Chapter 271

Cluster Randomization – Create Cluster Proportions Dataset

Introduction

A *cluster randomization trial* occurs when whole groups or *clusters* of individuals are treated together. Examples of such clusters are clinics, hospitals, cities, schools, or neighborhoods. In the two-group case, each cluster is randomized to receive a particular treatment. That is, all individuals in a cluster receive the same treatment. One way to analyze the data from such a design is to form the means of each cluster and then analyze those means using a two-sample t-test, an unequal-variance two-sample t-test, or a regression analysis. When the endpoint is a binary variable coded as a zero or one, the mean is the proportion.

This procedure creates a new dataset containing the cluster proportions from an original dataset containing information on individuals. This summarized dataset can then be analyzed further using t-tests or regression analysis.

Cluster-randomized trials are covered in several texts, including Hayes and Moulton (2017), Campbell and Walters (2014), Eldridge and Kerry (2012), Donner and Klar (2000), and Murray (1998).

Data Structure

A dataset analyzed by this procedure requires three variables: a categorical cluster variable, a categorical group variable, and a binary data variable that is coded with either a 0 (no) or a one (yes).

Here is an example of a dataset that can be successfully manipulated with this procedure. The Cluster column gives the cluster identification number. The Group column gives an identification number of the group to which each cluster belongs. All group values in a given cluster should be equal. A Data column (Outcome) gives the endpoint value for each individual. This example dataset is called **ClusRandProps**.

ClusRandProps Dataset (Subset)

Cluster	Group	Outcome
1	1	0
1	1	0
1	1	0
1	1	0
1	1	1
1	1	0
.	.	.
.	.	.
.	.	.

© NCSS, LLC. All Rights Reserved.

Example 1 – Creating a Summarized Dataset from the ClusRandProps Data

This section presents an example of how to summarize the data contained in the ClusRandProps dataset.

Setup

To run this example, complete the following steps:

1 Open the ClusRandProps example dataset

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

2 Specify the Cluster Randomization – Create Cluster Proportions Dataset procedure options

- Find and open the **Cluster Randomization – Create Cluster Proportions Dataset** 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	
Cluster Variable(s)	**Cluster**
Treatment Group Variable	**Group**
Primary Endpoint Variable(s)	**Outcome**
Store the Summary List in a New NCSS Data File	**Checked**
Output File Name	**%mydocs_NCSS%\Data\Cluster Proportions.NCSS**
Cluster Statistics Storage	**Store as Columns**
Automatically Reopen the Current Dataset after the Save Operation Completes	**Checked**

3 Run the procedure

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

© NCSS, LLC. All Rights Reserved.

Summary List Storage Information

Summary List Storage Information

Output Data File Name:	{NCSS Documents Folder}\Data\Cluster Proportions.NCSS
Original Raw Data File:	{Example Data Folder}\ClusRandProps.NCSS
Data Variable(s):	(1) Outcome
Group Variable(s):	(2) Cluster, Group
Summary Statistic(s):	(3) Count, Mean, Sum

This report shows where the new, summarized file is stored.

