
EXERCISE B: HOW TO DO RANDOM ASSIGNMENT USING MS EXCEL

TABLE OF CONTENTS

Introduction.....	1
Part 1: Simple Randomization.....	1
Part 2: Complete randomization.....	5
Part 3: stratified randomization.....	8

INTRODUCTION

Like most spreadsheet programs, MS Excel can generate random numbers on command. MS Excel has two native random-number-generating functions. The first, =RAND(), creates a *continuous* random number between 0 and 1—it could be any number of 9 decimal places between 0 and 1. The second, =RANDBETWEEN(*bottom*, *top*) creates *integers* between any two integer values within a range, where you specify the *bottom* and *top* of that range.

PART 1: SIMPLE RANDOMIZATION

Say we had a list of schools and we wanted to assign them to treatment or control based on a coin flip (heads = treatment and tails = control). We can do this by randomly generating the value of 0 or 1 using the RANDBETWEEN function, and choosing 0 and 1 as the range. We could then assign all schools with 0 to the control group, and all schools with 1 to the treatment group (or vice versa). This is equivalent to a coin flip where 0 represents tails and 1 represents control. Equivalently, we could produce a *continuous* random number for each observation and assign those with (say) random number greater than or equal to 0.5 to treatment and smaller than 0.5 to control.

The illustration below shows how to do this step-by-step.

We have a list of all schools

	A	B	C	D	E
1	SchoolID	SchoolName	Pre-test score	Random #	T-C
2	101	Babajipura G.M.M.Kumar shala No. 1	34.12		
3	103	Babajipura Kanya Shala No. 3	25.49		
4	107	Babajipura Mishra Shala No. 7	12.60		
5	108	Babajipura Mishra Shala No. 8	20.77		
6	112	Babajipura Marathi Mishra Shala No. 12	37.69		
7	113	Babajipura Kanya Shala No. 13	32.71		
8	114	Babajipura Mishra Shala No. 14	16.32		
9	117	Babajipura Kumar Shala No. 17	20.12		
10	118	Babajipura Mishra Shala No. 18	28.05		
11	119	Babajipura Mishra Shala No. 19	21.29		
12	120	Babajipura Mishra Shala No. 20	26.34		
13	121	Babajipura Mishra Shala No. 21	16.36		
14	125	Babajipura Kumar Shala No. 25	21.32		
15	126	Babajipura Kanya Shala No. 26	25.25		
16	127	Babajipura Mishra Shala No. 27	26.43		
17	128	Babajipura Mishra Shala No. 28	29.38		
18	130	Babajipura Hindi Mishra Shala No. 30	18.21		
19	131	Babajipura Mishra Shala No. 31	20.70		
20	132	Babajipura Mishra Shala No. 32	34.72		
21	201	Fatehpura Kumar Shala No. 1	30.04		
22	202	Fatehpura Mishra Shala No. 2	19.53		
23	209	Fatehpura Mishra Shala No. 9	25.63		
24	210	Fatehpura Kanya Shala No. 10	18.96		
25	211	Fatehpura Mishra Shala No. 11	21.11		
26	213	Fatehpura Kumar Shala No. 13	18.09		
27	215	Fatehpura Hindi Mishra Shala No. 15	23.27		
28	216	Fatehpura Mishra Shala No. 16	22.74		
29	218	Fatehpura Mishra Shala No. 18	15.08		
30	219	Fatehpura Mishra Shala No. 19	25.37		
31	301	N. Sayajiganj Mishra Shala No. 1 (center)	18.27		
32	303	N. Sayajiganj Marathi Mishra Shala No. 3	31.90		
33	305	Sayajiganj Mishra Shala No. 5	19.00		
34	306	Sayajiganj Kumar Shala No. 6	20.81		
35	307	Sayajiganj Mishra Shala No. 7	47.18		

Step 1: Assign a random number to each school

The function RAND () is Excel's basic random number generator. To use it, go to Column D and type

=RAND()

in each cell, adjacent to each school name. Or you can type this function in the top row (row 2) and simply copy and paste to the entire column, or click and drag.

	A	B	C	D	E
1	SchoolID	SchoolName	Pre-test score	Random #	T-C
2	101	Babajipura G.M.M.Kumar shala No. 1	34.12	=RAND()	
3	103	Babajipura Kanya Shala No. 3	25.49		
4	107	Babajipura Mishra Shala No. 7	12.60		
5	108	Babajipura Mishra Shala No. 8	20.77		
6	112	Babajipura Marathi Mishra Shala No. 12	37.69		
7	113	Babajipura Kanya Shala No. 13	32.71		
8	114	Babajipura Mishra Shala No. 14	16.32		
9	117	Babajipura Kumar Shala No. 17	20.12		

Typing =RAND() puts a 9-digit random number between 0 and 1 in the cell.

