Experiments 2: Clustering, blocking, noncompliance

LQRPS

Frederik Hjorth
fh@ifs.ku.dk
fghjorth.github.io
@fghjorth

Department of Political Science
University of Copenhagen

February 9th, 2017

1. Joe's paper
2. Clustering
3. Using covariates
   - Pre-treatment outcomes
   - Other covariates
   - Blocking
4. Noncompliance
   - Motivating ex.: Gerber & Green (2000)
   - Formal statement
5. Is voting contagious? Nickerson (2008)

Table 1 – Choice sets and treatment conditions in selective exposure experiment

Control		**T1**		**T2**	
Pro attitudinal 1 - *written*	Counter attitudinal 1 – *written*	**Pro attitudinal 1 - *viral***	Counter attitudinal 1 – *written*	Pro attitudinal 1 - *written*	**Counter attitudinal 1 – *viral***
Pro attitudinal 1 - *written*	Counter attitudinal 2 – *written*	Pro attitudinal 1 - *written*	Counter attitudinal 2 – *written*	Pro attitudinal 1 - *written*	Counter attitudinal 2 – *written*
Expert opinion - *written*	Fact on the issue - *written*	Expert opinion - *written*	Fact on the issue - *written*	Expert opinion - *written*	Fact on the issue - *written*

N.B. pro attitudinal means that the subject has a (somewhat) preference toward the party the statement comes from. It can thus be from a party from blue or red bloc depending on the subject's political affiliation.

Questions or comments?

The standard error for $\widehat{ATE}$:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{N-1}\{\frac{mVar(Y_{i0})}{N-m} + \frac{(N-m)Var(Y_{i1})}{m} + 2Cov(Y_{i0}, Y_{i1})\}} \quad (1)$$

Ways to reduce $SE(\widehat{ATE})$:

- $N \uparrow$
- $Var(Y_{i0})$ or $Var(Y_{i1}) \downarrow$
- $Cov(Y_{i0} Y_{i1}) \downarrow$

The standard error for $\widehat{ATE}$:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{N-1}\{\frac{mVar(Y_{i0})}{N-m} + \frac{(N-m)Var(Y_{i1})}{m} + 2Cov(Y_{i0}, Y_{i1})\}} \quad (1)$$

Ways to reduce $SE(\widehat{ATE})$:

- $N \uparrow$
- $Var(Y_{i0})$ or $Var(Y_{i1}) \downarrow$
- $Cov(Y_{i0} Y_{i1}) \downarrow$

The standard error for $\widehat{ATE}$:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{N-1}\{\frac{mVar(Y_{i0})}{N-m} + \frac{(N-m)Var(Y_{i1})}{m} + 2Cov(Y_{i0}, Y_{i1})\}} \quad (1)$$

Ways to reduce $SE(\widehat{ATE})$:

- $N \uparrow$
- $Var(Y_{i0})$ or $Var(Y_{i1}) \downarrow$
- $Cov(Y_{i0} Y_{i1}) \downarrow$

The standard error for $\widehat{ATE}$:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{N-1}\{\frac{mVar(Y_{i0})}{N-m} + \frac{(N-m)Var(Y_{i1})}{m} + 2Cov(Y_{i0}, Y_{i1})\}} \quad (1)$$

Ways to reduce $SE(\widehat{ATE})$:

- $N \uparrow$
- $Var(Y_{i0})$ or $Var(Y_{i1}) \downarrow$
- $Cov(Y_{i0} Y_{i1}) \downarrow$

The standard error for $\widehat{ATE}$:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{N-1}\{\frac{mVar(Y_{i0})}{N-m} + \frac{(N-m)Var(Y_{i1})}{m} + 2Cov(Y_{i0}, Y_{i1})\}} \quad (1)$$

Ways to reduce $SE(\widehat{ATE})$:

- $N \uparrow$
- $Var(Y_{i0})$ or $Var(Y_{i1}) \downarrow$
- $Cov(Y_{i0} Y_{i1}) \downarrow$

The standard error for $\widehat{ATE}$:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{N-1}\{\frac{mVar(Y_{i0})}{N-m} + \frac{(N-m)Var(Y_{i1})}{m} + 2Cov(Y_{i0}, Y_{i1})\}} \quad (1)$$

Ways to reduce $SE(\widehat{ATE})$:

- $N \uparrow$
- $Var(Y_{i0})$ or $Var(Y_{i1}) \downarrow$
- $Cov(Y_{i0} Y_{i1}) \downarrow$

In some cases we measure *outcome* at the individual level, but *assignment* occurs at the cluster level

- e.g. media markets, municipalities, classrooms
- → other examples?
- basic implication: does not produce bias, but reduces precision

In some cases we measure *outcome* at the individual level, but *assignment* occurs at the cluster level

- e.g. media markets, municipalities, classrooms
- $\rightarrow$ other examples?
- basic implication: does not produce bias, but reduces precision

In some cases we measure *outcome* at the individual level, but *assignment* occurs at the cluster level

- e.g. media markets, municipalities, classrooms
- $\rightarrow$ other examples?
- basic implication: does not produce bias, but reduces precision

In some cases we measure *outcome* at the individual level, but *assignment* occurs at the cluster level

- e.g. media markets, municipalities, classrooms
- $\rightarrow$ other examples?
- basic implication: does not produce bias, but reduces precision

$SE(\widehat{ATE})$ with k clusters:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\{\frac{mVar(\overline{Y}_{j0})}{N-m} + \frac{(N-m)Var(\overline{Y}_{j1})}{m} + 2Cov(\overline{Y}_{j0}, \overline{Y}_{j1})\}} \quad (2)$$

Ways to reduce $SE(\widehat{ATE})$:

- $k \uparrow$
- $N \uparrow$
- $Var(Y_{i0})$ or $Var(Y_{i1}) \downarrow$
- $Cov(Y_{i0} Y_{i1}) \downarrow$

