

# How to Easily Perform Stratified Sampling in Excel

Authored by  
**stats writer**

December 4, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Easily Perform Stratified Sampling in Excel*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=104701>

Stratified sampling is a highly effective statistical technique designed to guarantee that a research sample accurately reflects the diversity present within the larger population. Unlike simple random sampling, this approach involves dividing the population into non-overlapping subgroups, known as strata, based on shared characteristics (such as age, gender, location, or team). This methodology ensures proportional representation, which significantly improves the validity and generalizability of the study's conclusions.

To execute **stratified random sampling** efficiently using Excel, the process involves several key computational steps. We begin by organizing the data, identifying the strata, and then leveraging Excel's powerful built-in functions, specifically the RAND function, to assign random selection scores. The goal is to randomly draw a predefined number of members from each group. The following detailed tutorial outlines the precise, step-by-step methodology required to successfully implement this sophisticated sampling design.

The procedure relies heavily on accurate data manipulation and sorting capabilities within the spreadsheet environment. While the initial overview of the method often mentions advanced functions like COUNTIF for verification, the core mechanism focuses on pairing categorization (the strata) with truly random assignment scores. This allows us to select the top-ranked entries within each stratified group automatically after sorting.

In the field of **statistics**, drawing reliable conclusions about a large population typically necessitates the collection of data from a smaller, representative sample. The quality of the conclusions drawn is inherently linked to how well the sample reflects the original population structure.

This need for accuracy makes **stratified random sampling** a preferred choice over simpler methods. It involves deliberately dividing the population into mutually exclusive subgroups (strata) and then conducting a simple random selection within each stratum. This technique ensures that crucial subgroups are not underrepresented in the final data set.

This comprehensive, step-by-step example utilizes a practical scenario to illustrate the precise methodology required to perform stratified random sampling using the robust data management capabilities of Excel.

## Setting the Context and Inputting the Initial Dataset

To begin this exercise, we must first define our scenario and input the foundational **dataset** into a clean Excel worksheet. Our scenario involves a total population of basketball players (IDs 1 through 18) categorized across three distinct teams (Team A, Team B, and Team C). Each team represents a single, non-overlapping stratum for our sampling purposes.

The primary goal of this stratification exercise is to select a uniform **sample** size of two players from each of the three basketball teams. This ensures that the final sample of six players is perfectly balanced across the strata, preventing any single team from dominating the final results, thus maintaining internal validity. Input the data exactly as shown below, defining columns for Player ID and Team.

