

Chapter 127

# Simple Random Sampling with Group Assignment

## Introduction

The Simple Random Sampling with Group Assignment tool in **NCSS** can be used to quickly generate  $K$  independent random samples from a dataset, where each random sample has  $N$  items, and randomly assign the selected items to groups. Each random sample is generated without replacement. The user may create a sample based on the data in entire rows or simply sample values from a single column. The random sample selection data along with group assignments may be written to a report in the Output window or stored in user-specified columns of the Data Table.

Random numbers for sampling are generated using the Mersenne Twister algorithm. The user may enter a random seed to replicate previous sampling results or generate a random seed based on the computer's internal clock.

The Simple Random Sampling tool can be accessed from the Data or Tools menu on the Data window.

**This procedure uses random sorting to assign randomly selected values to groups. To assign groups using more complex randomization algorithms, use the Randomization Lists procedure after using the Simple Random Sampling procedure to generate the random sample.**

*Example – Result of randomly selecting 4 values from column Y and randomly assigning selected values to 2 groups ("Group 1" and "Group 2")*

Row	X	Y	Z	Random Sample (Binary Format)	Random Sample	Group	Random Sample (Collapsed Format)	Group (Collapsed)
1	10						1	Group 1
2	0	1		Selected	1	Group 1	9	Group 1
3	4	5		Selected	5	Group 2	3	Group 2
4	7	2		Not Selected			5	Group 2
5	3	3	6	Selected	3	Group 2		
6	2	2	3	Not Selected				
7	3		10					
8	8	10	2	Not Selected				
9								
10	4	9	8	Selected	9	Group 1		

## Random Sampling Options

The methods in this procedure that may be used to sample the dataset are described below.

### Dataset Sampling Type

This option specifies the method that will be used to sample the dataset. The options are

- **Randomly Select Rows from the Dataset**

In this sampling type, the data in all non-empty columns is used to determine eligibility for sampling. Only rows that are completely empty are ignored. Rows that have some missing values may still be selected in the random sample. This sampling type stores either row numbers (collapsed or expanded to the corresponding rows) or expanded binary selection values that indicate whether each non-empty row was selected.

*Example – Result of randomly selecting 4 rows (from 9 eligible rows) and assigning to 2 groups*

Row	X	Y	Z	Random Sample (Binary Format)	Random Sample	Group	Random Sample (Collapsed Format)	Group (Collapsed)
1	10			Selected	1	Group 1	1	Group 1
2	0	1		Selected	2	Group 2	7	Group 1
3	4	5		Not Selected			2	Group 2
4	7	2		Not Selected			10	Group 2
5	3	3	6	Not Selected				
6	2	2	3	Not Selected				
7	3		10	Selected	7	Group 1		
8	8	10	2	Not Selected				
9								
10	4	9	8	Selected	10	Group 2		

- **Randomly Select Values from a Column**

In this sampling type, only the data in the chosen column is used to determine eligibility for sampling. Only non-missing values in the column may be selected. This sampling type stores either actual values (collapsed or expanded to the corresponding rows) or expanded binary selection values that indicate whether each non-missing value was selected.

*Example – Result of randomly selecting 4 values from column Y (from 7 eligible, non-missing values) and assigning to 2 groups*

Row	X	Y	Z	Random Sample (Binary Format)	Random Sample	Group	Random Sample (Collapsed Format)	Group (Collapsed)
1	10						1	Group 1
2	0	1		Selected	1	Group 1	9	Group 1
3	4	5		Selected	5	Group 2	3	Group 2
4	7	2		Not Selected			5	Group 2
5	3	3	6	Selected	3	Group 2		
6	2	2	3	Not Selected				
7	3		10					
8	8	10	2	Not Selected				
9								
10	4	9	8	Selected	9	Group 1		

## Sample Allocation among Groups

This option is used along with the sample size to determine how many sampled items will be assigned to each group. If the chosen sample size is not divisible by the number of groups, some of the groups may have more or fewer items than expected.

The options are

- **Equal**

An equal number of items is assigned to each group.

