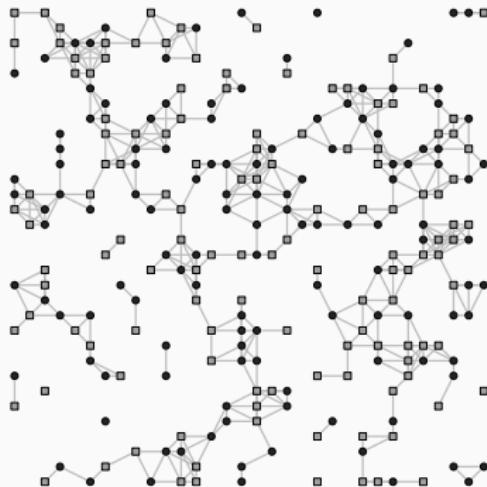


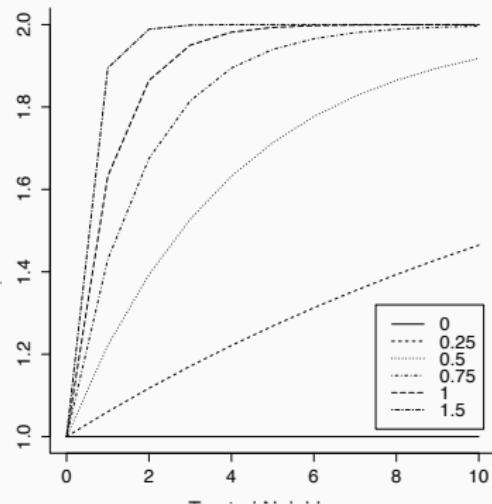
Figure: A simulated data set with 256 units and 512 connections. The $256/2 = 128$ treated units ($Z_i = 1$) are shown as filled circles and an equal number of control units ($Z_i = 0$) are shown as gray squares.

Imagine an experiment on a network: With a model of propagation

- The direct effect of treatment is β (it is a multiplicative effect).

$$\mathcal{H}(\mathbf{y}_0, \mathbf{z}, \beta, \tau) = \left[\beta + (1 - z_i)(1 - \beta) \exp(-\tau^2 \mathbf{z}^T \mathbf{S}) \right] \mathbf{y}_0 \quad (1)$$

$$\mathcal{H}(\mathbf{y}_z, \mathbf{0}, \beta, \tau) = \left[\beta + (1 - z_i)(1 - \beta) \exp(-\tau^2 \mathbf{z}^T \mathbf{S}) \right]^{-1} \mathbf{y}_z \equiv \mathbf{y}_0 \quad (2)$$

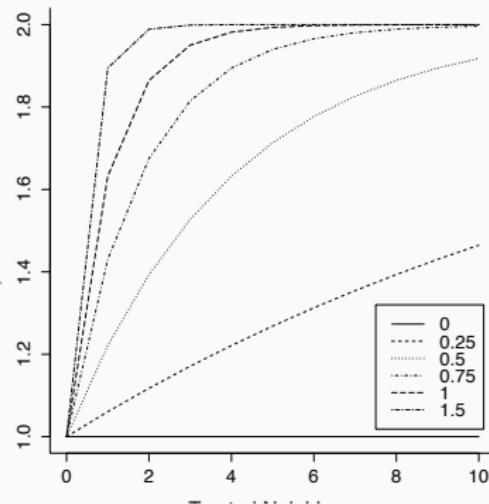


Imagine an experiment on a network: With a model of propagation

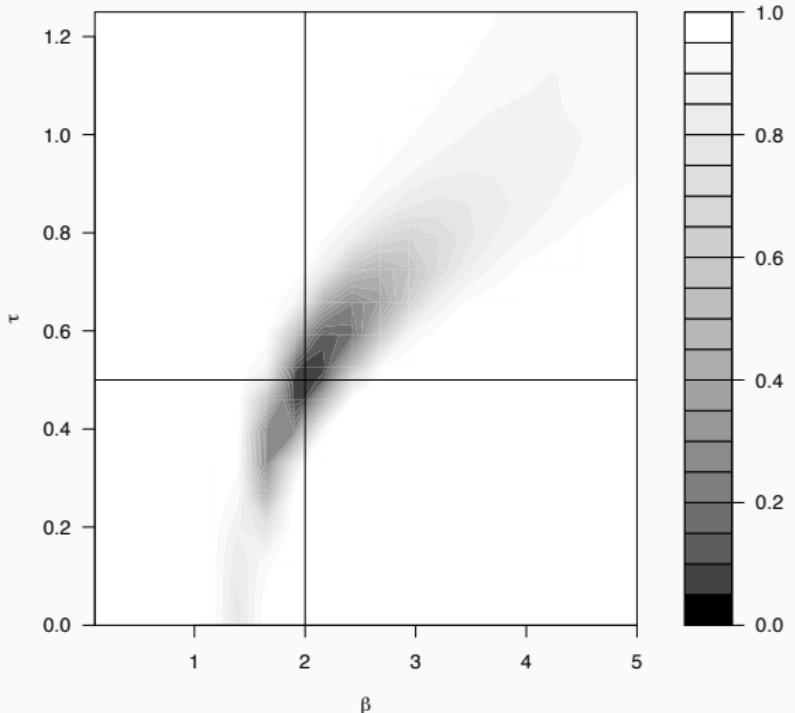
- The direct effect of treatment is β (it is a multiplicative effect).
- Treatment effects flows from treated to control units, increasing with number of treated neighbors, with rate of growth of effect τ .