	A	B	C	D	E
1	SchoolID	SchoolName	Pre-test score	Random #	T-C
2	101	Babajipura G.M.M.Kumar shala No. 1	34.12	0.0789009	
3	103	Babajipura Kanya Shala No. 3	25.49	0.8999008	
4	107	Babajipura Mishra Shala No. 7	12.60	0.4359626	
5	108	Babajipura Mishra Shala No. 8	20.77	0.1408828	
6	112	Babajipura Marathi Mishra Shala No. 12	37.69	0.9634844	
7	113	Babajipura Kanya Shala No. 13	32.71	0.2145561	
8	114	Babajipura Mishra Shala No. 14	16.32	0.2558066	
9	117	Babajipura Kumar Shala No. 17	20.12	0.0169244	
10	118	Babajipura Mishra Shala No. 18	28.05	0.0655376	
11	119	Babajipura Mishra Shala No. 19	21.29	0.2724011	
12	120	Babajipura Mishra Shala No. 20	26.34	0.7489921	
13	121	Babajipura Mishra Shala No. 21	16.36	0.0268576	
14	125	Babajipura Kumar Shala No. 25	21.32	0.0661789	
15	126	Babajipura Kanya Shala No. 26	25.25	0.6946606	
16	127	Babajipura Mishra Shala No. 27	26.43	0.5000895	
17	128	Babajipura Mishra Shala No. 28	29.38	0.642025	
18	130	Babajipura Hindi Mishra Shala No. 30	18.21	0.8219122	
19	131	Babajipura Mishra Shala No. 31	20.70	0.7963628	
20	132	Babajipura Mishra Shala No. 32	34.72	0.5042257	
21	201	Fatehpura Kumar Shala No. 1	30.04	0.9492957	
22	202	Fatehpura Mishra Shala No. 2	19.53	0.9989293	
23	209	Fatehpura Mishra Shala No. 9	25.63	0.2719192	
24	210	Fatehpura Kanya Shala No. 10	18.96	0.5246963	
25	211	Fatehpura Mishra Shala No. 11	21.11	0.2142812	
26	213	Fatehpura Kumar Shala No. 13	18.09	0.6100928	
27	215	Fatehpura Hindi Mishra Shala No. 15	23.27	0.8909558	
28	216	Fatehpura Mishra Shala No. 16	22.74	0.2995547	
29	218	Fatehpura Mishra Shala No. 18	15.08	0.2206103	

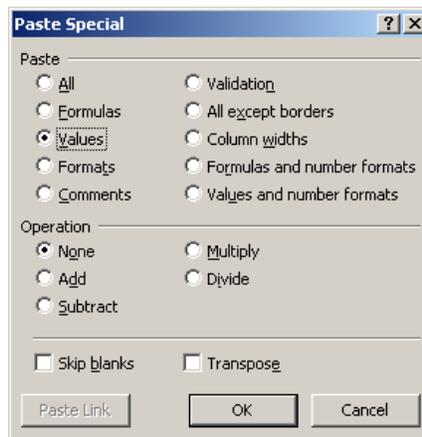
Step 2: Copy the cells in Column D, then paste the values over the same cells

The function =RAND() will re-randomize each time you make any changes to any other part of the spreadsheet. Excel does this because it recalculates all values with any change to any cell. (You can also induce recalculation, and hence re-randomization, by pressing the F9 key.)

Once we've generated our column of random numbers, we do not need to re-randomize. We already have a clean column of random values. To stop Excel from recalculating, you can replace the "functions" in this column with the "values".

To do this, highlight all values in Column D. Then right-click anywhere in the highlighted column, and choose "Copy".

Then, right-click anywhere in that column and choose "Paste Special." The "Paste Special" window will appear. Click on "Values".



Step 3: Assign treatment/control status for each group

Now use the IF function to assign schools to treatment and control. Go to column E and type

`=IF(D2>=0.5,"T","C")`

And click and drag (or copy and paste) to the rest of the column. This will enter a "T" for schools that have a random number greater than or equal to 0.5 and "C" for schools with random number less than 0.5.

