Randomized Controlled Trials
or
Experiments

Rajeev Dehejia

The fundamental problem, multiple units

Unit	Outcome under treatment	Outcome under control	Treatment effect
1	$\boxed{Y_{1,1}}$	$\cancel{Y_{0,1}}$	τ_1 *(Unobservable) treatment effect for Y_1*
2	$\cancel{Y_{1,2}}$	$\boxed{Y_{0,2}}$	τ_2
3	$\boxed{Y_{1,3}}$	$\cancel{Y_{0,3}}$	τ_3
4	$\boxed{Y_{1,4}}$	$\cancel{Y_{0,4}}$	τ_4
N	$\cancel{Y_{1,N}}$	$\boxed{Y_{0,N}}$	τ_N
Sum	$\mu_T = \Sigma Y_{1n}/N$	$\mu_C = \Sigma Y_{0n}/N$	$\tau = \Sigma \tau_n/N = \mu_T - \mu_C$ *(Unobservable) Population treatment effec*
	$\bar{x}_T = \sum_{i\in\{T\}} y_{1i}/N_T$	$\bar{x}_C = \sum_{i\in\{C\}} y_{01i}/N_C$	
	$E(\bar{x}_T) \neq \mu_T$	$E(\bar{x}_C) \neq \mu_C$	

Recall: The fundamental problem

$$
\begin{aligned}
\tau &= E(Y_{1i}) - E(Y_{0i}) \\
&= \Pr(T_i = 1)E(Y_{1i} \mid T_i = 1) + \Pr(T_i = 0)E(Y_{1i} \mid T_i = 0) \\
&\quad - \left[\Pr(T_i = 1)E(Y_{0i} \mid T_i = 1) + \Pr(T_i = 0)E(Y_{0i} \mid T_i = 0)\right]
\end{aligned}
$$

Angrist & Levy paper example:

- What we want to know: The effect of cash incentive on taking the Bagrut
- What we're missing: For the students who received the cash incentive (T=1), how many of them would have taken the Bagrut if they had not? (T=0)

Or selection bias (example)

How randomization helps (1)

Recall the selection problem when comparing the mean outcomes for the treated and the untreated:

$$\underbrace{E\left[Y \middle| T=1\right]-E\left[Y \middle| T=0\right]}_{\text{Difference in Means}}=E\left[Y_1 \middle| T=1\right]-E\left[Y_0 \middle| T=0\right]$$

$$=\underbrace{E\left[Y_1-Y_0 \middle| T=1\right]}_{ATET}+\underbrace{\left\{E\left[Y_0 \middle| T=1\right]-E\left[Y_0 \middle| T=0\right]\right\}}_{BIAS}$$

- Random assignment of units to the treatment forces the selection bias to be zero
- The treatment and control group will tend to be similar along all characteristics, including the potential outcomes under the control condition

How randomization helps (2)

Unit	Outcome under treatment		Outcome under control	Treatment effect
1	$\boxed{Y_{11}}$	Randomly assigned	$\cancel{Y_{01}}$	τ_1
2	$\cancel{Y_{12}}$		$\boxed{Y_{02}}$	τ_2
3	$\boxed{Y_{13}}$		$\cancel{Y_{03}}$	τ_3
4	$\boxed{Y_{14}}$		$\cancel{Y_{04}}$	τ_4
N	$\cancel{Y_{1N}}$		$\boxed{Y_{0N}}$	τ_N
Sum	$\mu_T = \Sigma Y_{1n}/N$		$\mu_C = \Sigma Y_{0n}/N$	$\tau = \Sigma \tau_n/N = \mu_T - \mu_C$
	$\bar{x}_T = \sum_{i \in \{T\}} y_{1i}/N_T$		$\bar{x}_C = \sum_{i \in \{C\}} y_{01i}/N_C$	
	$E(\bar{x}_T) = \mu_T$		$E(\bar{x}_C) = \mu_C$	

How randomization helps (3)

T_i=1 or 0, randomly assigned (assume N treated, N control)

$$
\begin{aligned}
E(\bar{x}_T) &= E\left(\tfrac{1}{N}\sum_{n\in T} Y_n\right) \\
&= E\left(\tfrac{1}{N}\sum_{n\in T,C} T_i Y_n\right) \\
&= \tfrac{1}{N}\left(\sum_{n\in T,C} Y_n E(T_i)\right) \\
&= \tfrac{1}{N}\left(\sum_{n\in T,C} \tfrac{1}{2} Y_n\right) \\
&= \tfrac{1}{2N}\left(\sum_{n\in T,C} Y_n\right) \\
&= \mu_T
\end{aligned}
$$

- Likewise with control and treatment effect.

