Chapter 272

Cluster Randomization – Create Cluster Rates 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 mean-rate of each cluster and then analyze those rates 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 and the elapsed time is also recorded, the cluster response rate can be calculated.

This procedure creates a new dataset containing the cluster rates 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 four variables: a categorical cluster variable, a categorical group variable, a binary data variable that is coded with either a 0 (no) or a one (yes), and a time variable that contains the elapsed time under which the subject is under study.

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. Data columns include a binary event indicator (Outcome) and the elapsed time under which each subject is observed. This example dataset is called **ClusRandRates**.

ClusRandRates Dataset (Subset)

Cluster	Group	Outcome	ETime
1	1	0	16
1	1	0	6
1	1	0	33
1	1	0	24
.	.	.	.
.	.	.	.
.	.	.	.

Example 1 – Creating a Summarized Dataset from the ClusRandRates Data

This section presents an example of how to analyze the data contained in the ClusRandRates dataset.

Setup

To run this example, complete the following steps:

1 Open the ClusRandRates example dataset

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

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

- Find and open the **Cluster Randomization – Create Cluster** Rates **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**
Event Outcome Variable	**Outcome**
Follow-Up Time Variable	**ETime** (*This is number of months of follow-up*)
Rate Scale Factor	**12000** (*This is 1000 x 12 and puts the rates to events per 1000 people per year*)
Store the Summary List in a New NCSS Data File	**Checked**
Output File Name	**%mydocs_NCSS%\Data\Cluster Rates.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.

Summary List Storage Information

Summary List Storage Information

Output Data File Name:	{NCSS Documents Folder}\Data\Cluster Rates.NCSS
Original Raw Data File:	{Example Data Folder}\ClusRandRates.NCSS
Data Variable(s):	(2) Outcome, ETime
Group Variable(s):	(2) Cluster, Group
Summary Statistic(s):	(4) Count, Mean, Sum, Rate

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

Summary List of Outcome

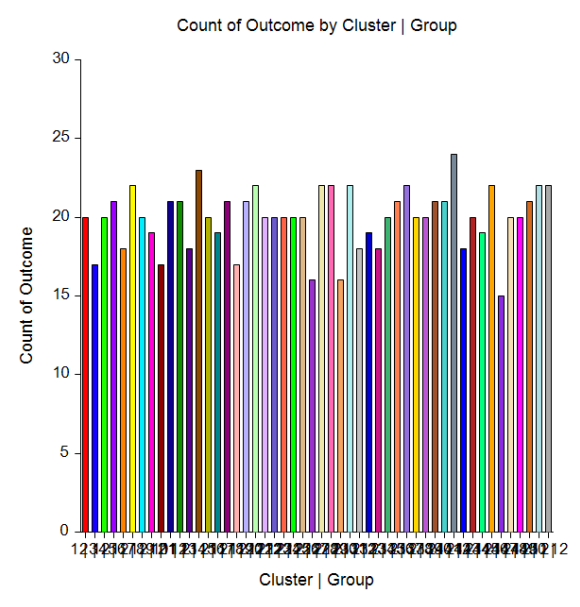
Summary List of Outcome

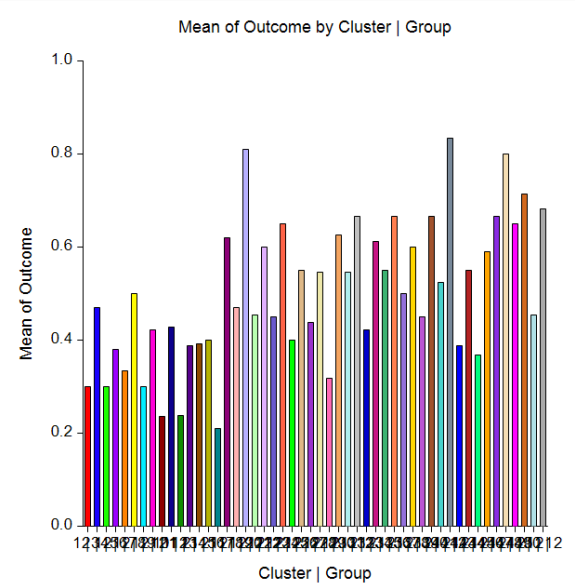
	Statistics for Outcome			
Cluster | Group	**Count**	**Mean**	**Sum**	**Rate**
1 | 1	20	0.3	6	188.5
2 | 2	17	0.4705882	8	347.8
3 | 1	20	0.3	6	179.6
4 | 2	21	0.3809524	8	222.2
5 | 1	18	0.3333333	6	225.0
6 | 2	22	0.5	11	285.7
7 | 1	20	0.3	6	181.8
8 | 2	19	0.4210526	8	246.2
9 | 1	17	0.2352941	4	132.6
10 | 2	21	0.4285714	9	236.8
11 | 1	21	0.2380952	5	150.8
12 | 2	18	0.3888889	7	247.8
13 | 1	23	0.3913043	9	220.4
14 | 2	20	0.4	8	225.9
15 | 1	19	0.2105263	4	106.2
16 | 2	21	0.6190476	13	399.0
17 | 1	17	0.4705882	8	310.7
18 | 2	21	0.8095238	17	445.4
19 | 1	22	0.4545455	10	275.9
20 | 2	20	0.6	12	394.5
21 | 1	20	0.45	9	309.5
22 | 2	20	0.65	13	370.5
23 | 1	20	0.4	8	290.9
24 | 2	20	0.55	11	370.8
25 | 1	16	0.4375	7	258.5
26 | 2	22	0.5454546	12	333.3
27 | 1	22	0.3181818	7	184.2
28 | 2	16	0.625	10	300.0
29 | 1	22	0.5454546	12	358.2
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