$SE(\widehat{ATE})$ with k clusters:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\{\frac{mVar(\overline{Y}_{j0})}{N-m} + \frac{(N-m)Var(\overline{Y}_{j1})}{m} + 2Cov(\overline{Y}_{j0}, \overline{Y}_{j1})\}} \quad (2)$$

Ways to reduce $SE(\widehat{ATE})$:

- $k \uparrow$
- $N \uparrow$
- $Var(Y_{i0})$ or $Var(Y_{i1}) \downarrow$
- $Cov(Y_{i0} Y_{i1}) \downarrow$

$SE(\widehat{ATE})$ with k clusters:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\left\{\frac{mVar(\overline{Y}_{j0})}{N-m} + \frac{(N-m)Var(\overline{Y}_{j1})}{m} + 2Cov(\overline{Y}_{j0}, \overline{Y}_{j1})\right\}} \quad (2)$$

Ways to reduce $SE(\widehat{ATE})$:

- $k \uparrow$
- $N \uparrow$
- $Var(Y_{i0})$ or $Var(Y_{i1}) \downarrow$
- $Cov(Y_{i0} Y_{i1}) \downarrow$

$SE(\widehat{ATE})$ with k clusters:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\{\frac{mVar(\overline{Y}_{j0})}{N-m} + \frac{(N-m)Var(\overline{Y}_{j1})}{m} + 2Cov(\overline{Y}_{j0}, \overline{Y}_{j1})\}} \quad (2)$$

Ways to reduce $SE(\widehat{ATE})$:

- $k \uparrow$
- $N \uparrow$
- $Var(Y_{i0})$ or $Var(Y_{i1}) \downarrow$
- $Cov(Y_{i0} Y_{i1}) \downarrow$

$SE(\widehat{ATE})$ with k clusters:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\{\frac{mVar(\overline{Y}_{j0})}{N-m} + \frac{(N-m)Var(\overline{Y}_{j1})}{m} + 2Cov(\overline{Y}_{j0}, \overline{Y}_{j1})\}} \quad (2)$$

Ways to reduce $SE(\widehat{ATE})$:

- $k \uparrow$
- $N \uparrow$
- $Var(Y_{i0})$ or $Var(Y_{i1}) \downarrow$
- $Cov(Y_{i0} Y_{i1}) \downarrow$

$SE(\widehat{ATE})$ with k clusters:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\{\frac{mVar(\overline{Y}_{j0})}{N-m} + \frac{(N-m)Var(\overline{Y}_{j1})}{m} + 2Cov(\overline{Y}_{j0}, \overline{Y}_{j1})\}} \quad (2)$$

Ways to reduce $SE(\widehat{ATE})$:

- $k \uparrow$
- $N \uparrow$
- $Var(Y_{i0})$ or $Var(Y_{i1}) \downarrow$
- $Cov(Y_{i0} Y_{i1}) \downarrow$

$SE(\widehat{ATE})$ with k clusters:

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\{\frac{mVar(\overline{Y}_{j0})}{N-m} + \frac{(N-m)Var(\overline{Y}_{j1})}{m} + 2Cov(\overline{Y}_{j0}, \overline{Y}_{j1})\}} \quad (2)$$

Ways to reduce $SE(\widehat{ATE})$:

- $k \uparrow$
- $N \uparrow$
- $Var(Y_{i0})$ or $Var(Y_{i1}) \downarrow$
- $Cov(Y_{i0} Y_{i1}) \downarrow$

We can calculate p-values (with or without cluster assignment) using *randomization inference*

1. simulate all (or large number of) possible configurations of treatment assignments
2. for each simulated configuration, estimate ATE
3. calculate p-value based on the estimated ATE's position in the distribution of simulated ATE's

Useful package for randomization inference: `ri`

We can calculate p-values (with or without cluster assignment) using *randomization inference*

1. simulate all (or large number of) possible configurations of treatment assignments
2. for each simulated configuration, estimate ATE
3. calculate p-value based on the estimated ATE's position in the distribution of simulated ATE's

Useful package for randomization inference: `ri`

We can calculate p-values (with or without cluster assignment) using *randomization inference*

1. simulate all (or large number of) possible configurations of treatment assignments
2. for each simulated configuration, estimate ATE
3. calculate p-value based on the estimated ATE's position in the distribution of simulated ATE's

Useful package for randomization inference: `ri`

We can calculate p-values (with or without cluster assignment) using *randomization inference*

1. simulate all (or large number of) possible configurations of treatment assignments
2. for each simulated configuration, estimate ATE
3. calculate p-value based on the estimated ATE's position in the distribution of simulated ATE's

Useful package for randomization inference: `ri`

We can calculate p-values (with or without cluster assignment) using *randomization inference*

1. simulate all (or large number of) possible configurations of treatment assignments
2. for each simulated configuration, estimate ATE
3. calculate p-value based on the estimated ATE's position in the distribution of simulated ATE's

Useful package for randomization inference: `ri`

We can calculate p-values (with or without cluster assignment) using *randomization inference*

1. simulate all (or large number of) possible configurations of treatment assignments
2. for each simulated configuration, estimate ATE
3. calculate p-value based on the estimated ATE's position in the distribution of simulated ATE's

Useful package for randomization inference: `ri`

Questions or comments?