How randomization helps (4) or How does this relate to regression?

$$
\begin{aligned}
Y_i &= T_i Y_{1i} + (1 - T_i) Y_{0i} \\
&= Y_{0i} + T_i (Y_{1i} - Y_{0i}) \\
&= \alpha + \beta_i T_i + \varepsilon_i, \quad \text{where } \varepsilon_i = Y_{0i} - \alpha
\end{aligned}
$$

- Random assignment guarantees

$$T_i \perp\!\!\!\perp Y_{1i}, Y_{0i}$$

- Hence that T_i is uncorrelated with ε_i.

Identification in randomized experiments

Randomization implies:

(Y_1, Y_0) independent of T, or $(Y_1, Y_0) \perp\!\!\!\perp T$.

We have that $E\left[Y_0 \middle| T=1\right] = E\left[Y_0 \middle| T=0\right]$ and therefore

$$\alpha_{ATET} = E\left[Y_1 - Y_0 \middle| T=1\right] = E\left[Y \middle| T=1\right] - E\left[Y \middle| T=0\right]$$

Also, we have that

$$\alpha_{ATE} = E\left[Y_1 - Y_0\right] = E\left[Y_1 - Y_0 \middle| T=1\right] = E\left[Y \middle| T=1\right] - E\left[Y \middle| T=0\right]$$

As a result,

$$\underbrace{E\left[Y \middle| T=1\right] - E\left[Y \middle| T=0\right]}_{\text{Difference in Means}} = \alpha_{ATE} = \alpha_{ATET}$$

Identification in randomized experiments

The identification result extends beyond average treatment effects.
Given random assignment $(Y_1, Y_0) \perp\!\!\!\perp T$:

$$F_{Y_0}(y) = \Pr(Y_0 \leq y) = \Pr(Y_0 \leq y \mid T = 0)$$
$$= \Pr(Y \leq y \mid T = 0)$$

Similarly,

$$F_{Y_1}(y) = \Pr(Y \leq y \mid T = 1).$$

So effect of the treatment at any quantile, $Q_\theta(Y_1) - Q_\theta(Y_0)$ is identified.

•Randomization identifies the whole marginal distributions of Y_0 and Y_1

•Does not identify the quantiles of the effect: $Q_\theta(Y_1 - Y_0)$ (the difference of quantiles is not the quantile of the difference).

Estimation in randomized experiments

$$\alpha_{ATE} = E[Y_1 - Y_0] = E[Y \mid T=1] - E[Y \mid T=0]$$

$$\hat{\alpha} = \bar{Y}_1 - \bar{Y}_0$$

$$\bar{Y}_1 = \frac{\sum Y_i \cdot T_i}{\sum T_i} = \frac{1}{N_1} \sum_{T_i=1} Y_i$$

$$\bar{Y}_0 = \frac{\sum Y_i \cdot (1 - T_i)}{\sum (1 - T_i)} = \frac{1}{N_0} \sum_{T_i=0} Y_i$$

$N_1 = \sum_i T_i$

α_{ATE}

Testing in large samples: Two sample t-test

From stats, we know that:

$$\frac{\hat{\alpha} - \alpha_{ATE}}{\sqrt{\frac{\hat{\sigma}_1^2}{N_1} + \frac{\hat{\sigma}_0^2}{N_0}}} \xrightarrow{d} N(0,1),$$

where $\hat{\sigma}_d^2 = \sum_{T_i=d} \left(Y_i - \bar{Y}_d\right)^2 (N_d - 1), \left(d \in \{0,1\}\right)$ Let

$$t = \frac{\hat{\alpha}}{\sqrt{\frac{\hat{\sigma}_1^2}{N_1} + \frac{\hat{\sigma}_0^2}{N_0}}}.$$

We reject the null hypothesis $H_0 : \alpha_{ATE} = 0$ against the alternative $H_1 : \alpha_{ATE} \neq 0$ at the 5% significance level if $|t| > 1.96$.

Example: Bertrand/Mullainathan

- Run an experiment in which they submitted c.v.'s to job applications.
- The treatment was that on an otherwise identical c.v., they changed randomized a "white sounding" and "black sounding" name.

What is a white/black sounding name?

Appendix Table 1
First Names Used in Experiment[a]

White Female			**African American Female**		
Name	$\frac{L(W)}{L(B)}$	Perception White	Name	$\frac{L(B)}{L(W)}$	Perception Black
Allison	∞	0.926	Aisha	209	0.97
Anne	∞	0.962	Ebony	∞	0.9
Carrie	∞	0.923	Keisha	116	0.93
Emily	∞	0.925	Kenya	∞	0.967
Jill	∞	0.889	Lakisha	∞	0.967
Laurie	∞	0.963	Latonya	∞	1
Kristen	∞	0.963	Latoya	∞	1
Meredith	∞	0.926	Tamika	284	1
Sarah	∞	0.852	Tanisha	∞	1
Fraction of all births:			Fraction of all births		
3.8%			7.1%		