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

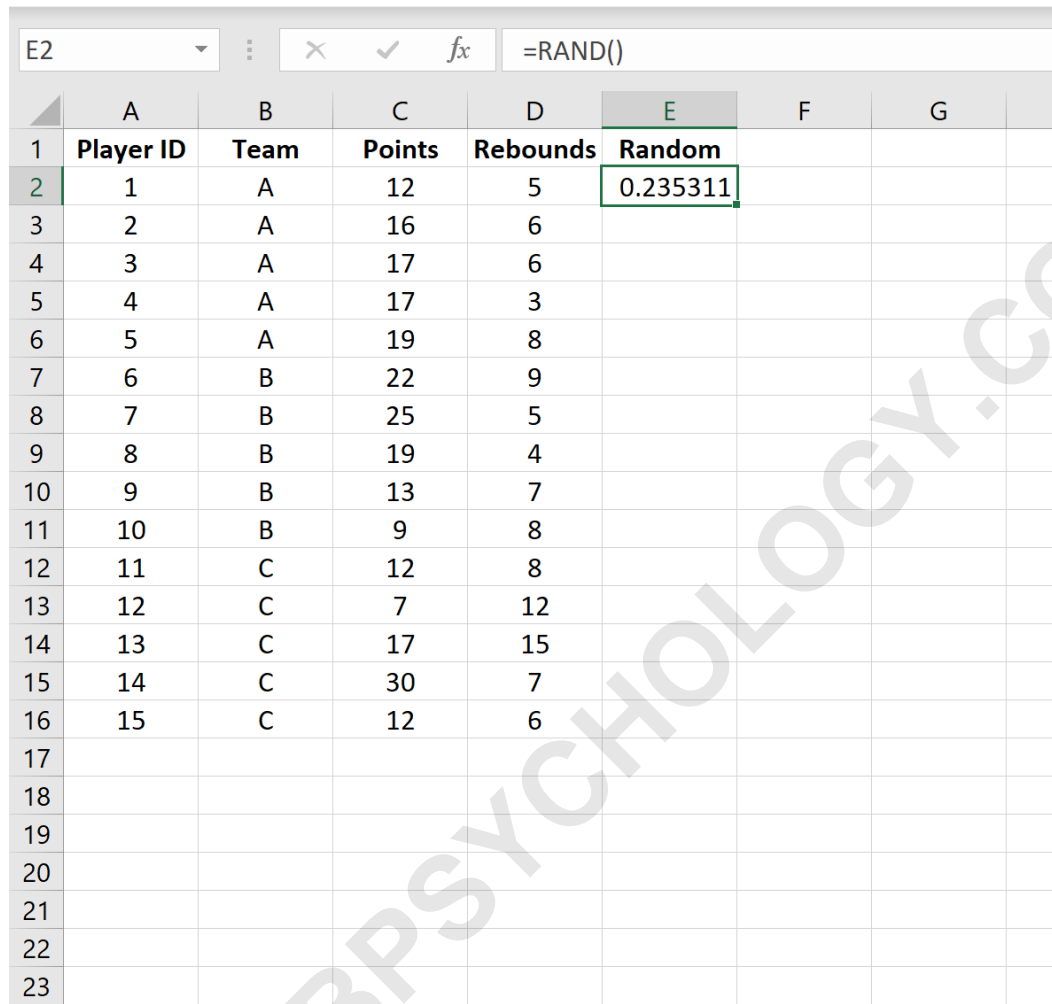
Careful organization of the initial data is paramount before proceeding to randomization. Ensure that the team labels are consistent, as these categories will be used later as the primary key for sorting and selecting the final stratified sample.

## Step 2: Assigning Random Scores Using the RAND Function

The next critical phase in performing **stratified sampling** is introducing true randomness to the selection process within each stratum. This is achieved by generating a unique, continuous random score for every entry in the population. Create a new column adjacent to your data, naming it **Random**.

In the first cell of this new column (e.g., cell E2, assuming your data starts in Row 2), input the

basic Excel randomization formula: **=RAND()**. This specific function returns a floating-point, uniformly distributed random real number greater than or equal to 0 and less than 1. This number acts as the randomization key for selection.



The screenshot shows an Excel spreadsheet with the following data:

|    | A                | B           | C             | D               | E             | F | G |
|----|------------------|-------------|---------------|-----------------|---------------|---|---|
| 1  | <b>Player ID</b> | <b>Team</b> | <b>Points</b> | <b>Rebounds</b> | <b>Random</b> |   |   |
| 2  | 1                | A           | 12            | 5               | 0.235311      |   |   |
| 3  | 2                | A           | 16            | 6               |               |   |   |
| 4  | 3                | A           | 17            | 6               |               |   |   |
| 5  | 4                | A           | 17            | 3               |               |   |   |
| 6  | 5                | A           | 19            | 8               |               |   |   |
| 7  | 6                | B           | 22            | 9               |               |   |   |
| 8  | 7                | B           | 25            | 5               |               |   |   |
| 9  | 8                | B           | 19            | 4               |               |   |   |
| 10 | 9                | B           | 13            | 7               |               |   |   |
| 11 | 10               | B           | 9             | 8               |               |   |   |
| 12 | 11               | C           | 12            | 8               |               |   |   |
| 13 | 12               | C           | 7             | 12              |               |   |   |
| 14 | 13               | C           | 17            | 15              |               |   |   |
| 15 | 14               | C           | 30            | 7               |               |   |   |
| 16 | 15               | C           | 12            | 6               |               |   |   |
| 17 |                  |             |               |                 |               |   |   |
| 18 |                  |             |               |                 |               |   |   |
| 19 |                  |             |               |                 |               |   |   |
| 20 |                  |             |               |                 |               |   |   |
| 21 |                  |             |               |                 |               |   |   |
| 22 |                  |             |               |                 |               |   |   |
| 23 |                  |             |               |                 |               |   |   |

