Chapter 270

Cluster Randomization – Create Cluster Means 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.

This procedure creates a new dataset containing the cluster means 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 numeric data variable.

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 (Pulse) gives the endpoint value for each individual. This example dataset is called **ClusRandMeans**.

ClusRandMeans Dataset (Subset)

Cluster	Group	Pulse
1	1	60
1	1	56
1	1	58
1	1	58
1	1	64
1	1	54
1	1	51
.	.	.
.	.	.
.	.	.

Example 1 – Creating a Summarized Dataset from the ClusRandMeans Data

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

Setup

To run this example, complete the following steps:

1 Open the ClusRandMeans example dataset

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

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

- Find and open the **Cluster Randomization – Create Cluster Means 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)	**Pulse**
Store the Summary List in a New NCSS Data File	**Checked**
Output File Name	**%mydocs_NCSS%\Data\Cluster Means.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 Means.NCSS
Original Raw Data File:	{Example Data Folder}\ClusRandMeans.NCSS
Data Variable(s):	(1) Pulse
Group Variable(s):	(2) Cluster, Group
Summary Statistic(s):	(4) Count, Mean, SD, Sum

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

Summary List of Pulse

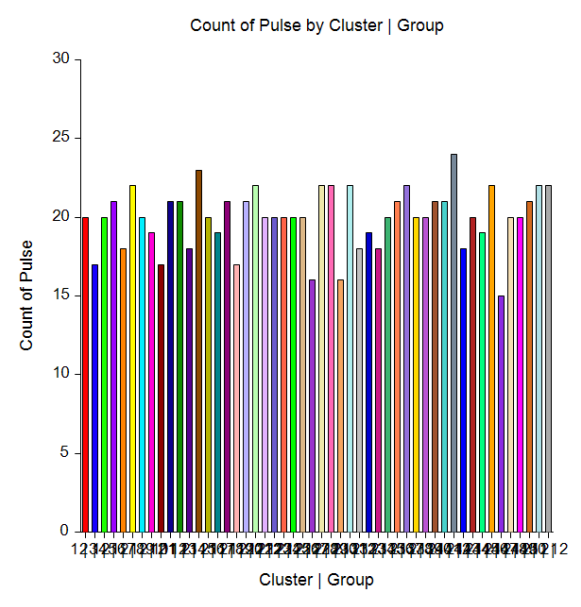
Summary List of Pulse

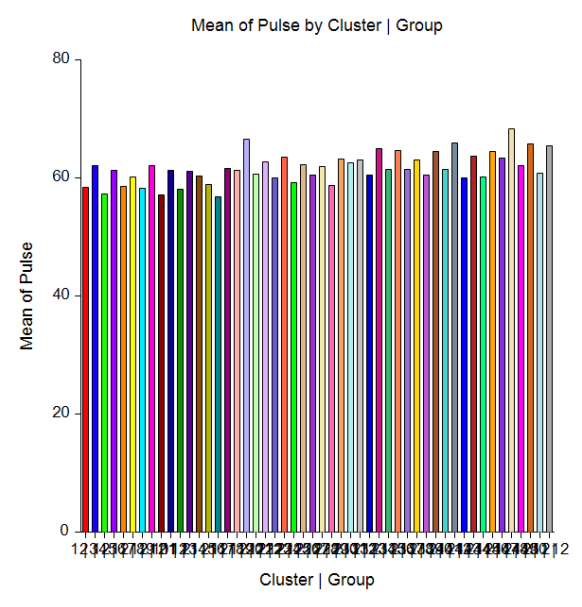
	Statistics for Pulse			
Cluster | Group	**Count**	**Mean**	**SD**	**Sum**
1 | 1	20	58.45	5.633313	1169
2 | 2	17	62.11765	7.72077	1056
3 | 1	20	57.3	5.731721	1146
4 | 2	21	61.28571	7.590407	1287
5 | 1	18	58.5	6.195349	1053
6 | 2	22	60.22727	7.444601	1325
7 | 1	20	58.25	6.972691	1165
8 | 2	19	62.15789	7.537043	1181
9 | 1	17	57.17647	6.521458	972
10 | 2	21	61.33333	6.821535	1288
11 | 1	21	58.09524	4.805255	1220
12 | 2	18	61.16667	4.937849	1101
13 | 1	23	60.26087	6.195434	1386
14 | 2	20	58.85	7.154498	1177
15 | 1	19	56.84211	5.080061	1080
16 | 2	21	61.52381	7.540683	1292
17 | 1	17	61.29412	5.763731	1042
18 | 2	21	66.52381	5.51923	1397
19 | 1	22	60.68182	6.342497	1335
20 | 2	20	62.65	6.310184	1253
21 | 1	20	60.05	4.18613	1201
22 | 2	20	63.5	7.366353	1270
23 | 1	20	59.25	6.568465	1185
24 | 2	20	62.3	4.974673	1246
25 | 1	16	60.5	6.377042	968
26 | 2	22	61.95454	7.370968	1363
27 | 1	22	58.72727	7.579307	1292
28 | 2	16	63.25	5.322906	1012
29 | 1	22	62.5	6.076888	1375
30 | 2	18	63.05556	7.487026	1135
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