	A	B	C	D	E	F	G
1	SchoolID	SchoolName	Pre-test score	Random #	T-C		
2	215	Fatehpura Hindi Mishra Shala No. 15	23.27	0.4280308	=IF(D2>=0.5,"T","C")		
3	511	Shaher Vibhag Mishra Shala No. 11	22.37	0.4977787	C		
4	341	Sayajiganj Kanya Shala No. 41	24.57	0.5068542	T		
5	632	Wadi Kanya Shala No. 32	30.52	0.6071675	T		
6	514	Shaher Vibhag Mishra Shala No. 14	25.29	0.9297094	T		
7	626	Wadi Hindi Mishra Shala No. 26	41.10	0.3811165	C		
8	345	Sayajiganj Mishra Shala No. 45	20.33	0.0250151	C		
9	210	Fatehpura Kanya Shala No. 10	18.96	0.3442701	C		
10	622	Wadi Mishra Shala No. 22	21.90	0.0106587	C		
11	101	Babajipura G.M.M.Kumar shala No. 1	34.12	0.8055242	T		
12	315	Sayajiganj Hindi Mishra Shala No. 15	28.60	0.9751691	T		
13	313	Sayajiganj Mishra Shala No. 13	27.39	0.9867175	T		

Your list of schools has now been randomly assigned to treatment and control!

Is the number of schools in in both groups the same? We also have the average pre-test scores for each school. Does the average pre-test score look balanced between the two groups?

	Treatment	Control	Difference
Number of Schools	63	59	-4
Average Pre-test Score	25.62	26.92	1.30

Note, however, that the number of schools in treatment and control will vary each time you re-randomize, as will the average pre-test score. To check this, repeat step 1, but this time instead of copy pasting values, press the F9 key to re-randomize. Re-randomize 10 times and see what happens to the number of schools and the average pre-test score in each group.

Does the number of schools change when you re-randomize? Does the average pre-test score look balanced every time you re-randomize?

Try the above steps using the RANDBETWEEN() function instead of the RAND() function. Do you expect significantly different results? How does the “IF” function change?

PART 2: COMPLETE RANDOMIZATION

Say we had a list of schools and wanted to assign exactly half of them to treatment and half to control

Step 1: Assign a random number to each school

Go to Column D and type:

```
=RAND()
```

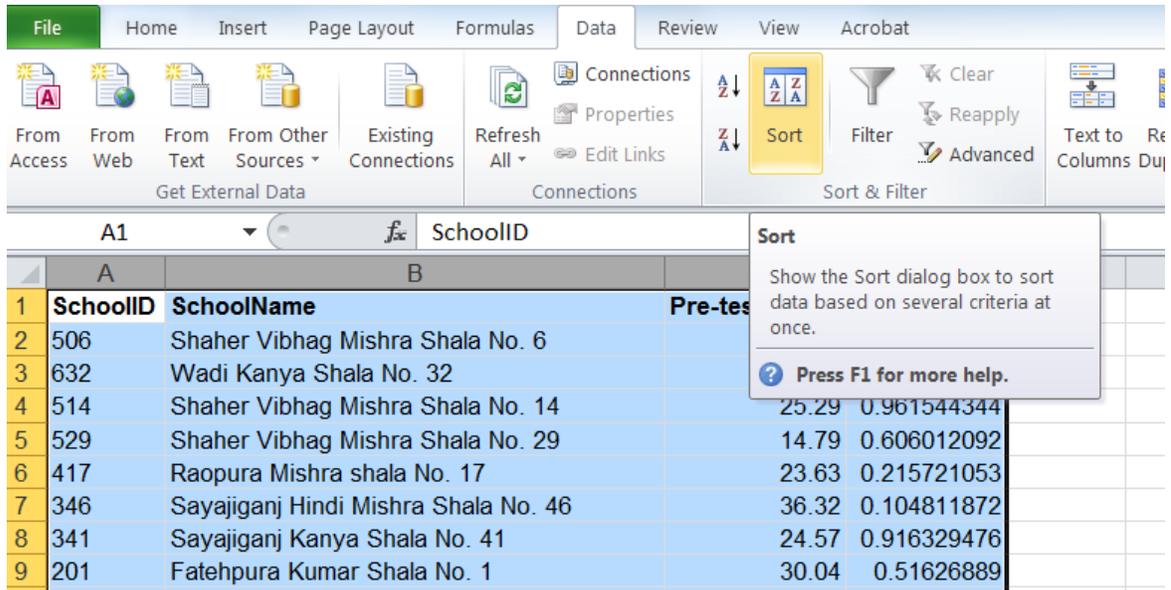
And click and drag (or copy and paste) to the entire column.

