7: Eksperimenter I

Videregående kvantitative metoder i studiet af politisk adfærd

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

Institut for Statskundskab
Københavns Universitet

24. oktober 2016

1 Formalia

2 Opsamling fra sidst

3 Mere om potential outcomes framework

4 Randomisering i praksis

5 Faldgruber ved eksperimentelle designs

6 Case: Gerber & Green (2000)

7 Kig fremad

Frederik Hjorth   Institut for Statskundskab København Universitet

Uge	Dato	Tema	Litteratur	Case
1	5/9	Introduktion til R	Imai kap 1	
2	12/9	Regression I: OLS	GH kap 3, MM kap 2	Gilens & Page (2014)
3	26/9	Regression II: Paneldata	GH kap 11	Larsen et al. (2016)
4	29/9	Regression III: Multileveldata, interaktioner	GH kap 12	Berkman & Plutzer
5	3/10	Introduktion til kausal inferens	Hariri (2012), Samii (2016)	
6	10/10	Matching	Justesen & Klemmensen (2014)	Ladd & Lenz (2009)
	17/10	*Efterårsferie*		

Formalia
○●○

Opsamling
○○○

Mere om POF
○○○○○○○○

Randomisering i praksis
○○

Faldgruber
○○○○

Case: Gerber & Green (2000)
○○○

Kig fremad
○○

Uge	Dato	Tema	Litteratur	Case
	17/10	*Efterårsferie*		
7	24/10	Eksperimenter I	MM kap 1, GG kap 1+2	Gerber et al. (2008)
8	31/10	Eksperimenter II	GG kap 3+4+5	Gerber & Green (2000)
9	10/11	Instrumentvariable	MM kap 3	Arunachalam & Watso
10	14/11	Regressionsdiskontinuitetsdesigns	MM kap 4	Eggers & Hainmueller
11	21/11	Difference-in-difference designs	MM kap 5	Enos (2016)
12	28/11	'Big data' og maskinlæring	Grimmer (2015), Varian (2014)	
13	5/12	Scraping af data fra online-kilder	MRMN kap 9	
14	12/12	Tekst som data	Grimmer & Stewart (2013), Imai kap 5	

Spørgsmål?

- introduktion til potential outcomes framework
- problemer ved ekstrapolation og modelafhængighed
- common support
- propensity score matching
- coarsened exact matching
- balance checks
- case: Ladd & Lenz (2009)

Centrale tendenser i midterms:

- mere omfangsrig end estimeret
- stor grundighed i opgaveløsningen
- positivt: styr på fortolkninger, implementering i R
- negativt: lidt meget pensumrepetition ctr. overvejelser knyttet til problemstillingen

Formalia
000

Opsamling
00●

Mere om POF
00000000

Randomisering i praksis
00

Faldgruber
0000

Case: Gerber & Green (2000)
000

Kig fremad
00

Spørgsmål?

Motiverende eksempel: NHIS ($N \approx 18.600$)

TABLE 1.1
Health and demographic characteristics of insured and uninsured
couples in the NHIS

	Husbands			Wives		
	Some HI (1)	No HI (2)	Difference (3)	Some HI (4)	No HI (5)	Difference (6)
	A. Health					
Health index	4.01 [.93]	3.70 [1.01]	.31 (.03)	4.02 [.92]	3.62 [1.01]	.39 (.04)
	B. Characteristics					
Nonwhite	.16	.17	−.01 (.01)	.15	.17	−.02 (.01)
Age	43.98	41.26	2.71 (.29)	42.24	39.62	2.62 (.30)
Education	14.31	11.56	2.74 (.10)	14.44	11.80	2.64 (.11)
Family size	3.50	3.98	−.47	3.49	3.93	−.43

To MIT-studerende, **K**huzdar & **M**aria

$$Y_{1K} - Y_{0K} = 4 - 3 = 1 \tag{1}$$

$$Y_{1M} - Y_{0M} = 5 - 5 = 0 \tag{2}$$

Fuldt potential outcomes schedule:

	Khuzdar	Maria
Y_{0i}	3	5
Y_{1i}	4	5
D_i	1	0
Y_i	4	5
$Y_{1i} - Y_{0i}$	1	0

Observerede outcomes:

	Khuzdar	Maria
Y_{0i}	?	5
Y_{1i}	4	?

$$\rightarrow \bar{Y}_1 - \bar{Y}_0 = 4 - 5 = -1$$

Den direkte sammenligning afspejler både ATE hos de treatede + selection bias:

$$Y_K - Y_M = Y_{1K} - Y_{0M} \tag{3}$$
$$= Y_{1K} - Y_{0K} + Y_{0K} - Y_{0M} \tag{4}$$
$$= 1 + (-2) \tag{5}$$
$$= -1 \tag{6}$$