White Male			**African American Male**		
Name	$\frac{L(W)}{L(B)}$	Perception White	Name	$\frac{L(B)}{L(W)}$	Perception Black
Brad	∞	1	Darnell	∞	0.967
Brendan	∞	0.667	Hakim	∞	0.933
Geoffrey	∞	0.731	Jamal	257	0.967
Greg	∞	1	Jermaine	90.5	1
Brett	∞	0.923	Kareem	∞	0.967
Jay	∞	0.926	Leroy	44.5	0.933
Matthew	∞	0.888	Rasheed	∞	0.931
Neil	∞	0.654	Tremayne	∞	0.897
Todd	∞	0.926	Tyrone	62.5	0.900
Fraction of all births:			Fraction of all births		
1.7%			3.1%		

Example: Bertrand/Mullainathan

Table 1
Mean Callback Rates By Racial Soundingness of Names [a]

	Callback Rate for White Names	*Callback Rate for African American Names*	*Ratio*	*Difference (p-value)*
Sample:				
All sent resumes	**9.65%** [2435]	**6.45%** [2435]	**1.50**	**3.20%** (0.0000)
Chicago	**8.06%** [1352]	**5.40%** [1352]	**1.49**	**2.66%** (0.0057)
Boston	**11.63%** [1083]	**7.76%** [1083]	**1.50**	**4.05%** (0.0023)
Females	**9.89%** [1860]	**6.63%** [1886]	**1.49**	**3.26%** (0.0003)
Females in administrative jobs	**10.46%** [1358]	**6.55%** [1359]	**1.60**	**3.91%** (0.0003)
Females in sales jobs	**8.37%** [502]	**6.83%** [527]	**1.22**	**1.54%** (0.3523)
Males	**8.87%** [575]	**5.83%** [549]	**1.52**	**3.04%** (0.0513)

Estimating treatment effects: A regression approach

- Run regression:

$$y_i = \beta_0 + \beta_1 T_i + \varepsilon$$

```
reg y_var x_var
```

The OLS estimator of β_1 is an unbiased estimator of the causal effect of T on y.

- Actually in this case computationally identical to difference in means!
- Hence can read off estimate of treatment effect from coefficient on T.
- Approach easily generalizes to where T is not binary.
- Also gives estimate of standard error.

Computing standard errors

- Unless told otherwise regression package will compute standard errors assuming errors are homoskedastic.
- Even if only interested in effect of treatment on mean *T* may affect other aspects of distribution, e.g. variance
- This will cause heteroskedasticity
- Heteroskedasticity does not make OLS regression coefficients inconsistent but does make OLS standard errors inconsistent (hypothesis tests wrong!)

‘Robust’ standard errors

- Also called:
  - Huber standard errors
  - White standard errors
  - Heteroskedastic-consistent standard errors

Including other regressors

- Can get consistent estimate of treatment effect without worrying about other variables
- Reason is that randomization ensures no problem of omitted variables bias
- But there are reasons to include other regressors:
  - Improved efficiency
  - Check for randomization
    - If randomization worked, then average values for all variables except for treatment should be roughly equal in T and C
  - Improve randomization
  - Control for conditional randomization
  - Heterogeneity in treatment effects
    - Can bring useful insights to like on differential effects for different subgroups

Uses of regressors 1: Improved efficiency

- Don’t just want consistent estimate of causal effect – also want low standard error (or high precision or efficiency).
- Standard formula for standard error of OLS estimate of β is $\sigma^2(X'X)^{-1}$
- σ^2 comes from variance of residual in regression – $(1-R^2)$* Var(y)
- `In stata, SE’s are directly to the right of parameter estimates in regression output`

Uses of regressors 2: Check for successful randomization

- Randomization balances observed but also unobserved characteristics between treatment and control group
- Can check random assignment with respect to observed covariates, X, using so called “balance tests” (e.g., t-tests) to see if distributions of the covariates are the same in the treatment and control groups
- X are pre-treatment variables that are measured prior to treatment assignment (i.e., at the “baseline”)
- Use difference in means or regressions.

Uses of regressors 2 (cont'd)

Three approaches:

•For each covariate, do a difference in means and t-test between treatment and control group

•For each covariate, *X*, run the regression:

$$X_i = \alpha + \beta T_i + \varepsilon_i$$

```
reg covariate1 treatment_var
```

•Run a probit of the form

$$\Pr(T_i\!=\!1) = \alpha + \beta_1 X_{1i} + \ldots + \beta_2 X_{2i+\ldots}\ \varepsilon_i$$

And test hypothesis $\beta_1=\beta_2=\ldots=\beta_n=0$

Uses of regressions 3:
Heterogeneity in treatment effects

- So far have assumed causal (treatment) effect the same for everyone
- No good reason to believe this
- This can be tackled either using differences in means or regressions (latter has advantages)
- Start with case of no other regressors:

$$y_i = \beta_0 + \beta_{1i} T_i + \varepsilon_i$$

- Random assignment implies T independent of β_{1i}
- Sometimes called random coefficients model

