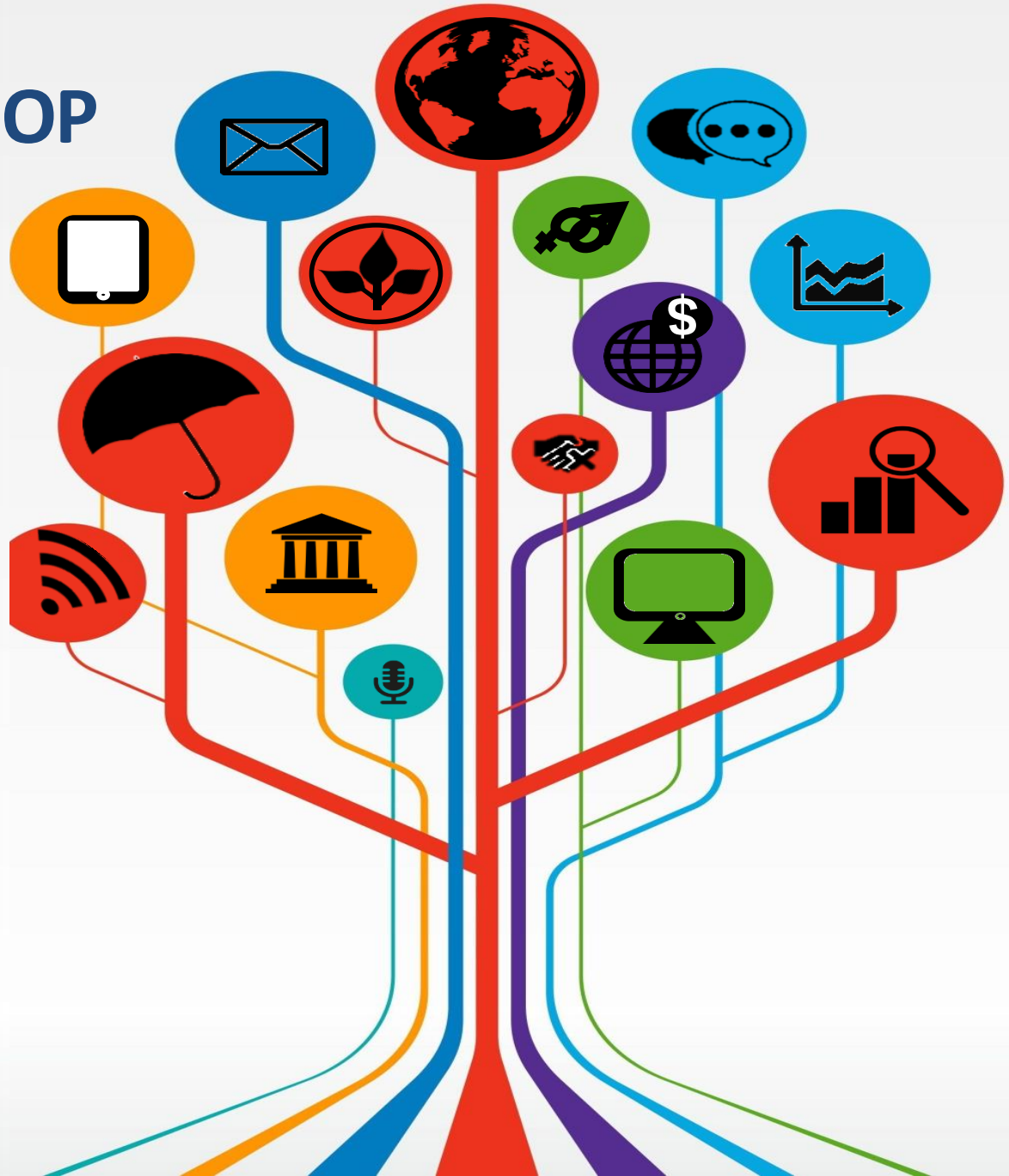
FIELD COORDINATOR WORKSHOP

Manage Successful Impact Evaluations

18 - 22 JUNE 2018
WASHINGTON, DC



i2i DIME
TRANSFORM DEVELOPMENT

WORLD BANK GROUP

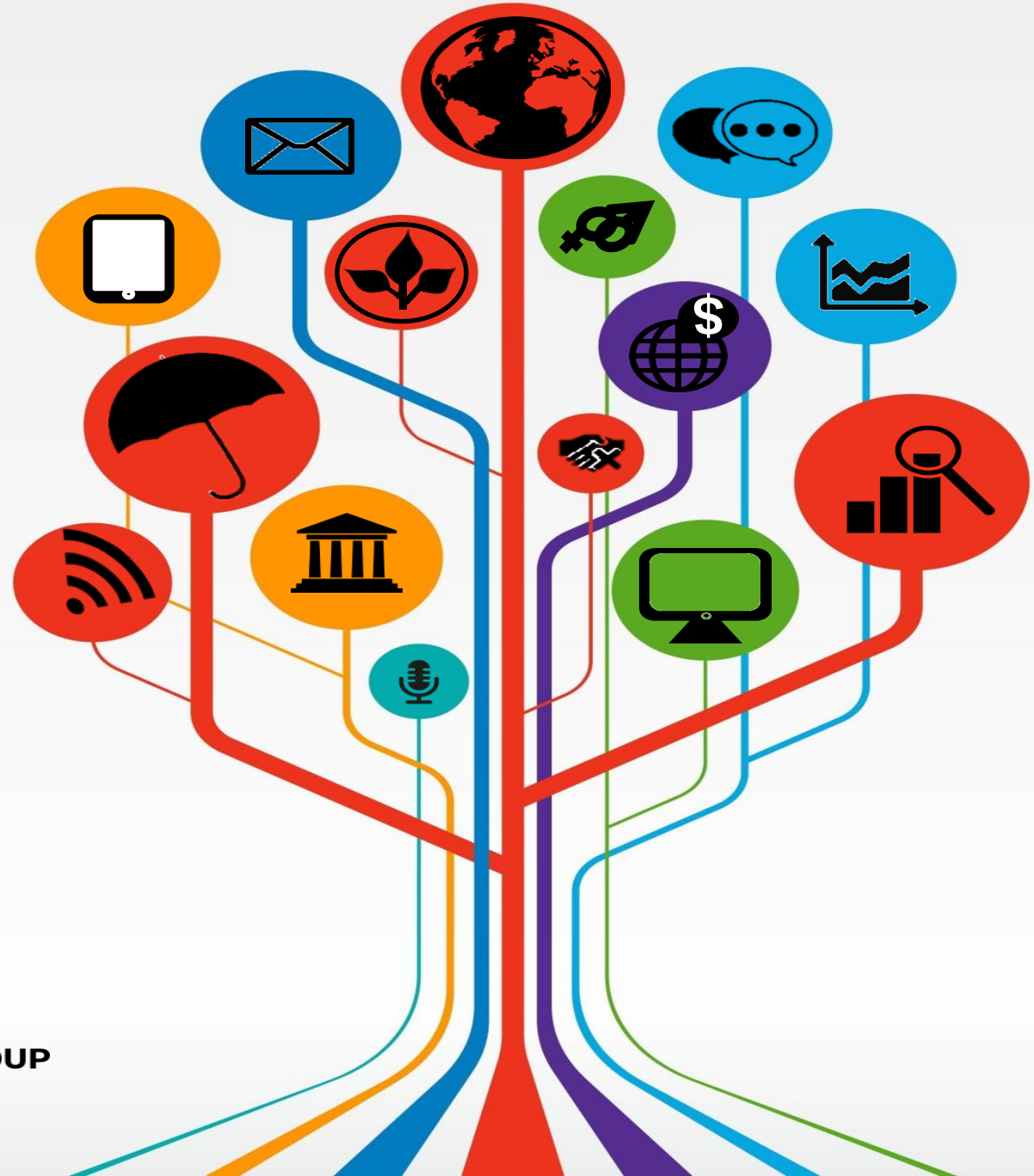
Reproducible Randomization

Stata Track 2

Prepared by **DIME Analytics**
DIMEAnalytics_Internal_Use_Only@worldbank.org

Presented by **Benjamin Daniels and Alexander Spevack**
bdaniels@worldbank.org | *aspevack@worldbank.org*

http://www.worldbank.org/en/research/dime
June 20, 2018



i2i DIME
TRANSFORM DEVELOPMENT

WORLD BANK GROUP

Why do we randomize?

- We want to assign the intervention of our projects so that the control group is as similar as possible to the treatment group as possible
  - This is called a balanced treatment assignment
  - Randomization is the most common tool to achieve that
  - With randomization, we know that the treatment is uncorrelated with the outcome ex ante

- Each observation needs to have the same probability to end up in the each treatment arm, and all groups need to be statistically similar

Background

- Pure random assignment guarantees that the treatment and control groups, on average, will have identical characteristics
- In any particular random allocation, the two groups will differ along some dimensions, with the probability that such differences are large, falling with sample
  - Confounders
  - Efficiency (e.g. unequal group size)
- Small-scale experiments (30 to 100 observations)
- What options exist and what to do?

Review: Balance tries to establish a counterfactual

What we know:

Good randomization methods produce an unbiased estimate for treatment *in the average randomization*.

What we don't know:

How far from the "average" randomization is this particular draw *in terms of counterfactual outcome equivalence*?

Variable	(1) Control Mean/SE	(2) Treatment Mean/SE	T-test Difference (1)-(2)
Age in years	42.880	42.126	0.754
	(1.746)	(0.535)	
Respondent is male	0.538	0.479	0.059
	(0.050)	(0.008)	
Years of schooling	10.930	10.838	0.092
	(0.171)	(0.183)	
Respondent is employed	0.835	0.892	-0.057
	(0.060)	(0.041)	
