

How do I perform cluster sampling in Excel?

Authored by
stats writer

May 6, 2024

RECOMMENDED CITATION

stats writer (2024). *How do I perform cluster sampling in Excel?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=143221>

Cluster sampling is a statistical method used to collect data by dividing a population into smaller groups or clusters and then selecting a random sample from each cluster. This sampling technique is commonly used in situations where the population is large and geographically dispersed. To perform cluster sampling in Excel, you can use the random number generator function to select a random sample from each cluster. This can be achieved by creating a list of cluster numbers and using the "RAND" function to generate random numbers. The corresponding data for the selected clusters can then be collected and analyzed. This method ensures that each cluster has an equal chance of being included in the sample, making it a reliable and efficient way to obtain data from a large population.

Perform Cluster Sampling in Excel (Step-by-Step)

In statistics, we often take samples from a population and use the data from the sample to draw conclusions about the population as a whole.

One commonly used sampling method is cluster sampling, in which a population is split into clusters and all members of *some* clusters are chosen to be included in the sample.

The following step-by-step example shows how to perform cluster sampling in Excel.

Step 1: Enter the Data

First, let's enter the following dataset into Excel:

	A	B	C	D	E	F	G
1	Player ID	Team	Points	Rebounds			
2	1	A	17	6			
3	2	A	19	8			
4	3	A	17	3			
5	4	A	16	6			
6	5	B	12	5			
7	6	B	13	7			
8	7	B	22	9			
9	8	B	19	4			
10	9	C	9	8			
11	10	C	25	5			
12	11	C	12	6			
13	12	C	12	8			
14	13	D	30	7			
15	14	D	7	12			
16	15	D	17	15			
17	16	D	21	6			
18	17	E	23	7			
19	18	E	24	7			
20	19	E	8	4			
21	20	E	17	9			
22							
23							
24							

Next, we'll perform cluster sampling in which we randomly select two teams and choose to include every player from those two teams in the final sample.

Step 2: Find Unique Values

Next, type in `=UNIQUE(B2:B21)` to produce an array of unique values from the Team column:

F2		✕		✓		fx		=UNIQUE(B2:B21)	
	A	B	C	D	E	F	G		
1	Player ID	Team	Points	Rebounds		Unique			
2	1	A	17	6		A			
3	2	A	19	8		B			
4	3	A	17	3		C			
5	4	A	16	6		D			
6	5	B	12	5		E			
7	6	B	13	7					
8	7	B	22	9					
9	8	B	19	4					
10	9	C	9	8					
11	10	C	25	5					
12	11	C	12	6					
13	12	C	12	8					
14	13	D	30	7					
15	14	D	7	12					
16	15	D	17	15					
17	16	D	21	6					
18	17	E	23	7					
19	18	E	24	7					
20	19	E	8	4					
21	20	E	17	9					
22									
23									
24									

Next, we'll type an integer (starting at 1) next to each unique team name:

	A	B	C	D	E	F	G	H	
1	Player ID	Team	Points	Rebounds		Unique			
2	1	A	17	6		A	1		
3	2	A	19	8		B	2		
4	3	A	17	3		C	3		
5	4	A	16	6		D	4		
6	5	B	12	5		E	5		
7	6	B	13	7					
8	7	B	22	9					
9	8	B	19	4					
10	9	C	9	8					
11	10	C	25	5					
12	11	C	12	6					
13	12	C	12	8					
14	13	D	30	7					
15	14	D	7	12					
16	15	D	17	15					
17	16	D	21	6					
18	17	E	23	7					
19	18	E	24	7					
20	19	E	8	4					
21	20	E	17	9					
22									
23									
24									

Step 3: Select Random Clusters

Next, we'll type **=RANDBETWEEN(G2, G6)** to randomly select one of the integers from the list:

=RANDBETWEEN(G2, G6)						
D	E	F	G	H	I	J
ebounds		Unique				
6		A	1		5	
8		B	2			
3		C	3			
6		D	4			
5		E	5			
7						
9						
4						
8						
5						
6						
8						
7						
12						
15						
6						
7						
7						
4						
9						

=RANDBETWEEN(G2, G6)						
D	E	F	G	H	I	J
Rebounds		Unique				
6		A	1		3	
8		B	2			
3		C	3			
6		D	4			
5		E	5			
7						
9						
4						
8						
5						
6						
8						
7						
12						
15						
6						
7						
7						
4						
9						

We can see that the value 3 was randomly selected. The team associated with this value is team C, which represents the second team we'll include in our final sample.

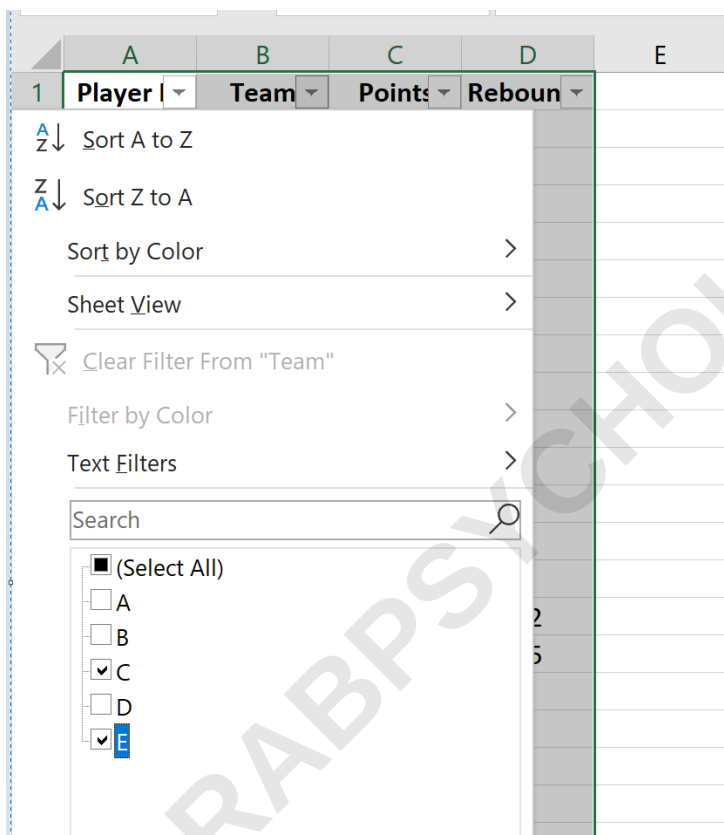
Step 4: Filter the Final Sample

The final sample will simply include all players who belong to either team C or team E.

To filter for just these players, highlight all of the data.

Then click the **Data** tab along the top ribbon and then click the **Filter** button within the **Sort & Filter** group.

When the filter appears above each column, click the dropdown arrow next to the **Team** column and check the boxes next to teams **C** and **E** only:



Once you click **OK**, the dataset will be filtered to only show players on team **C** or team **E**:

	A	B	C	D	E
1	Player ▾	Team ▾	Points ▾	Reboun ▾	
10	9	C	9	8	
11	10	C	25	5	
12	11	C	12	6	
13	12	C	12	8	
18	17	E	23	7	
19	18	E	24	7	
20	19	E	8	4	
21	20	E	17	9	
22					
23					
24					
25					
26					
27					
28					
29					

This represents our final sample.

Our cluster sampling is complete because we randomly chose two teams and included each player from those two teams in our final sample.

The following tutorials explain how to select other types of samples from a population using Excel:

How to Perform Stratified Sampling in Excel