After successfully inputting the **=RAND()** formula into the first cell, extend this formula down through all remaining rows of the dataset. The most efficient way to do this in Excel is to use the fill handle: hover over the bottom-right corner of the cell containing the formula until the cursor transforms into a thin black cross ( + ) and double-click. This action instantly populates the entire column with unique random scores, matching the height of your initial data range.

|    | A                | B           | C             | D               | E             | F | G |
|----|------------------|-------------|---------------|-----------------|---------------|---|---|
| 1  | <b>Player ID</b> | <b>Team</b> | <b>Points</b> | <b>Rebounds</b> | <b>Random</b> |   |   |
| 2  | 1                | A           | 12            | 5               | 0.720303      |   |   |
| 3  | 2                | A           | 16            | 6               | 0.699921      |   |   |
| 4  | 3                | A           | 17            | 6               | 0.437896      |   |   |
| 5  | 4                | A           | 17            | 3               | 0.644913      |   |   |
| 6  | 5                | A           | 19            | 8               | 0.48128       |   |   |
| 7  | 6                | B           | 22            | 9               | 0.242952      |   |   |
| 8  | 7                | B           | 25            | 5               | 0.987388      |   |   |
| 9  | 8                | B           | 19            | 4               | 0.419407      |   |   |
| 10 | 9                | B           | 13            | 7               | 0.202296      |   |   |
| 11 | 10               | B           | 9             | 8               | 0.983954      |   |   |
| 12 | 11               | C           | 12            | 8               | 0.504556      |   |   |
| 13 | 12               | C           | 7             | 12              | 0.934046      |   |   |
| 14 | 13               | C           | 17            | 15              | 0.95429       |   |   |
| 15 | 14               | C           | 30            | 7               | 0.640191      |   |   |
| 16 | 15               | C           | 12            | 6               | 0.401028      |   |   |
| 17 |                  |             |               |                 |               |   |   |
| 18 |                  |             |               |                 |               |   |   |
| 19 |                  |             |               |                 |               |   |   |
| 20 |                  |             |               |                 |               |   |   |
| 21 |                  |             |               |                 |               |   |   |
| 22 |                  |             |               |                 |               |   |   |
| 23 |                  |             |               |                 |               |   |   |

### Step 3: Freezing the Random Scores via Paste Values

A crucial aspect of using the RAND function in Excel is its volatile nature. By default, **=RAND()** recalculates and generates a completely new value every time the spreadsheet is opened, a cell is edited, or the sheet is recalculated. If we do not freeze these values, our random selection will change dynamically during the sorting process, invalidating our **stratified sampling** results.

To convert the volatile formulas into static numerical data, highlight the entire **Random** column. Copy the data (Ctrl+C or right-click > Copy). Immediately after copying, right-click on the selection again and choose the **Paste Special** option, specifically selecting **Values** (often denoted by the clipboard icon with '123' on it).

This action replaces the dynamic **=RAND()** formula with the fixed numerical output it generated at that precise moment. Once this step is complete, the random scores assigned to each player ID will remain constant, ensuring a stable foundation for the subsequent sorting and selection process. This stability is essential for replicability in statistical work.