Step 2: Copy the cells in Column D, then paste the values over the same cells

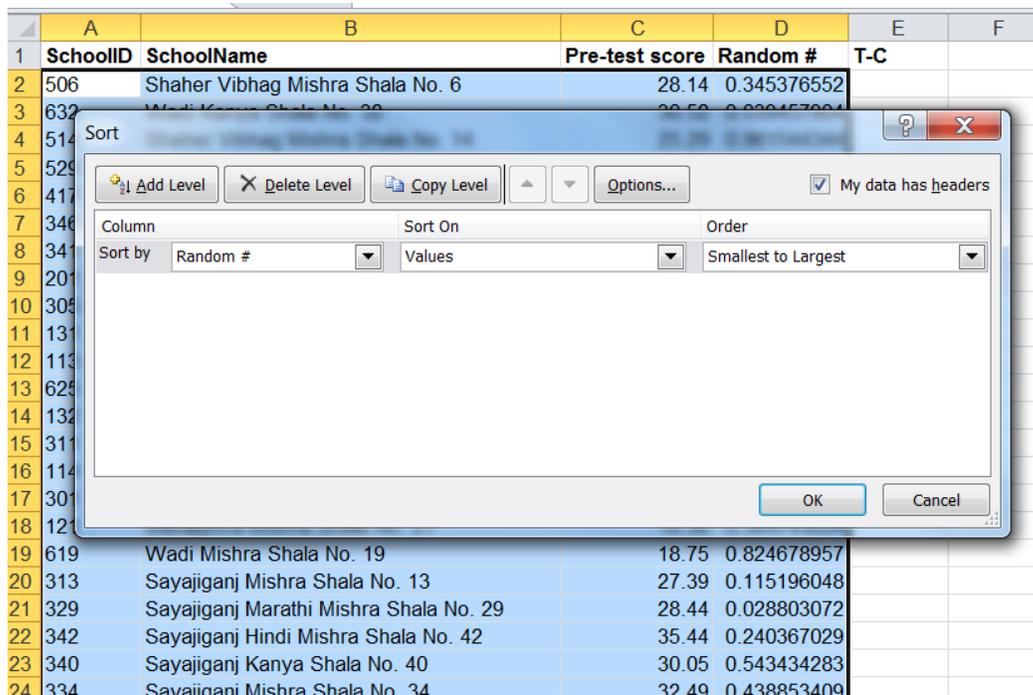
Highlight all values in Column D. Then right-click anywhere in the highlighted column, and choose “Copy”. Then, right-click anywhere in that column and choose “Paste Special.”

Step 3: Sort the columns in either descending or ascending order of Column D

Highlight columns A, B, C and D. In the data tab, press the “Sort” button:



A Sort box will pop up.



In the “Sort by” column, select “Random #.” Click OK. Doing this sorts the list by the random number in ascending or descending order, whichever you chose.

There! You have a randomly sorted list.

	A	B	C	D	E
1	SchoolID	SchoolName	Pre-test score	Random #	T-C
2	409	Raopura Marathi Mishra Shala No. 9	35.15	0.009954911	
3	323	Sayajiganj Mishra Shala No. 23	24.18	0.014434557	
4	510	Shaher Vibhag Mishra Shala No. 10	42.07	0.021591901	
5	529	Shaher Vibhag Mishra Shala No. 29	14.79	0.02215352	
6	611	Wadi Marathi Mishra Shala No. 11	36.31	0.027430569	
7	307	Sayajiganj Mishra Shala No. 7	47.18	0.041933781	
8	503	Shaher Vibhag Kanya Shala No. 3	22.15	0.043774186	
9	401	Raopura Mishra Shala No. 1	26.21	0.048703656	
10	326	Sayajiganj Hindi Mishra Shala No. 26	33.54	0.063108687	
11	411	Raopura Kumar Shala No. 11	20.62	0.06761041	
12	127	Babajipura Mishra Shala No. 27	26.43	0.073245278	
13	638	Wadi Hindi Mishra Shala No. 38	32.19	0.073990044	
14	119	Babajipura Mishra Shala No. 19	21.29	0.076187741	
15	132	Babajipura Mishra Shala No. 32	34.72	0.076497148	
16	322	Sayajiganj Mishra Shala No. 22	12.76	0.087452794	
17	128	Babajipura Mishra Shala No. 28	29.38	0.088030307	
18	618	Wadi Kumar Shala No. 18	23.77	0.093828679	
19	349	Sayajiganj Mishra Shala No. 49	40.16	0.10107943	
20	637	Wadi Mishra Shala No. 37	25.26	0.109629264	
21	311	Sayajiganj Mishra Shala No. 11	26.24	0.110998693	
22	501	Shaher Vibhag Mishra Shala No. 1	27.31	0.127617726	
23	639	Wadi Mishra Shala No. 39	20.49	0.127796284	
24	508	Shaher Vibhag Mishra Shala No. 8	22.19	0.134358453	
25	348	Sayajiganj Mishra Shala No. 48	28.23	0.140845208	
26	338	Sayajiganj Kanya Shala No. 38	29.22	0.144582844	
27	617	Wadi Marathi Mishra Shala No. 17	35.67	0.172061028	
28	213	Fatehpura Kumar Shala No. 13	18.09	0.17400346	
29	314	Sayajiganj Kumar Shala No. 14	22.93	0.182726341	
30	623	Wadi Kanya Shala No. 23	28.91	0.186591564	
31	624	Wadi Mishra Shala No. 24	21.52	0.208067391	
32	344	Sayajiganj Mishra Shala No. 44	25.91	0.209175575	
33	633	Wadi Kanya Shala No. 33	29.51	0.227463469	