$$\mathcal{H}(\mathbf{y}_0, \mathbf{z}, \beta, \tau) = [\beta + (1 - z_i)(1 - \beta) \exp(-\tau^2 \mathbf{z}^T \mathbf{S})] \mathbf{y}_0 \quad (1)$$

$$\mathcal{H}(\mathbf{y}_z, \mathbf{0}, \beta, \tau) = [\beta + (1 - z_i)(1 - \beta) \exp(-\tau^2 \mathbf{z}^T \mathbf{S})]^{-1} \mathbf{y}_z \equiv \mathbf{y}_0 \quad (2)$$



Learning about a causal model from an experiment on a network



Plot shows the proportion of p -values less than .05 for randomization tests of joint hypotheses about τ and β . Darker values mean less rejection. Truth is at $\tau = .5, \beta = 2$. All tests reject the null hypothesis more than 5% of the time at $\alpha = .05$. All simulations using permutation-based

An Agent-Based Causal Model Example

A causal model relates potential outcomes to each other and a research design relates potential outcomes to observed data (and to sources of uncertainty). For examples:

- For no effects at all: $y_{i,0} = H(Y_i, Z_i, \tau_i) = Y_i$.

In fact any function that produces vectors of \mathbf{y}_0 could be used to represent these kinds of causal models.

So: why not an agent-based model?

An Agent-Based Causal Model Example

A causal model relates potential outcomes to each other and a research design relates potential outcomes to observed data (and to sources of uncertainty). For examples:

- For no effects at all: $y_{i,0} = H(Y_i, Z_i, \tau_i) = Y_i$.
- For constant, additive effects: $y_{i,0} = H(Y_i, Z_i, \tau_i) = Y_i - Z_i\tau$.

In fact any function that produces vectors of \mathbf{y}_0 could be used to represent these kinds of causal models.

So: why not an agent-based model?

An Agent-Based Causal Model Example

A causal model relates potential outcomes to each other and a research design relates potential outcomes to observed data (and to sources of uncertainty). For examples:

- For no effects at all: $y_{i,0} = H(Y_i, Z_i, \tau_i) = Y_i$.
- For constant, additive effects: $y_{i,0} = H(Y_i, Z_i, \tau_i) = Y_i - Z_i\tau$.
- For vector valued outcomes in a network with nonlinear propagation of causal effects:

$$\mathbf{y}_0 = H(\mathbf{y}_z, \mathbf{0}, \beta, \tau) = (\beta + (1 - z_i)(1 - \beta) \exp(-\tau^2 \mathbf{z}^T \mathbf{S}))^{-1} \mathbf{y}_z \quad (7)$$

In fact any function that produces vectors of \mathbf{y}_0 could be used to represent these kinds of causal models.

So: why not an agent-based model?

An agent-based model of electoral fraud in Ghana

- Party agents are in charge of registering voters (honestly and dishonestly). They mobilize potential voters (for example, in buses). They get paid for fraud (in part).
- Party agents want to register as many people using as few resources as possible (and with as little risk as possible). They know that many voters in Ghana (where the political parties are strongly associated with particular ethnicities):
 - Prefer to have a co-ethnic in office who is more likely to favor them than a non-co-ethnic politician
 - Believe that co-ethnic leaders matter for local public goods
 - Anticipate a close election, citizens may not report registration fraud
- So, agents may target ethnically-homogeneous areas where it's less likely they'll be reported
- Alternatively: potential reporting by ordinary citizens may not be a concern, and distances/resources may be a more important factor.

Formal decision-theoretic models

- k The total number of agents.
- t The total number of 'ticks' or time periods, in which agents can visit ELAs.
- τ The number of false registrants an agent can add to an unobserved ELA.

Distance-minimizing model In each 'tick,' agents go to the nearest ELA by road distance (start at ELA nearest the most others); if they encounter an observer, immediately move to the nearest ELA from there. Cannot revisit ELAs. Implies starting at ELA that are close to others.

Ethnic homogeneity-seeking model In each 'tick,' agents only consider moving to an ELA with $F \leq \alpha$ where F is ethnic fractionalization and α is a percentile of F within a constituency. Among these, move to the closest ELA by road distance, and move again if encounter observer. Implies starting at ELA with lowest F .