Monthly earnings (number of minimum wages)	1.582	1.491	0.091
	(0.094)	(0.067)	
Average commuting distance	18.241	11.737	6.504***
	(1.078)	(0.233)	
N	158	167	
Clusters	6	6	
F-test of joint significance (F-stat)			9.892***
F-test, number of observations			325

Notes: The value displayed for t-tests are the differences in the means across the groups. The value displayed for F-tests are the F-statistics. Standard errors are clustered at variable neighborhood. The covariate variable stratum is included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Pure RCTs are balanced – in the *average* assignment

- No matter what statistics you are going by, if a (conditional) means difference has p=0.05, a difference of that size will be realized in 5% of draws. That's the definition.
- The *average* randomized treatment assignment produces statistically identical groups across treatments
- But, a p=0.05 tail draws happen one time in 20
- You only get *one draw* on your randomization!
- We'd like balance tests to show ours was a "good" one, so that "real" differences must be due to treatment assignment

Balance → unbiasedness?
Need a model and an assumption

- **Model – Asymptotically normal OLS**
  - T-statistics of "closeness" to the average randomization depend on large-sample properties
- **Assumption – no omitted variables in regression**
  - $Y_i = \alpha + \beta X_i + \varepsilon_i$
  - Unless all x's in X are in the model, balance on observables doesn't imply that expected outcomes for treated and untreated are equal
  - If you are wrong, biases are of unknown magnitude and direction depending on the omitted variable's actual correlation with treatment and outcome variables (classical OVB)
  - In small samples especially, can't guarantee that treatment randomization produces zero expected differences for all omitted covariates

What we often do: block/stratify the experiment

- Meets our core criteria:
  - Randomization process is fixed over real data
  - Counterfactuals (zero treatment effect) has a distribution that can be simulated exactly
- Allows use of “randomization inference” statistics that do not depend on extrapolation from model