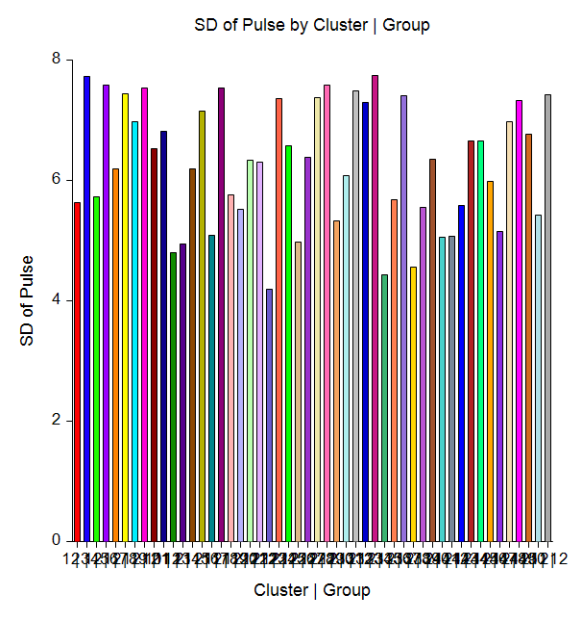
This report displays count, mean, standard deviation, and sum of the Pulse variable for each cluster.

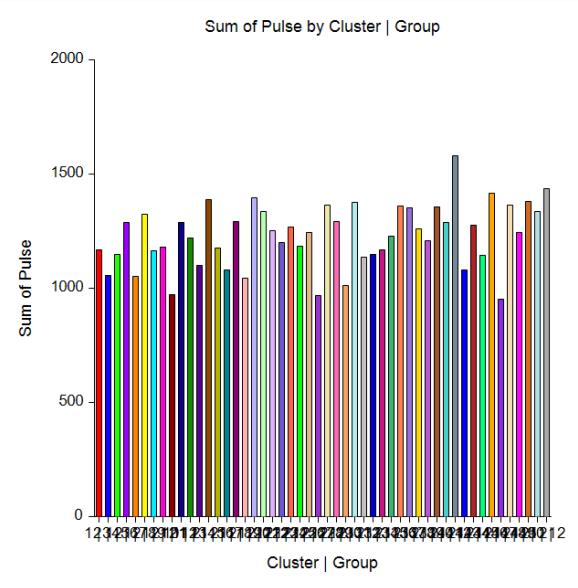
Plots of Each Statistic for Pulse

Plots of Each Statistic for Pulse









This report displays the statistics cluster-by-cluster. Of particular interest are the means and whether the standard deviations (SD's) can be assumed to be equal.