Summary List of Outcome

Summary List of Outcome

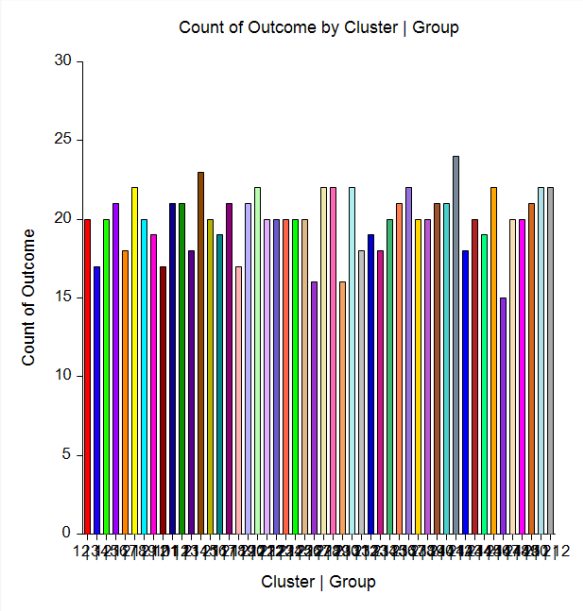
	Statistics for Outcome		
Cluster | Group	**Count**	**Proportion**	**Sum**
1 | 1	20	0.3	6
2 | 2	17	0.4705882	8
3 | 1	20	0.3	6
4 | 2	21	0.3809524	8
5 | 1	18	0.3333333	6
6 | 2	22	0.5	11
7 | 1	20	0.3	6
8 | 2	19	0.4210526	8
9 | 1	17	0.2352941	4
10 | 2	21	0.4285714	9
11 | 1	21	0.2380952	5
12 | 2	18	0.3888889	7
13 | 1	23	0.3913043	9
14 | 2	20	0.4	8
15 | 1	19	0.2105263	4
16 | 2	21	0.6190476	13
17 | 1	17	0.4705882	8
18 | 2	21	0.8095238	17
19 | 1	22	0.4545455	10
20 | 2	20	0.6	12
21 | 1	20	0.45	9
22 | 2	20	0.65	13
23 | 1	20	0.4	8
24 | 2	20	0.55	11
25 | 1	16	0.4375	7
26 | 2	22	0.5454546	12
27 | 1	22	0.3181818	7
28 | 2	16	0.625	10
29 | 1	22	0.5454546	12
30 | 2	18	0.6666667	12
.	.	.	.
.	.	.	.
.	.	.	.

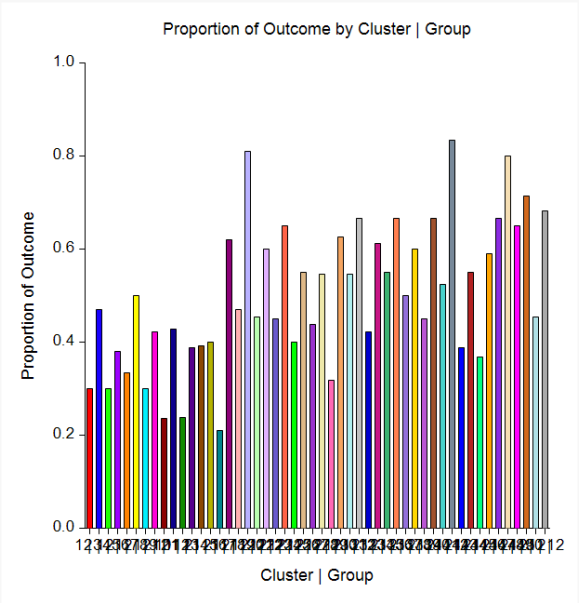
This report displays count, proportion (mean), and sum of the Outcome variable for each cluster.

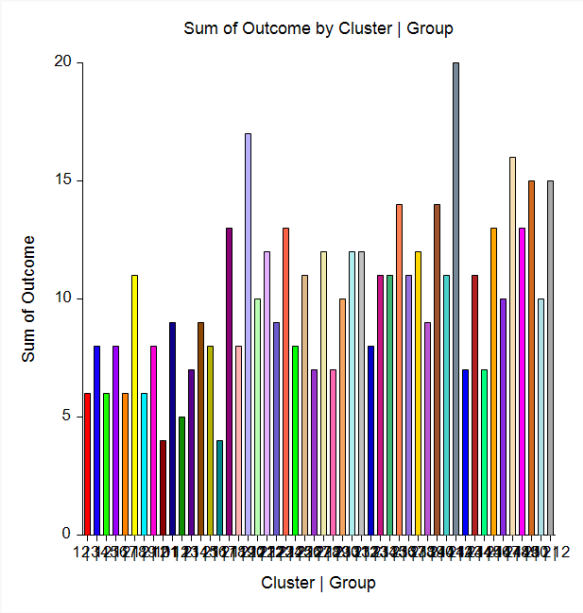
© NCSS, LLC. All Rights Reserved.