*Example – Result of randomly selecting 6 rows and assigning to 2 groups with equal group allocation*

Row	X	Y	Z	Random Sample (Binary Format)	Random Sample	Group	Random Sample (Collapsed Format)	Group (Collapsed)
1	10			Selected	1	Group 1	1	Group 1
2	0	1		Selected	2	Group 1	2	Group 1
3	4	5		Not Selected			7	Group 1
4	7	2		Selected	4	Group 2	4	Group 2
5	3	3	6	Not Selected			8	Group 2
6	2	2	3	Not Selected			10	Group 2
7	3		10	Selected	7	Group 1		
8	8	10	2	Selected	8	Group 2		
9								
10	4	9	8	Selected	10	Group 2		

- **Custom**

The relative group allotments are determined by user-entered allocation ratios. The first value is the ratio for the first group, the second for the second group, and so on. The values need not sum to one since they will be scaled to relative percentages, so to achieve 1:2:1 group allocation among 3 groups you could enter 1 2 1, 0.25 0.5 0.25, or 25 50 25. All would give the same results for the group allocation percentages.

*Example – Result of randomly selecting 6 rows and assigning to 2 groups with 1:2 custom group allocation*

Row	X	Y	Z	Random Sample (Binary Format)	Random Sample	Group	Random Sample (Collapsed Format)	Group (Collapsed)
1	10			Selected	1	Group 1	1	Group 1
2	0	1		Selected	2	Group 2	7	Group 1
3	4	5		Not Selected			2	Group 2
4	7	2		Selected	4	Group 2	4	Group 2
5	3	3	6	Not Selected			8	Group 2
6	2	2	3	Not Selected			10	Group 2
7	3		10	Selected	7	Group 1		
8	8	10	2	Selected	8	Group 2		
9								
10	4	9	8	Selected	10	Group 2		

## Random Sample Storage

The random sample may be stored on the Data Table using the options on the Storage tab. The various methods that may be used to store the selected data are described below.

### Selection Data to Store

This option specifies how selection data will be stored in the Data Table. The options are

- **Binary Selection Indicators (Dummy Values)**

The stored selection data will appear as binary indicator (i.e., “dummy”) values. The random sample data retains all information contained in other columns of the dataset.

You may enter custom values for the binary indicators of selected and not selected.

*Example – Result of randomly selecting 4 rows with Selected = “1” and Not Selected = “0” and assigning to 2 groups with equal group allocation*

Row	X	Y	Z	Random Sample (Binary Format)	Group
1	10			1	Group 1
2	0	1		1	Group 2
3	4	5		0	
4	7	2		0	
5	3	3	6	0	
6	2	2	3	0	
7	3		10	1	Group 1
8	8	10	2	0	
9					
10	4	9	8	1	Group 2

*Example – Result of randomly selecting 4 values from column Y with Selected = “Yes” and Not Selected = “[Blank]” and assigning to 2 groups with equal group allocation*

Row	X	Y	Z	Random Sample (Binary Format)	Group
1	10				
2	0	1		Yes	Group 1
3	4	5		Yes	Group 2
4	7	2			
5	3	3	6	Yes	Group 2
6	2	2	3		
7	3		10		
8	8	10	2		
9					
10	4	9	8	Yes	Group 1

Simple Random Sampling with Group Assignment

- Selected Rows/Values**

The stored selection data (either row numbers or actual data values depending on your selection for Dataset Sampling Type) will appear in place on corresponding rows. The random sample data retains all information contained in other columns of the dataset.

*Example – Result of randomly selecting 4 rows and assigning to 2 groups with equal group allocation*

Row	X	Y	Z	Random Sample	Group
1	10			1	Group 1
2	0	1		2	Group 2
3	4	5			
4	7	2			
5	3	3	6		
6	2	2	3		
7	3		10	7	Group 1
8	8	10	2		
9					
10	4	9	8	10	Group 2

*Example – Result of randomly selecting 4 values from column Y and assigning to 2 groups with equal group allocation*

Row	X	Y	Z	Random Sample	Group
1	10				
2	0	1		1	Group 1
3	4	5		5	Group 2
4	7	2			
5	3	3	6	3	Group 2
6	2	2	3		
7	3		10		
8	8	10	2		
9					
10	4	9	8	9	Group 1

Simple Random Sampling with Group Assignment

- Selected Rows/Values (Collapsed)**

The stored selection data (either row numbers or actual data values depending on your selection for Dataset Sampling Type) will appear in collapsed format. The stored data may be sorted if desired with sorting controlled by the options on the Reports tab. The stored data will appear the same as that displayed in the Random Sample List(s) in the output report.