1 Joe's paper

2 Clustering

3 Using covariates
- Pre-treatment outcomes
- Other covariates
- Blocking

4 Noncompliance

5 Is voting contagious? Nickerson (2008)

1 Joe's paper

2 Clustering

3 Using covariates
- Pre-treatment outcomes
- Other covariates
- Blocking

4 Noncompliance

5 Is voting contagious? Nickerson (2008)

Special type of covariate: pre-treatment observations of the outcome

- allows for measuring outcome as *change* in variable of interest
- i.e. in lieu of difference-in-means, *difference-in-differences* estimator (cf. yesterday)
- when pre-treatment covariates are correlated with potential outcomes → increase in precision

Special type of covariate: pre-treatment observations of the outcome

- allows for measuring outcome as *change* in variable of interest
- i.e. in lieu of difference-in-means, *difference-in-differences* estimator (cf. yesterday)
- when pre-treatment covariates are correlated with potential outcomes $\rightarrow$ increase in precision

Special type of covariate: pre-treatment observations of the outcome

- allows for measuring outcome as *change* in variable of interest
- i.e. in lieu of difference-in-means, *difference-in-differences* estimator (cf. yesterday)
- when pre-treatment covariates are correlated with potential outcomes → increase in precision

Special type of covariate: pre-treatment observations of the outcome

- allows for measuring outcome as *change* in variable of interest
- i.e. in lieu of difference-in-means, *difference-in-differences* estimator (cf. yesterday)
- when pre-treatment covariates are correlated with potential outcomes $\rightarrow$ increase in precision

Regression of Y_i on d_i and covariate X_i:

$$Y_i = Y_{i0}(1 - d_i) + Y_{i1}d_i = a + bd_i + cX_i + (u_i - cX_i) \tag{3}$$

$\rightarrow$ if d_i predicts Y_i, residual term $\downarrow \rightarrow \sigma_{\hat{b}} \downarrow$

Regression of Y_i on d_i and covariate X_i:

$$Y_i = Y_{i0}(1 - d_i) + Y_{i1}d_i = a + bd_i + cX_i + (u_i - cX_i) \tag{3}$$

$\rightarrow$ if d_i predicts Y_i, residual term $\downarrow \rightarrow \sigma_{\hat{b}} \downarrow$

Regression of Y_i on d_i and covariate X_i:

$$Y_i = Y_{i0}(1 - d_i) + Y_{i1}d_i = a + bd_i + cX_i + (u_i - cX_i) \tag{3}$$

$\rightarrow$ if d_i predicts Y_i, residual term $\downarrow$ $\rightarrow \sigma_{\hat{b}} \downarrow$

Regression of Y_i on d_i and covariate X_i:

$$Y_i = Y_{i0}(1 - d_i) + Y_{i1}d_i = a + bd_i + cX_i + (u_i - cX_i) \tag{3}$$

$\rightarrow$ if d_i predicts Y_i, residual term $\downarrow \rightarrow \sigma_{\hat{b}} \downarrow$

But: covariate inclusion also increases 'researcher degrees of freedom'

»This type of analysis introduces an element of discretion in terms of what results are reported. Perhaps unconsciously, the researcher may settle on a regression model that makes the estimated ATE look impressive or interesting, a decision rule that jeopardizes the unbiasedness of the estimator.« (105)

Table 1. Explaining support for socially protective policies with physiological reactions to threatening images. Results of ordinary least squares (OLS) regression with support for socially protective policies (possible range from 0 to 18), with higher numbers indicating attitudes more supportive of policies thought to protect the social unit regressed on five explanatory variables: gender (0 = male; 1 = female), age (in years), education (six categories ranging from "did not finish high school" to "college degree plus"), income (six categories ranging from an annual salary of less than $20,000 to an annual salary of more than $100,000), and changes in skin conductance level (SCL) occasioned by the viewing of threatening images. Descriptive statistics on the variables and further discussion of the regression techniques are available in the SOM. *$P < 0.05$, two-tailed *t* test.

Variable	Unstandardized coefficient (SE)	Standardized coefficient
SCL	92.2* (29.03)	0.377
Income	-0.395 (0.471)	-0.10
Education	-1.63* (0.465)	-0.42
Age	0.19 (0.10)	0.235
Gender	-2.34 (1.3)	-0.20
Constant	-353* (193)	
N	46	
Adj. R-square	0.37	

Table 2. Explaining support for socially protective policies with physiological reactions to nonthreatening images. Results of regression (OLS) with support for socially protective policies regressed on five explanatory variables. Variables are the same as those described for Table 1 except that skin conductance (SCL) is the change in skin conductance occasioned by the viewing of nonthreatening images. Descriptive statistics and further discussion of the regression techniques are available in the SOM. *$P < 0.05$, two-tailed *t* test.

Variable	Unstandardized coefficient (SE)	Standardized coefficient
SCL	-1.8 (35.08)	-0.007
Income	-0.438 (0.533)	-0.115
Education	-1.57* (0.53)	-0.408
Age	0.165 (0.11)	0.204
Gender	-2.23 (1.52)	-0.196
Constant	-304* (217)	
N	46	
Adj. R-square	0.21	

But: covariate inclusion also increases 'researcher degrees of freedom'

»This type of analysis introduces an element of discretion in terms of what results are reported. Perhaps unconsciously, the researcher may settle on a regression model that makes the estimated ATE look impressive or interesting, a decision rule that jeopardizes the unbiasedness of the estimator.« (105)

Table 1. Explaining support for socially protective policies with physiological reactions to threatening images. Results of ordinary least squares (OLS) regression with support for socially protective policies (possible range from 0 to 18), with higher numbers indicating attitudes more supportive of policies thought to protect the social unit regressed on five explanatory variables: gender (0 = male; 1 = female), age (in years), education (six categories ranging from "did not finish high school" to "college degree plus"), income (six categories ranging from an annual salary of less than $20,000 to an annual salary of more than $100,000), and changes in skin conductance level (SCL) occasioned by the viewing of threatening images. Descriptive statistics on the variables and further discussion of the regression techniques are available in the SOM. *$P < 0.05$, two-tailed *t* test.

Variable	Unstandardized coefficient (SE)	Standardized coefficient
SCL	92.2* (29.03)	0.377
Income	-0.395 (0.471)	-0.10
Education	-1.63* (0.465)	-0.42
Age	0.19 (0.10)	0.235
Gender	-2.34 (1.3)	-0.20
Constant	-353* (193)	
N	46	
Adj. R-square	0.37	

Table 2. Explaining support for socially protective policies with physiological reactions to nonthreatening images. Results of regression (OLS) with support for socially protective policies regressed on five explanatory variables. Variables are the same as those described for Table 1 except that skin conductance (SCL) is the change in skin conductance occasioned by the viewing of nonthreatening images. Descriptive statistics and further discussion of the regression techniques are available in the SOM. *$P < 0.05$, two-tailed *t* test.