Uniformity: model without observers

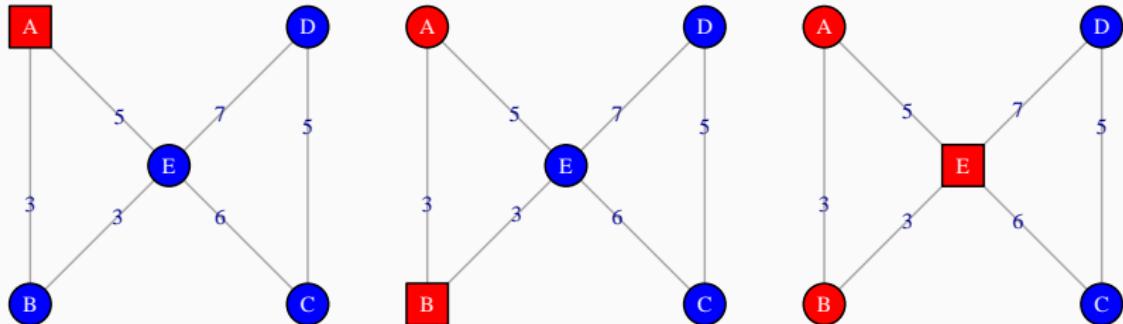


Figure 1: Agent movement rules when no observers are encountered. Squares indicate the agent's current location. Red ELAs are visited. Blue ELAs are not yet visited. From left to right: 1) $t = 0$, the agent starts at A, 2) agent selects B as closest ELA, 3) Agent moves to E in final period.

Experiment: model with observers

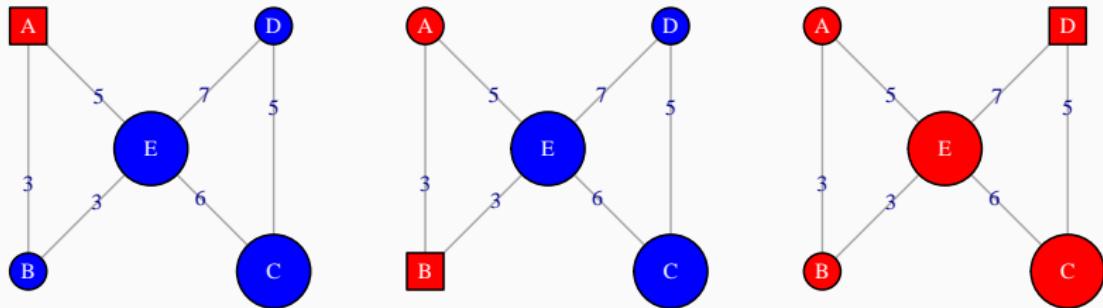


Figure 2: Agent movement rules when observers are present. Squares indicate the agent's current location. Red ELAs are visited. Blue ELAs are not yet visited. The large circles indicate observer ELAs. From left to right: 1) $t = 0$, the agent starts at **A**, 2) agent selects **B** as closest ELA, 3) Agent moves to **E**, but as an observer is present, immediately moves to **C**, again encounters an observer, and finally stops at **D**.

Assessing competing models of party agents

Which combinations of parameters would have been surprising given the observed data?

- Pick a set of values for the 4 parameters k , τ , t , and α to determine a path for our agents through the road network. Each set of parameters generates one sharp hypothesis as an output of the agent-based model.

Assessing competing models of party agents

Which combinations of parameters would have been surprising given the observed data?

- Pick a set of values for the 4 parameters k , τ , t , and α to determine a path for our agents through the road network. Each set of parameters generates one sharp hypothesis as an output of the agent-based model.
- p -value records the information our data and design provide against these hypotheses given the test statistic (here the Kolmogorov-Smirnov test statistic).

Assessing competing models of party agents

Which combinations of parameters would have been surprising given the observed data?

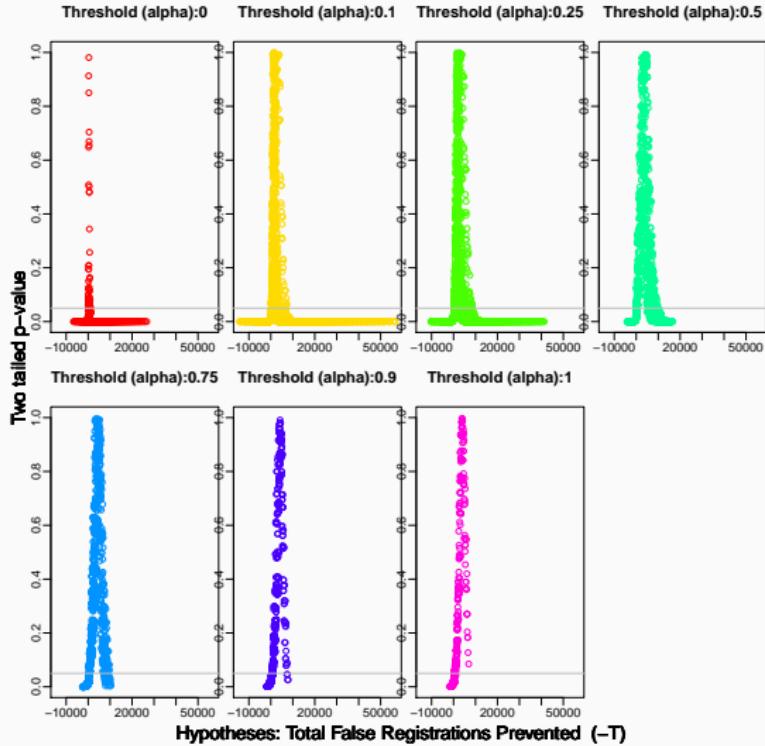
- Pick a set of values for the 4 parameters k , τ , t , and α to determine a path for our agents through the road network. Each set of parameters generates one sharp hypothesis as an output of the agent-based model.
- p -value records the information our data and design provide against these hypotheses given the test statistic (here the Kolmogorov-Smirnov test statistic).
- Q: How to interpret a 4d confidence region?