## Step 4: Multi-Level Sorting by Strata and Random Score

With the random scores assigned and frozen, the next step involves using Excel's powerful multi-level sorting feature to group the strata and then rank the members within those groups based on the random scores. Highlight the entire dataset, including all columns (Player ID, Team, and Random). Navigate to the **Data** tab on the Excel ribbon and click the **Sort** button found within the **Sort & Filter** group.

In the resulting dialog box, you must define two distinct sorting levels. The first level must be the categorization variable--in this case, **Team**. Set the order to sort from A to Z (or smallest to largest). This action successfully clusters all members of Team A together, followed by Team B, and finally Team C.

The second, and equally crucial, level of sorting is applied within each team cluster. Set the second sort level to the **Random** score column. Crucially, sort this column from **Smallest to Largest**. When the sort operation is executed, the players within Team A will be randomly ordered based on their score, with the lowest scores placed at the top of the group, and the same process will be repeated independently for Team B and Team C.

|    | A                | B           | C             | D               | E             | F | G | H | I | J | K | L | M | N |
|----|------------------|-------------|---------------|-----------------|---------------|---|---|---|---|---|---|---|---|---|
| 1  | <b>Player ID</b> | <b>Team</b> | <b>Points</b> | <b>Rebounds</b> | <b>Random</b> |   |   |   |   |   |   |   |   |   |
| 2  | 1                | A           | 12            | 5               | 0.720303      |   |   |   |   |   |   |   |   |   |
| 3  | 2                | A           | 16            | 6               | 0.699921      |   |   |   |   |   |   |   |   |   |
| 4  | 3                | A           | 17            | 6               | 0.437896      |   |   |   |   |   |   |   |   |   |
| 5  | 4                | A           | 17            | 3               | 0.644913      |   |   |   |   |   |   |   |   |   |
| 6  | 5                | A           | 19            | 8               | 0.48128       |   |   |   |   |   |   |   |   |   |
| 7  | 6                | B           | 22            | 9               | 0.242952      |   |   |   |   |   |   |   |   |   |
| 8  | 7                | B           | 25            | 5               | 0.987388      |   |   |   |   |   |   |   |   |   |
| 9  | 8                | B           | 19            | 4               | 0.419407      |   |   |   |   |   |   |   |   |   |
| 10 | 9                | B           | 13            | 7               | 0.202296      |   |   |   |   |   |   |   |   |   |
| 11 | 10               | B           | 9             | 8               | 0.983954      |   |   |   |   |   |   |   |   |   |
| 12 | 11               | C           | 12            | 8               | 0.504556      |   |   |   |   |   |   |   |   |   |
| 13 | 12               | C           | 7             | 12              | 0.934046      |   |   |   |   |   |   |   |   |   |
| 14 | 13               | C           | 17            | 15              | 0.95429       |   |   |   |   |   |   |   |   |   |
| 15 | 14               | C           | 30            | 7               | 0.640191      |   |   |   |   |   |   |   |   |   |
| 16 | 15               | C           | 12            | 6               | 0.401028      |   |   |   |   |   |   |   |   |   |
| 17 |                  |             |               |                 |               |   |   |   |   |   |   |   |   |   |
| 18 |                  |             |               |                 |               |   |   |   |   |   |   |   |   |   |
| 19 |                  |             |               |                 |               |   |   |   |   |   |   |   |   |   |
| 20 |                  |             |               |                 |               |   |   |   |   |   |   |   |   |   |
| 21 |                  |             |               |                 |               |   |   |   |   |   |   |   |   |   |
| 22 |                  |             |               |                 |               |   |   |   |   |   |   |   |   |   |
| 23 |                  |             |               |                 |               |   |   |   |   |   |   |   |   |   |
| 24 |                  |             |               |                 |               |   |   |   |   |   |   |   |   |   |
| 25 |                  |             |               |                 |               |   |   |   |   |   |   |   |   |   |
| 26 |                  |             |               |                 |               |   |   |   |   |   |   |   |   |   |
| 27 |                  |             |               |                 |               |   |   |   |   |   |   |   |   |   |

This dual-sort process mathematically implements the definition of stratified random sampling. By sorting first by the stratum and then by the randomized score, we have simultaneously isolated our groups and randomly ranked all members within them, setting the stage for the final selection. Click **OK** to apply the sorting order.