Plots of Each Statistic for Outcome

Plots of Each Statistic for Outcome







This report displays the statistics cluster-by-cluster.

© NCSS, LLC. All Rights Reserved.

New Cluster Proportions Dataset

You can open the new Cluster Proportions dataset by using the File menu on the Data Window. The following dataset will appear.

Cluster	Group	Outcome_Count	Outcome_Proportion	Outcome_Sum
1	1	20	0.3	6
2	2	17	0.470588235294118	8
3	1	20	0.3	6
4	2	21	0.380952380952381	8
5	1	18	0.333333333333333	6
6	2	22	0.5	11
7	1	20	0.3	6
8	2	19	0.421052631578947	8
9	1	17	0.235294117647059	4
10	2	21	0.428571428571429	9
11	1	21	0.238095238095238	5
12	2	18	0.388888888888889	7
13	1	23	0.391304347826087	9
14	2	20	0.4	8
15	1	19	0.210526315789474	4
16	2	21	0.619047619047619	13
17	1	17	0.470588235294118	8
18	2	21	0.80952380952381	17
19	1	22	0.454545454545455	10
20	2	20	0.6	12
21	1	20	0.45	9
22	2	20	0.65	13
23	1	20	0.4	8
24	2	20	0.55	11
25	1	16	0.4375	7
26	2	22	0.545454545454545	12
27	1	22	0.318181818181818	7
28	2	16	0.625	10
29	1	22	0.545454545454545	12
30	2	18	0.666666666666667	12
31	1	19	0.421052631578947	8
32	2	18	0.611111111111111	11
33	1	20	0.55	11
34	2	21	0.666666666666667	14
35	1	22	0.5	11
36	2	20	0.6	12
37	1	20	0.45	9
38	2	21	0.666666666666667	14
39	1	21	0.523809523809524	11
40	2	24	0.833333333333333	20
41	1	18	0.388888888888889	7
42	2	20	0.55	11
43	1	19	0.368421052631579	7
44	2	22	0.590909090909091	13
45	1	15	0.666666666666667	10
46	2	20	0.8	16
47	1	20	0.65	13
48	2	21	0.714285714285714	15
49	1	22	0.454545454545455	10
50	2	22	0.681818181818182	15

This dataset can now be analyzed using the Two-Sample T-Test procedure in which the two groups are defined by the Group column and the Response is the Outcome_Proportion column. We suggest that the Randomization test, the Mann-Whitney U test, and/or the Aspin-Welch Unequal-Variance T-Test be used to test for significance.

© NCSS, LLC. All Rights Reserved.

Example 1b – Analyzing the Summarized Dataset

This section continues the analysis begun with Example 1 by analyzing the summarized dataset, **Cluster Proportions**, using the Two-Sample T-Test procedure.

Setup

To run this example, complete the following steps:

1 Open the Cluster Proportions dataset that you just created in Example 1

- From the File menu of the NCSS Data window, select **Cluster Proportions** in the list of recent datasets.

  or

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

2 Specify the Two-Sample T-Test procedure options

- Find and open the **Two-Sample T-Test** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1b** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab	
Data Input Type	**Response Variable(s) and Group Variable(s)**
Response Variable(s)	**Outcome_Proportion**
Group Variable(s)	**Group**

Reports Tab	
Descriptive Statistics and Confidence Intervals of Each Group	**Checked**
Confidence Interval of μ1 - μ2	**Checked**
Equal-Variance T-Test	**Checked**
Unequal-Variance T-Test	**Checked**
Randomization Test	**Checked**
Mann-Whitney U Test (Wilcoxon Rank-Sum Test)	**Checked**
Exact Test	**Checked**
Normal Approximation Test	**Checked**
Normal Approximation Test with Continuity Correction	**Checked**
Tests of Assumptions	**Checked**

Plots Tab	
Probability Plot	**Checked**
Box Plot	**Checked**

© NCSS, LLC. All Rights Reserved.

