

Chapter 125

Simple Random Sampling

Introduction

The Simple Random Sampling tool in **NCSS** can be used to quickly generate K independent random samples from a dataset, where each random sample has N items. 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 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.

Example – Result of randomly selecting 4 values from column Y

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

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) from a dataset

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

- 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)

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

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”

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

Example – Result of randomly selecting 4 values from column Y with Selected = “Yes” and Not Selected = “[Blank]”

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

Simple Random Sampling

- 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

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

Example – Result of randomly selecting 4 values from column Y

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

Simple Random Sampling

- 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 with stored selection data not sorted

Row	X	Y	Z	Random Sample (Collapsed Format)
1	10			1
2	0	1		7
3	4	5		2
4	7	2		10
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 selected value and with original row numbers stored in an adjacent column

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

This example will demonstrate how to take a simple random sample of 30 rows from a dataset. 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 will use a random seed of 60677.

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 procedure options

- Find and open the **Simple Random Sampling** 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.

Variables 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)

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

Storage: The random sample selection data was stored in column 21 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

Selected Item	Selected Row
---------------	--------------

1	15
2	18
3	23
4	24
5	25
6	29
7	32
8	33
9	34
10	37
11	40
12	43
13	52
14	54
15	59
16	66
17	69
18	74
19	78
20	86
.	.
.	.
.	.

This report lists in collapsed format the 30 rows that were selected.

Stored Random Sample Data

Row	Sample_1
1	0
2	0
3	0
.	.
.	.
15	1
16	0
17	0
18	1
.	.
.	.
145	0
146	1
147	0
148	0
149	0
150	1

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. You can now set a filter on this column 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

* 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

This example will demonstrate how to take a simple random sample from a column in a dataset. 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 value.

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

This example will also demonstrate how to randomly assign selected values to groups with sequential balance using the Block Randomization method in the Randomization Lists procedure.

If you want to randomly assign the selected values to G groups without regard for sequential balance over time, you can accomplish this in one step using the Simple Random Sampling with Group Assignment procedure.

Setup

To run this part of the 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 procedure options

- Find and open the **Simple Random Sampling** 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	Enter the Percent of Rows/Values to Select
Percent of Rows/Values to Select.....	10
Dataset Sampling Type	Randomly Select Values from a Column
Column to Sample	Price
Random Seed.....	22077 (for reproducibility)
Reports Tab	
Sort Random Sample List(s).....	Checked
Sort Type	By Selected Row/Value

Simple Random Sampling

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 Sales Price
Number of Random Sample Replicates	1
User-Entered Random Seed*	22077

Storage: The random sample selection data was stored in columns 22 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. 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 and 22.

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 Sales Price
 User-Entered Random Seed 22077

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

This report lists in collapsed format the 15 values from Price (10% of all values) that were selected along with their corresponding original row numbers. The report is sorted by value.

Stored Random Sample Data

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

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.

Simple Random Sampling

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

* 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.

Assigning Randomly Selected Values to Groups using the Randomization Lists Procedure

Continuing with this example, suppose you want to assign the 15 randomly selected values to groups with sequential balance. This can be accomplished using the Block Randomization method in the Randomization Lists procedure.

If you want to randomly assign the selected values to G groups without regard for sequential balance over time, you can accomplish this in one step using the Simple Random Sampling with Group Assignment procedure.

The purpose of this example is to demonstrate how the Randomization Lists procedure may be used in conjunction with the Simple Random Sampling procedure to assign actual sampled items to groups. The Randomization Lists procedure contains additional randomization algorithms for assigning items to groups that are not available in the Simple Random Sampling with Group Assignment procedure and may be of interest to the user.

For reproducibility, we will use a random seed of 60502 in the Randomization Lists procedure.

Setup

To run this part of the example, complete the following steps:

4 Specify the Randomization Lists procedure options

- Find and open the **Randomization Lists** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 5** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Randomization Tab

Randomization Algorithm.....	Block Randomization
Random Seed.....	60502 (for reproducibility)
Block Size Multiplier(s)	1 2
Block-Subject Allocation	Random
Constrain the actual block allocation... ..	Unchecked
Search for a randomization list... ..	Unchecked
Maximum Search Iterations	1000

Simple Random Sampling

Sample Size Entry Type **Enter the Overall Total Sample Size for the Study**
 Overall Total Sample Size **15**
 Number of Treatment Groups **3**
 Title or Name **Group**
 Assign Equal Sample Sizes to All Treatment Groups.... **Checked**
 Treatment Group 1 Label..... **A**
 Treatment Group 2 Label..... **B**
 Treatment Group 3 Label..... **C**

Centers (Strata) Tab

Number of Centers (Strata) **1 (Not Stratified)**
 Number of Additional Factors **0**

List Options Tab

Include a Column of Sequence Numbers **Checked**
 Include a Column of Subject ID's..... **Unchecked**
 Include a Column of Block Identifiers..... **Checked**
 Include a Column of Combined Stratification Codes..... **Unchecked**
 Include a Column of Abbreviated Treatment Codes **Unchecked**
 Include a Column of Unique Randomization Codes **Unchecked**

Reports Tab

Show Summary **Checked**
 Show Randomization List **Checked**
 Show Randomization List Details **Unchecked**
 Show References **Checked**

Storage Tab

Store Randomization List..... **Checked**
 Store Randomization List Data Starting In..... **<Empty>**
 Update output column names... **Checked**

5 Run the procedure

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

Simple Random Sampling

Randomization List Summary

Randomization List Summary

Randomization Algorithm	Block Randomization
User-Entered Random Seed	60502
Block Size Multiplier(s)	1 2
Block Size(s)	3 6
Total Number of Blocks	3

Overall Total Sample Size	15
Number of Treatment Groups	3

Group	Group Code	Sample Size	Allocation %	
			Actual	Target
A	A	5	33.33%	33.33%
B	B	5	33.33%	33.33%
C	C	5	33.33%	33.33%

This report displays the summary of the randomization list, including block information.

Randomization List

Randomization List

Randomization Algorithm	Block Randomization (Block Sizes = 3 6)
User-Entered Random Seed	60502

Sequence	Block	Group
1	1	A
2	1	A
3	1	C
4	1	C
5	1	B
6	1	B
7	2	C
8	2	A
9	2	C
10	2	B
11	2	A
12	2	B
13	3	A
14	3	C
15	3	B

The report lists the group to which each item (denoted by Sequence) should be assigned.

Simple Random Sampling

Stored Randomization List Data

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

In columns 21 and 22 of the Data Table on the Data Window, you'll find the random sample data stored in collapsed format from the Simple Random Sampling procedure. In columns 23 through 25 you'll find the stored list from the Randomization Lists procedure, including the assigned group for each selected value.