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

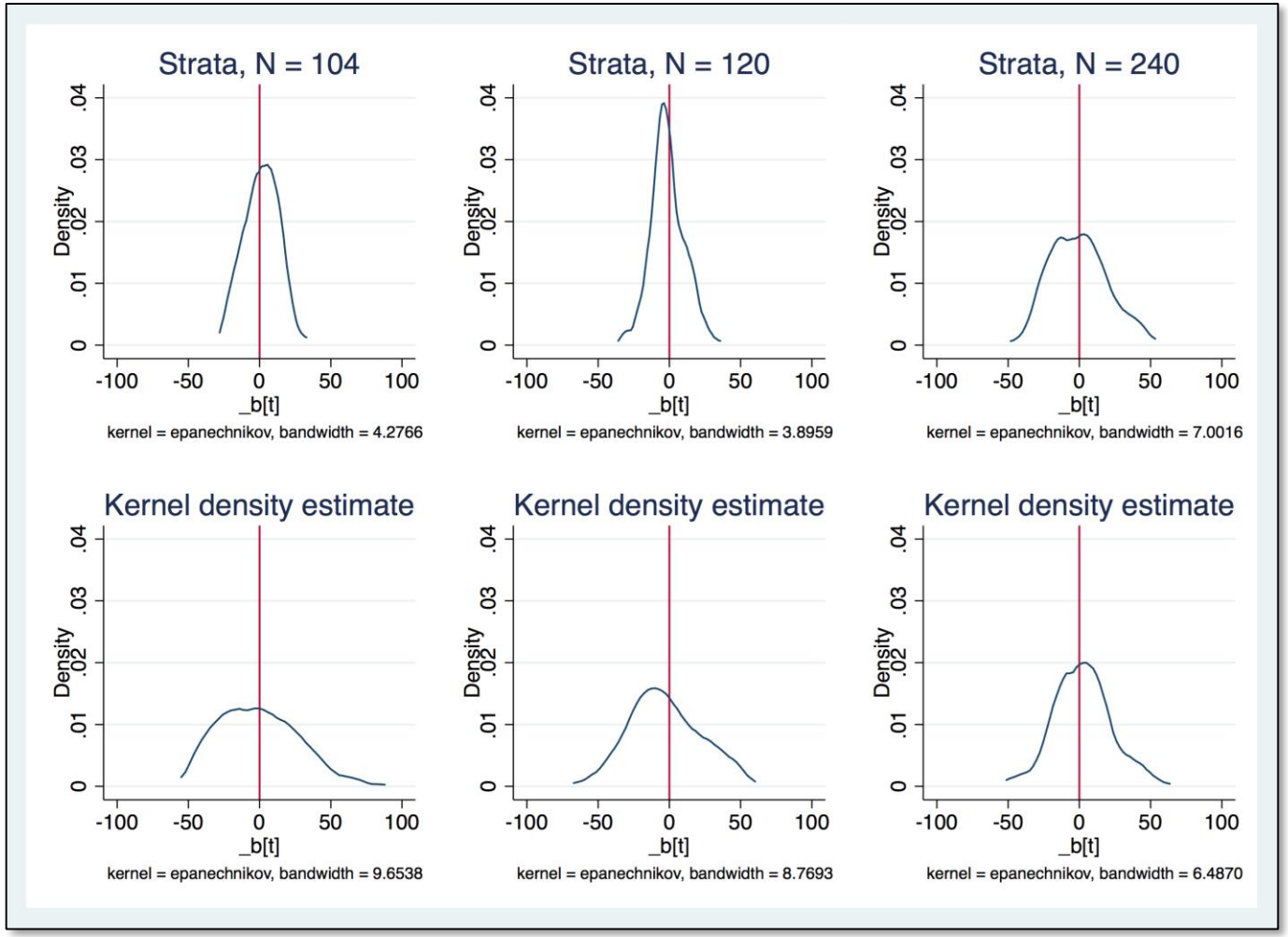
$\rightsquigarrow$ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

(King & Nielsen 2016)

Especially valuable when N is small:

Key idea:

Reducing the number of possible randomizations to those where the treatment and control are intentionally similar reduces the possible variation in estimate of the true treatment effect



Moderate sample size reduces the advantage, but it depends critically on underlying dispersion of outcomes (Bruhn and McKenzie 2009)

Types of Randomization

- Simple
  - Each individual person is assigned either Treatment or Control with some probability
- Cluster
  - Randomization happens on group level (village, school etc.) and all individuals in a group are assigned the same treatment
- Stratified
  - Sub-sets of similar observations (rich/poor, male/female etc.) are determined in advance, and randomization is done separately within each sub-set
  - Guarantees that equal number of similar observations (rich/poor, male/female etc.) are assigned to be each treatment and control
- Pairwise
  - Extreme form of stratification
  - All units are matched to make pairs that are as similar as possible and one unit from each pair is assigned to be T or C

Methods of randomization

- Good:
  - Field Based
  - Stata, especially via *programming* a randomization command
  - R, Python, or other replicable software

- Never use:
  - Excel – and other non-replicable software
  - Excel and many other software has random generators, but they do not allow a controlled randomization

Randomization: Field vs Computer

Field-based

- Public lottery, drawing numbers, etc.
- Pros
  - Ubiquitous, well-accepted
  - Can be done publicly thus transparent
- Cons
  - Not replicable
  - Can’t be done for large sample
  - Difficult to manage any a complex randomization with for example stratification

Computer-based

```
** Redo the stable sort. This is the third requirement
*  for a replicable randomization. (We do not need to
*  set verion and seed again)
sort student_id

** Start identical to the randomization above. Create
*  a random varaible and sort the observations on it.
gen      rand_multi = runiform()
sort     rand_multi

*Create the rank and tot_obs var. See above for exlinations.
gen rank_multi      = _n
gen tot_obs_multi   = _N

** Create a the treatment variable and assign a quarter
*  of the observations to each treatment.
gen      treatment_multi = 0                                  //Set all to 0
replace treatment_multi = 1 if rank_multi > 1 * tot_obs_multi/4 //Set the upper three quarters to 1
replace treatment_multi = 2 if rank_multi > 2 * tot_obs_multi/4 //Set the upper half to 2
replace treatment_multi = 3 if rank_multi > 3 * tot_obs_multi/4 //Set the upper quarter to 3
```

- STATA, R, Python, etc.
- Pros
  - Fully replicable
  - Easy to set up complex randomizations
- Cons
  - Might seem to beneficiaries and project staff as if it is NOT transparent

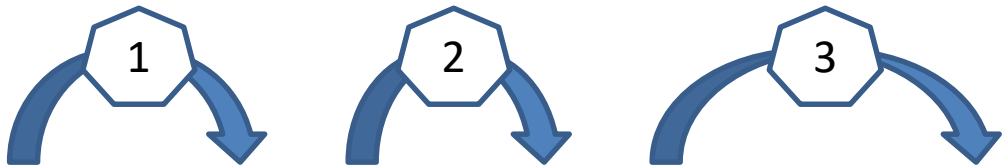
Prepare randomization in Stata

- Define randomization design
  - Unit of randomization (cluster)
  - Number & size of treatment arms
  - Stratification
  - Variables of interest to test balance
- Obtain a list of observations to be randomized
- Randomize and document using Stata!