Example: Bertrand/Mullainathan

- Different treatment effect for high and low quality CVs detected via difference in means:

	Low	High	Ratio	Difference (p-value)
White Names	**8.50%**	**10.79%**	**1.27**	**2.29%**
	[1212]	[1223]		(0.0557)
African American Names	**6.19%**	**6.70%**	**1.08**	**0.51%**
	[1212]	[1223]		(0.6084)

Observable heterogeneity

- Potential outcomes notation:
  - Outcome if in control group:

$$Y_{0i} = \alpha_0 + \beta_0 X_i + \varepsilon_{0i}$$

  - Outcome if in treatment group:

$$Y_{1i} = \alpha_1 + \beta_1 X_i + \varepsilon_{1i}$$

- Treatment effect is $(Y_{1i} - Y_{0i})$ and can be written as:

$$(Y_{1i} - Y_{0i}) = (\beta_1 - \beta_0)X_i + \varepsilon_{1i} - \varepsilon_{0i}$$

- Note treatment effect has observable and unobservable component
- Can estimate as:
  - Two separate equations
  - One single equation

Combining treatment and control groups into single regression

- Recall: $$Y_i = T_i Y_{1i} + (1 - T_i) Y_{0i}$$
- Combining outcomes equations leads to:

$$
\begin{aligned}
Y_i &= T_i(\alpha_1 + \beta_1 X_i + \varepsilon_{1i}) + (1 - T_i)(\alpha_0 + \beta_0 X_i + \varepsilon_{0i}) \\
&= \alpha_0 + (\alpha_1 - \alpha_0)T_i + \beta_0 X_i + (\beta_1 - \beta_0)T_i X_i + \varepsilon_{0i} + (\varepsilon_{1i} - \varepsilon_{0i})T_i \\
&= a + bT_i + cT_i X_i + dX_i + e_i
\end{aligned}
$$

```
Reg y_var trt_var trt*co_var co_var
```

- We can do this! Just a regression of Y on treatment indicator, covariate, and interaction.
- Advantages: easy correction for heteroskedasticity and can incorporate many covariates.

Experimental design

- The idea is to design an experiment in a way that the data have useful properties in analysis.
- Begin with randomization of treatment.
- But more broadly you can control the design to take into account what you want eventually to do with the data.
- “Design away your problems, rather than assume them away…”

Experimental design (1): Conditional randomization

- Rather than just randomize the treatment between treatment and control, you randomize within subgroups defined on covariates.
- Simple example: Bertrand-Mullainathan within low and high quality c.v.'s randomize on race.

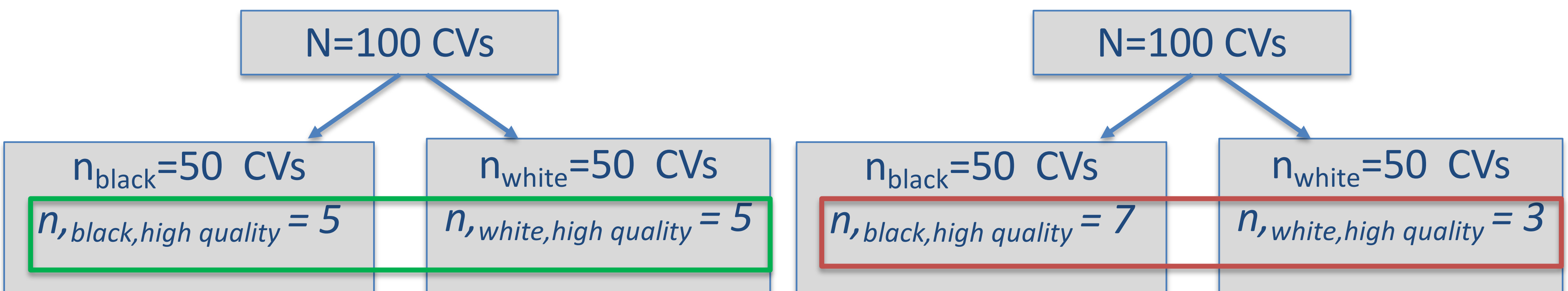
Experimental design (1): Conditional randomization

- Why do this? On average, you will the right mix of treatment and control within each covariate cell…*On average*.
- But in small cells, just by randomization you may not get balance.
- E.g., 100 observations, 90 high quality c.v.'s, 10 low quality. Randomize 50-50 white and black names. By chance you might get 5 and 5 in the low quality, but could get 3 and 7.
- Instead force the randomization to balance within these cells.

Example data: Tabular format

CV ID #	Treatment (1= Black, 0= white)	Quality (1= high, 0=low)
CV_1	0	0
CV_2	1	1
CV_3	1	1
CV_4	0	1
CV_5	1	1
....CV_100	0	0

Good randomization **Bad randomization**



Experimental design (1): Conditional randomization

- When to do it? When you know that ex post you will be interested in that split of the data (and when some of the subgroups could be small).
- How to deal with it? Either run difference in means separately within each group or…
- You guessed it, run a regression.
- Note: if the probability of treatment differs across cells, you *must* use one of the two methods above – can't just compare overall difference in means.