Report Options (*in the Toolbar*)

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

3 Run the procedure

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

Two-Sample Test Report

Descriptive Statistics

Variable	Count	Mean	Standard Deviation of Data	Standard Error of Mean	T*	95% LCL of Mean	95% UCL of Mean
Group=1	25	0.4143283	0.1202122	0.02404244	2.0639	0.3647071	0.4639495
Group=2	25	0.5908215	0.1293741	0.02587481	2.0639	0.5374185	0.6442245

Descriptive Statistics for the Median

Variable	Count	Median	95% LCL of Median	95% UCL of Median
Group=1	25	0.4210526	0.3333333	0.4545455
Group=2	25	0.6	0.5454546	0.6666667

Two-Sided Confidence Interval for μ1 - μ2

Variance Assumption	DF	Mean Difference	Standard Error	T*	95% LCL of Difference	95% UCL of Difference
Equal	48	-0.1764932	0.0353206	2.0106	-0.24751	-0.1054763
Unequal	47.74	-0.1764932	0.0353206	2.0109	-0.2475199	-0.1054665

Equal-Variance T-Test

Alternative Hypothesis	Mean Difference	Standard Error	T-Statistic	DF	Prob Level	Reject H0 at α = 0.05?
μ1 - μ2 ≠ 0	-0.1764932	0.0353206	-4.9969	48	0.00001	Yes

Aspin-Welch Unequal-Variance T-Test (This is a key report)

Alternative Hypothesis	Mean Difference	Standard Error	T-Statistic	DF	Prob Level	Reject H0 at α = 0.05?
μ1 - μ2 ≠ 0	-0.1764932	0.0353206	-4.9969	47.74	0.00001	Yes

© NCSS, LLC. All Rights Reserved.

Randomization Tests

Alternative Hypothesis: |μ1 - μ2| ≠ 0. This is a Two-Sided Test.
Number of Monte Carlo samples: 10000
Computer-Generated Random Seed: 4585361

Variance Assumption	Prob Level	Reject H0 at α = 0.05?
Equal Variance	0.00020	Yes
Unequal Variance	0.00020	Yes

Mann-Whitney U or Wilcoxon Rank-Sum Test for Difference in Location (This is another a key report)

Group Details

Variable	Mann-Whitney U	Sum of Ranks (W)	Mean of W	Std Dev of W
Group=1	105	430	637.5	51.50416
Group=2	520	845	637.5	51.50416

Number of Sets of Ties = 13, Multiplicity Factor = 168

Test Results

Test Type	Alternative Hypothesis	Z-Value	Prob Level	Reject H0 at α = 0.05?
Exact*	Location Diff. ≠ 0			
Normal Approximation	Location Diff. ≠ 0	-4.0288	0.00006	Yes
Normal Approx. with C.C.	Location Diff. ≠ 0	-4.0191	0.00006	Yes

* The Exact Test is provided only when there are no ties and the sample size is ≤ 20 in both groups.

Tests of the Normality Assumption for Group=1

Normality Test	Test Statistic	Prob Level	Reject H0 of Normality at α = 0.05?
Shapiro-Wilk	0.9718	0.68992	No
Skewness	0.6258	0.53143	No
Kurtosis	-0.0891	0.92904	No
Omnibus (Skewness or Kurtosis)	0.3996	0.81890	No

