

Chapter 128

Stratified Random Sampling with Group Assignment

Introduction

In stratified random sampling, samples are drawn from a population that has been partitioned into subpopulations (or strata) based on shared characteristics (e.g., gender, age, location, etc.). Each individual stratum is sampled independently of all other strata. The Stratified Random Sampling with Group Assignment tool in **NCSS** can be used to quickly generate K independent stratified random samples from a dataset, where each random sample has N items divided among H strata, and randomly assign the selected items to G 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 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 Stratified 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 Stratified Random Sampling procedure to generate the random sample.

Stratified Random Sampling with Group Assignment

Example – Result of randomly selecting 8 rows from a dataset with S1 as stratification column, sample allocation among strata proportional to stratum size, and random assignment to 2 groups (“Group 1” and “Group 2”)

Row	S1	S2	X	Y	Z	Random Sample (Binary Format)	Group
1	A	1	43			Not Selected	
2	A	1	2	3		Selected	Group 2
3	A	1	40	66		Selected	Group 2
4	A	1	92		30	Not Selected	
5	A	1	45	0	12	Not Selected	
6	A	1	29	46	44	Not Selected	
7	A	2	57	44		Selected	Group 1
8	A	2	48	12	38	Not Selected	
9	A	2	20		2	Not Selected	
10	A	2	27			Not Selected	
11	A	2	28	84	98	Not Selected	
12	A	2	35	97	86	Selected	Group 1
13	B	1	3	34		Not Selected	
14	B	1	78		66	Not Selected	
15	B	1	52	58	59	Not Selected	
16	B	2	16	18	51	Selected	Group 1
17	B	2	3			Not Selected	
18	B	2	73	61		Selected	Group 2
19	C	1	79	55	88	Not Selected	
20	C	1	95	65	28	Selected	Group 2
21	C	1	67			Not Selected	
22	C	2	61	38	26	Not Selected	
23	C	2	25		31	Selected	Group 1
24	C	2	13	79	99	Not Selected	

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 6 rows from a dataset with S1 as the stratification column, sample allocation among strata proportional to stratum size, and assigning to 2 groups

Row	S1	X	Y	Z	Random Sample (Binary Format)	Random Sample	Group
1	A	43			Not Selected		
2	A	2	3		Selected	2	Group 1
3	A	40	66		Selected	3	Group 2
4	A	92		30	Not Selected		
5	A	45	0	12	Not Selected		
6	A	29	46	44	Not Selected		
7	A	57	44		Selected	7	Group 1
8	A	48	12	38	Selected	8	Group 2
9	B	3	34		Not Selected		
10							
11	B	78		66	Not Selected		
12	B	52	58	59	Selected	12	Group 2
13	B	16	18	51	Selected	13	Group 1

Stratified Random Sampling with Group Assignment

- **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 6 values from column Y with S1 as the stratification column, sample allocation among strata proportional to stratum size, and assigning to 2 groups

Row	S1	X	Y	Z	Random Sample (Binary Format)	Random Sample	Group
1	A	43					
2	A	2	3		Not Selected		
3	A	40	66		Selected	66	Group 1
4	A	92		30			
5	A	45	0	12	Selected	0	Group 1
6	A	29	46	44	Selected	46	Group 2
7	A	57	44		Selected	44	Group 2
8	A	48	12	38	Not Selected		
9	B	3	34		Selected	34	Group 1
10							
11	B	78		66			
12	B	52	58	59	Selected	58	Group 2
13	B	16	18	51	Not Selected		

Stratification Columns

You may select up to 8 categorical data columns for stratification. The strata will be created as the combination of all unique values in the stratification columns. For example, if stratification column 1 has three unique values (A, B, and C) and stratification column 2 has two unique values (1 and 2), then the six resulting strata would be A1, A2, B1, B2, C1, and C2.

Random samples are generated independently within each stratum. Any rows for which any of the stratification values are missing will not be eligible for sampling.