Variable	Unstandardized coefficient (SE)	Standardized coefficient
SCL	-1.8 (35.08)	-0.007
Income	-0.438 (0.533)	-0.115
Education	-1.57* (0.53)	-0.408
Age	0.165 (0.11)	0.204
Gender	-2.23 (1.52)	-0.196
Constant	-304* (217)	
N	46	
Adj. R-square	0.21	

But: covariate inclusion also increases 'researcher degrees of freedom'

»This type of analysis introduces an element of discretion in terms of what results are reported. Perhaps unconsciously, the researcher may settle on a regression model that makes the estimated ATE look impressive or interesting, a decision rule that jeopardizes the unbiasedness of the estimator.« (105)

Table 1. Explaining support for socially protective policies with physiological reactions to threatening images. Results of ordinary least squares (OLS) regression with support for socially protective policies (possible range from 0 to 18), with higher numbers indicating attitudes more supportive of policies thought to protect the social unit regressed on five explanatory variables: gender (0 = male; 1 = female), age (in years), education (six categories ranging from "did not finish high school" to "college degree plus"), income (six categories ranging from an annual salary of less than $20,000 to an annual salary of more than $100,000), and changes in skin conductance level (SCL) occasioned by the viewing of threatening images. Descriptive statistics on the variables and further discussion of the regression techniques are available in the SOM. *P < 0.05, two-tailed *t* test.

Variable	Unstandardized coefficient (SE)	Standardized coefficient
SCL	92.2* (29.03)	0.377
Income	-0.395 (0.471)	-0.10
Education	-1.63* (0.465)	-0.42
Age	0.19 (0.10)	0.235
Gender	-2.34 (1.3)	-0.20
Constant	-353* (193)	
N	46	
Adj. R-square	0.37	

Table 2. Explaining support for socially protective policies with physiological reactions to nonthreatening images. Results of regression (OLS) with support for socially protective policies regressed on five explanatory variables. Variables are the same as those described for Table 1 except that skin conductance (SCL) is the change in skin conductance occasioned by the viewing of nonthreatening images. Descriptive statistics and further discussion of the regression techniques are available in the SOM. *P < 0.05, two-tailed *t* test.

Variable	Unstandardized coefficient (SE)	Standardized coefficient
SCL	-1.8 (35.08)	-0.007
Income	-0.438 (0.533)	-0.115
Education	-1.57* (0.53)	-0.408
Age	0.165 (0.11)	0.204
Gender	-2.23 (1.52)	-0.196
Constant	-304* (217)	
N	46	
Adj. R-square	0.21	

But: covariate inclusion also increases 'researcher degrees of freedom'

»This type of analysis introduces an element of discretion in terms of what results are reported. Perhaps unconsciously, the researcher may settle on a regression model that makes the estimated ATE look impressive or interesting, a decision rule that jeopardizes the unbiasedness of the estimator.« (105)

Table 1. Explaining support for socially protective policies with physiological reactions to threatening images. Results of ordinary least squares (OLS) regression with support for socially protective policies (possible range from 0 to 18), with higher numbers indicating attitudes more supportive of policies thought to protect the social unit regressed on five explanatory variables: gender (0 = male; 1 = female), age (in years), education (six categories ranging from "did not finish high school" to "college degree plus"), income (six categories ranging from an annual salary of less than $20,000 to an annual salary of more than $100,000), and changes in skin conductance level (SCL) occasioned by the viewing of threatening images. Descriptive statistics on the variables and further discussion of the regression techniques are available in the SOM. *P < 0.05, two-tailed *t* test.

Variable	Unstandardized coefficient (SE)	Standardized coefficient
SCL	92.2* (29.03)	0.377
Income	-0.395 (0.471)	-0.10
Education	-1.63* (0.465)	-0.42
Age	0.19 (0.10)	0.235
Gender	-2.34 (1.3)	-0.20
Constant	-353* (193)	
N	46	
Adj. R-square	0.37	

Table 2. Explaining support for socially protective policies with physiological reactions to nonthreatening images. Results of regression (OLS) with support for socially protective policies regressed on five explanatory variables. Variables are the same as those described for Table 1 except that skin conductance (SCL) is the change in skin conductance occasioned by the viewing of nonthreatening images. Descriptive statistics and further discussion of the regression techniques are available in the SOM. *P < 0.05, two-tailed *t* test.

Variable	Unstandardized coefficient (SE)	Standardized coefficient
SCL	-1.8 (35.08)	-0.007
Income	-0.438 (0.533)	-0.115
Education	-1.57* (0.53)	-0.408
Age	0.165 (0.11)	0.204
Gender	-2.23 (1.52)	-0.196
Constant	-304* (217)	
N	46	
Adj. R-square	0.21	

But: covariate inclusion also increases 'researcher degrees of freedom'

»This type of analysis introduces an element of discretion in terms of what results are reported. Perhaps unconsciously, the researcher may settle on a regression model that makes the estimated ATE look impressive or interesting, a decision rule that jeopardizes the unbiasedness of the estimator.« (105)

Table 1. Explaining support for socially protective policies with physiological reactions to threatening images. Results of ordinary least squares (OLS) regression with support for socially protective policies (possible range from 0 to 18), with higher numbers indicating attitudes more supportive of policies thought to protect the social unit regressed on five explanatory variables: gender (0 = male; 1 = female), age (in years), education (six categories ranging from "did not finish high school" to "college degree plus"), income (six categories ranging from an annual salary of less than $20,000 to an annual salary of more than $100,000), and changes in skin conductance level (SCL) occasioned by the viewing of threatening images. Descriptive statistics on the variables and further discussion of the regression techniques are available in the SOM. *P < 0.05, two-tailed *t* test.