Assessing competing models of party agents

Which combinations of parameters would have been surprising given the observed data?

- Pick a set of values for the 4 parameters k , τ , t , and α to determine a path for our agents through the road network. Each set of parameters generates one sharp hypothesis as an output of the agent-based model.
- p -value records the information our data and design provide against these hypotheses given the test statistic (here the Kolmogorov-Smirnov test statistic).
- Q: How to interpret a 4d confidence region?
- Our approach for now: Focus on a composite quantity, 'total fraud', $T = \sum_{i=1}^n (Y_i - y_{i0})$, where y_{i0} is the number of registrations implied by the model under the uniformity trial (inspired by Rosenbaum's attributable effects, $A = \sum_i Z_i \tau_i$). If the minimum p -value for all hypotheses that make up a given T is greater than .05, then this T is in the confidence set.

Testing the models



Notes: $\alpha = 0$ (only the most ethnically homogeneous ELAs available for a visit), $\alpha = .1$ (only 10% of ELAs available to agent), ..., $\alpha = .9$ (nearly all ELAs available to agent).

A General Testing-Based Causal Inference Algorithm

1. Write a model converting uniformity trial potential outcomes (like \mathbf{y}_0 or simply $y_{i,0}$) into observed data (like \mathbf{y}_z or simply Y_i): What is the mechanism by which the treatment changes the outcomes of the units? (This is a structural model of potential outcomes. It could be an agent-based model.)

A General Testing-Based Causal Inference Algorithm

1. Write a model converting uniformity trial potential outcomes (like \mathbf{y}_0 or simply $y_{i,0}$) into observed data (like \mathbf{y}_z or simply Y_i): What is the mechanism by which the treatment changes the outcomes of the units? (This is a structural model of potential outcomes. It could be an agent-based model.)
2. Solve for \mathbf{y}_0 . What adjustment of the observed data does this model imply?
 $H(\mathbf{y}_z, \mathbf{0}, \theta_0) = \mathbf{y}_0$ like $y_{i,0} = Y_i + Z\tau$ for the simple constant, additive effects model.

A General Testing-Based Causal Inference Algorithm

1. Write a model converting uniformity trial potential outcomes (like \mathbf{y}_0 or simply $y_{i,0}$) into observed data (like \mathbf{y}_z or simply Y_i): What is the mechanism by which the treatment changes the outcomes of the units? (This is a structural model of potential outcomes. It could be an agent-based model.)
2. Solve for \mathbf{y}_0 . What adjustment of the observed data does this model imply?
 $H(\mathbf{y}_z, \mathbf{0}, \theta_0) = \mathbf{y}_0$ like $y_{i,0} = Y_i + Z\tau$ for the simple constant, additive effects model.
3. Select a test statistic that is effect-increasing in all relevant dimensions (like the sum of squared residuals test statistic or the KS-test statistic for certain models, or, I conjecture, an energy-statistic).

A General Testing-Based Causal Inference Algorithm

1. Write a model converting uniformity trial potential outcomes (like \mathbf{y}_0 or simply $y_{i,0}$) into observed data (like \mathbf{y}_z or simply Y_i): What is the mechanism by which the treatment changes the outcomes of the units? (This is a structural model of potential outcomes. It could be an agent-based model.)
2. Solve for \mathbf{y}_0 . What adjustment of the observed data does this model imply?
 $H(\mathbf{y}_z, \mathbf{0}, \theta_0) = \mathbf{y}_0$ like $y_{i,0} = Y_i + Z\tau$ for the simple constant, additive effects model.
3. Select a test statistic that is effect-increasing in all relevant dimensions (like the sum of squared residuals test statistic or the KS-test statistic for certain models, or, I conjecture, an energy-statistic).
4. Compute p -values for substantively meaningful range of θ . Or calculate boundaries of regions. (Perhaps collapse or aggregate the rejection regions to aid interpretation.)

Conclusion

Conclusion

- Assumptions of “no interference” are not inherently necessary for statistical inference about counterfactual causal quantities. The sharp null hypothesis of no effects is also a causal model of no interference. (And now we have average causal effects defined on networks too — the average effect of having one treated neighbor, etc..)