The number of items sampled from each stratum is controlled by the *Sample Allocation among Strata* and *Sample Size* options.

In the example below, S1 (with values A, B, and C) is entered as the stratification column. There are twice as many A's as B's and C's. Since the sample allocation among strata is proportional to sample size, there are twice as many values selected from stratum A than from B and C.

Stratified Random Sampling with Group Assignment

Example – Result of randomly selecting 8 rows from a dataset with S1 as stratification column, sample allocation among strata proportional to stratum size, and random assignment to 2 groups

Row	S1	S2	X	Y	Z	Random Sample (Binary Format)	Group
1	A	1	43			Not Selected	
2	A	1	2	3		Selected	Group 2
3	A	1	40	66		Selected	Group 2
4	A	1	92		30	Not Selected	
5	A	1	45	0	12	Not Selected	
6	A	1	29	46	44	Not Selected	
7	A	2	57	44		Selected	Group 1
8	A	2	48	12	38	Not Selected	
9	A	2	20		2	Not Selected	
10	A	2	27			Not Selected	
11	A	2	28	84	98	Not Selected	
12	A	2	35	97	86	Selected	Group 1
13	B	1	3	34		Not Selected	
14	B	1	78		66	Not Selected	
15	B	1	52	58	59	Not Selected	
16	B	2	16	18	51	Selected	Group 1
17	B	2	3			Not Selected	
18	B	2	73	61		Selected	Group 2
19	C	1	79	55	88	Not Selected	
20	C	1	95	65	28	Selected	Group 2
21	C	1	67			Not Selected	
22	C	2	61	38	26	Not Selected	
23	C	2	25		31	Selected	Group 1
24	C	2	13	79	99	Not Selected	

Stratified Random Sampling with Group Assignment

In the example below, S1 (with values A, B, and C) and S2 (with values 1 and 2) are entered as the stratification columns. There are 6 resulting strata (A1, A2, B1, B2, C1, and C2). With equal sample allocation among strata, two values are selected from each stratum when randomly sampling 12 rows.

Example – Result of randomly selecting 12 rows from the dataset with S1 and S2 as stratification variables, equal sample allocation among strata, and random assignment to 2 groups

Row	S1	S2	X	Y	Z	Random Sample (Binary Format)	Group
1	A	1	43			Not Selected	
2	A	1	2	3		Selected	Group 1
3	A	1	40	66		Selected	Group 2
4	A	1	92		30	Not Selected	
5	A	1	45	0	12	Not Selected	
6	A	1	29	46	44	Not Selected	
7	A	2	57	44		Selected	Group 2
8	A	2	48	12	38	Not Selected	
9	A	2	20		2	Not Selected	
10	A	2	27			Not Selected	
11	A	2	28	84	98	Not Selected	
12	A	2	35	97	86	Selected	Group 1
13	B	1	3	34		Selected	Group 2
14	B	1	78		66	Not Selected	
15	B	1	52	58	59	Selected	Group 1
16	B	2	16	18	51	Selected	Group 1
17	B	2	3			Not Selected	
18	B	2	73	61		Selected	Group 2
19	C	1	79	55	88	Not Selected	
20	C	1	95	65	28	Selected	Group 1
21	C	1	67			Selected	Group 2
22	C	2	61	38	26	Not Selected	
23	C	2	25		31	Selected	Group 1
24	C	2	13	79	99	Selected	Group 2

Sample Allocation among Strata

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

The options are

- **Equal**

An equal number of items is selected from each stratum.

In the example below, S1 (with values A, B, and C) is entered as the stratification column. Even though there are twice as many A's as B's and C's three rows are selected from each stratum since the sample allocation among strata is set to "Equal".