Because your list is randomly sorted, it is completely random whether schools are in the top half of the list, or the bottom half. Therefore, if you assign the top half to the treatment group and the bottom half to the control group, your schools have been “randomly assigned.”

Step 4: Assign treatment/control status for each group

There are two ways to do this. To do this manually, in column E, type “T” for the first half of the rows (rows 2–63) and for the second half of the rows (rows 62–123), type “C”. You can also do this by using the IF and MEDIAN functions. In Column E type:

```
=IF(D2<=MEDIAN($D$2:$D$123),"T","C")
```

And click and drag (or copy and paste) to the entire column. This will enter a “T” for schools that are below or at the median of the random number and a “C” for schools that are above it.

	A	B	C	D	E	F	G
1	SchoolID	SchoolName	Pre-test score	Random #	T-C		
2	409	Raopura Marathi Mishra Shala No. 9	35.15	0.009954911	=IF(ROW()<=62,"T","C")		
3	323	Sayajiganj Mishra Shala No. 23	24.18	0.014434557			
4	510	Shaher Vibhag Mishra Shala No. 10	42.07	0.021591901			
5	529	Shaher Vibhag Mishra Shala No. 29	14.79	0.02215352			
6	611	Wadi Marathi Mishra Shala No. 11	36.31	0.027430569			
7	307	Sayajiganj Mishra Shala No. 7	47.18	0.041933781			
8	503	Shaher Vibhag Kanya Shala No. 3	22.15	0.043774186			
9	401	Raopura Mishra Shala No. 1	26.21	0.048703656			
10	326	Sayajiganj Hindi Mishra Shala No. 26	33.54	0.063108687			
11	411	Raopura Kumar Shala No. 11	20.62	0.06761041			
12	127	Babajipura Mishra Shala No. 27	26.43	0.073245278			
13	638	Wadi Hindi Mishra Shala No. 38	32.19	0.073990044			
14	119	Babajipura Mishra Shala No. 19	21.29	0.076187741			
15	132	Babajipura Mishra Shala No. 32	34.72	0.076497148			
16	322	Sayajiganj Mishra Shala No. 22	12.76	0.087452794			
17	128	Babajipura Mishra Shala No. 28	29.38	0.088030307			
18	618	Wadi Kumar Shala No. 18	23.77	0.093828679			
19	349	Sayajiganj Mishra Shala No. 49	40.16	0.10107943			
20	637	Wadi Mishra Shala No. 37	25.26	0.109629264			
21	311	Sayajiganj Mishra Shala No. 11	26.24	0.110998693			
22	501	Shaher Vibhag Mishra Shala No. 1	27.31	0.127617726			
23	639	Wadi Mishra Shala No. 39	20.49	0.127796284			
24	508	Shaher Vibhag Mishra Shala No. 8	22.19	0.134358453			
25	348	Sayajiganj Mishra Shala No. 48	28.23	0.140845208			

Now select columns A through E and re-sort your list back in order of “SchoolID.” You’ll see that your schools have been randomly assigned to treatment and control groups.

Is the number of schools in both groups the same? Does the average pre-test score look balanced between the two groups?