Ex 1: Basic Randomization in Stata

- We have 10 students and we want half of them to be treatment and control
- What is our randomization rule?
- How do we randomize in Stata such that it is random but replicable?

Simplified example of randomization



HH_ID
AAAA
BBBB
CCCC
DDDD
EEEE
FFFF
GGGG
HHHH
IIII
JJJJ

HH_ID	rand
AAAA	.10
BBBB	.50
CCCC	1.0
DDDD	.20
EEEE	.80
FFFF	.90
GGGG	.70
HHHH	.40
IIII	.60
JJJJ	.30

HH_ID	rand
AAAA	.10
DDDD	.20
JJJJ	.30
HHHH	.40
BBBB	.50
IIII	.60
GGGG	.70
EEEE	.80
FFFF	.90
CCCC	1.0

HH_ID	rand	tmt
AAAA	.10	0
DDDD	.20	0
JJJJ	.30	0
HHHH	.40	0
BBBB	.50	0
IIII	.60	1
GGGG	.70	1
EEEE	.80	1
FFFF	.90	1
CCCC	1.0	1

1. Start with a list of observations you want to randomize. Generate random numbers.

2. Sort the observations after these random number. The order of the observations are now randomly sorted.

3. Assign 0 (control) to the first half of the observations, and assign 1 (treatment) to the second half.

The 3 rules of replicable randomization

- We should make our randomization to be replicable and get the same results for research transparency

- Three rules of replicable randomization in Stata:
  1. Same Stata version
  2. Same seed
  3. Sample ordered in same way

- The next slides explains the meaning of these rules and why it matters

Rule 1: Same Stata version

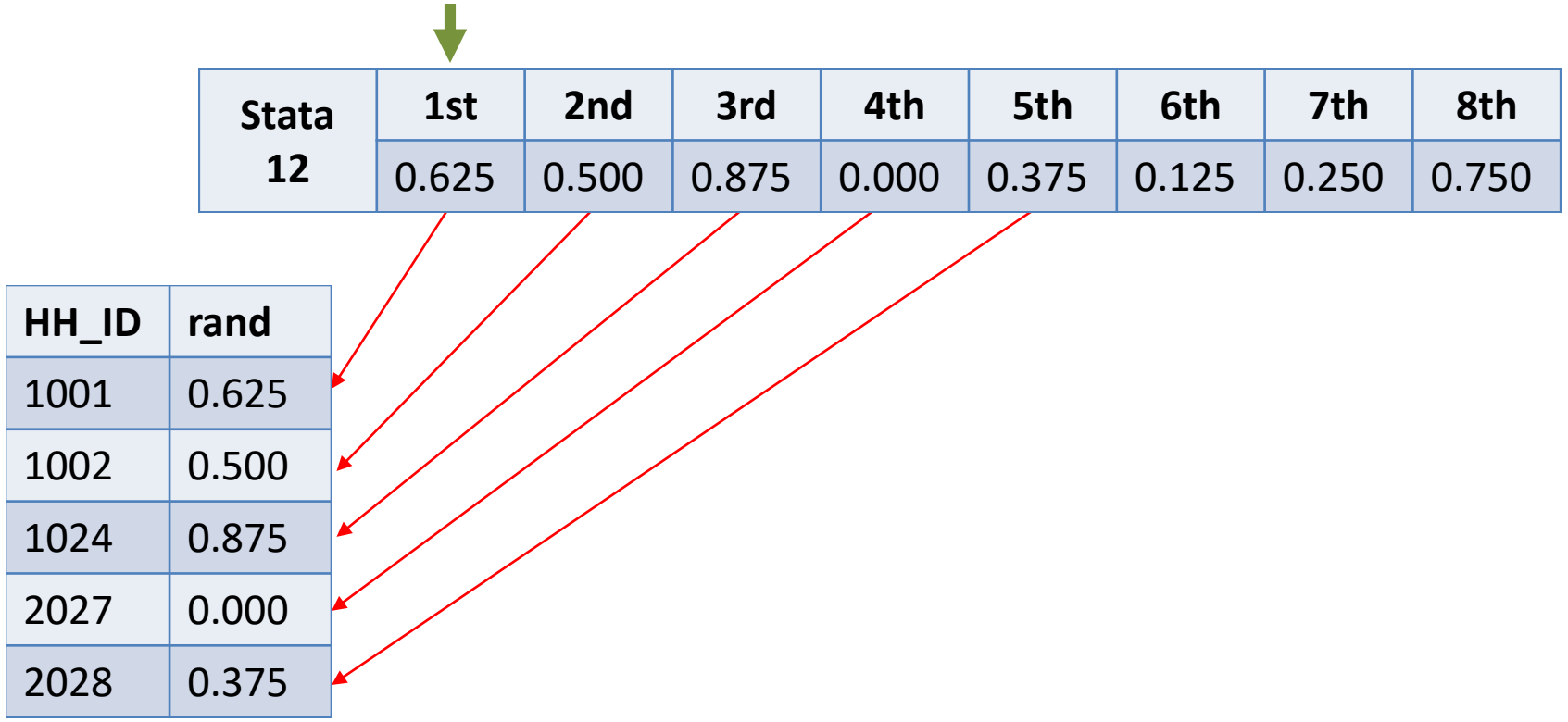
Stata 11	1st	2nd	3rd	4th	5th	6th	7th	8th
	0.625	0.000	0.125	0.375	0.250	0.875	0.500	0.750