Example – Result of randomly selecting 9 rows from the dataset with S1 as the stratification column, Sample Allocation among Strata = "Equal", and random assignment to 2 groups

Row	S1	S2	X	Y	Z	Random Sample (Binary Format)	Group
1	A	1	43			Not Selected	
2	A	1	2	3		Selected	Group 2
3	A	1	40	66		Not Selected	
4	A	1	92		30	Not Selected	
5	A	1	45	0	12	Not Selected	
6	A	1	29	46	44	Not Selected	
7	A	2	57	44		Selected	Group 2
8	A	2	48	12	38	Not Selected	
9	A	2	20		2	Not Selected	
10	A	2	27			Not Selected	
11	A	2	28	84	98	Not Selected	
12	A	2	35	97	86	Selected	Group 1
13	B	1	3	34		Not Selected	
14	B	1	78		66	Not Selected	
15	B	1	52	58	59	Selected	Group 1
16	B	2	16	18	51	Selected	Group 2
17	B	2	3			Selected	Group 1
18	B	2	73	61		Not Selected	
19	C	1	79	55	88	Not Selected	
20	C	1	95	65	28	Not Selected	
21	C	1	67			Selected	Group 1
22	C	2	61	38	26	Selected	Group 2
23	C	2	25		31	Selected	Group 2
24	C	2	13	79	99	Not Selected	

Stratified Random Sampling with Group Assignment

- Proportional to Stratum Size**

The relative strata sample size allotments are determined by the size of each stratum. The relative sample size is larger for larger strata.

In the example below, S1 (with values A, B, and C) is entered as the stratification column. There are twice as many A's as B's and C's. Since the sample allocation among strata is proportional to sample size, there are twice as many values selected from stratum A than from B and C.

Example – Result of randomly selecting 8 rows from the dataset with S1 as the stratification column, sample allocation among strata proportional to stratum size, and random assignment to 2 groups

Row	S1	S2	X	Y	Z	Random Sample (Binary Format)	Group
1	A	1	43			Not Selected	
2	A	1	2	3		Selected	Group 2
3	A	1	40	66		Selected	Group 2
4	A	1	92		30	Not Selected	
5	A	1	45	0	12	Not Selected	
6	A	1	29	46	44	Not Selected	
7	A	2	57	44		Selected	Group 1
8	A	2	48	12	38	Not Selected	
9	A	2	20		2	Not Selected	
10	A	2	27			Not Selected	
11	A	2	28	84	98	Not Selected	
12	A	2	35	97	86	Selected	Group 1
13	B	1	3	34		Not Selected	
14	B	1	78		66	Not Selected	
15	B	1	52	58	59	Not Selected	
16	B	2	16	18	51	Selected	Group 1
17	B	2	3			Not Selected	
18	B	2	73	61		Selected	Group 2
19	C	1	79	55	88	Not Selected	
20	C	1	95	65	28	Selected	Group 2
21	C	1	67			Not Selected	
22	C	2	61	38	26	Not Selected	
23	C	2	25		31	Selected	Group 1
24	C	2	13	79	99	Not Selected	

Stratified Random Sampling with Group Assignment

• **Custom**

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

In the example below, S1 (with values A, B, and C) is entered as the stratification column. There are twice as many A's as B's and C's. The custom sample allocation among strata is "1 2 1", such that there are twice as many rows sampled from stratum B (4) as from strata A and C (each with 2).

Example – Result of randomly selecting 8 rows from the dataset with S1 as the stratification column, sample allocation among strata set to custom (1 2 1), and random assignment to 2 groups

Row	S1	S2	X	Y	Z	Random Sample (Binary Format)	Group
1	A	1	43			Not Selected	
2	A	1	2	3		Not Selected	
3	A	1	40	66		Not Selected	
4	A	1	92		30	Not Selected	
5	A	1	45	0	12	Not Selected	
6	A	1	29	46	44	Not Selected	
7	A	2	57	44		Selected	Group 2
8	A	2	48	12	38	Not Selected	
9	A	2	20		2	Not Selected	
10	A	2	27			Not Selected	
11	A	2	28	84	98	Not Selected	
12	A	2	35	97	86	Selected	Group 1
13	B	1	3	34		Not Selected	
14	B	1	78		66	Not Selected	
15	B	1	52	58	59	Selected	Group 2
16	B	2	16	18	51	Selected	Group 1
17	B	2	3			Selected	Group 2
18	B	2	73	61		Selected	Group 1
19	C	1	79	55	88	Not Selected	
20	C	1	95	65	28	Selected	Group 2
21	C	1	67			Not Selected	
22	C	2	61	38	26	Not Selected	
23	C	2	25		31	Selected	Group 1
24	C	2	13	79	99	Not Selected	