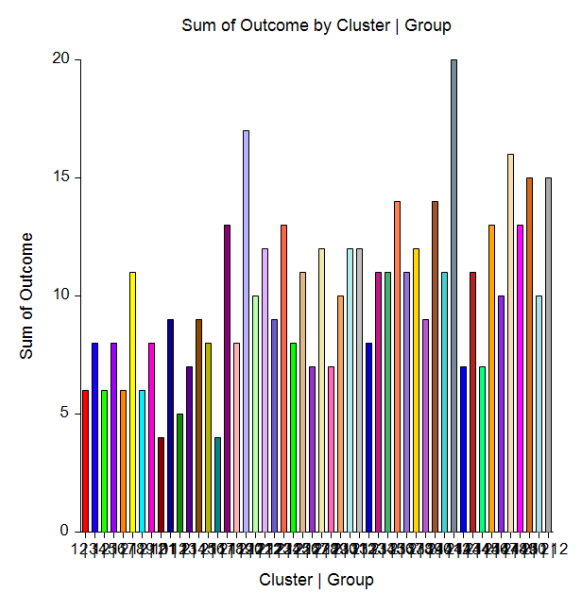
This report displays count, mean, sum, and rate of the Outcome variable for each cluster.

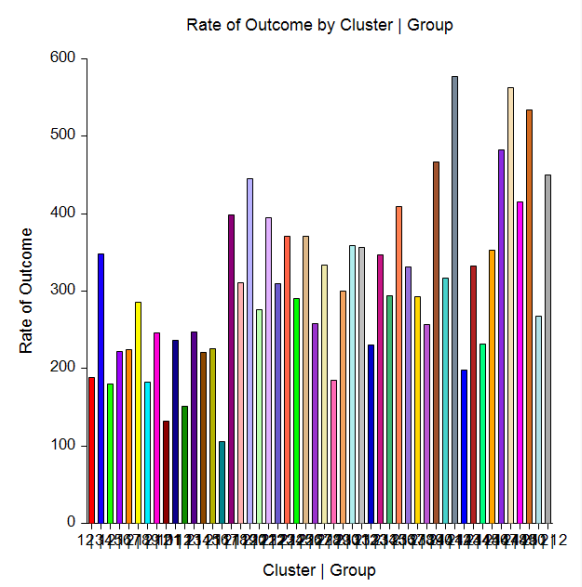
Plots of Each Statistic for Outcome

Plots of Each Statistic for Outcome









This report displays the statistics cluster-by-cluster.