Experimental design (1): Uses of regression (4) / conditional randomization

- Example. Suppose your divide your data into 4 cells: low & high quality c.v.'s (Q=0,1) , and junior & senior jobs (S=0,1).
- Within each cell you randomize black/white names with 50% probability.
- Now your regression would be:

$$Y_i = a + bT_i + cQ_i + dS_i + e_i$$

or

$$Y_i = a + bT_i + cQ_i + dS_i + gT_iS_i + hT_iQ_i + kT_iQ_iS_i + jQ_iS_i + e_i$$

Experimental design (1): Uses of regression (4) / conditional randomization

What does this mean?

$$Y_i = a + bT_i + cQ_i + dS_i + e_i$$

Probability of getting a call back = constant + effect of having a black sounding name + effect of having a high-quality CV + effect of applying to a senior job + error

$$Y_i = a + bT_i + cQ_i + dS_i + gT_iS_i + hT_iQ_i + kT_iQ_iS_i + jQ_iS_i + e_i$$

Same as above, but with interaction terms (allows for heterogeneous effects… would you expect that the quality of a CV would have a differential impact on the probability of getting a call back for a junior vs. senior job?

Experimental design (1): Conditional randomization

- First regression computes an average treatment effect over the entire population. Controlling for *S* and *Q* increases efficiency but isn't needed.
- Second regression computes a different treatment effect in each cell. How?

$$Y_i = a + bT_i + cQ_i + dS_i + gT_iS_i + hT_iQ_i + kT_iQ_iS_i + jQ_iS_i + e_i$$

- $\alpha_{ATE} | S=1, Q=1 = E(Y|S=1,Q=1,T=1) - E(Y|S=1,Q=1,T=0)$
  $= a+b+c+d+g+h+k+j-(a+c+d+j) = b+g+h+k$
- $\alpha_{ATE}|S=1,Q=0 = a+b+d+g-(a+d) = b+g$
- $\alpha_{ATE}|S=0,Q=1 = a+b+c+h-(a+c) = b+h$
- $\alpha_{ATE}|S=0,Q=0 = a+b - (a) = b$

Experimental design (1): Conditional randomization

- If you wanted to average these 4 subtreatment effects into one average how would you do it?
- Weight each one by:
  - The proportion of the population in each cell.
  - The proportion of the treated population in each cell.
  - In the current example these are the same, but won't always be.

Example: Angrist-Lavy

- Angrist and Lavy investigate the effective on financial incentives on high school matriculation.
- The high-end of a big problem in developing countries (developed ones too...)
- In developing countries we tend to care more about literacy and elementary school, but interesting idea.

Example: Angrist and Lavy

- A small pilot program selected individual students within schools for treatment, with treatment status determined by previous test scores and a partially randomized cutoff for low socioeconomic status.
- In a larger follow-up program, entire schools were randomly selected for treatment and the program operated with the cooperation of principals and teachers.
- The results suggest the Achievement Awards program that randomized treatment at the school level raised matriculation rates, while the student-based program did not.

The design

- The population of 1302 seniors enrolled in the 1999-2000 school year into three groups on the basis of the **number of Bagrut subject tests they had taken previously** and their **maximum score on these tests.**
- We estimated Logit regressions with information from the previous cohort of students to predict the probability of Bagrut certification as a function of these two variables, denoted here by p_{1i} for student i.
- **All** students with a very low probability of Bagrut attainment (p_{1i}<0.053) were offered the opportunity to earn a bonus. It was inexpensive and politically expedient to offer bonuses to this group, about 15 percent of enrolled seniors in the Southern cohort.
- Students with a very high probability of success were excluded; in particular, we did not offer bonuses to 612 students with p_{1i}>0.66, about half of seniors.

The design (cont'd)

- The remaining 491 students were potentially eligibile.
- Treatment was assigned to these students as a function of family size and father's education, with students of lower socioeconomic status more likely to be in the treatment group.
- We used the previous cohort of seniors to estimate the probability a student would obtain a Bagrut certificate as a function of family size and father's schooling, denoted p_{2i}:

$$T_{ij} = 1[p_{2i} < q_{.22}(j)(1-Z_i) + q_{.7}(j)Z_i]$$

where $q_{.22}(j)$ and $q_{.7}(j)$ are the .22 and .7 quantiles of the p_{2i} distribution in school j.

- So Z_i is like a randomized encouragement variable.

