

Excel Guide: Randomization

This document will walk you through how to randomize individuals in a dataset into a control group and a treatment group using Microsoft Excel.¹ For this exercise, we are randomizing for a hypothetical evaluation to see how a modified version of an email affects the number of recipients who click on a hyperlink in the email. In this hypothetical evaluation, treatment individuals would receive the modified email and control individuals would receive the business-as-usual email.

Randomization

To randomly assign email recipients into two groups, we will use Excel's *rand()* function to generate a random number for each row, sort the spreadsheet based on the randomly generated numbers, and assign the first half of the rows to the control group and the other half to the treatment group.

What will you learn how to do?

1. Assign random values
 - a. Create "rand1" column
 - b. Enter randomization command `=rand()` and fill down the "rand1" column
 - c. Repeat 1a and 1b to create and fill "rand2" column
2. Convert randomization command outputs to values
 - a. Copy "rand1" and "rand2" columns
 - b. Paste as values
 - c. Delete original "rand1" and "rand2" columns
3. Sort the dataset
4. Assign treatment values
 - a. Create "treatment" column
 - b. Select the first half of the rows and assign with "0" in the "treatment" column
 - c. Select the second half of rows and assign with "1" in "treatment" column

Step 1) Assign random values

Step 1a. Create "rand1" column. On the spreadsheet that includes trial participants' email addresses, add a new column titled "rand1".

Step 1b. Enter randomization command `=rand()` and fill down the "rand1" column. On the following row, type the command: `=rand()`

Figure 1. Entering the randomization command

¹ This guide was created using Microsoft Excel 2021 Version 16.57 and the keyboard commands are defined for Windows keyboards. Locations of Excel functions may differ depending on the version of Excel that you are using, and keyboard commands will differ for other keyboards.

	A	B
1	Email	rand1
2	klm012@gmail.com	=RAND()
3	hij789@hotmail.com	
4	abd123@gmail.com	
5	efg456@yahoo.com	

After pressing Enter, you will see that a decimal number between 0 and 1 is randomly generated. Now, fill this command down the entire column to assign a random number for each of the rows. You can do so manually or using keyboard commands.

Manual fill: Select the cell where you have entered the command (in this case, B2) and drag the fill handle  (the dot that appears in the bottom right corner) down to the final row. (Note: The randomization function recalculates whenever a worksheet is opened or changed.)

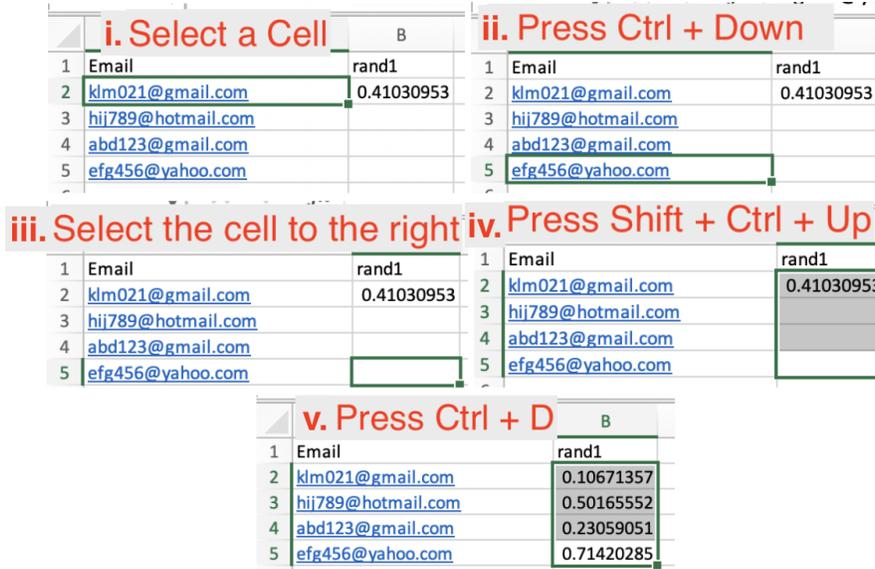
Figure 2. Manually filling down the first randomization column

	A	B		A	B	
1	Email	rand1		1	Email	rand1
2	klm012@gmail.com	0.6064885		2	klm012@gmail.com	0.0496934
3	hij789@hotmail.com			3	hij789@hotmail.com	0.7574394
4	abd123@gmail.com			4	abd123@gmail.com	0.1248301
5	efg456@yahoo.com			5	efg456@yahoo.com	0.2576304
				6		

Fill using keyboard commands. If the dataset is so large that it is difficult to manually scroll down to the end, you can also fill this command down the entire column using keyboard commands. (See Figure 3 below.)

