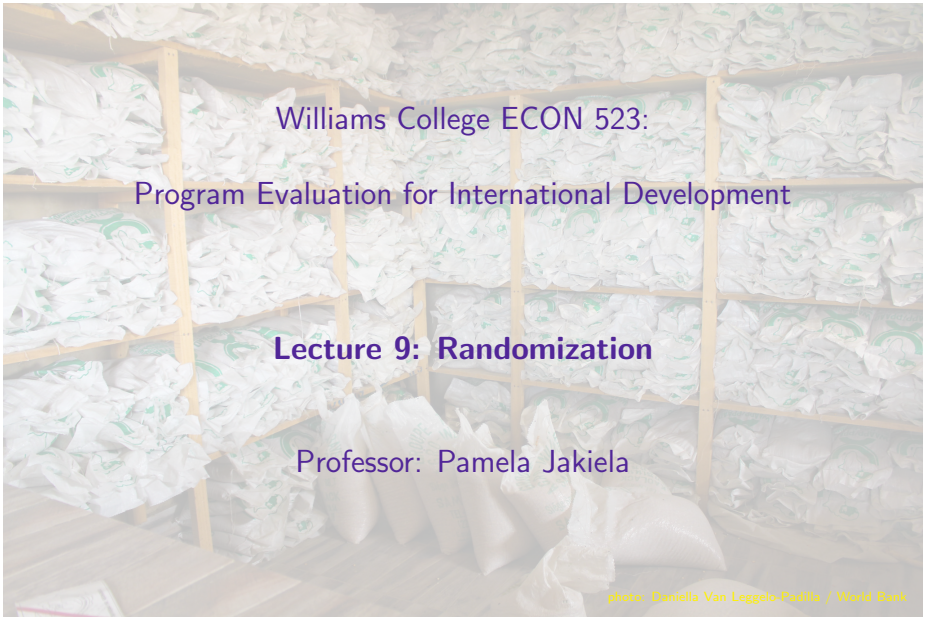


Williams College ECON 523:

Program Evaluation for International Development

Lecture 9: Randomization

Professor: Pamela Jakiela

photo: Daniella Van Leggelo-Padilla / World Bank

The Mechanics of Random Assignment

Random Assignment in Three Steps

Randomly assigning treatment using a computer program involves three steps:

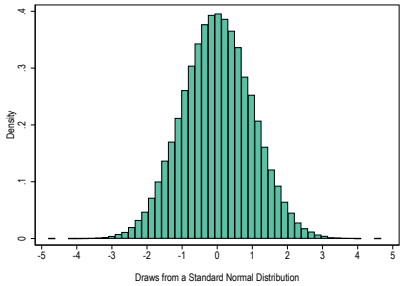
1. Assign each eligible unit a (pseudo-)random number
2. Sort the units based on the random numbers
3. Assign treatment based on the random sort order

Step 0: Obtain a List of Eligible Units

	name
1	Collins
2	Damien
3	Darcy
4	Ethan
5	Gamal
6	Gibson
7	Graham
8	Karisa
9	Kelvis
10	Michael
11	Miora
12	Nèmè
13	Oliver
14	Oyuka
15	Raquel

Step 1: Assign Each Eligible Unit a Random Number

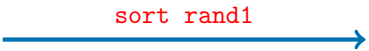
```
clear all
use names.dta
gen rand1 = rnormal()
```



	name	rand1
1	Collins	-1.562377
2	Damien	.3883744
3	Darcy	1.117409
4	Ethan	1.239225
5	Gamal	-.5867468
6	Gibson	2.226561
7	Graham	.0776411
8	Karisa	-.0444348
9	Kelvis	1.062407
10	Michael	.8970407
11	Miora	-.1408677
12	Nèmè	-.6115746

Step 2: Sort Units by Random Number

	name	rand1
1	Collins	-1.562377
2	Damien	.3883744
3	Darcy	1.117409
4	Ethan	1.239225
5	Gamal	-.5867468
6	Gibson	2.226561
7	Graham	.0776411
8	Karisa	-.0444348
9	Kelvis	1.062407
10	Michael	.8970407
11	Miora	-.1408677
12	Nèmè	-.6115746



	name	rand1
1	Collins	-1.562377
2	Vanya	-1.015693
3	Raquel	-.6689101
4	Rubana	-.6267059
5	Nèmè	-.6115746
6	Gamal	-.5867468
7	Raïssa	-.5758651
8	Oyuka	-.2037489
9	Miora	-.1408677
10	Karisa	-.0444348
11	Graham	.0776411
12	Tatia	.2315856

Step 3: Assign Treatment Based on Random Sort Order

```
egen treatment = seq(), from(0) to(1)
```

	name	rand1	treatment
1	Collins	-1.562377	0
2	Vanya	-1.015693	1
3	Raquel	-.6689101	0
4	Rubana	-.6267059	1
5	Nèmè	-.6115746	0
6	Gamal	-.5867468	1
7	Raïssa	-.5758651	0
8	Oyuka	-.2037489	1
9	Miora	-.1408677	0
10	Karisa	-.0444348	1
11	Graham	.0776411	0
12	Tatia	.2315856	1

```
clear all
use names.dta
gen rand1 = rnormal()
sort rand1
egen treatment = seq(), from(0) to(1)
```

Step 3: Assign Treatment Based on Random Sort Order

```
egen treatment = seq(), from(1) to(3)
```

	name	rand1	treatment
1	Collins	-1.562377	1
2	Vanya	-1.015693	2
3	Raquel	-.6689101	3
4	Rubana	-.6267059	1
5	Nèmè	-.6115746	2
6	Gamal	-.5867468	3
7	Raïssa	-.5758651	1
8	Oyuka	-.2037489	2
9	Miora	-.1408677	3
10	Karisa	-.0444348	1
11	Graham	.0776411	2
12	Tatia	.2315856	3

```
clear all
use names.dta
gen rand1 = rnormal()
sort rand1
egen treatment = seq(), from(0) to(1)
```