Table 1: Experimental Design for the Pilot Demonstration

Range for p_{1i}	Range for p_{2i}	Threshold for p_{2i} Low $q_{.22}$	Threshold for p_{2i} High $q_{.7}$	Offered Bonus No	Offered Bonus Yes	Row Totals
		A. All-Treated Sample (p_{1i}<.053)				
$[0, q_{.15}]$		--	--	0	146	
		B. Eligible Sample (.053<p_{1i}<.67)				
$[q_{.15}, q_{.53}]$	$[0, q_{.22}]$	59	64	0	123	123
	$[q_{.22}, q_{.7}]$	127	125	127	125	252
	$[q_{.7}, 1]$	56	58	114	0	114
	Column Totals	241	248	242	247	489
		C. No-treated Sample (p_{1i}>.67)				
$[q_{.53}, 1]$		--	--	612	0	

Table 3: Reduced Form Effects in the Pilot Experiment (Eligible Sample)

	All Eligible Pupils				Jewish Eligible Pupils	
	No Covariates	School Covs p_{2i}	School f.e. p_{2i}	p_{1i}, sex, School f.e., P_{2i}	No Covariates	School f.e. $p2i$
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
Offered	0.521	0.531	0.535	0.535	0.503	0.526
	(0.039)	(0.030)	(0.028)	(0.028)	(0.041)	(0.030)
Received Bagrut	-0.003	0.005	0.001	-0.017	0.013	0.014
	(0.043)	(0.042)	(0.042)	(0.039)	(0.045)	(0.044)

School-level design

- They then do an assignment at the school level, 40 schools half treated.
- In some sense N=40, so don't expect balance of all student characteristics.
- But do matched pairs based on lagged values of the Bagrut rate.

Table 6. Grouped Estimates for the Schools Experiment

		Unweighted			Weighted		
		No controls	Sch Cov	Sch Cov + Pair	No controls	Sch Cov	Sch Cov + Pair
Sample	Mean	(1)	(2)	(3)	(4)	(5)	(6)
		A. 2001 Sample					
1. All Pairs (39 Schools; 3828 Pupils)	0.245	0.075	0.078	0.082	0.048	0.057	0.056
		(0.063)	(0.059)	(0.059)	(0.050)	(0.049)	(0.050)
		[0.062]	[0.057]	[0.038]	[0.047]	[0.047]	[0.033]

Experimental design 2: Matched-pair design

- Start with a similar motivation as stratification. Your treatment and comparison groups are heterogeneous, but you want a similar set of characteristics in the treatment and comparison groups.
- Randomization deals with this if enough units.
- But if not, and if too many characteristics to stratify?
- Matched-pair (-triple,…) design.
  - Group obs together in pairs that are similar, and assign one of each pair to each treatment.

Example: Duflo-Saez

- Duflo and Saez investigate the role of social interactions in retirement saving decisions.
- Question: people tend not to go to information sessions (and incidently make suboptimal choices).
- Randomly encourage them to. But the treatment could be being encouraged yourself or indirect (having someone in your department encouraged).

The setup

- The university we study has approximately 12,500 employees. About a quarter of the employees are faculty members. Our study was limited to nonfaculty employees only. The university provides retirement benefits to its employees through a traditional mandatory pension plan, but employees can also voluntarily contribute to a complementary Tax Deferred Account (TDA) 403(b) plan. Every employee can contribute to the 403(b) plan any percentage of their salary up to the IRS limit ($10,500 per year for each individual in 2001). The university does not match contributions. Employees can choose how to invest their contributions in any combination of four different vendors.
- Each year, the university organizes a benefits fair where all employees are invited to come and learn about the different kinds of benefits (such as health benefits, retirement benefits, etc.) provided by the university. The fair is held on two consecutive days in early November in two different locations, each one close to the two separate main university campuses. About one week before the fair, every employee receives a letter through the university mail system inviting her to attend the fair. This letter also provides a brief description of the event. At the same time, under separate cover, every employee receives a packet describing in detail university benefits along with enrollment forms. November is "open enrollment" month during which each employee may change her benefits choices by submitting the enrollment form. If the employee does not send back the form, her benefits choices are automatically carried over from the previous year. However, employees are free to enroll in the TDA or change their contribution level or investment decision at any time throughout the year.
- In both locations, the fair is held in a large hotel reception room. There are a large number of stands representing the university Benefits Office, and the various health and retirement benefits service providers. The university Benefits Office offers information on all benefits through direct conversation with Benefits Office staff present at the fair, and through a number of information pamphlets freely available at their stand. The benefits Office also provides information on how the other stands at the fair are organized. These other stands are run by each of the specialized service providers. For example, each of the mutual fund vendors has a stand at which they provide information about the TDA plan and the specific services they offer within that plan. The fair also offers individuals the chance to use a specially designed computer program to analyze their specific situation. Employees are free to come any time during the three and a half hours during which the fair is held, and visit any number of stands they want.