Stata 12	1st	2nd	3rd	4th	5th	6th	7th	8th
	0.625	0.500	0.875	0.000	0.375	0.125	0.250	0.750

Stata 13	1st	2nd	3rd	4th	5th	6th	7th	8th
	0.250	0.625	0.125	0.750	0.875	0.500	0.000	0.375

- Stata has pre-calculated list of random numbers. However, each Stata version has a different list of random numbers
- Any list is equally good for our randomization, but we need to use only one list to get the same result. You can pick one by specifying which Stata version to use in randomization (older version can be used, but not newer)
- In reality these lists are billions of items long, instead of 8 as in the example above

Rule 1: Same Stata version



Stata 12	1st	2nd	3rd	4th	5th	6th	7th	8th
	0.625	0.500	0.875	0.000	0.375	0.125	0.250	0.750

HH_ID	rand
1001	0.625
1002	0.500
1024	0.875
2027	0.000
2028	0.375

- Stata goes through the lists and assigns the 1st value to the first observation, 2nd to the second observation, etc.

Rule 2: Same seed

Stata 12	1st	2nd	3rd	4th	5th	6th	7th	8th
	0.625	0.500	0.875	0.000	0.375	0.125	0.250	0.750

HH_ID	rand
1001	0.375
1002	0.125
1024	0.250
2027	0.750
2028	0.625

- Setting the seed change the starting place in the list
- Stata has its default seed at the time of launching
- You should set the seed to certain number to make it replicable as well as unique to your study.

Rule 3: Sample ordered in same way

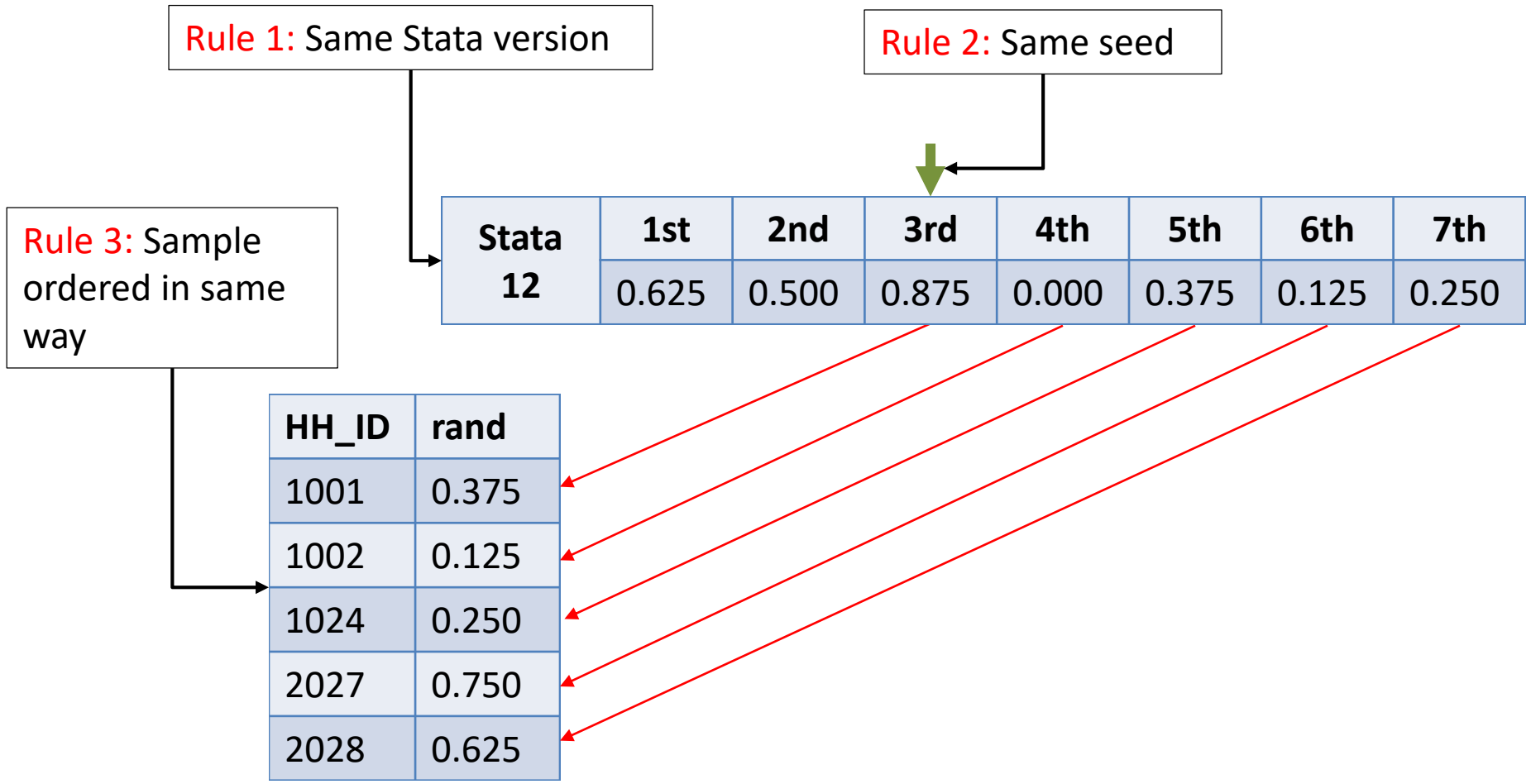
Stata 12	1st	2nd	3rd	4th	5th	6th	7th	8th
	0.625	0.500	0.875	0.000	0.375	0.125	0.250	0.750

Ascending sort	
HH_ID	**rand**
1001	0.875
1002	0.000
1003	0.375

Descending sort	
HH_ID	**rand**
1003	0.875
1002	0.000
1001	0.375

- The two data sets to the left are sorted differently, and while the same random numbers were picked, the result of the randomization is different.
- Adding and removing observations can change the sort order

The 3 rules of replicable randomization



3 rules of replicable randomization

1. Same Stata version
   - Guarantees the same list
   - See commands [*ieboilstart*] or [*version*] in Stata
2. Same seed
   - Guarantees the same starting point in the list
   - Choose a random number outside Stata (ex. random.org)
   - Large number (6-digit or longer) recommended
   - See command [*set seed*].
3. Sample ordered in same way
   - Guarantees that the same observation gets the same random number from the list
   - Sort the data in a way that will remain constant even if someone else change the sort order of the data set you are using (use [*isid varlist, sort*], NOT [*sort varlist, stable*])
   - Be aware that changing the data set, adding or removing observations, changes the sort order and therefore the entire randomization!