Note that the number of schools in treatment and control will remain the same each time you re-randomize. This is because you are making sure that you always assign half of them to treatment and half to control. To check this, repeat step 1, but this time instead of copy pasting values, press the F9 key to re-randomize. Notice that the formula in column E will automatically recalculate the median each time and re-assign treatment and control status. Re-randomize 10 times and see what happens to the number of schools and the average pre-test score in each group.

Does the number of schools change when you re-randomize? Does the average pre-test score look balanced every time you re-randomize?

PART 3: STRATIFIED RANDOMIZATION

Stratification is the process of dividing a sample into groups, and then randomly assigning individuals within each group to the treatment and control. The reasons for doing this are rather technical. One reason for stratifying is that it ensures subgroups are balanced, making it easier to perform certain subgroup analyses. For example, if you want to test the effectiveness on a new education program separately for schools where children are taught in Hindi versus schools where children are taught in Gujarati, you can stratify by “language of instruction” and ensure that there are an equal number of schools of each language type in the treatment and control groups.

We have our list of schools and potential “strata”

Mechanically, the only difference in random sorting is that instead of simply sorting by the random number, you would first sort by language, and then the random number. Obviously, the first step is to ensure you have the variables by which you hope to stratify.

Step 1: Assign a random number to each school

Go to Column F and type:

```
=RAND()
```

And click and drag (or copy and paste) to the entire column.

Step 2: Copy the cells in Column F, then paste the values over the same cells

Highlight all values in Column F. Then right-click anywhere in the highlighted column, and choose “Copy”. Then, right-click anywhere in that column and choose “Paste Special.”

Step 3: Sort by strata and then by random number

Assuming you have all the variables you need, you can now click “Sort” in the data tab. The Sort window will pop up. Sort by “Language.” Press the “Add Level” button. Then select “Random #”.

	A	B	C	D	E	F	G
1	SchoolID	SchoolName	Pre-test score	Language	Gender	Random #	T-C
2	603	Wadi Mishra Shala No. 3	14.76	Gujarati	Co-ed	0.069825549	
3	625	Wadi Kumar Shala No. 25	28.90	Gujarati	Boys	0.32824482	
4	107						605
5	349						508
6	425						405
7	308						966
8	502						479
9	624						304
10	317						088
11	524						532
12	405						123
13	632						909
14	619						627
15	639						043
16	506						528
17	118						568
18	602						671
19	131						886
20	408	Raopura Kanya Shala No. 8	25.03	Gujarati	Girls	0.836292353	
21	316	Sayajiganj Kanya Shala No. 16	25.08	Gujarati	Girls	0.570278699	
22	417	Raopura Mishra shala No. 17	23.63	Gujarati	Co-ed	0.730099722	
23	601	Wadi Mishra Shala No. 1	36.44	Gujarati	Co-ed	0.085829273	
24	529	Shaheer Vibhag Mishra Shala No. 29	14.79	Gujarati	Co-ed	0.496626789	
25	637	Wadi Mishra Shala No. 37	25.26	Gujarati	Co-ed	0.903677425	
26	128	Bahainura Mishra Shala No. 28	29.38	Gujarati	Co-ed	0.285842087	

Step 4: Assign treatment/control status for each group

There are two ways to do this. To do this manually, in column G, within each languages category, type “T” for the first half of the rows, and “C” for the second half. You can also do this by using the IF and MEDIAN functions. In Column G type and instead of just hitting ENTER, hit CTRL+SHIFT+ENTER in order to tell Excel that it’s an array formula:

```
=IF(F2<MEDIAN(IF($D$2:$D$123=D2,$F$2:$F$123)), "T", "C")
```

Click and drag (or copy and paste) to the entire column. This will enter a “T” for schools that are below or at the median of the random number and a “C” for schools that are above it for *each* language category.

Is the total number of schools in both groups the same? Is the number of schools for each language category for both groups the same? Does the average pre-test score look balanced between the two groups?

Note that the total number of schools and the number of schools for each language category in treatment and control will remain the same each time you re-randomize. To check this, repeat step 1, but this time instead of copy pasting values, press the F9 key to re-randomize. Notice that the formula in column E will automatically recalculate the median for each category every time and re-assign treatment and control status. Re-randomize 10 times and see what happens to the number of schools and the average pre-test score in each group.

Does the total number of schools change when you re-randomize? Does the number of schools for each language category change? Does the average pre-test score look balanced every time you re-randomize?