The design

- The university organizes the annual fair in order to disseminate information about benefits and to help its employees make better decisions. The benefits office of the university realizes that the participation rate among staff (34 percent) is too low compared with that at other universities, and suspects that this may be due to lack of information.
- In order to identify the causal effect of fair attendance on TDA enrollment, we set up an "encouragement design," by promising a random subset of employees a small amount of money for attending the fair. In order to shed light on social effects within departments, we sent those letters only within a random subset of departments. There are thus two distinct treatments in our experiment: receiving the letter, and being in the same department as someone who receives a letter.
- In the first step, we randomly selected two-thirds of the departments of the university (220 out of a total of 330) as follows. In order to maximize the power of the experiment (in a context in which we know there are strong department effects), we first matched departments according to their size (i.e., number of employees) and participation rate in the TDA before the fair. We separated departments into deciles of participation rates among the staff. Each decile contains 33 departments. We then ranked them by size within each decile, and formed groups of three departments by putting three consecutive departments on these lists in the same triplet. Within each of these triplets, we randomly selected two departments to be part of the group of treated departments. From now on, we denote by the dummy variable D the treatment status of departments. We have $D = 1$ in treated departments and $D = 0$ in control departments.
- In the second step, within each of the treated departments, any individual not enrolled as of August 2000, was selected with a probability of one-half. This treatment group is composed of 2039 individuals. From now on, we denote by the dummy variable L the selection status of individuals. We refer to this group as the Treated individuals and denote them by 11 ($D = 1$ for Treated department and $L = 1$ for being selected). The group formed by the employees in the treated departments who were not selected contains 2129 individuals and is denoted by 10 ($D = 1$ for Treated department and $L = 0$ for not being selected). In total, there are 4168 individuals in the treated departments. The control group is formed by employees in the control departments where no treatments were selected; it contains 2043 individuals and is denoted by 00 ($D = 0$ and $L = 0$).
- One week before the fair, we sent a letter via university mail to the 2039 employees in the treatment group 11. The letter reminded them of the fair and informed them that they would receive a check for $20 from us if they were to come to the fair and register at our desk.

Descriptive statistics serve as randomization checks, inform the interpretation of findings, and inform the generalizability of findings

TABLE I

DESCRIPTIVE STATISTICS, BY GROUPS

	Treated departments			
	All (group $D = 1$)	Treated (group $D = 1$, $L = 1$)	Untreated (group $D = 1$, $L = 0$)	Untreated departments (group $D = 0$)
	(1)	(2)	(3)	(4)
PANEL A: BACKGROUND CHARACTERISTICS				
TDA participation before	0.010	0.009	0.011	0.012
the fair (Sept. 2000)	(.0015)	(.0021)	(.0022)	(.0024)
Observations	4168	2039	2129	2043
Sex (fraction male)	0.398	0.400	0.396	0.418
	(.0076)	(.0109)	(.0107)	(.011)
Years of service	5.898	5.864	5.930	6.008
	(.114)	(.161)	(.16)	(.157)
Annual salary	38,547	38,807	38,297	38,213
	(304)	(438)	(422)	(416)
Age	38.3	38.4	38.2	38.7
	(.17)	(.24)	(.24)	(.24)
Observations	4126	2020	2106	2018
PANEL B: FAIR ATTENDANCE (REGISTRATION DATA)				
Fair attendance rate among	0.214	0.280	0.151	0.049
non-TDA enrollees	(.0064)	(.01)	(.0078)	(.0048)
Observations	4126	2020	2106	2018
Fair attendance rate for all	0.192			0.063
staff employees	(.0132)			(.0103)
Observations	6687			3311
PANEL C: TDA PARTICIPATION (ADMINISTRATIVE DATA)				
TDA participation rate after	0.049	0.045	0.053	0.040
4.5 months	(.0035)	(.0049)	(.0051)	(.0045)
Observations	3726	1832	1894	1861
TDA participation rate after	0.088	0.089	0.088	0.075
11 months	(.005)	(.0071)	(.007)	(.0065)
Observations	3246	1608	1638	1633

Experimental design 3: Cross-cutting RCT designs

- You can randomize not only one treatment, but two (or more) orthogonally.
- Example from Yuva experiment:

	No prize	Uniform prize	Choice prize
EU classrooms			
Fixed theory	23	23	24
Incremental theory	23	23	24

Experimental design 3: Cross-cutting RCT designs

- There are two reasons to do this: get two experiments in one.
- Or you may be interested in the interactions of the treatments.

	No prize	Uniform prize	Choice prize
EU classrooms			
Fixed theory	23	23	24
Incremental theory	23	23	24

Example: Karlan and Zinman

- They are looking at the effect of interest rates in microfinance borrowing.
- So the randomly assign different borrowers different interest rates.
- But they also randomize other things.

Example: Karlan-Zinman

the trusted way to borrow cash

30 October 2003

A low rate for you.

Congratulations! As a valued client, you are now eligible for a low interest rate on your next cash loan from This is a limited time offer, so please come in by 30 November 2003 to take advantage of this offer.

You can use this cash to pay for school, or for anything else you want.