When sampling values from a column, the original row numbers of selected values may also be stored in an adjacent column to retain a reference to information contained in other columns of the dataset.

*Example – Result of randomly selecting 4 rows and assigning to 2 groups with equal group allocation and stored selection data sorted by original row number*

Row	X	Y	Z	Random Sample (Collapsed Format)	Group (Collapsed)
1	10			1	Group 1
2	0	1		2	Group 2
3	4	5		7	Group 1
4	7	2		10	Group 2
5	3	3	6		
6	2	2	3		
7	3		10		
8	8	10	2		
9					
10	4	9	8		

*Example – Result of randomly selecting 4 values from column Y with stored selection data sorted by assigned group then by selected value and with corresponding row numbers stored in an adjacent column*

Row	X	Y	Z	Original Row Number	Random Sample (Collapsed Format)	Group (Collapsed)
1	10			2	1	Group 1
2	0	1		10	9	Group 1
3	4	5		5	3	Group 2
4	7	2		3	5	Group 2
5	3	3	6			
6	2	2	3			
7	3		10			
8	8	10	2			
9						
10	4	9	8			

# Example 1 – Selecting a Simple Random Sample of Rows from the Resale Dataset with Random Assignment to 3 Groups

This example will demonstrate how to take a simple random sample of 30 rows from a dataset and randomly assign the selected rows to 3 groups (A, B, or C) with equal group allocation. The sample selection data will be stored as indicator or dummy values (0's and 1's) in the Data Table on the Data Window, with a 1 indicating that the row was selected and a 0 indicating that it was not selected.

This example will also cover how to save the randomly selected rows to a new subset data file using the filter.

For reproducibility, we'll use a random seed of 60677.

**This example uses random sorting to assign randomly selected values to groups. To assign groups using more complex randomization algorithms, use the Randomization Lists procedure after using the Simple Random Sampling procedure to generate the random sample.**

## Setup

To run this example, complete the following steps:

### 1 Open the Resale example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Resale** and click **OK**.

### 2 Specify the Simple Random Sampling with Group Assignment procedure options

- Find and open the **Simple Random Sampling with Group Assignment** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Random Sampling Tab

---

Sample Size Entry Type .....**Enter the Number of Rows/Values to Select**  
 Number of Rows/Values to Select .....**30**  
 Dataset Sampling Type .....**Randomly Select Rows from the Dataset**  
 Random Seed.....**60677** (for reproducibility)  
 Number of Groups to Assign.....**3**  
 Group Names .....**Custom**  
     Custom Group Names .....**A B C**  
 Sample Allocation among Groups .....**Equal**

### 3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

## Random Sampling Summary Report

### Simple Random Sampling Summary

Number of Rows Selected	30 of 150 (20%)
Dataset Sampling Type	Rows Randomly Selected
Number of Random Sample Replicates	1
User-Entered Random Seed*	60677
Number of Groups Assigned	3
Sample Allocation among Groups	Equal
Group Names	"A" "B" "C"

Storage: The random sample selection data was stored in columns 21 through 22 of the Data Table.

\* You can use this random seed to replicate this random sample, but the original dataset, number of random samples to generate, and sampling options must be the same.

This report gives a summary of the random sample that was generated. The sample size of 30 corresponds to 20% of all 150 available rows with data. The random sample itself was stored in the Data Table in column 21.