Conclusion

- Assumptions of “no interference” are not inherently necessary for statistical inference about counterfactual causal quantities. The sharp null hypothesis of no effects is also a causal model of no interference. (And now we have average causal effects defined on networks too — the average effect of having one treated neighbor, etc..)
- Counterfactual causal inference focuses on comparisons of (and functions of) partially observed outcomes (“potential outcomes”). Averages of those outcomes are often an intuitive and useful estimand and also fairly easy to estimate with data. Averages are not the only way to learn about what we do not observe from what we do observe.

Conclusion

- Assumptions of “no interference” are not inherently necessary for statistical inference about counterfactual causal quantities. The sharp null hypothesis of no effects is also a causal model of no interference. (And now we have average causal effects defined on networks too — the average effect of having one treated neighbor, etc..)
- Counterfactual causal inference focuses on comparisons of (and functions of) partially observed outcomes (“potential outcomes”). Averages of those outcomes are often an intuitive and useful estimand and also fairly easy to estimate with data. Averages are not the only way to learn about what we do not observe from what we do observe.
- Models of effects can specify flexible theoretical models of propagation over networks (including algorithmic models). Structural causal models are possible to specify and test.

Conclusion

- Assumptions of “no interference” are not inherently necessary for statistical inference about counterfactual causal quantities. The sharp null hypothesis of no effects is also a causal model of no interference. (And now we have average causal effects defined on networks too — the average effect of having one treated neighbor, etc..)
- Counterfactual causal inference focuses on comparisons of (and functions of) partially observed outcomes (“potential outcomes”). Averages of those outcomes are often an intuitive and useful estimand and also fairly easy to estimate with data. Averages are not the only way to learn about what we do not observe from what we do observe.
- Models of effects can specify flexible theoretical models of propagation over networks (including algorithmic models). Structural causal models are possible to specify and test.
- In this talk, focusing on randomized-experiments, randomization justified both statistical inference (p -values, confidence intervals) and causal inference. But the specification of causal models would be the same whether or not the research design is randomized. Statistical inference (testing and estimation) requires more work to justify and assess.

References

References i

- Aronow, Peter M and Cyrus Samii (2013). "Estimating average causal effects under interference between units". In: arXiv preprint arXiv:1305.6156.
- Baird, Sarah et al. (2014). "Designing experiments to measure spillover effects". Unpublished manuscript.
- Barnard, J. et al. (2003). "Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City". In: Journal of the American Statistical Association 98.462, pp. 299–324.
- Bowers, Jake, Bruce A Desmarais, et al. (2018). "Models, methods and network topology: Experimental design for the study of interference". In: Social Networks 54, pp. 196–208.
- Bowers, Jake, Mark Fredrickson, and Peter M Aronow (2016). "Research Note: A more powerful test statistic for reasoning about interference between units". In: Political Analysis 24.3, pp. 395–403.
- Bowers, Jake, Mark M. Fredrickson, and Costas Panagopoulos (2013). "Reasoning about Interference Between Units: A General Framework". In: Political Analysis 21.1, pp. 97–124.
- Caughey, Devin, Allan Dafoe, and Luke Miratrix (July 2016). "Beyond the Sharp Null: Permutation Tests Actually Test Heterogeneous Effects".

References ii

- Gerber, Alan S and Donald P Green (2012). Field experiments: Design, analysis, and interpretation. WW Norton.
- Hahn, P Richard, Jared S Murray, and Carlos M Carvalho (2020). "Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects". In: Bayesian Analysis.
- Liu, Lan and Michael G Hudgens (2014). "Large sample randomization inference of causal effects in the presence of interference". In: Journal of the american statistical association 109.505, pp. 288–301.
- Rosenbaum, Paul and Jeffrey H Silber (2008). "Aberrant effects of treatment". In: Journal of the American Statistical Association 103.481, pp. 240–247.
- Rosenbaum, Paul R, Richard N Ross, and Jeffrey H Silber (2007). Minimum Distance Matched Sampling With Fine Balance in an Observational Study of Treatment
- Sinclair, Betsy, Margaret McConnell, and Donald P Green (2012). "Detecting spillover effects: Design and analysis of multilevel experiments". In: American Journal of Political Science 56.4, pp. 1055–1069.

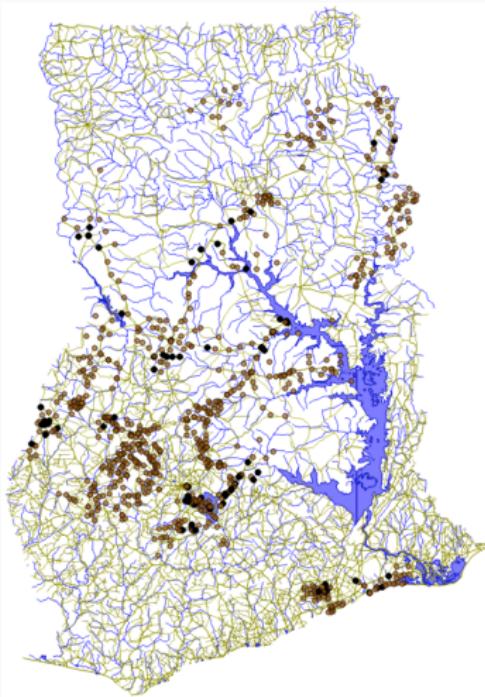
References iii

Toulis, Panos and Edward Kao (2013). "Estimation of causal peer influence effects". In: International Conference on Machine Learning, pp. 1489–1497.