## Step 5: Extracting the Final Stratified Sample

Following the comprehensive sorting operation, the selection of the final **stratified sampling** subset becomes a straightforward visual extraction. Since we defined that we required a sample size of two players per team, the members chosen are simply the first two entries appearing under each respective stratum (Team A, Team B, and Team C) after the multi-level sort.

The critical insight here is that because the data was sorted first by Team and then by the random score (smallest to largest), the top entries within each team bracket represent the players with the highest random selection priority, ensuring both stratification and randomness were preserved.

|    | A         | B    | C      | D        | E        | F | G |
|----|-----------|------|--------|----------|----------|---|---|
| 1  | Player ID | Team | Points | Rebounds | Random   |   |   |
| 2  | 3         | A    | 17     | 6        | 0.437896 |   |   |
| 3  | 5         | A    | 19     | 8        | 0.48128  |   |   |
| 4  | 4         | A    | 17     | 3        | 0.644913 |   |   |
| 5  | 2         | A    | 16     | 6        | 0.699921 |   |   |
| 6  | 1         | A    | 12     | 5        | 0.720303 |   |   |
| 7  | 9         | B    | 13     | 7        | 0.202296 |   |   |
| 8  | 6         | B    | 22     | 9        | 0.242952 |   |   |
| 9  | 8         | B    | 19     | 4        | 0.419407 |   |   |
| 10 | 10        | B    | 9      | 8        | 0.983954 |   |   |
| 11 | 7         | B    | 25     | 5        | 0.987388 |   |   |
| 12 | 15        | C    | 12     | 6        | 0.401028 |   |   |
| 13 | 11        | C    | 12     | 8        | 0.504556 |   |   |
| 14 | 14        | C    | 30     | 7        | 0.640191 |   |   |
| 15 | 12        | C    | 7      | 12       | 0.934046 |   |   |
| 16 | 13        | C    | 17     | 15       | 0.95429  |   |   |
| 17 |           |      |        |          |          |   |   |
| 18 |           |      |        |          |          |   |   |
| 19 |           |      |        |          |          |   |   |
| 20 |           |      |        |          |          |   |   |
| 21 |           |      |        |          |          |   |   |
| 22 |           |      |        |          |          |   |   |
| 23 |           |      |        |          |          |   |   |

The final resulting sample includes the following player IDs, validating that exactly two individuals were selected from each stratum, fulfilling the requirements for proportional stratified random sampling:

**Team A:** Player IDs 3 and 5

**Team B:** Player IDs 9 and 6

**Team C:** Player IDs 15 and 11

This concludes the technical process for selecting a stratified sample in Excel. The method guarantees both the random selection of individuals and the controlled representation of all predefined subgroups within the larger population.

## Summary of the Stratified Selection Methodology

The process of performing stratified sampling in a spreadsheet environment like Excel simplifies a complex statistical requirement into a manageable series of steps. The efficacy of this method stems from the combination of group definition (strata) and objective randomization provided by the RAND function. This is paramount when conducting research where key demographics or characteristics must be proportionally represented.

The main benefit of using stratification is the reduction of sampling error, particularly when the variable of interest is strongly correlated with the stratifying variable (e.g., performance correlated with team). By ensuring adequate representation, we enhance the precision of our estimates concerning the entire population.

## Related Sampling Techniques in Excel

While **stratified sampling** is ideal for heterogeneous populations requiring specific subgroup representation, other sampling techniques may be more appropriate depending on the data structure and research goals. The following tutorials explore alternative methods for extracting representative samples from a population using Excel's statistical tools:

[How to Perform Cluster Sampling in Excel](#)