## Random Sample List Report

### Simple Random Sample 1 of 1

Number of Rows Selected	30 of 150 (20%)
Dataset Sampling Type	Rows Randomly Selected
User-Entered Random Seed	60677
Number of Rows Assigned to Each Group	"A" (10), "B" (10), "C" (10)

Selected Item	Selected Row	Assigned Group
1	15	C
2	18	B
3	23	A
4	24	C
5	25	A
6	29	B
7	32	C
8	33	A
9	34	A
10	37	C
.	.	.
.	.	.
.	.	.

This report lists in collapsed format the 30 rows that were selected along with their random group assignments.



## Stored Random Sample Data

Row	Sample_1	Group_1
1	0	
2	0	
3	0	
.	.	
.	.	
15	1	C
16	0	
17	0	
18	1	B
.	.	
.	.	
146	1	A
147	0	
148	0	
149	0	
150	1	B

In column 21 of the Data Table on the Data Window (now named "Sample\_1") you'll find the random sample stored using binary indicator (dummy) values. In column 22 (now named "Group\_1") you'll find the group assignment for each selected row. You can now set a filter on one or both columns to restrict further analyses to just this sampled set or for saving or exporting the random subset to a new file.

The newly created Note for column "Sample\_1" contains the following information about the sampled data.

Simple Random Sample Storage Column 1 of 1

---

Number of Rows Selected: 30 of 150 (20%)  
 Dataset Sampling Type: Rows Randomly Selected  
 User-Entered Random Seed\*: 60677

Number of Groups Assigned: 3  
 Sample Allocation among Groups: Equal  
 Number of Rows Assigned to Each Group: "A" (10), "B" (10), "C" (10)

\* You can use this random seed to replicate this random sample, but the original dataset, number of random samples to generate, and sampling options must be the same.

---

## Saving the Random Sample Subset to a New Data File using the Data Filter

To save the random sample data to a new subset file using the data filter, do the following:

### 4 Set a filter on column 21

- Click the **Filter button** in the toolbar to load the **Filter Editor**.
- Click **New List Condition**, select **column 21 "Sample\_1"** from the column selection window, and click **OK**.
- Check only the value **1** in the list under **Sample\_1 Condition** and click **OK**.
- The dataset will now be filtered with only the randomly selected rows highlighted.

### 5 Save the filtered data to a new file

- On the Data Window menu, select **File** then **Save Subset As**.
- Under Rows to Save, select Save Only Those Rows that Pass the Active Filter.
- Leave Columns to Save set to Save All Columns.
- Click **OK** to save the subset to a new data file.
- Name the new subset data file and click Save to save it.
- Open the saved data file by selecting it from the **File** menu.
- The subset file contains only the randomly selected rows and can be used for further analyses.

## Example 2 – Selecting a Simple Random Sample of 10% of Values from a Column in the Resale Dataset with 1:2 Allocated Random Assignment to 2 Groups

This example will demonstrate how to take a simple random sample from a column in a dataset and randomly assign the selected values to 2 groups (“Group 1” and “Group 2”) with 1:2 group allocation. We’ll select 10% of the values in the column called Price. The sample selection data will be stored in collapsed format in the Data Table with row numbers also stored for reference to the original dataset. The resulting sample data will be sorted by assigned group and then by value.

For reproducibility, we will use a random seed of 22077.

**This example uses random sorting to assign randomly selected values to groups. To assign groups using more complex randomization algorithms, use the Randomization Lists procedure after using the Simple Random Sampling procedure to generate the random sample.**

### Setup

To run this example, complete the following steps:

**1 Open the Resale example dataset**

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Resale** and click **OK**.