Variable	Unstandardized coefficient (SE)	Standardized coefficient
SCL	92.2* (29.03)	0.377
Income	-0.395 (0.471)	-0.10
Education	-1.63* (0.465)	-0.42
Age	0.19 (0.10)	0.235
Gender	-2.34 (1.3)	-0.20
Constant	-353* (193)	
N	46	
Adj. R-square	0.37	

Table 2. Explaining support for socially protective policies with physiological reactions to nonthreatening images. Results of regression (OLS) with support for socially protective policies regressed on five explanatory variables. Variables are the same as those described for Table 1 except that skin conductance (SCL) is the change in skin conductance occasioned by the viewing of nonthreatening images. Descriptive statistics and further discussion of the regression techniques are available in the SOM. *P < 0.05, two-tailed *t* test.

Variable	Unstandardized coefficient (SE)	Standardized coefficient
SCL	-1.8 (35.08)	-0.007
Income	-0.438 (0.533)	-0.115
Education	-1.57* (0.53)	-0.408
Age	0.165 (0.11)	0.204
Gender	-2.23 (1.52)	-0.196
Constant	-304* (217)	
N	46	
Adj. R-square	0.21	

1 Joe's paper

2 Clustering

3 Using covariates
- Pre-treatment outcomes
- Other covariates
- Blocking

4 Noncompliance

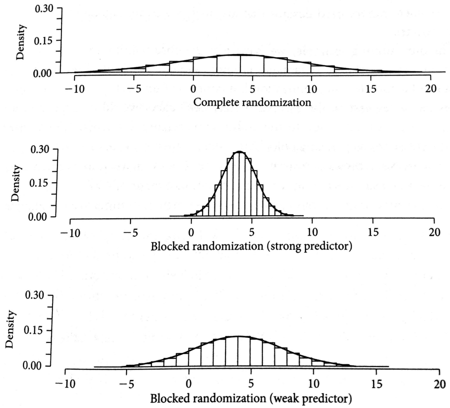
5 Is voting contagious? Nickerson (2008)

Blocking on covariate X helps when:

- N is relatively small
- X is unbalanced across experimental conditions
- X strongly predicts Y

Useful R package for block random assignment: `randomizr`

FIGURE 4.2
Comparison of sampling distributions based on completely randomized and block randomized designs

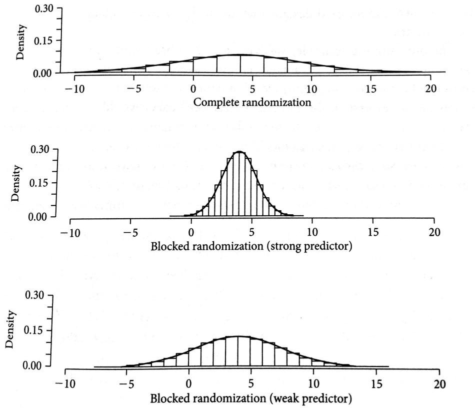


Blocking on covariate X helps when:

- N is relatively small
- X is unbalanced across experimental conditions
- X strongly predicts Y

Useful R package for block random assignment: `randomizr`

FIGURE 4.2
Comparison of sampling distributions based on completely randomized and block randomized designs

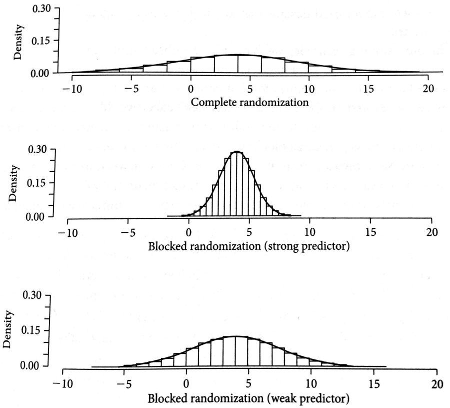


Blocking on covariate X helps when:

- N is relatively small
- X is unbalanced across experimental conditions
- X strongly predicts Y

Useful R package for block random assignment: `randomizr`

FIGURE 4.2
Comparison of sampling distributions based on completely randomized and block randomized designs

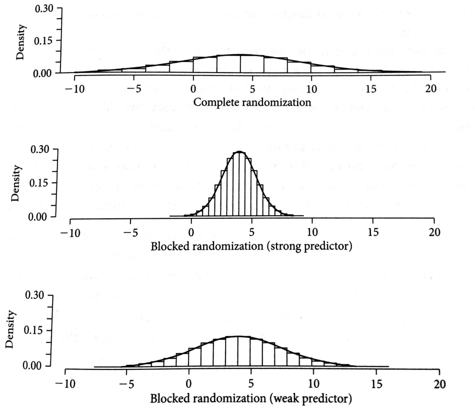


Blocking on covariate X helps when:

- N is relatively small
- X is unbalanced across experimental conditions
- X strongly predicts Y

Useful R package for block random assignment: `randomizr`

FIGURE 4.2
Comparison of sampling distributions based on completely randomized and block randomized designs

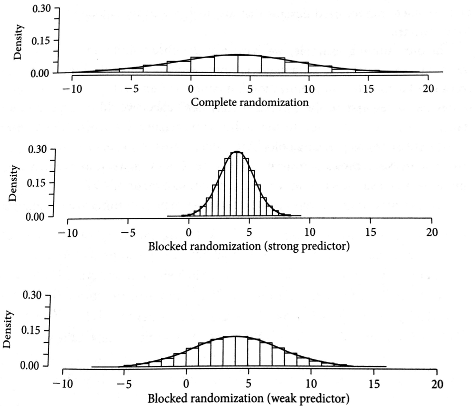


Blocking on covariate X helps when:

- N is relatively small
- X is unbalanced across experimental conditions
- X strongly predicts Y