- i. Select a cell in a filled column (in this case, “Email”).
- ii. Press Ctrl + Down arrow. This brings you to the final row of your dataset.
- iii. Select the cell to the right of the final cell of the “Email” column. This will be the final cell of the “rand1” column.
- iv. Press Shift + Ctrl + Up arrow. This selects from the final cell up to the closest filled cell. In this case, B2 which contains our randomization command.
- v. Fill down the command by pressing Ctrl + D.

Figure 3. Filling down the first randomization column using keyboard commands



Step 1c. Repeat 1a and 1b to create and fill “rand2” column. Create another column titled “rand2” and repeat the process performed for the “rand1” column.

The final data should contain two columns of randomly generated numbers.

Figure 4. Creating a second randomization column

	A	B	C
1	Email	rand1	rand2
2	klm012@gmail.com	0.5810108	0.227079
3	hij789@hotmail.com	0.72175	0.3546413
4	abd123@gmail.com	0.7540408	0.0100365
5	efg456@yahoo.com	0.2549857	0.8285137

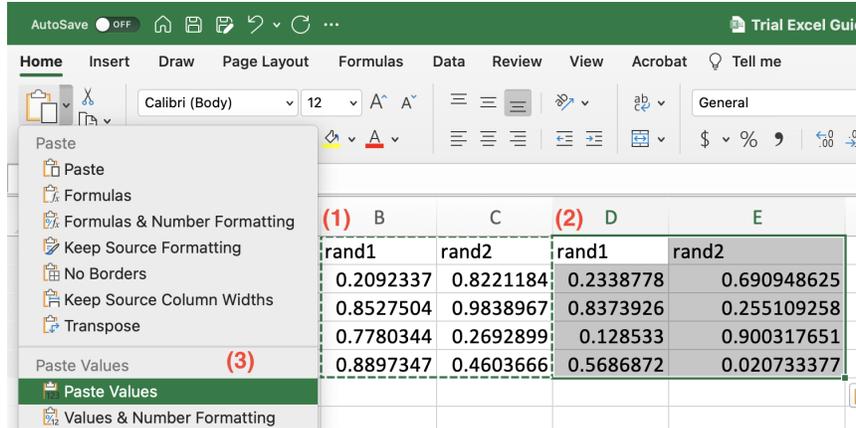
Step 2) Convert randomization command outputs to values

Step 2a. Copy “rand1” and “rand2” columns. Before we sort the spreadsheet, we will first copy the “rand1” and “rand2” columns (1).

Step 2b. Paste as values. Select cell D1 (2), and paste your selection in the adjacent columns, using “Paste Values” (3). You can find the “Paste Values” option by clicking the down arrow next to the clipboard in the Home tab. (See Figure 5 below.)

Once we paste the two columns, we see that the original columns’ values changed. This is because Excel is continuously performing the *rand()* function to generate new random numbers, but sorting the spreadsheet would be difficult with constantly changing numbers. By having the new columns with static values, we can then sort the spreadsheet and check whether sorting worked properly.

Figure 5. Pasting randomization outputs as values

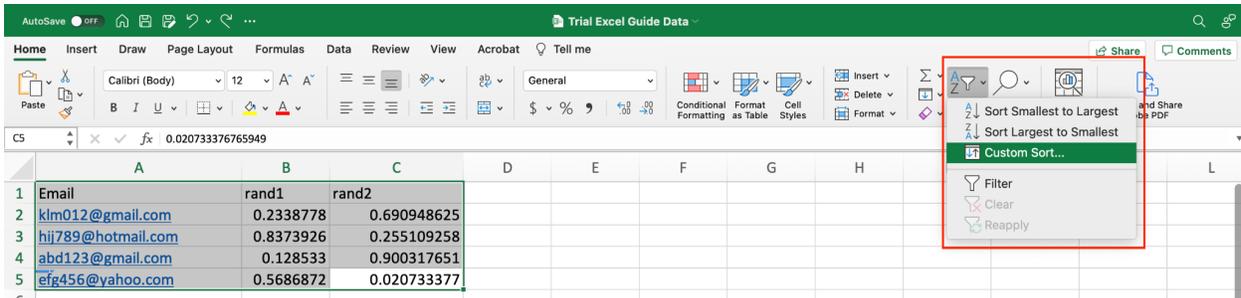


Step 2c. Delete the original “rand1” and “rand2” columns. We can now delete the original “rand1” and “rand2” columns containing the *rand()* function (columns B and C above). Do not delete the new columns with the values.