Sample Allocation among Groups

This option is used along with the sample size and sample allocation among strata to determine how many sampled items will be assigned to each group within each stratum. 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 8 rows from the dataset with S1 as the stratification column, sample allocation among strata proportional to stratum size, and random assignment to 2 groups with equal allocation among groups

Row	S1	S2	X	Y	Z	Random Sample (Binary Format)	Group
1	A	1	43			Not Selected	
2	A	1	2	3		Selected	Group 2
3	A	1	40	66		Selected	Group 2
4	A	1	92		30	Not Selected	
5	A	1	45	0	12	Not Selected	
6	A	1	29	46	44	Not Selected	
7	A	2	57	44		Selected	Group 1
8	A	2	48	12	38	Not Selected	
9	A	2	20		2	Not Selected	
10	A	2	27			Not Selected	
11	A	2	28	84	98	Not Selected	
12	A	2	35	97	86	Selected	Group 1
13	B	1	3	34		Not Selected	
14	B	1	78		66	Not Selected	
15	B	1	52	58	59	Not Selected	
16	B	2	16	18	51	Selected	Group 1
17	B	2	3			Not Selected	
18	B	2	73	61		Selected	Group 2
19	C	1	79	55	88	Not Selected	
20	C	1	95	65	28	Selected	Group 2
21	C	1	67			Not Selected	
22	C	2	61	38	26	Not Selected	
23	C	2	25		31	Selected	Group 1
24	C	2	13	79	99	Not Selected	

Stratified Random Sampling with Group Assignment

- 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 9 rows from the dataset with S1 as the stratification column, equal sample allocation among strata, and random assignment to 2 groups with custom (1 2) allocation among groups such that there are twice as many assigned to Group 2 as are assigned to Group 1

Row	S1	S2	X	Y	Z	Random Sample (Binary Format)	Group
1	A	1	43			Not Selected	
2	A	1	2	3		Selected	Group 2
3	A	1	40	66		Not Selected	
4	A	1	92		30	Not Selected	
5	A	1	45	0	12	Not Selected	
6	A	1	29	46	44	Not Selected	
7	A	2	57	44		Selected	Group 2
8	A	2	48	12	38	Not Selected	
9	A	2	20		2	Not Selected	
10	A	2	27			Not Selected	
11	A	2	28	84	98	Not Selected	
12	A	2	35	97	86	Selected	Group 1
13	B	1	3	34		Not Selected	
14	B	1	78		66	Not Selected	
15	B	1	52	58	59	Not Selected	
16	B	2	16	18	51	Selected	Group 1
17	B	2	3			Selected	Group 2
18	B	2	73	61		Selected	Group 2
19	C	1	79	55	88	Not Selected	
20	C	1	95	65	28	Selected	Group 2
21	C	1	67			Not Selected	
22	C	2	61	38	26	Not Selected	
23	C	2	25		31	Selected	Group 1
24	C	2	13	79	99	Selected	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 6 rows from a dataset with S1 as the stratification column, sample allocation among strata proportional to stratum size with Selected = “1” and Not Selected = “0”, and random assignment to 2 groups

Row	S1	X	Y	Z	Random Sample (Binary Format)	Group
1	A	43			0	
2	A	2	3		1	Group 1
3	A	40	66		1	Group 2
4	A	92		30	0	
5	A	45	0	12	0	
6	A	29	46	44	0	
7	A	57	44		1	Group 1
8	A	48	12	38	1	Group 2
9	B	3	34		0	
10	B	78		66	0	
11	B	52	58	59	1	Group 2
12	B	16	18	51	1	Group 1