Useful R package for block random assignment: `randomizr`

FIGURE 4.2
Comparison of sampling distributions based on completely randomized and block randomized designs

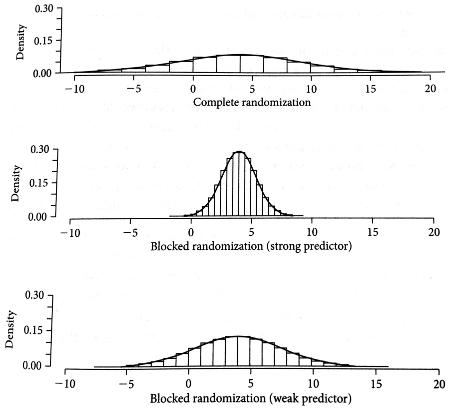


Blocking on covariate X helps when:

- N is relatively small
- X is unbalanced across experimental conditions
- X strongly predicts Y

Useful R package for block random assignment: `randomizr`

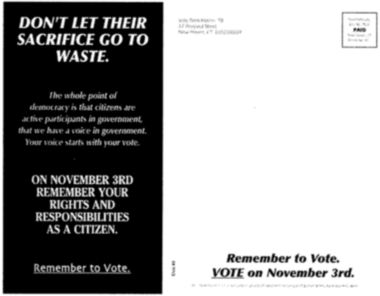
FIGURE 4.2
Comparison of sampling distributions based on completely randomized and block randomized designs



Questions or comments?

FIGURE 2. Picture Side

FIGURE 1. Text Side



> »to find the treatment effect, subtract the turnout rate of the control group from the turnout rate of the experimental group and divide this difference by the observed "contact rate," which is 28%. Using this formula, we find that personal contact raises the probability of turnout by 8.7 percentage points« (658)

»to find the treatment effect, subtract the turnout rate of the control group from the turnout rate of the experimental group and divide this difference by the observed "contact rate," which is 28%. Using this formula, we find that personal contact raises the probability of turnout by 8.7 percentage points« (658)

»Despite limitations, this experiment provides important new clues in the ongoing mystery of why turnout has declined even as the average age and education of the population has risen. A certain segment of the electorate tends not to vote unless encouraged to do so through face-to-face contact. As voter mobilization grows more impersonal, fewer people receive this kind of encouragement. This point is of great practical significance for those who seek to reverse the declining trend in turnout. Many of the recent policy innovations designed to encourage voter participation (e.g., absentee balloting) focus on reducing the costs of voting. Our findings suggest the importance of focusing as well on the personal connection between voters and the electoral process.« (662)

»Despite limitations, this experiment provides important new clues in the ongoing mystery of why turnout has declined even as the average age and education of the population has risen. A certain segment of the electorate tends not to vote unless encouraged to do so through face-to-face contact. As voter mobilization grows more impersonal, fewer people receive this kind of encouragement. This point is of great practical significance for those who seek to reverse the declining trend in turnout. Many of the recent policy innovations designed to encourage voter participation (e.g., absentee balloting) focus on reducing the costs of voting. Our findings suggest the importance of focusing as well on the personal connection between voters and the electoral process.« (662)

»Despite limitations, this experiment provides important new clues in the ongoing mystery of why turnout has declined even as the average age and education of the population has risen. A certain segment of the electorate tends not to vote unless encouraged to do so through face-to-face contact. As voter mobilization grows more impersonal, fewer people receive this kind of encouragement. This point is of great practical significance for those who seek to reverse the declining trend in turnout. Many of the recent policy innovations designed to encourage voter participation (e.g., absentee balloting) focus on reducing the costs of voting. Our findings suggest the importance of focusing as well on the personal connection between voters and the electoral process.« (662)

»Despite limitations, this experiment provides important new clues in the ongoing mystery of why turnout has declined even as the average age and education of the population has risen. A certain segment of the electorate tends not to vote unless encouraged to do so through face-to-face contact. As voter mobilization grows more impersonal, fewer people receive this kind of encouragement. This point is of great practical significance for those who seek to reverse the declining trend in turnout. Many of the recent policy innovations designed to encourage voter participation (e.g., absentee balloting) focus on reducing the costs of voting. Our findings suggest the importance of focusing as well on the personal connection between voters and the electoral process.« (662)

»Despite limitations, this experiment provides important new clues in the ongoing mystery of why turnout has declined even as the average age and education of the population has risen. A certain segment of the electorate tends not to vote unless encouraged to do so through face-to-face contact. As voter mobilization grows more impersonal, fewer people receive this kind of encouragement. This point is of great practical significance for those who seek to reverse the declining trend in turnout. Many of the recent policy innovations designed to encourage voter participation (e.g., absentee balloting) focus on reducing the costs of voting. Our findings suggest the importance of focusing as well on the personal connection between voters and the electoral process.« (662)

»Despite limitations, this experiment provides important new clues in the ongoing mystery of why turnout has declined even as the average age and education of the population has risen. A certain segment of the electorate tends not to vote unless encouraged to do so through face-to-face contact. As voter mobilization grows more impersonal, fewer people receive this kind of encouragement. This point is of great practical significance for those who seek to reverse the declining trend in turnout. Many of the recent policy innovations designed to encourage voter participation (e.g., absentee balloting) focus on reducing the costs of voting. Our findings suggest the importance of focusing as well on the personal connection between voters and the electoral process.« (662)

1 Joe's paper

2 Clustering

3 Using covariates

4 Noncompliance
- Motivating ex.: Gerber & Green (2000)
- Formal statement

5 Is voting contagious? Nickerson (2008)

Conceptually: under one-sided noncompliance, two types of subjects

- compliers: $d_i(z = 1) = 1$
- never-takers: $d_i(z = 1) = 0$

$\rightarrow$ after treatment, three groups:

1. treated compliers
2. non-treated compliers
3. never-takers

Conceptually: under one-sided noncompliance, two types of subjects

- compliers: $d_i(z=1)=1$
- never-takers: $d_i(z=1)=0$

$\rightarrow$ after treatment, three groups:

1. treated compliers
2. non-treated compliers
3. never-takers

Conceptually: under one-sided noncompliance, two types of subjects

- compliers: $d_i(z=1)=1$
- never-takers: $d_i(z=1)=0$

$\rightarrow$ after treatment, three groups:

1. treated compliers
2. non-treated compliers
3. never-takers

Conceptually: under one-sided noncompliance, two types of subjects

- compliers: $d_i(z = 1) = 1$
- never-takers: $d_i(z = 1) = 0$

$\rightarrow$ after treatment, three groups:

1. treated compliers
2. non-treated compliers
3. never-takers

Conceptually: under one-sided noncompliance, two types of subjects

- compliers: $d_i(z=1)=1$
- never-takers: $d_i(z=1)=0$

$\rightarrow$ after treatment, three groups:

1. treated compliers
2. non-treated compliers
3. never-takers

Conceptually: under one-sided noncompliance, two types of subjects

- compliers: $d_i(z = 1) = 1$
- never-takers: $d_i(z = 1) = 0$

$\rightarrow$ after treatment, three groups:

1. treated compliers
2. non-treated compliers
3. never-takers

Conceptually: under one-sided noncompliance, two types of subjects

- compliers: $d_i(z=1)=1$
- never-takers: $d_i(z=1)=0$

$\rightarrow$ after treatment, three groups:

1. treated compliers
2. non-treated compliers
3. never-takers

For each subject i we define:

$$ITT_{i,D} \equiv d_i(1) - d_i(0) \tag{4}$$

$$ITT_{i,Y} \equiv Y_i(1) - Y_i(0) \tag{5}$$

CACE is equal to the relation of $\overline{ITT_{i,Y}}$ to $\overline{ITT_{i,D}}$:

$$CACE = \frac{ITT}{ITT_D} \tag{6}$$

For each subject i we define:

$$ITT_{i,D} \equiv d_i(1) - d_i(0) \tag{4}$$

$$ITT_{i,Y} \equiv Y_i(1) - Y_i(0) \tag{5}$$

CACE is equal to the relation of $\overline{ITT_{i,Y}}$ to $\overline{ITT_{i,D}}$:

$$CACE = \frac{ITT}{ITT_D} \tag{6}$$

For each subject i we define:

$$ITT_{i,D} \equiv d_i(1) - d_i(0) \tag{4}$$

$$ITT_{i,Y} \equiv Y_i(1) - Y_i(0) \tag{5}$$

CACE is equal to the relation of $\overline{ITT_{i,Y}}$ to $\overline{ITT_{i,D}}$:

$$CACE = \frac{ITT}{ITT_D} \tag{6}$$

For each subject i we define:

$$ITT_{i,D} \equiv d_i(1) - d_i(0) \tag{4}$$

$$ITT_{i,Y} \equiv Y_i(1) - Y_i(0) \tag{5}$$

CACE is equal to the relation of $\overline{ITT_{i,Y}}$ to $\overline{ITT_{i,D}}$:

$$CACE = \frac{ITT}{ITT_D} \tag{6}$$

For each subject i we define:

$$ITT_{i,D} \equiv d_i(1) - d_i(0) \tag{4}$$

$$ITT_{i,Y} \equiv Y_i(1) - Y_i(0) \tag{5}$$

CACE is equal to the relation of $\overline{ITT_{i,Y}}$ to $\overline{ITT_{i,D}}$:

$$CACE = \frac{ITT}{ITT_D} \tag{6}$$

Consider this potential outcomes schedule:

TABLE 5.1

Hypothetical schedule of potential outcomes assuming one-sided noncompliance

Observation	$Y_i(d=0)$	$Y_i(d=1)$	$d_i(z=0)$	$d_i(z=1)$	Type
1	4	6	0	1	Complier
2	2	8	0	0	Never-Taker
3	1	5	0	1	Complier
4	5	7	0	1	Complier
5	6	10	0	1	Complier
6	2	10	0	0	Never-Taker
7	6	9	0	1	Complier
8	2	5	0	1	Complier
9	5	9	0	0	Never-Taker

- what is the ATE?
- what is the ITT?
- what is the CACE?

Consider this potential outcomes schedule:

TABLE 5.1
Hypothetical schedule of potential outcomes assuming one-sided noncompliance

Observation	$Y_i(d=0)$	$Y_i(d=1)$	$d_i(z=0)$	$d_i(z=1)$	Type
1	4	6	0	1	Complier
2	2	8	0	0	Never-Taker
3	1	5	0	1	Complier
4	5	7	0	1	Complier
5	6	10	0	1	Complier
6	2	10	0	0	Never-Taker
7	6	9	0	1	Complier
8	2	5	0	1	Complier
9	5	9	0	0	Never-Taker

- what is the ATE?
- what is the ITT?
- what is the CACE?

Consider this potential outcomes schedule:

TABLE 5.1
Hypothetical schedule of potential outcomes assuming one-sided noncompliance

Observation	$Y_i(d=0)$	$Y_i(d=1)$	$d_i(z=0)$	$d_i(z=1)$	Type
1	4	6	0	1	Complier
2	2	8	0	0	Never-Taker
3	1	5	0	1	Complier
4	5	7	0	1	Complier
5	6	10	0	1	Complier
6	2	10	0	0	Never-Taker
7	6	9	0	1	Complier
8	2	5	0	1	Complier
9	5	9	0	0	Never-Taker

- what is the ATE?
- what is the ITT?
- what is the CACE?