M. mere generel notation i GG:

$$E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] =$$
$$E[Y_{1i} - Y_{0i}|D_i = 1] + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] \quad (7)$$

når treatment randomiseres er Y_0i uafhængig af D_i:

$$E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] = 0 \quad (8)$$

Succesfuld randomisering kan efterprøves m. *balance tests*

TABLE 1.3

Demographic characteristics and baseline health in the RAND HIE

	Means	Differences between plan groups			
	Catastrophic plan (1)	Deductible − catastrophic (2)	Coinsurance − catastrophic (3)	Free − catastrophic (4)	Any insurance − catastrophic (5)
A. Demographic characteristics					
Female	.560	−.023 (.016)	−.025 (.015)	−.038 (.015)	−.030 (.013)
Nonwhite	.172	−.019 (.027)	−.027 (.025)	−.028 (.025)	−.025 (.022)
Age	32.4 [12.9]	.56 (.68)	.97 (.65)	.43 (.61)	.64 (.54)
Education	12.1 [2.9]	−.16 (.19)	−.06 (.19)	−.26 (.18)	−.17 (.16)
Family income	31,603 [18,148]	−2,104 (1,384)	970 (1,389)	−976 (1,345)	−654 (1,181)
Hospitalized last year	.115	.004	−.002	.001	.001

Spørgsmål?

Gerber & Greens procedure:

»First, determine N, the number of subjects in your experiment, and m, the number of subjects who will be allocated to the treatment group. Second, set a random number 'seed' using a statistics package, so that your random numbers may be reproduced by anyone who cares to replicate your work. Third, generate a random number for each subject. Fourth, sort the subjects by the random numbers in ascending order. Finally, classify the first m observations as the treatment group.« (37)

Spørgsmål?

To kritiske antagelser om potential outcomes:

1. excludability
2. non-interferens (SUTVA)

Ad (1):

Lad $Y_i(z, d)$ være potential outcome for treatment assignment $z_i = z$ og faktisk treatment status $d_i = d$

Eksklusionsrestriktionsantagelsen: $Y_i(1, d) = Y_i(0, d)$

Ad (2):

Lad $Y_i(\mathbf{z}, \mathbf{d})$ være PO for Y_i for den fulde mængde af assignments og treatments

Under non-interferens: $Y_i(\mathbf{z}, \mathbf{d}) = Y_i(z, d)$

Spørgsmål?

Formalia
○○○

Opsamling
○○○

Mere om POF
○○○○○○○○

Randomisering i praksis
○○

Faldgruber
○○○○

Case: Gerber & Green (2000)
●○○

Kig fremad
○○

TABLE 2.   Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N of Individuals	191,243	38,218	38,204	38,218	38,201

Formalia
ooo

Opsamling
ooo

Mere om POF
oooooooo

Randomisering i praksis
oo

Faldgruber
oooo

Case: Gerber & Green (2000)
o●o

Kig fremad
oo

Neighbors mailing

30423-3 ||| || || | |||

For more information: (517) 351-1975
email: ctov@grebner.com
Practical Political Consulting
P. O. Box 6249
East Lansing, MI 48826

PRSRT STD
U.S. Postage
PAID
Lansing, MI
Permit # 444

ECRLOT **C050
THE JACKSON FAMILY
9999 MAPLE DR
FLINT MI 48507

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for
years, but it only seems to get worse. This year, we're taking a new approach.
We're sending this mailing to you and your neighbors to publicize who does and
does not vote.

The chart shows the names of some of your neighbors, showing which have voted in
the past. After the August 8 election, we intend to mail an updated chart. You
and your neighbors will all know who voted and who did not.
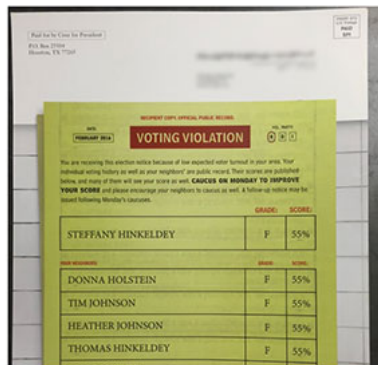
DO YOUR CIVIC DUTY — VOTE!

MAPLE DR	Aug 04	Nov 04	Aug 06
9995 JOSEPH JAMES SMITH	Voted	Voted	_____
9995 JENNIFER KAY SMITH		Voted	_____
9997 RICHARD B JACKSON		Voted	_____
9999 KATHY MARIE JACKSON		Voted	_____

Formalia
○○○

Opsamling
○○○

Mere om POF
○○○○○○○○

Randomisering i praksis
○○

Faldgruber
○○○○

Case: Gerber & Green (2000)
○○●

Kig fremad
○○

Næste gang:

- eksperimenter II
- fokus: cluster random assignment, covariate adjustment, noncompliance
- case: Gerber & Green (2000)

Tak for i dag!