Making Treatment Assignments Reproducible

Start every treatment assignment do file by setting the seed: `set seed 8675309`

	name	rand1	treatment
1	Karisa	-2.196844	1
2	Vanya	-1.476004	2
3	Kelvis	-.962465	3
4	Graham	-.9155009	1
5	Oyuka	-.587773	2
6	Collins	-.4868504	3
7	Gibson	-.4059047	1
8	Victoria	-.2623108	2
9	Tatia	-.2230017	3
10	Gamal	-.1391217	1
11	Miora	-.0306005	2
12	Rubana	-.0137603	3

```
clear all
set seed 1234
use names.dta
gen rand1 = rnormal()
sort rand1
egen treatment = seq(), from(0) to(1)
```

Randomizing at the Right Level

SUTVA

The **Stable Unit Treatment Value Assumption (SUTVA)**:
Potential outcomes of individual i do not depend on another unit j's treatment assignment

When is SUTVA likely to be violated?

- SUTVA violations matter when we anticipate detectable spillovers unto other eligible units

Conceptually, you should design your RCT so that you are randomizing at a high enough scale to avoid (serious) SUTVA violations (e.g. at the village or school rather than the child level)

Cluster-Randomized Trials

An RCT is **cluster-randomized** if treatment assignment occurs at a higher level than outcomes

- Example: extension agent assigned to village, but farmer-level data on inputs and crops

When to cluster, and how to choose the level at which to assign treatments:

- You cannot assign treatments at a level below the level of data collection
- You should not assign treatments at a level that will lead to SUTVA violations
- You should consider compliance logistics: who needs to implement treatment assignments?
  - Teachers, doctors, tax collectors, etc. have different objectives and limited attention
- Are there political or social reasons to cluster treatment assignments?

For statistical power reasons, you want to randomize at the lowest level that works

Cluster-Randomized Trials: How To

The mechanics of random assignment are the same in cluster-randomized

- Instead of a list of individual units, you need a list of eligible clusters to randomize

It is (usually) important to construct the list of eligible individuals prior to randomization

- Treatment may influence individuals' willingness to participate in the study
- Constructing lists of individuals/units to survey post-treatment may introduce selection
  - Example: a school meals program may bring additional students to school, or increase attendance; so you should not collect data by surveying students at (T and C) schools
  - Example: a tax compliance intervention may make firms less welling to be surveyed

Stratification

The Expected Level of Imbalance

Thought experiment:
You randomly assign treatment in a large sample, and then test 100 variables to see if they treatment and control group means are different – how many variables will be imbalanced?

Some imbalances (between treatment and control) matter more than others

- To enforce balance on important covariates, we typically **stratify** treatment assignments
- Intuitively, stratification is like running separate RCTs within each stratum
- In practice, we first sort by the stratification variables, then assign treatment

Stratification in Practice

Sort by stratification variable(s) or strata number and then by random number

	name	cde	rand1
1	Collins	1	-.4868504
2	Damien	1	-1.476004
3	Darcy	1	-.0109115
4	Ethan	0	-.0137603
5	Gamal	1	.5155788
6	Gibson	1	-.1391217
7	Graham	0	1.287317
8	Karisa	1	-.587773
9	Kelvis	1	-.0306005
10	Michael	0	-2.196844
11	Miora	1	-.9155009
12	Nèmè	1	-.2230017

	name	cde	rand1
1	Michael	0	-2.196844
2	Ethan	0	-.0137603
3	Vanya	0	.3223809
4	Raquel	0	.3661239
5	Graham	0	1.287317
6	Damien	1	-1.476004
7	Raïssa	1	-.962465
8	Miora	1	-.9155009
9	Karisa	1	-.587773
10	Collins	1	-.4868504
11	Victoria	1	-.4059047
12	Oyuka	1	-.2623108

Stratification in Practice

Stratification is equivalent to running many small within-strata randomized trials

	name	cde	hehim	rand1
1	Vanya	0	0	.3223809
2	Raquel	0	0	.3661239
3	Michael	0	1	-2.196844
4	Ethan	0	1	-.0137603
5	Graham	0	1	1.287317
6	Raïssa	1	0	-.962465
7	Miora	1	0	-.9155009
8	Victoria	1	0	-.4059047
9	Oyuka	1	0	-.2623108
10	Nèmè	1	0	-.2230017
11	Rubana	1	0	.2532923
12	Tatia	1	0	2.284249

You can only stratify by discrete variables

The number of observations per strata must be at least as large as the number of treatments

Must have data on all stratification variables for all units being assigned to treatment

Reasons to Stratify

1. To enforce balance in terms of important covariates (e.g. baseline outcomes)
2. To enforce balance for fairness (or other political/feasibility) reasons
3. To enable tests of treatment effect heterogeneity
4. To increase statistical power (by explain variation in outcomes of interest)
   - You need to know what predicts your outcomes in cross-sectional data

Random Assignment in Practice: Questions and Takeaways

In the research design stage, ask yourself:

- What are you randomizing, and at at what level?
- Do you have a list of eligible units, and, if not, how will you create one?
- What are your stratification variables, and do you have data on them for all units?

Once you've collected the data described above, randomization is easy:

1. Set the seed(!), and then assign each eligible unit a pseudo-random number
2. Sort by stratum, and then by random number within each stratum
3. Assign treatments by counting off (0/1 or 1/2/3/etc.)
4. Check for balance, and make a balance check table