Example – Result of randomly selecting 6 values from column Y with S1 as the stratification column, sample allocation among strata proportional to stratum size with Selected = “Yes” and Not Selected = “[Blank]”, and random assignment to 2 groups

Row	S1	X	Y	Z	Random Sample (Binary Format)	Group
1	A	43				
2	A	2	3			
3	A	40	66		Yes	Group 1
4	A	92		30		
5	A	45	0	12	Yes	Group 1
6	A	29	46	44	Yes	Group 2
7	A	57	44		Yes	Group 2
8	A	48	12	38		
9	B	3	34		Yes	Group 1
10	B	78		66		
11	B	52	58	59	Yes	Group 2
12	B	16	18	51		

Stratified 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 6 rows from a dataset with S1 as the stratification column, sample allocation among strata proportional to stratum size, and random assignment to 2 groups

Row	S1	X	Y	Z	Random Sample	Group
1	A	43				
2	A	2	3		2	Group 1
3	A	40	66		3	Group 2
4	A	92		30		
5	A	45	0	12		
6	A	29	46	44		
7	A	57	44		7	Group 1
8	A	48	12	38	8	Group 2
9	B	3	34			
10	B	78		66		
11	B	52	58	59	11	Group 2
12	B	16	18	51	12	Group 1

Example – Result of randomly selecting 6 values from column Y with S1 as the stratification column, sample allocation among strata proportional to stratum size, and random assignment to 2 groups

Row	S1	X	Y	Z	Random Sample	Group
1	A	43				
2	A	2	3			
3	A	40	66		66	Group 1
4	A	92		30		
5	A	45	0	12	0	Group 1
6	A	29	46	44	46	Group 2
7	A	57	44		44	Group 2
8	A	48	12	38		
9	B	3	34		34	Group 1
10	B	78		66		
11	B	52	58	59	58	Group 2
12	B	16	18	51		

Stratified 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 6 rows from a dataset with S1 as the stratification column, sample allocation among strata proportional to stratum size, the random sample not sorted, and random assignment to 2 groups

Row	S1	X	Y	Z	Value from S1 (Collapsed Format)	Random Sample (Collapsed Format)	Group (Collapsed Format)
1	A	43			A	7	Group 1
2	A	2	3		A	2	Group 1
3	A	40	66		A	3	Group 2
4	A	92		30	A	8	Group 2
5	A	45	0	12	B	12	Group 1
6	A	29	46	44	B	11	Group 2
7	A	57	44				
8	A	48	12	38			
9	B	3	34				
10	B	78		66			
11	B	52	58	59			
12	B	16	18	51			

Example – Result of randomly selecting 6 values from column Y with S1 as the stratification column, sample allocation among strata proportional to stratum size, output sorted by assigned group and then by selected value, with corresponding stratification values and row numbers stored in adjacent columns, and random assignment to 2 groups

Row	S1	X	Y	Z	Value from S1 (Collapsed Format)	Original Row Number	Random Sample (Collapsed Format)	Group (Collapsed Format)
1	A	43			A	5	0	Group 1
2	A	2	3		A	3	66	Group 1
3	A	40	66		A	7	44	Group 2
4	A	92		30	A	6	46	Group 2
5	A	45	0	12	B	9	34	Group 1
6	A	29	46	44	B	11	58	Group 2
7	A	57	44					
8	A	48	12	38				
9	B	3	34					
10	B	78		66				
11	B	52	58	59				
12	B	16	18	51				

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

This example will demonstrate how to take a stratified random sample of 30 rows (across all strata) from a dataset and randomly assign the selected rows to 3 groups (A, B, or C) with equal group allocation. The sampling will be stratified by State, with equal sample allocation among strata. 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 102203.