Consider this potential outcomes schedule:

TABLE 5.1

Hypothetical schedule of potential outcomes assuming one-sided noncompliance

Observation	$Y_i(d=0)$	$Y_i(d=1)$	$d_i(z=0)$	$d_i(z=1)$	Type
1	4	6	0	1	Complier
2	2	8	0	0	Never-Taker
3	1	5	0	1	Complier
4	5	7	0	1	Complier
5	6	10	0	1	Complier
6	2	10	0	0	Never-Taker
7	6	9	0	1	Complier
8	2	5	0	1	Complier
9	5	9	0	0	Never-Taker

- what is the ATE?
- what is the ITT?
- what is the CACE?

Under one-sided noncompliance, a direct comparison of treated vs. nontreated estimates:

$$CACE + \{E[Y_i(d=0)|D_i(1)=1] - E[Y_i(d=0)|D_i(1)=0]\}(1 - ITT_D) \tag{7}$$

$\rightarrow$ bias if compliers and never-takers have unequal untreated potential outcomes

Under one-sided noncompliance, a direct comparison of treated vs. nontreated estimates:

$$CACE + \{E[Y_i(d=0)|D_i(1)=1] - E[Y_i(d=0)|D_i(1)=0]\}(1 - ITT_D) \quad (7)$$

→ bias if compliers and never-takers have unequal untreated potential outcomes

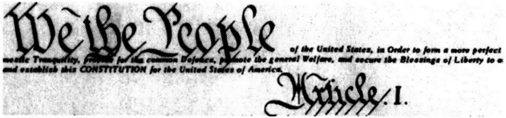
Under one-sided noncompliance, a direct comparison of treated vs. nontreated estimates:

$$CACE + \{E[Y_i(d = 0)|D_i(1) = 1] - E[Y_i(d = 0)|D_i(1) = 0]\}(1 - ITT_D) \tag{7}$$

$\rightarrow$ bias if compliers and never-takers have unequal untreated potential outcomes

Questions or comments?

Experimental conditions:



Women waited 144 years for the right to vote.

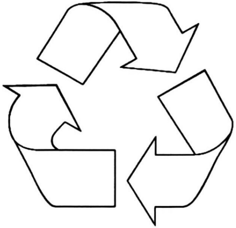
African-Americans waited 94 years for the right to vote and another 94 years to make that right meaningful.

All you had to do was turn 18.

Make your voice heard.

Vote Tuesday, September 10th.

Think recycling doesn't matter?



One million tons of aluminum containers are thrown away each year.

Americans throw away enough aluminum every three months to rebuild our entire commercial air fleet.

Making new aluminum cans from used cans takes 95 percent less energy and 20 recycled cans can be made with the energy needed to produce one can using virgin ore.

The energy required to replace the aluminum cans thrown away in 2001 is roughly the equivalent of 16 million gallons of crude oil: enough to meet the electricity needs of all the homes in Chicago, Dallas, Detroit, San Francisco, and Seattle combined.

Please do your part and recycle!

Possible outcomes

TABLE 1. Possible Outcomes under placebo protocol

		Probability of Event Occurring	Voting Rate of Answerer	Voting Rate of Person Who Did Not Answer Door
GOTV	Door Answered	π	$\mu_1 + T$	$\mu_2 + S$
	No Answer	$1 - \pi$	N.A.[a]	μ_3
Recycling	Door Answered	π	μ_1	μ_2
	No Answer	$1 - \pi$	N.A.	μ_3

[a] N.A. = Not applicable.

$$\alpha = \frac{S}{T}$$

$$T = \overline{V}_{Ga} - \overline{V}_{Ra}$$

$$S = \overline{V}_{G\tilde{a}} - \overline{V}_{R\tilde{a}}$$

$$\alpha = \frac{S}{T}$$

$$T = \overline{V}_{Ga} - \overline{V}_{Ra}$$

$$S = \overline{V}_{G\tilde{a}} - \overline{V}_{R\tilde{a}}$$

$$\alpha = \frac{S}{T}$$

$$T = \overline{V}_{Ga} - \overline{V}_{Ra}$$

$$S = \overline{V}_{G\tilde{a}} - \overline{V}_{R\tilde{a}}$$

Effect estimates

TABLE 3. Treatment Effect among Contacted Households

	Denver		Minneapolis		Pooled	
	Direct	Secondary	Direct	Secondary	Direct	Secondary
Percent Voting in GOTV Group	47.7% (3.0)	42.4% (2.9)	27.1% (3.1)	23.6% (3.0)		
Percent Voting in Recycling Group	39.1% (2.9)	36.9% (2.9)	16.2% (2.7)	17.3% (2.7)		
Estimated Treatment Effect	**8.6% (4.2)**	**5.5% (4.1)**	**10.9% (4.1)**	**6.4% (4.1)**	**9.8% (2.9)**	**6.0% (2.9)**
P-Value	0.02	0.09	<0.01	0.06	<0.01	0.02

Note. Numbers in parentheses represent standard errors. P-values test the one-tailed hypothesis. Pooled estimates are weighted averages of results for both cities.

$\rightarrow$ pooled estimated $\alpha = \frac{6}{9.8} = 61.2\%$

Effect estimates

TABLE 3. Treatment Effect among Contacted Households

	Denver		Minneapolis		Pooled	
	Direct	Secondary	Direct	Secondary	Direct	Secondary
Percent Voting in GOTV Group	47.7% (3.0)	42.4% (2.9)	27.1% (3.1)	23.6% (3.0)		
Percent Voting in Recycling Group	39.1% (2.9)	36.9% (2.9)	16.2% (2.7)	17.3% (2.7)		
Estimated Treatment Effect	**8.6% (4.2)**	**5.5% (4.1)**	**10.9% (4.1)**	**6.4% (4.1)**	**9.8% (2.9)**	**6.0% (2.9)**
P-Value	0.02	0.09	<0.01	0.06	<0.01	0.02

Note. Numbers in parentheses represent standard errors. P-values test the one-tailed hypothesis. Pooled estimates are weighted averages of results for both cities.

$\rightarrow$ pooled estimated $\alpha = \frac{6}{9.8} = 61.2\%$

See you tomorrow!