**2 Specify the Simple Random Sampling with Group Assignment procedure options**

- Find and open the **Simple Random Sampling with Group Assignment** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Random Sampling Tab	
Sample Size Entry Type .....	<b>Enter the Percent of Rows/Values to Select</b>
Percent of Rows/Values to Select.....	<b>10</b>
Dataset Sampling Type .....	<b>Randomly Select Values from a Column</b>
Column to Sample .....	<b>Price</b>
Random Seed.....	<b>22077</b> (for reproducibility)
Number of Groups to Assign.....	<b>2</b>
Group Names .....	<b>Custom</b>
Custom Group Names .....	<b>"Group 1" "Group 2"</b>
Sample Allocation among Groups .....	<b>Custom</b>
Custom Sample Allocation.....	<b>1 2</b>

Simple Random Sampling with Group Assignment

Reports Tab

---

Sort Random Sample List(s)..... **Checked**  
 Sort Type ..... **By Assigned Group then by Selected Row/Value**

Storage Tab

---

Store the Random Sample(s) in the Data Table ..... **Checked**  
 Selection Data to Store ..... **Selected Rows/Values (Collapsed)**  
 Store the original row number of each selected..... **Checked**  
 value

Report Options (*in the Toolbar*)

---

Variable Labels ..... **Column Names**

**3 Run the procedure**

- Click the **Run** button to perform the calculations and generate the output.

---

## Random Sampling Summary Report

**Simple Random Sampling Summary**

---

Number of Values Selected	15 of 150 (10%)
Dataset Sampling Type	Values Randomly Selected from Price
Number of Random Sample Replicates	1
User-Entered Random Seed*	22077
Number of Groups Assigned	2
Sample Allocation among Groups	Custom (1 2)
Group Names	"Group 1" "Group 2"

Storage: The random sample selection data was stored in columns 21 through 23 of the Data Table.

---

\* You can use this random seed to replicate this random sample, but the original dataset, number of random samples to generate, and sampling options must be the same.

This report gives a summary of the random sample that was generated by selecting values from Price, along with the group summary information. A sample size of 15 corresponds to 10% of all 150 available rows with data. The random sample itself was stored in the Data Table in columns 21 to 23.

## Random Sample List Report

### Simple Random Sample 1 of 1

---

Number of Values Selected	15 of 150 (10%)
Dataset Sampling Type	Values Randomly Selected from Price
User-Entered Random Seed	22077
Number of Values Assigned to Each Group	"Group 1" (5), "Group 2" (10)

---

Selected Item	Original Row	Value Selected from Price	Assigned Group
1	71	53300	Group 1
2	20	101800	Group 1
3	64	141400	Group 1
4	129	318500	Group 1
5	40	357700	Group 1
6	82	122500	Group 2
7	138	133400	Group 2
8	45	151200	Group 2
9	81	164000	Group 2
10	119	195400	Group 2
11	83	202300	Group 2
12	32	208900	Group 2
13	74	220600	Group 2
14	140	233200	Group 2
15	105	341000	Group 2

---

This report lists the 15 values from Price (10% of all values) that were selected along with their corresponding original row numbers and group assignments. The report is sorted first by assigned group and then by value.

## Stored Random Sample Data

Row	Row_1	Sample_1	Group_1
1	71	53300	Group 1
2	20	101800	Group 1
3	64	141400	Group 1
4	129	318500	Group 1
5	40	357700	Group 1
6	82	122500	Group 2
7	138	133400	Group 2
8	45	151200	Group 2
9	81	164000	Group 2
10	119	195400	Group 2
11	83	202300	Group 2
12	32	208900	Group 2
13	74	220600	Group 2
14	140	233200	Group 2
15	105	341000	Group 2

---

## Simple Random Sampling with Group Assignment

In columns 21 and 22 of the Data Table on the Data Window, you'll find the random sample data stored in collapsed format. The stored data includes the sampled values ("Sample\_1") and the original row numbers ("Row\_1") of the values selected from Price. The randomly assigned groups are stored in column 23 ("Group\_1").

The newly created Note for column "Sample\_1" contains the following information about the sampled data.

## Simple Random Sample Storage Column 1 of 1

-----  
Number of Values Selected: 15 of 150 (10%)  
Dataset Sampling Type: Values Randomly Selected from Price  
User-Entered Random Seed\*: 22077

Number of Groups Assigned: 2  
Sample Allocation among Groups: Custom (1 2)  
Number of Values Assigned to Each Group: "Group 1" (5), "Group 2" (10)

\* You can use this random seed to replicate this random sample, but the original dataset, number of random samples to generate, and sampling options must be the same.