New Cluster Means Dataset

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

Cluster	Group	Pulse_Count	Pulse_Mean	Pulse_SD	Pulse_Sum
1	1	20	58.45	5.63331257133099	1169
2	2	17	62.1176470588235	7.72077030597641	1056
3	1	20	57.3	5.73172151966121	1146
4	2	21	61.2857142857143	7.59040748012159	1287
5	1	18	58.5	6.19534929936775	1053
6	2	22	60.2272727272727	7.44460058848545	1325
7	1	20	58.25	6.97269109115208	1165
8	2	19	62.1578947368421	7.53704302388275	1181
9	1	17	57.1764705882353	6.52145779444335	972
10	2	21	61.3333333333333	6.82153452921946	1288
11	1	21	58.0952380952381	4.80525505987728	1220
12	2	18	61.1666666666667	4.9378490196822	1101
13	1	23	60.2608695652174	6.19543373860831	1386
14	2	20	58.85	7.15449803307424	1177
15	1	19	56.8421052631579	5.08006078151233	1080
16	2	21	61.5238095238095	7.5406833086866	1292
17	1	17	61.2941176470588	5.76373040966474	1042
18	2	21	66.5238095238095	5.51923045015379	1397
19	1	22	60.6818181818182	6.34249735729332	1335
20	2	20	62.65	6.31018391591177	1253
21	1	20	60.05	4.18612998012798	1201
22	2	20	63.5	7.36635309327057	1270
23	1	20	59.25	6.56846533385884	1185
24	2	20	62.3	4.97467269486673	1246
25	1	16	60.5	6.37704215656966	968
26	2	22	61.9545454545455	7.37096797111267	1363
27	1	22	58.7272727272727	7.57930652803325	1292
28	2	16	63.25	5.32290647422377	1012
29	1	22	62.5	6.07688830147234	1375
30	2	18	63.0555555555556	7.48702581507207	1135
31	1	19	60.4210526315789	7.29014395128171	1148
32	2	18	64.9444444444444	7.74195766565098	1169
33	1	20	61.4	4.42956573848712	1228
34	2	21	64.7142857142857	5.67576300723398	1359
35	1	22	61.5	7.41138120923296	1353
36	2	20	63.1	4.55261636748295	1262
37	1	20	60.45	5.54858919954101	1209
38	2	21	64.4761904761905	6.35310197949826	1354
39	1	21	61.3809523809524	5.05446525832546	1289
40	2	24	65.875	5.0760520434513	1581
41	1	18	60	5.58358940003966	1080
42	2	20	63.75	6.64811013905851	1275
43	1	19	60.1578947368421	6.66052348538876	1143
44	2	22	64.4090909090909	5.9893484096598	1417
45	1	15	63.4	5.15197604253525	951
46	2	20	68.25	6.98023525466917	1365
47	1	20	62.15	7.32174123667602	1243
48	2	21	65.7142857142857	6.76862509777914	1380
49	1	22	60.7272727272727	5.41762395803265	1336
50	2	22	65.3636363636364	7.41649034276517	1438

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 Pulse_Mean 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 1a – Analyzing the Summarized Dataset

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

Setup

To run this example, complete the following steps:

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

- From the File menu of the NCSS Data window, select **Cluster Means** in the list of recent datasets.
  or
- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Cluster Means** 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 1a** 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)	**Pulse_Mean**
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 T-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	59.9786	1.716245	0.3432491	2.0639	59.27017	60.68703
Group=2	25	63.29973	2.141713	0.4283426	2.0639	62.41567	64.18378

Descriptive Statistics for the Median

Variable	Count	Median	95% LCL of Median	95% UCL of Median
Group=1	25	60.26087	58.72727	60.72727
Group=2	25	63.1	62.11765	64.47619

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	-3.321125	0.5489056	2.0106	-4.424773	-2.217476
Unequal	45.82	-3.321125	0.5489056	2.0131	-4.426129	-2.21612

Equal-Variance T-Test

Alternative Hypothesis	Mean Difference	Standard Error	T-Statistic	DF	Prob Level	Reject H0 at α = 0.05?
μ1 - μ2 ≠ 0	-3.321125	0.5489056	-6.0504	48	0.00000	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	-3.321125	0.5489056	-6.0504	45.82	0.00000	Yes

Randomization Tests

Alternative Hypothesis: |μ1 - μ2| ≠ 0. This is a Two-Sided Test.
Number of Monte Carlo samples: 10000
Computer-Generated Random Seed: 4392681

Variance Assumption	Prob Level	Reject H0 at α = 0.05?
Equal Variance	0.00010	Yes
Unequal Variance	0.00010	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	67	392	637.5	51.53882
Group=2	558	883	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	-4.7634	0.00000	Yes
Normal Approx. with C.C.	Location Diff. ≠ 0	-4.7537	0.00000	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	0.9708	0.66618	No
Skewness	-0.2911	0.77097	No
Kurtosis	-0.5891	0.55582	No
Omnibus (Skewness or Kurtosis)	0.4317	0.80584	No