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 Stratified 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 Stratified Random Sampling with Group Assignment procedure options

- Find and open the **Stratified 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.....	102203 (for reproducibility)
Stratification Column(s)	State
Sample Allocation among Strata.....	Equal
Number of Groups to Assign.....	3
Group Names	Custom
Custom Group Names	A B C
Sample Allocation among Groups	Equal

Stratified Random Sampling with Group Assignment

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

Stratified 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*	102203
Stratification Column	State
Number of Strata	2
Sample Allocation among Strata	Equal
Number of Rows Selected by Stratum	State="Nev": 15 of 88 (17.045%) State="Vir": 15 of 62 (24.194%)
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. 15 of 88 rows were selected from State = "Nev" and 15 of 62 rows were selected from State = "Vir". The random sample itself and group assignments were stored in the Data Table in columns 21 and 22.

Random Sample List Report

Stratified Random Sample 1 of 1

Number of Rows Selected	30 of 150 (20%)
Dataset Sampling Type	Rows Randomly Selected
User-Entered Random Seed	102203
Stratification Column	State
Number of Rows Assigned to Each Group	"A" (10), "B" (10), "C" (10)

Selected Item	Value from Stratification Column State	Selected Row	Assigned Group
1	Nev	4	A
2	Nev	7	C
3	Nev	35	C
4	Nev	44	C
5	Nev	48	B
6	Nev	80	A
7	Nev	101	A
8	Nev	104	B
9	Nev	105	B
10	Nev	107	B
.	.	.	.
.	.	.	.
.	.	.	.

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

Stored Random Sample Data

Row	State	Sample_1	Group_1
1	Nev	0	
2	Nev	0	
3	Vir	1	B
4	Nev	1	A
5	Nev	0	
6	Nev	0	
7	Nev	1	C
.	.	.	.
.	.	.	.
.	.	.	.
149	Nev	1	C
150	Vir	0	

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.

Stratified Random Sampling with Group Assignment

The newly created Note for column "Sample_1" contains the following information about the sampled data.

Stratified 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*: 102203

Stratification Column: State
Number of Strata: 2
Sample Allocation among Strata: Equal
Number of Rows Selected by Stratum: State="Nev": 15 of 88 (17.045%); State="Vir": 15 of 62 (24.194%)

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 Stratified Random Sample of 20% 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 stratified 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 20% of the values in the column called Price. The sampling will be stratified by State, with equal sample allocation among strata. The sample selection data will be stored in collapsed format in the Data Table with stratification values and row numbers also stored for reference to the original dataset. The resulting sample data will be sorted by assigned group and then by value within each stratum.

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 Stratified 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 Stratified Random Sampling with Group Assignment procedure options

- Find and open the **Stratified 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	Enter the Percent of Rows/Values to Select
Percent of Rows/Values to Select.....	20
Dataset Sampling Type	Randomly Select Values from a Column
Column to Sample	Price
Random Seed.....	22077 (for reproducibility)
Stratification Column(s)	State
Sample Allocation among Strata.....	Equal
Number of Groups to Assign.....	2
Group Names	Custom
Custom Group Names	"Group 1" "Group 2"
Sample Allocation among Groups	Custom
Custom Sample Allocation.....	1 2

Stratified Random Sampling with Group Assignment

Reports Tab

Sort Random Sample List(s) within Strata **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 stratification values **Checked**
 Store the original row number of each selected **Checked**
 value

Report Options (*in the Toolbar*)

Variable Labels **Column Names**

Random Sampling Summary Report

Stratified Random Sampling Summary

Number of Values Selected	30 of 150 (20%)
Dataset Sampling Type	Values Randomly Selected from Price
Number of Random Sample Replicates	1
User-Entered Random Seed*	22077
Stratification Column	State
Number of Strata	2
Sample Allocation among Strata	Equal
Number of Values Selected by Stratum	State="Nev": 15 of 88 (17.045%) State="Vir": 15 of 62 (24.194%)
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 24 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 30 corresponds to 20% of all 150 available rows with data. The random sample itself was stored in the Data Table in columns 21 through 24.