Step 3) Sort the dataset

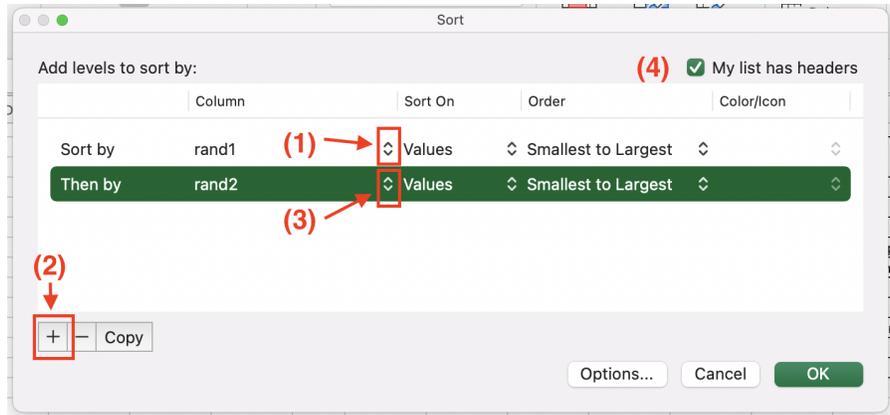
We will then custom sort the entire dataset. You can find the “Custom Sort” option under the “Sort & Filter” menu in the Home tab.

Figure 6. Accessing the “Custom Sort” function



Select to sort first by “rand1” (1), click “+” to add another sort level (2), then sort by “rand2” (3). Make sure “My list has headers” is checked (4). (See Figure 7 below.)

Figure 7. Sorting by the random values



Once we press “OK”, we see that the order of email addresses changed primarily based on the ordering of “rand1”. While the values of “rand1” almost always determine how the rows are ordered, we include “rand2” in case Excel generates two or more random numbers for “rand1” that are exactly the same. In such a rare case, ordering would be impossible, and “rand2” would determine how these rows are sorted.

Figure 8. Output of the custom sort

	A	B	C
1	Email	rand1	rand2
2	abd123@gmail.com	0.128533	0.900317651
3	klm012@gmail.com	0.2338778	0.690948625
4	efg456@yahoo.com	0.5686872	0.020733377
5	hij789@hotmail.com	0.8373926	0.255109258

Step 4) Assign treatment values

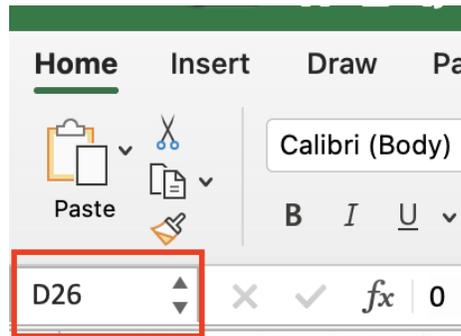
Step 4a. Create “treatment” column. Once the sorting is completed, we are ready to assign treatment status. Add a new column titled “treatment”.

Step 4b. Select the first half of the rows and assign with “0” in the “treatment” column. You can select the first half of the rows manually or you can use a similar strategy using keyboard commands as described in Step 1b.

- i. Enter “0” into the first cell below the “treatment” column title. In this case, D2.
- ii. Select a cell in any of your filled columns (in this case, “Email”, “rand1”, or rand2.) and select the entire column by clicking the column letter.
- iii. Note the count in the bottom right of the Excel workbook. From this count, subtract 1 to account for the row containing the column title. Divide this value by 2.
 - Ex: Count is 51. After subtracting by 1 and dividing by 2, the new value is 25.

- iv. Add 1 to this value. In the cell selection box, type in the column letter and this value. This will bring you to the halfway point of this column. (See Figure 9.)
 - *Ex: After adding 1, the new value is 26. Type D26 into the cell selection box. Press enter and this will bring you to D26.*
- v. Press Shift + Ctrl + Up arrow to select up to the cell where you entered “0”.
- vi. Press Ctrl + D to fill this selection of cells with “0”
- vii. Press Shift Ctrl + Down arrow to bring you to the final row filled with “0”.
- viii. Enter “1” into the cell below this row.

Figure 9. Cell Selection Box



Step 4c. Select the second half of rows and assign with “1” in the “treatment” column. Repeat above steps to fill the second half of rows with “1”.

Now, randomization is complete, and we are ready to launch the email trial. Participants with a treatment of value of “0” will receive the business-as-usual email, and those with a value of “1” will receive the modified version.

Figure 10. Assigned treatment values

	A	B	C	D
1	Email	rand1	rand2	treatment
2	abd123@gmail.com	0.128533	0.900317651	0
3	klm012@gmail.com	0.2338778	0.690948625	0
4	efg456@yahoo.com	0.5686872	0.020733377	1
5	hij789@hotmail.com	0.8373926	0.255109258	1