Basic Randomization in Stata

- We have 100 students and we want half of them to be treatment and control
- What are our randomization rules?
- Let’s see an example of a replicable randomization under the given sample and design

Three Rules of Randomization in Stata

Set version →

Set seed →

Unique sort →

```
* Setup

    clear all
    ieboilstart, v(12.0)
    `r(version)'

     di `c(version)'

     set seed 580917 // ← generated on Random.org at Timestamp: 2018-06-09 22:20:43 UTC

* Load the data

    use "${Lab5_dtInt}/hh_roster.dta", clear

        isid hh_id mem_id , sort
```

Basic Randomization in Stata

Rule 3: Sort sample

Generate random number and sort

Variables used in assignment

Assign to treatment if rank is less then half

Label the variable

```
** Stable sort. This is the third rule for a
*  replicable randomization.
sort student_id

** Generate a variable with a random number for all
*  observations and sort the observations after that
*  number.
gen      rand = runiform()
sort     rand

** Create one variable with the rank on the random
*  number. And a varaible with the total number of
*  observations.
gen rank    = _n
gen tot_obs = _N

** Create the treatment variable. Change the value
*  to 1 if the rank is more than half the number
*  of total observations in the data set.
gen     treatment = 0
replace treatment = 1 if rank > tot_obs/2

*Create a label docuementing the treatment variable
label define            treat_lab 0 "Control" 1 "Treatment"
label values treatment  treat_lab
```

Multi-arm randomization in Stata

- Multiple treatment arms; we have 100 students and we want one quarter of them to be control and three treatment arms with one quarter in each

```
** Redo the stable sort. This is the third requirement
*  for a replicable randomization. (We do not need to
*  set verion and seed again)
sort student_id

** Start identical to the randomization above. Create
*  a random varaible and sort the observations on it.
gen     rand_multi = runiform()
sort    rand_multi

*Create the rank and tot_obs var. See above for exlinations.
gen rank_multi      = _n
gen tot_obs_multi   = _N

** Create a the treatment variable and assign a quarter
*  of the observations to each treatment.
gen     treatment_multi = 0                                     //Set all to 0
replace treatment_multi = 1 if rank_multi > 1 * tot_obs_multi/4 //Set the upper three quarters to 1
replace treatment_multi = 2 if rank_multi > 2 * tot_obs_multi/4 //Set the upper half to 2
replace treatment_multi = 3 if rank_multi > 3 * tot_obs_multi/4 //Set the upper quarter to 3
```

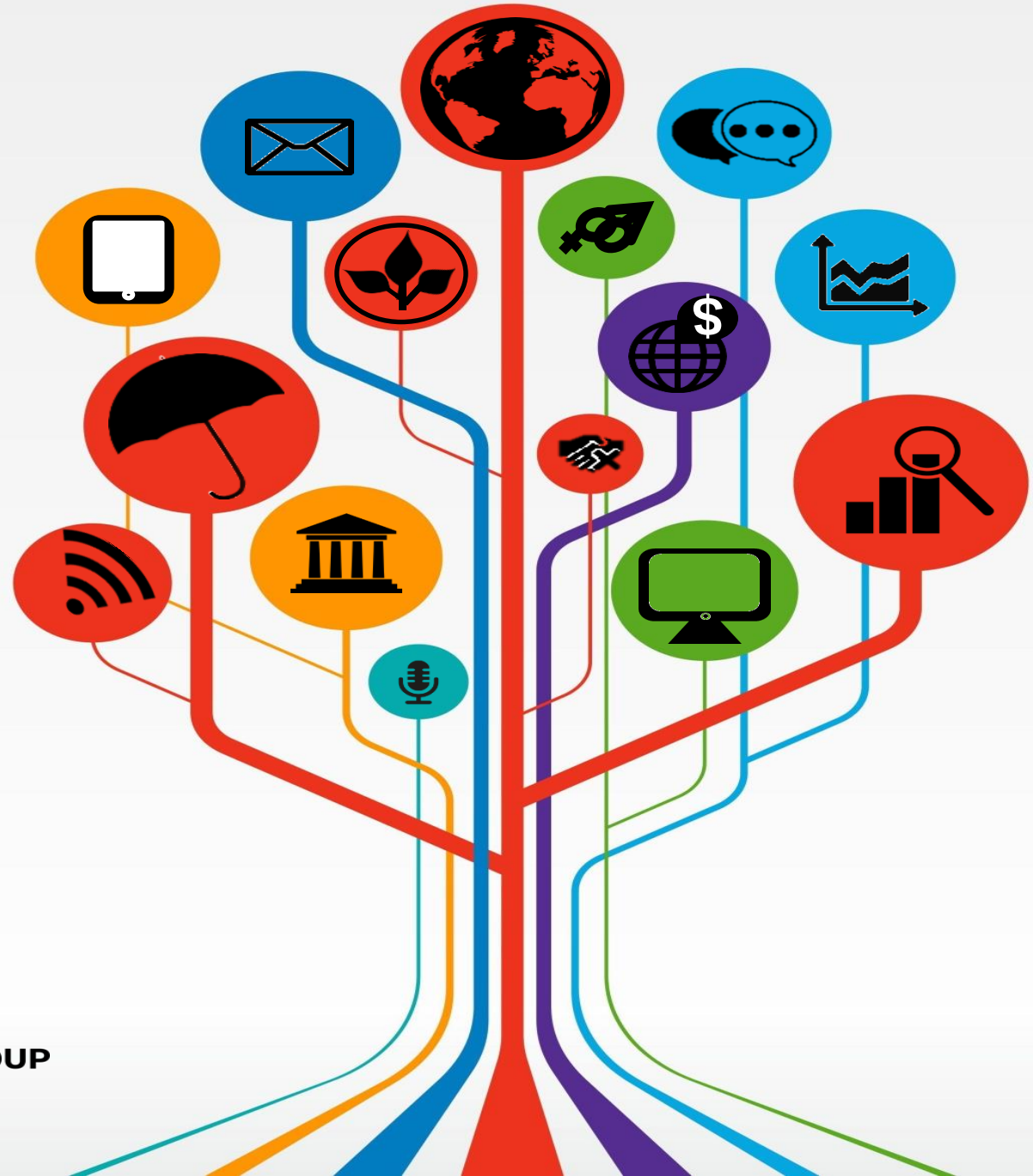
Exercise will show you how to use [*xtile*] for these steps