Random Sample List Report

Stratified Random Sample 1 of 1

Number of Values Selected	30 of 150 (20%)
Dataset Sampling Type	Values Randomly Selected from Price
User-Entered Random Seed	22077
Stratification Column	State
Number of Values Assigned to Each Group	"Group 1" (10), "Group 2" (20)

Selected Item	Value from Stratification Column State	Original Row	Value Selected from Price	Assigned Group
1	Nev	71	53300	Group 1
2	Nev	138	133400	Group 1
3	Nev	123	150100	Group 1
4	Nev	108	198000	Group 1
5	Nev	29	238900	Group 1
6	Nev	79	26300	Group 2
7	Nev	57	102200	Group 2
8	Nev	6	147500	Group 2
9	Nev	49	167000	Group 2
10	Nev	84	183300	Group 2
11	Nev	140	233200	Group 2
12	Nev	139	247000	Group 2
13	Nev	129	318500	Group 2
14	Nev	77	360100	Group 2
15	Nev	8	395700	Group 2
16	Vir	70	119200	Group 1
17	Vir	47	165900	Group 1
18	Vir	119	195400	Group 1
19	Vir	122	241800	Group 1
20	Vir	98	247500	Group 1
21	Vir	55	32900	Group 2
22	Vir	20	101800	Group 2
23	Vir	69	141200	Group 2
24	Vir	61	158400	Group 2
25	Vir	32	208900	Group 2
26	Vir	74	220600	Group 2
27	Vir	54	241300	Group 2
28	Vir	86	291700	Group 2
29	Vir	116	328300	Group 2
30	Vir	34	335300	Group 2

This report lists the 30 values from Price (20% of all values) that were selected along with their corresponding strata, original row numbers, and group assignments. The report is sorted by assigned group and then by value.

Stored Random Sample Data

Row	Strat_Col_1_State	Row_1	Sample_1	Group_1
1	Nev	71	53300	Group 1
2	Nev	138	133400	Group 1
3	Nev	123	150100	Group 1
4	Nev	108	198000	Group 1
5	Nev	29	238900	Group 1
6	Nev	79	26300	Group 2
7	Nev	57	102200	Group 2
8	Nev	6	147500	Group 2
9	Nev	49	167000	Group 2
10	Nev	84	183300	Group 2
11	Nev	140	233200	Group 2
12	Nev	139	247000	Group 2
13	Nev	129	318500	Group 2
14	Nev	77	360100	Group 2
15	Nev	8	395700	Group 2
16	Vir	70	119200	Group 1
17	Vir	47	165900	Group 1
18	Vir	119	195400	Group 1
19	Vir	122	241800	Group 1
20	Vir	98	247500	Group 1
21	Vir	55	32900	Group 2
22	Vir	20	101800	Group 2
23	Vir	69	141200	Group 2
24	Vir	61	158400	Group 2
25	Vir	32	208900	Group 2
26	Vir	74	220600	Group 2
27	Vir	54	241300	Group 2
28	Vir	86	291700	Group 2
29	Vir	116	328300	Group 2
30	Vir	34	335300	Group 2

In columns 21 through 23 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"), the stratification values ("Strat_Col_1_State"), and the original row numbers ("Row_1") of the values selected from Price. The randomly assigned groups are stored in column 24 ("Group_1").

The newly created Note for column "Sample_1" contains the following information about the sampled data.

Stratified Random Sample Storage Column 1 of 1

 Number of Values Selected: 30 of 150 (20%)
 Dataset Sampling Type: Values Randomly Selected from Price
 User-Entered Random Seed*: 22077

Stratification Column: State
 Number of Strata: 2
 Sample Allocation among Strata: Equal
 Number of Values Selected by Stratum: State="Nev": 15 of 88 (17.045%); State="Vir": 15 of 62 (24.194%)

Number of Groups Assigned: 2
 Sample Allocation among Groups: Custom (1 2)
 Number of Values Assigned to Each Group: "Group 1" (10), "Group 2" (20)

* 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.