New Cluster Rates Dataset

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

Cluster	Group	Outcome_Count	Outcome_Mean	Outcome_Sum	Outcome_Rate	ETime_Count	ETime_Mean	ETime_Sum
1	1	20	0.3	6	188.48167539267	20	19.1	382
2	2	17	0.470588235294118	8	347.826086956522	17	16.2352941176471	276
3	1	20	0.3	6	179.551122194514	20	20.05	401
4	2	21	0.380952380952381	8	222.222222222222	21	20.5714285714286	432
5	1	18	0.333333333333333	6	225	18	17.7777777777778	320
6	2	22	0.5	11	285.714285714286	22	21	462
7	1	20	0.3	6	181.818181818182	20	19.8	396
8	2	19	0.421052631578947	8	246.153846153846	19	20.5263157894737	390
9	1	17	0.235294117647059	4	132.596685082873	17	21.2941176470588	362
10	2	21	0.428571428571429	9	236.842105263158	21	21.7142857142857	456
11	1	21	0.238095238095238	5	150.753768844221	21	18.952380952381	398
12	2	18	0.388888888888889	7	247.787610619469	18	18.8333333333333	339
13	1	23	0.391304347826087	9	220.408163265306	23	21.304347826087	490
14	2	20	0.4	8	225.882352941176	20	21.25	425
15	1	19	0.210526315789474	4	106.194690265487	19	23.7894736842105	452
16	2	21	0.619047619047619	13	398.976982097187	21	18.6190476190476	391
17	1	17	0.470588235294118	8	310.679611650485	17	18.1764705882353	309
18	2	21	0.80952380952381	17	445.414847161572	21	21.8095238095238	458
19	1	22	0.454545454545455	10	275.862068965517	22	19.7727272727273	435
20	2	20	0.6	12	394.520547945205	20	18.25	365
21	1	20	0.45	9	309.45558739255	20	17.45	349
22	2	20	0.65	13	370.546318289786	20	21.05	421
23	1	20	0.4	8	290.909090909091	20	16.5	330
24	2	20	0.55	11	370.786516853933	20	17.8	356
25	1	16	0.4375	7	258.461538461538	16	20.3125	325
26	2	22	0.545454545454545	12	333.333333333333	22	19.6363636363636	432
27	1	22	0.318181818181818	7	184.210526315789	22	20.7272727272727	456
28	2	16	0.625	10	300	16	25	400
29	1	22	0.545454545454545	12	358.208955223881	22	18.2727272727273	402
30	2	18	0.666666666666667	12	356.435643564356	18	22.4444444444444	404
31	1	19	0.421052631578947	8	230.215827338129	19	21.9473684210526	417
32	2	18	0.611111111111111	11	346.456692913386	18	21.1666666666667	381
33	1	20	0.55	11	293.986636971047	20	22.45	449
34	2	21	0.666666666666667	14	409.756097560976	21	19.5238095238095	410
35	1	22	0.5	11	330.827067669173	22	18.1363636363636	399
36	2	20	0.6	12	293.279022403259	20	24.55	491
37	1	20	0.45	9	257.142857142857	20	21	420
38	2	21	0.666666666666667	14	466.666666666667	21	17.1428571428571	360
39	1	21	0.523809523809524	11	317.307692307692	21	19.8095238095238	416
40	2	24	0.833333333333333	20	576.923076923077	24	17.3333333333333	416
41	1	18	0.388888888888889	7	198.581560283688	18	23.5	423
42	2	20	0.55	11	332.493702770781	20	19.85	397
43	1	19	0.368421052631579	7	232.044198895028	19	19.0526315789474	362
44	2	22	0.590909090909091	13	352.941176470588	22	20.0909090909091	442
45	1	15	0.666666666666667	10	481.927710843373	15	16.6	249
46	2	20	0.8	16	563.049853372434	20	17.05	341
47	1	20	0.65	13	414.893617021277	20	18.8	376
48	2	21	0.714285714285714	15	534.124629080119	21	16.047619047619	337
49	1	22	0.454545454545455	10	267.857142857143	22	20.3636363636364	448
50	2	22	0.681818181818182	15	450	22	18.1818181818182	400

The Outcome_Rate variable is the variable of interest which gives the cluster event rates per 1000 individuals.

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_Rate 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.

Example 1c – Analyzing the Summarized Dataset

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

Setup

To run this example, complete the following steps:

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

- From the File menu of the NCSS Data window, select **Cluster Rates** in the list of recent datasets.
  or
- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Cluster Rates** 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 1c** 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_Rate**
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**

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	255.895	87.01437	17.40288	2.0639	219.9773	291.8128
Group=2	25	364.3253	100.7155	20.1431	2.0639	322.752	405.8987

Descriptive Statistics for the Median

Variable	Count	Median	95% LCL of Median	95% UCL of Median
Group=1	25	257.1429	198.5816	293.9866
Group=2	25	352.9412	300	398.977

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	-108.4303	26.61963	2.0106	-161.9527	-54.90796
Unequal	47.01	-108.4303	26.61963	2.0117	-161.9818	-54.87879

Equal-Variance T-Test

Alternative Hypothesis	Mean Difference	Standard Error	T-Statistic	DF	Prob Level	Reject H0 at α = 0.05?
μ1 - μ2 ≠ 0	-108.4303	26.61963	-4.0733	48	0.00017	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	-108.4303	26.61963	-4.0733	47.01	0.00018	Yes

Randomization Tests

Alternative Hypothesis: |μ1 - μ2| ≠ 0. This is a Two-Sided Test.
Number of Monte Carlo samples: 10000
Computer-Generated Random Seed: 4717079