- Similar to Ex 2, but we assign the students to 4 groups

Other common randomization

- There are other issues which make randomization more complicated, such as
  - Groups of different sizes: For example, assigning 40% to control and 30% to each of the two treatment arms.
  - Clustering: Randomization is assigned to “clusters” of units at once, such as a household-level treatment
  - Stratification: Randomization within each subgroup in sample: male/female, rich/poor, regions, etc.
  - Odd total sample size: The number of observations is not evenly divisible with the number of treatment arms in each strata.
- What kind of code do these require to be random and reproducible?
- [*randtreat*] is a user-written Stata command which can help handle all the issues above. We will practice this command in the exercise.

HOW TO RUN A RANDOMIZATION CEREMONY



i2i DIME
TRANSFORM DEVELOPMENT

WORLD BANK GROUP

Stratified Randomization

- Ensures randomization achieves a balance re: several dimensions
- **Example:** draw for the World Cup: 32 qualified countries grouped into 4 urns / strata according to FIFA ranking and assigned to 8 different groups

FIFA WORLD CUP
RUSSIA 2018

But first . . . Cluster or stratum?

- **Cluster:** Units that are grouped and may share similar characteristics; sometimes used as the level of randomization (i.e., treatment assigned v. individual students).
  - The advantage of cluster randomization is that it reduces the likelihood of treatment contamination, but requires more participants to achieve the same statistical power.
- **Strata:** Groups made up of individuals sharing the same characteristics (i.e., females, males) from which a random sample is drawn.
  - Ensures balance across treatment groups with respect to stratifying variables. Makes it possible to draw inferences about outcomes within each group.

Source: Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. 2016. Impact Evaluation in Practice, second edition. Washington, DC: Inter-American Development Bank and World Bank.

Examples of clusters

Table 15.1 Examples of Clusters

Benefit	Level at which benefits are assigned (cluster)	Unit at which outcome is measured
Cash transfers	Village	Households
Malaria treatment	School	Individuals
Training program	Neighborhood	Individuals

Source: Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. 2016. Impact Evaluation in Practice, second edition. Washington, DC: Inter-American Development Bank and World Bank.

Stratified Randomization

BOMBOS

POTE 1	POTE 2	POTE 3	POTE 4
Rusia	España	Dinamarca	Serbia
Alemania	Perú	Islandia	Nigeria
Brasil	Suiza	Costa Rica	Australia
Portugal	Inglaterra	Suecia	Japón
Argentina	Colombia	Túnez	Marruecos
Bélgica	México	Egipto	Panamá
Polonia	Uruguay	Senegal	República de Corea
Francia	Croacia	RI de Irán	Arabia Saudí

Stratified Randomization

Guiding Principles

(Bruhn and McKenzie,2009)

- Maximum number of strata depends on the sample size, the expected size of each strata and the importance of the stratification factors.
- According to simulations, between 22 and 48 strata for a sample size of 100 seems appropriate (individual level randomization).
- Select covariates that are strongly related to the outcomes of interest, such as geographic region dummies, and variables for which subgroup analysis is desired.

Example: Savings, Grants, Training

- In the context of new rural roads, we study the impact of complementary interventions for increasing productivity:
- 1) mechanisms for facilitating access to capital for investment . . .

1. Productive investment grants
USD 200

2. Simple saving accounts

3. Savings accounts + matched grants of up to USD 200 (constant)

- 2) developing the soft skills necessary to be a successful entrepreneur.

4. Individual aspirational training

5. Simple bank accounts + matched grants + individual aspirational training

Example: Savings, Grants, Training

- We want to randomize into the following 6 arms:

204 HH	815 HH	296 HH	342 HH	350 HH	795 HH
Treatment 1 Productive investment grant	Treatment 2a Savings account for productive investment	Treatment 2b Matched saving grant	Treatment 3 Aspirational training delivered to the household head	Treatment 4 Aspirational training delivered to the household head and matched grant	Control

- Which variables should we stratify by?
  - Proximity to a rehabilitated road (yes, no)
  - Income (high, medium, low)
  - Gender of HH head (male, female)
  - Municipality (10 different municipalities)
- How many strata do we have?

Why is this important?

- **Transparency, transparency, transparency.**
- There can be no doubt as to whether treatment was assigned to one household in lieu of another due to favoritism.
- Especially, when benefits include something like cash.
- This procedure is designed to be conducted in the company of government counterparts.
- Conducting the randomization ceremony and documenting it ensure there is no doubt as to whether assignment was random.

Why is this important?

Selección del beneficiario: El entrenador explicará al beneficiario que ha sido seleccionado sobre la base de muchos productores y que esta acción traerá beneficios para él y su familia.

> **ENTRENADOR:** *Si usted se acuerda, usted participó en una encuesta en el año 2015 en el contexto de una investigación del MTI. Ha sido seleccionado al azar para participar en un proceso de capacitación con la idea de mejorar su conocimiento para hacer negocios, producir mejor e incrementar sus ganancias. Ha sido seleccionado de varias personas, lo cual es motivo de regocijo para todos.*

- During visits to offer the training, we made it extremely clear that participants were chosen randomly.