Appendix

A Simple Design-based Estimation Approach

Voter Registration in Ghana 2008



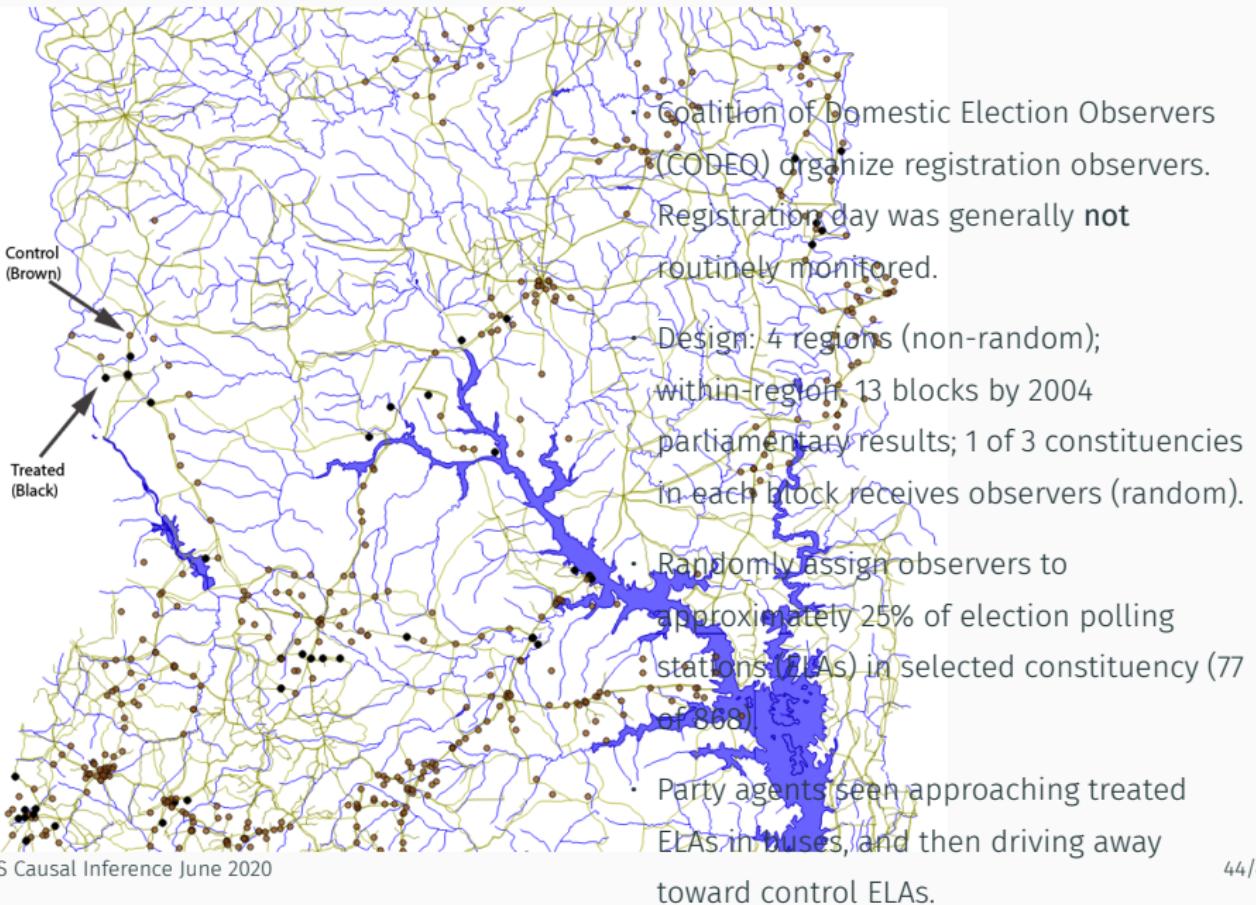
- Presidential and parliamentary elections in December 2008.
- 13 day voter registration exercise in August 2008.
- Estimated 800,000 people newly eligible to vote, but **2 million** new voters registered.
- Term-limited president, election expected to be very close. Decided by less than 50,000 votes out of more than 9 million votes cast.