Variance Assumption	Prob Level	Reject H0 at α = 0.05?
Equal Variance	0.00040	Yes
Unequal Variance	0.00040	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	125	450	637.5	51.53882
Group=2	500	825	637.5	51.53882

Number of Sets of Ties = 0, Multiplicity Factor = 0

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	-3.6380	0.00027	Yes
Normal Approx. with C.C.	Location Diff. ≠ 0	-3.6283	0.00029	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	0.9700	0.64424	No
Skewness	1.4380	0.15042	No
Kurtosis	0.9444	0.34498	No
Omnibus (Skewness or Kurtosis)	2.9598	0.22766	No

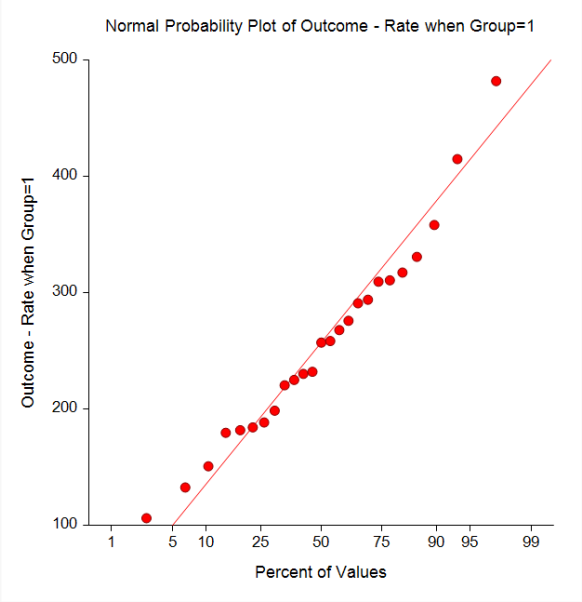
Tests of the Normality Assumption for Group=2

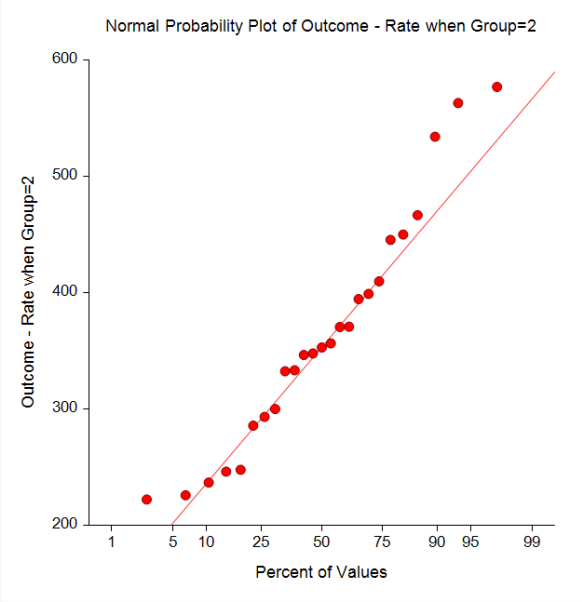
Normality Test	Test Statistic	Prob Level	Reject H0 of Normality at α = 0.05?
Shapiro-Wilk	0.9465	0.20882	No
Skewness	1.2547	0.20959	No
Kurtosis	-0.0444	0.96456	No
Omnibus (Skewness or Kurtosis)	1.5762	0.45470	No

Tests of the Equal Variance Assumption

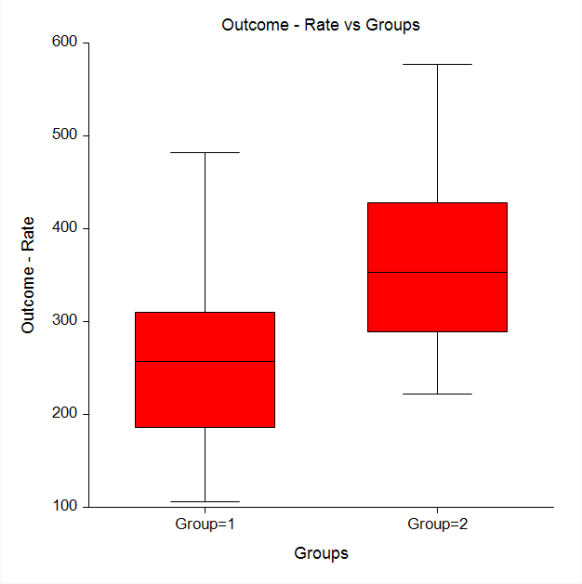
Equal-Variance Test	Test Statistic	Prob Level	Reject H0 of Equal Variances at α = 0.05?
Variance-Ratio	1.3397	0.47913	No
Modified-Levene	0.3253	0.57111	No

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.