Randomization Procedure

- Assignment of 2,802 households to the 5 treatment groups and the control group
- Each HH has the following probabilities of being selected to receive each treatment:
  - T1: Productive bonus: 200/2802 (7.1%) *(= 0)*
  - T2: Simple savings account: 800 / 2,802 (28.6%) *(= 1)*
  - T3: Account plus paired subsidy: 300 / 2,802 (10.7%) *(= 2)*
  - T4: Training "personal initiative": 350 / 2,802 (12.5%) *(= 3)*
  - T5: Account plus grant + training: 350 / 2,802 (12.5%) *(= 4)*
  - T6: Control: 800 / 2,802 (28.6%) *(= 5)*
- The last two homes are assigned at random to T1-T6.
- Then we allocate 200/301 other homes "just in case."

Background

- Random number tables have been used for a long time. Nowadays, **computational** random number generators are employed.
- Stata's random-number generation functions, such as *uniform(),* follow deterministic algorithms that are pseudorandom-number functions.
- The sequences these functions produce are determined by the *seed*.
- **Different seeds produce different sequences**

Example: unsorted

	id	strata	random_num	ordering
1	1070	1	.2201574	3
2	1077	1	.8268973	11
3	1080	1	.6892915	8
4	1085	1	.9497825	14
5	1088	1	.9280429	13
6	1098	1	.9141315	12
7	1099	1	.144082	2
8	1101	1	.6358364	4
9	1102	1	.6706831	7
10	1103	1	.0067598	1
11	1105	1	.737991	10
12	1106	1	.6626937	6

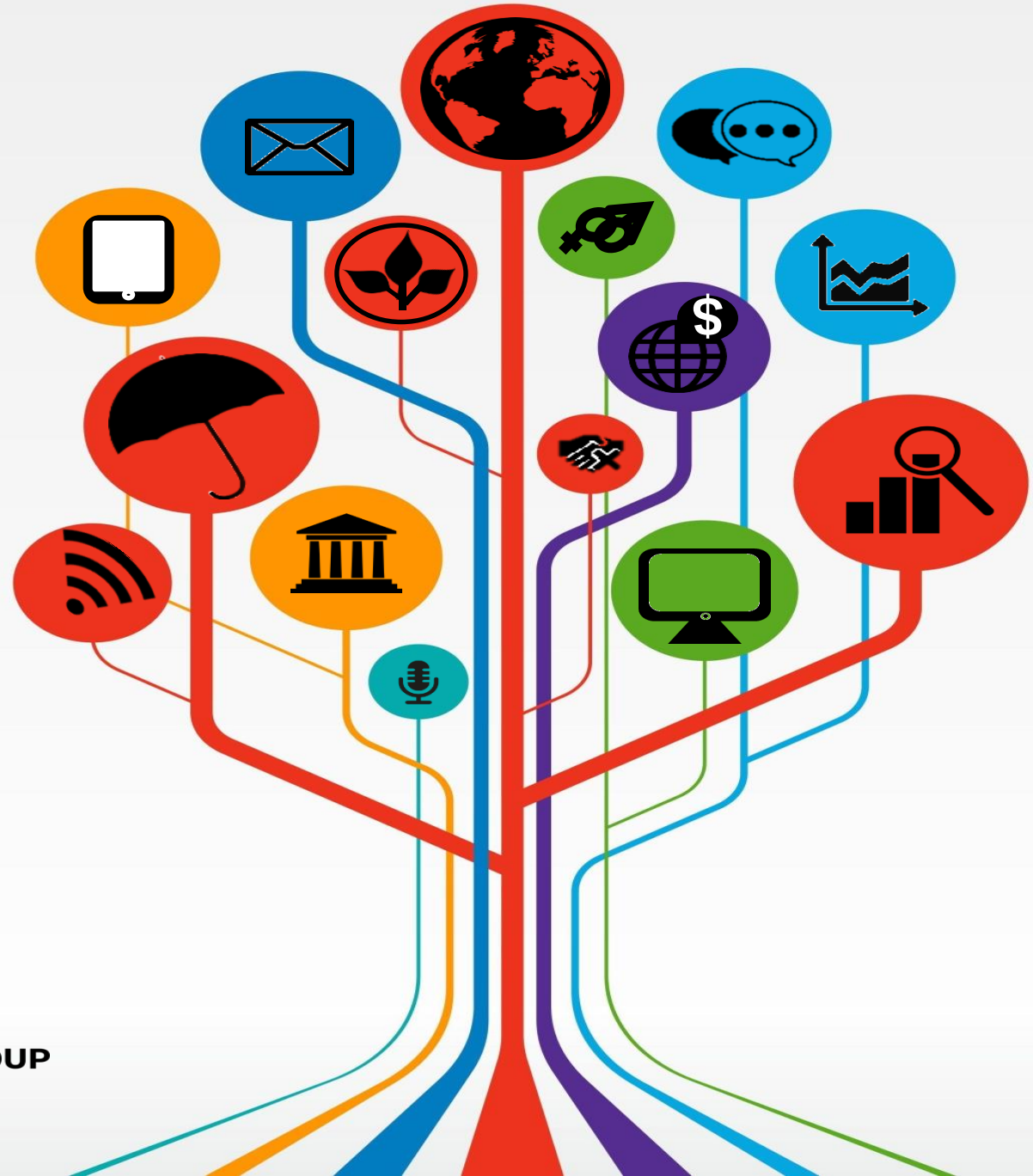
Example: HHs ordered according to *ordering*

	id	strata	random_num	ordering	treatment
1	1103	1	.0067598	1	T
2	1099	1	.144082	2	T
3	1070	1	.2201574	3	T
4	1101	1	.6358364	4	T
5	1115	1	.6527964	5	T
6	1106	1	.6626937	6	T
7	1102	1	.6706831	7	T
8	1080	1	.6892915	8	C
9	1112	1	.7302711	9	C
10	1105	1	.737991	10	C
11	1077	1	.8268973	11	C
12	1098	1	.9141315	12	C

Randomization in Practice

- We randomly draw the seed three times:
  1. Test #1
  2. Test #2
  3. Actual Randomization
- Show that each time randomization outcome is different one from the next.
- To make this easy, we will each time show results for a single household and compare afterwards.

Thank you!



i2i DIME
TRANSFORM DEVELOPMENT

WORLD BANK GROUP