Voter Registration in Ghana 2008



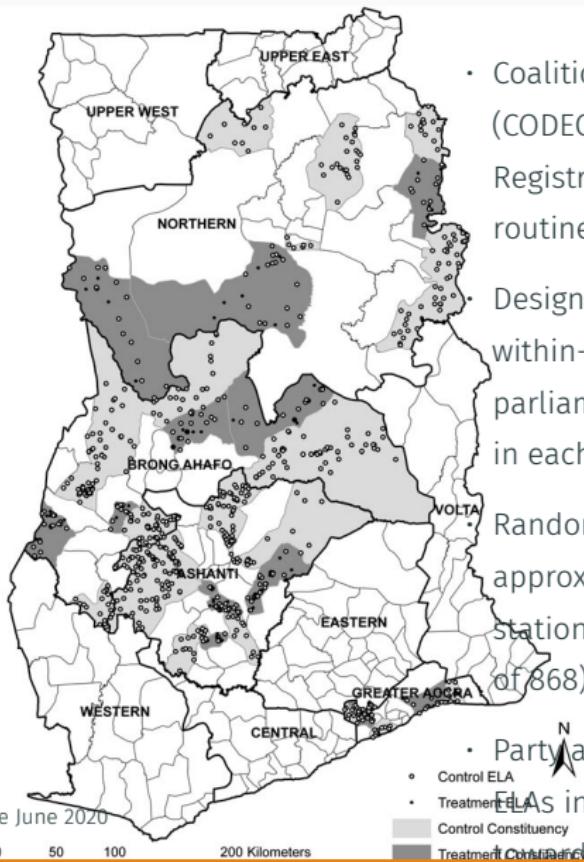
- Coalition of Domestic Election Observers (CODEO) organize registration observers. Registration day was generally **not** routinely monitored.
- Design: 4 regions (non-random); within-region, 13 blocks by 2004 parliamentary results; 1 of 3 constituencies in each block receives observers (random).
- Randomly assign observers to approximately 25% of election polling stations (ELAs) in selected constituency (77 of 868).
- Party agents seen approaching treated ELAs in buses, and then driving away toward control ELAs.

Voter Registration in Ghana 2008



Voter Registration in Ghana 2008

FIGURE 1 Ghana, with Treatment and Control Constituencies and Electoral Areas



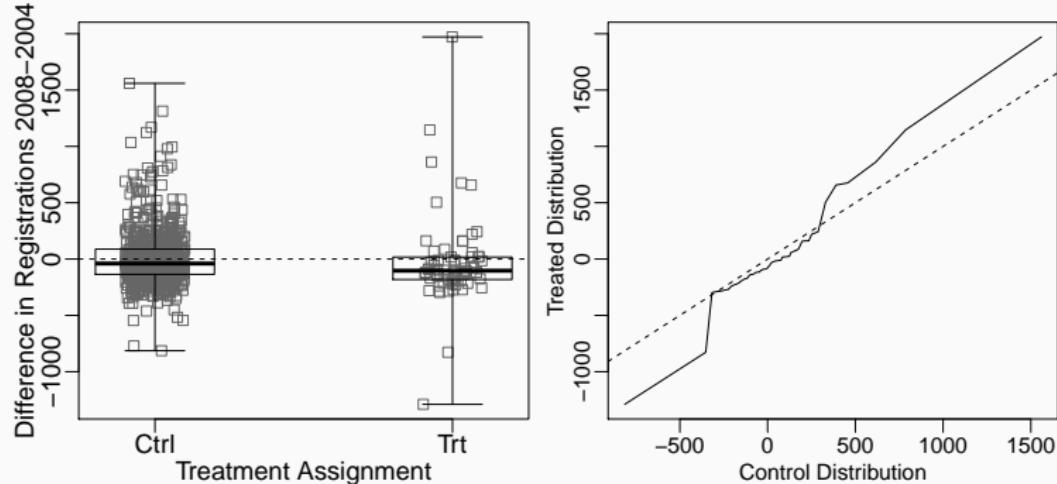
- Coalition of Domestic Election Observers (CODEO) organize registration observers. Registration day was generally **not** routinely monitored.

- Design: 4 regions (non-random); within-region, 13 blocks by 2004 parliamentary results; 1 of 3 constituencies in each block receives observers (random).

Randomly assign observers to approximately 25% of election polling stations (ELAs) in selected constituency (77 of 868).

- Party agents seen approaching treated ELAs in buses, and then driving away from control ELAs

Assessing the sharp null hypothesis of no effects.



Q: What is the probability of seeing as large an observed difference between the treated and control groups, if the observers had no effect at all — recalling that no effect means no interference as well as no other effect?

A: $p = 0.018$ (using a mean-difference test-statistic).

Approaches for going beyond the sharp-null of no effects

Estimation:

- Use design to isolate units

Testing:

Approaches for going beyond the sharp-null of no effects

Estimation:

- Use design to isolate units
- Or weight average differences by model of propagation / spillover (Aronow and Samii 2013; Toulis and Kao 2013)

Testing:

Approaches for going beyond the sharp-null of no effects

Estimation:

- Use design to isolate units
- Or weight average differences by model of propagation / spillover (Aronow and Samii 2013; Toulis and Kao 2013)

Testing:

- Assess implications of models of network-propagation effects (Bowers, Desmarais, et al. 2018; Bowers, M. Fredrickson, and Aronow 2016; Bowers, M. M. Fredrickson, and Panagopoulos 2013)

Approaches for going beyond the sharp-null of no effects

Estimation:

- Use design to isolate units
- Or weight average differences by model of propagation / spillover (Aronow and Samii 2013; Toulis and Kao 2013)

Testing:

- Assess implications of models of network-propagation effects (Bowers, Desmarais, et al. 2018; Bowers, M. Fredrickson, and Aronow 2016; Bowers, M. M. Fredrickson, and Panagopoulos 2013)
- Invert hypothesis tests comparing levels/ranks of treatment outcomes to the uniformity trial (P. R. Rosenbaum, Ross, and Silber 2007).