Tests of the Normality Assumption for Group=2

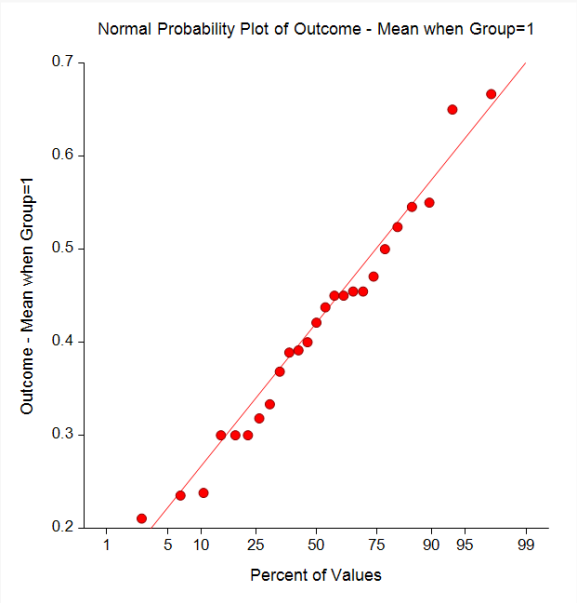
Normality Test	Test Statistic	Prob Level	Reject H0 of Normality at α = 0.05?
Shapiro-Wilk	0.9577	0.37051	No
Skewness	0.0860	0.93150	No
Kurtosis	-0.6137	0.53939	No
Omnibus (Skewness or Kurtosis)	0.3841	0.82528	No

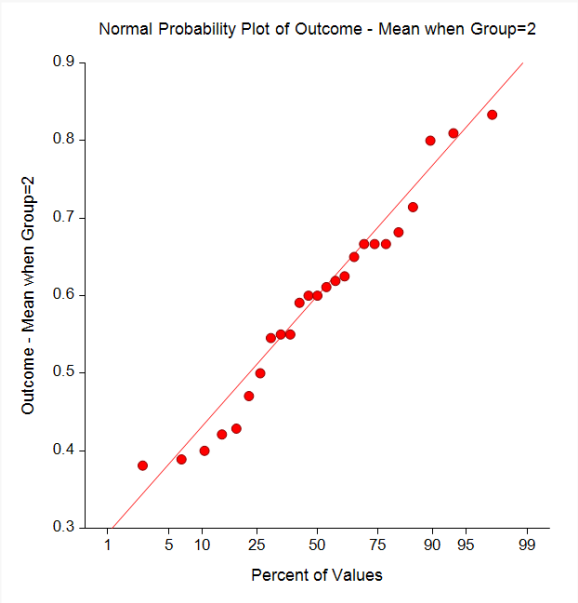
© NCSS, LLC. All Rights Reserved.

Tests of the Equal Variance Assumption

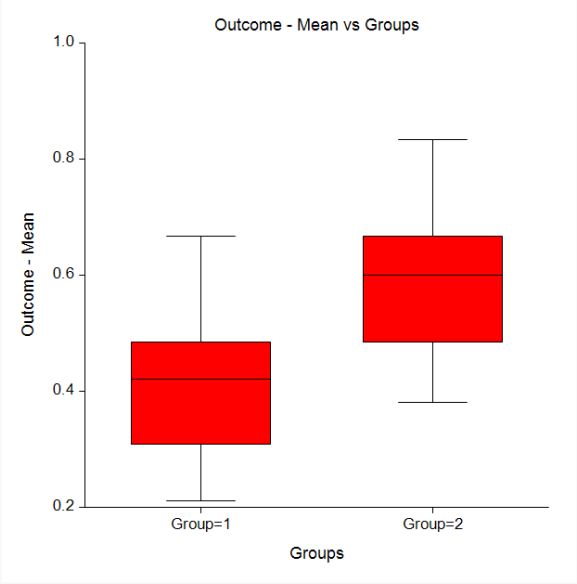
Equal-Variance Test	Test Statistic	Prob Level	Reject H0 of Equal Variances at α = 0.05?
Variance-Ratio	1.1582	0.72189	No
Modified-Levene	0.0775	0.78193	No

Probability Plots





Box Plots



This report displays the results of the various tests. The probability plots let you assess the validity of the normality assumptions. The box plots show the separation between the groups.

© NCSS, LLC. All Rights Reserved.