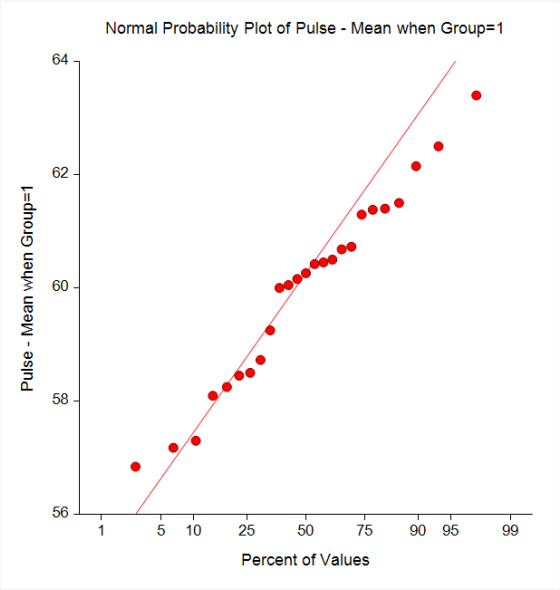
Tests of the Normality Assumption for Group=2

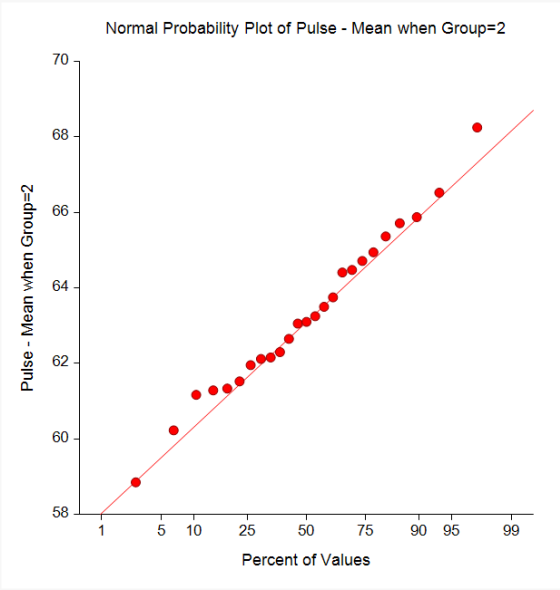
Normality Test	Test Statistic	Prob Level	Reject H0 of Normality at α = 0.05?
Shapiro-Wilk	0.9898	0.99490	No
Skewness	0.5264	0.59860	No
Kurtosis	0.3194	0.74941	No
Omnibus (Skewness or Kurtosis)	0.3791	0.82732	No

Tests of the Equal Variance Assumption

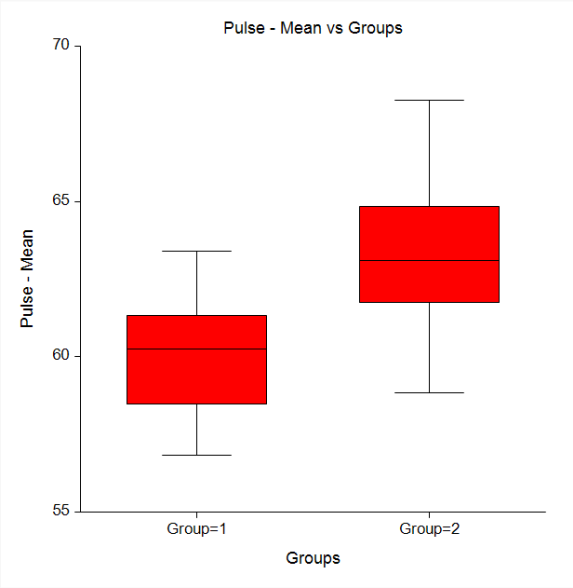
Equal-Variance Test	Test Statistic	Prob Level	Reject H0 of Equal Variances at α = 0.05?
Variance-Ratio	1.5573	0.28488	No
Modified-Levene	1.0372	0.31357	No

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.