Estimation restricting interference by design

Imagine that $Z_i \in \{U, C, T\}$ where T is treatment (election observers), C is control with possible spillover and U is “uniformity trial” or control with no possible spillover.

Thus, if you have isolated units and randomization (such that all units have positive probability of $Z_i \in \{U, C, T\}$) we have $y_{i,T}$, $y_{i,C}$, and $y_{i,U}$ for each unit.¹

¹The two-level design (Sinclair, McConnell, and Green 2012). See also Gerber and Green (2012, Chap 8) or generalized saturation design (Baird et al. 2014). Liu and Hudgens (2014) for some nice theory.

Estimation restricting interference by design

Imagine that $Z_i \in \{U, C, T\}$ where T is treatment (election observers), C is control with possible spillover and U is “uniformity trial” or control with no possible spillover.

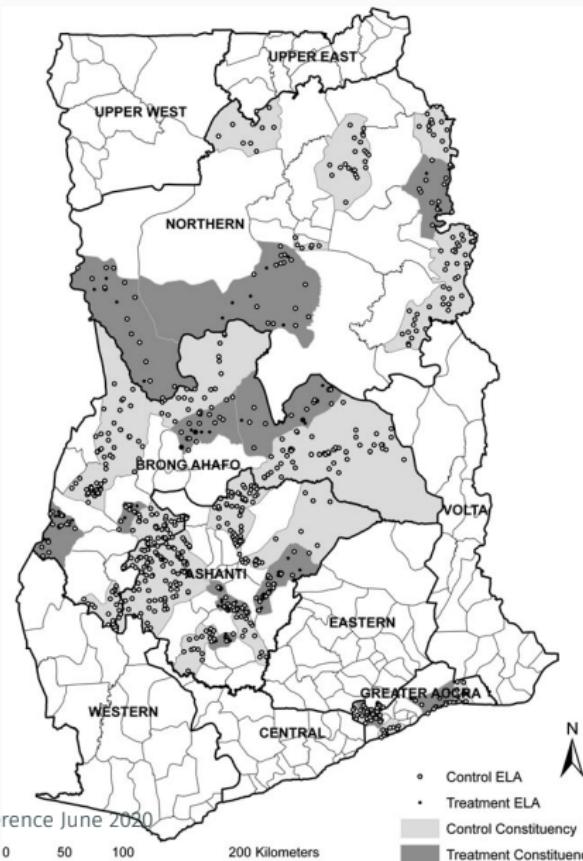
Thus, if you have isolated units and randomization (such that all units have positive probability of $Z_i \in \{U, C, T\}$) we have $y_{i,T}$, $y_{i,C}$, and $y_{i,U}$ for each unit.¹

And you can define and estimate $\bar{\tau}_{\text{spillover}} = \bar{y}_C - \bar{y}_U$ or $\bar{\tau}_{\text{Direct Effect}} = \bar{y}_T - \bar{y}_U$ etc..

¹The two-level design (Sinclair, McConnell, and Green 2012). See also Gerber and Green (2012, Chap 8) or generalized saturation design (Baird et al. 2014). Liu and Hudgens (2014) for some nice theory.

Estimation restricting interference by design

FIGURE 1 Ghana, with Treatment and Control Constituencies and Electoral Areas



Estimation restricting interference by design

How would we use this data to estimate direct, indirect, or spillover effects?

```
table(Z = ELAs.df$tela, TrtRegion = ELAs.df$NSF_Const_registration_Treat)

TrtRegion
Z      0   1
0 556 172
1    0  68

tmpdat <- group_by(ELAs.df, block, ZLv2 = tela, ZLv1 = NSF_Const_registration_Treat) %>%
  summarise(
    barYb = round(mean(reg2008ELA - reg2004ELA), 5),
    nb = n(), nTb = sum(tela),
    barY08 = mean(reg2008ELA),
    barY04 = mean(reg2004ELA)
  )
tmpdat

# A tibble: 34 x 8
# Groups:   block, ZLv2 [22]
  block ZLv2 ZLv1 barYb    nb    nTb barY08 barY04
  <int> <int> <int> <dbl> <int> <int> <dbl> <dbl>
1     1     0     0  356.    72     0  1326.  970.
2     1     0     1 2814     4     0 17305. 14491.
3     1     1     1 1226     2     2 12401. 11175
4     2     0     0  242.    77     0  1428.  1186.
5     2     0     1 282.     9     0  1435.  1152.
6     3     1     1 1368     5     5 1697.  1329.
7     3     0     0  290.    52     0  1702. 1412.
```