Enjoy low monthly repayments with this offer! Here is one example of a loan you can get under this offer:

Interest Rate	Loan Amount	Loan Term	Monthly Repayment
10.50%	R2000.00	4 Months	R710.00

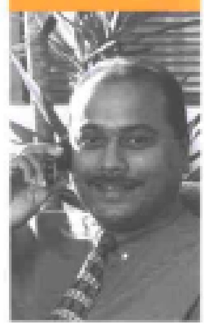
LOAN AVAILABILITY SUBJECT TO TERMS & CONDITIONS

Loans available in other amounts. There are no hidden costs. What you see is what you pay.

If you borrow elsewhere you will pay R360.00 more in total on a R2000.00, 4 month loan.

How to apply:

Bring your ID book and latest payslip to your usual branch, by **30 November 2003** and ask for



Area Manager

P.S. Unfortunately, if you have already taken a loan since the date this letter was issued, you do not qualify for this offer. Comparison based on a competitor's interest rate of 15% per month.

Example: Karlan-Zinman

- They randomize the interest rate, the term of the loan, complexity of the offer (show one example of a loan or two), and the picture of the person on the letter.

the trusted way to borrow cash

25 September 2003

A low rate for you.

Congratulations! You are now eligible for a special interest rate on a cash loan from y. This is a limited time offer, so please come in by 31 October 2003

You can use this cash to pay off a more expensive debt, or for anything else you want.

Enjoy low monthly repayments with this offer! For example:

Interest Rate	Loan Amount	Loan Term	Monthly Repayment
3.99%	R500	4 Months	R144.95
3.99%	R1000	4 Months	R289.90
3.99%	R2000	4 Months	R579.80
3.99%	R4000	4 Months	R1159.60

LOAN AVAILABILITY SUBJECT TO TERMS & CONDITIONS

Loans available in other amounts. There are no hidden costs. What you see is what you pay.

If you borrow from us you will pay R840.40 less in total on a R1000.00, 4 month loan.

How to apply:

Bring your ID book and latest payslip to your usual branch, by **31 October 2003** and ask for !

Customer Consultant

P.S. Unfortunately, if you have already taken a loan since the date this letter was issued, you do not qualify for this offer. Comparison based on a competitor's interest rate of 35%.

Example: Karlan-Zinman

- So they got multiple experiments out of the same RCT.
- One paper on interest rates.
- Another looking at the effect of the psychological manipulations.

Recent Papers

- **COMMUNICATING TAX PENALTIES TO DELINQUENT TAXPAYERS: EVIDENCE FROM A FIELD EXPERIMENT**- Taylor Cranor, Jacob Goldin, Tatiana Homonoff, and Lindsay Moore; National Tax Journal, June 2020, 73 (2), 331–360; https://doi.org/10.17310/ntj.2020.2.02
- **A randomized control trial evaluating the effects of police body-worn cameras;** David Yokum, Anita Ravishankar, and Alexander Coppock; PNAS May 21, 2019 116 (21) 10329-10332; first published May 7, 2019 https://doi.org/10.1073/pnas.1814773116
- **Breaking Routine for Energy Savings: An Appliance-level Analysis of Small Business Behavior under Dynamic Prices;** Jiyong Eom, Frank A. Wolak; NBER Working Paper No. 27263; Issued in May 2020

Applications and RCTs in the "real world"

- Ideal circumstances for running RCTs: You have control over randomization, assignment, data collection, implementation checks, SUTVA is reasonable to assume, and no exogenous shocks to broader life will happen (e.g. a pandemic….)
- Spotting well-designed RCTs: Results are replicable, limitations and generalizability are acknowledged, several robustness checks

Applications and RCTs in the “real world”

- Who conducts RCTs
  - Social scientists in academia
  - Tech companies (UX, user behavior)
  - Government offices (OMB, Mayor’s Office of Economic Opportunity)
  - Research consulting firms (Mathematica, MDRC, Ideas42, the Behavioral Insights Team…)
  - Medical academia/clinical trials…

Problems with experiments

- Expense
  - Though some firms are specializing in low-cost evaluations
- Ethical Issues
- Threats to Internal Validity
  - Failure to follow experiment
  - Experimental effects (Hawthorne effects)
- Threats to External Validity
  - Non-representative program
  - Non-representative sample
  - Scale effects

Conclusions on experiments

- Are 'gold standard' of empirical research
- Are becoming more common
- Not enough of them to keep us busy
- Study of non-experimental data can deliver useful knowledge
- Some issues similar, others different

What we'll cover in recitation

- ✓ Stata code and interpretation of output for RCTs
- ✓ Notation review
- ✓ Overview of readings/ discussion of methods and key take-aways (come prepared with questions!)
- ✓ Running basic descriptive statistics
- ✓ More discussion on RCTs in the field/employers that require this evaluation experience

Email Michelle at dimarm01@nyu.edu with questions beforehand if desired