

Chapter 532

GEE Tests for Multiple Means in a Cluster-Randomized Design

Introduction

This module calculates the power for testing for differences among the group means from **continuous**, correlated data from a cluster-randomized design that are analyzed using the GEE method.

GEE is different from mixed models in that it does not require the full specification of the joint distribution of the measurements, as long as the marginal mean model is correctly specified. Estimation consistency is achieved even if the correlation matrix is incorrect. For clustered designs such as those discussed here, GEE assumes a *compound symmetric* (CS) correlation structure.

The outcomes are averaged at the cluster level. The precision of the experiment is increased by increasing the number of subjects per cluster as well as the number of clusters.

Missing Values

This procedure allows you to specify the proportion of subjects that are missing by the end of the study.

Technical Details

Theory and Notation

Technical details are given in Ahn, Heo, and Zhang (2015), chapter 4, section 4.4.4, pages 116-119 and in Zhang and Ahn (2013).

Suppose we want to compare the means of G groups. Further suppose each of these groups consists of K_g ($g = 1, \dots, G$) clusters for a total of K clusters. Each cluster provides outcomes for M subjects.

The mean of y_{gki} is modeled by

$$y_{gki} = \mu_g + \epsilon_{gki}$$

where

y_{gki} is the response from subject i in cluster k in group g , with variance σ^2 ,

r_g is the proportion of subjects in group g ,

GEE Tests for Multiple Means in a Cluster-Randomized Design

μ_g is group mean ($g = 1, \dots, G$),

ϵ_{gki} is a zero-mean error term with variance σ^2 .

In this procedure, the primary interest is to test $H_0: \mu_1 = \dots = \mu_G$ against the alternative that at least one mean is different.

GEE is used to estimate and test hypotheses about these group means. H_0 is rejected with a type I error α if $\mathbf{B}'\mathbf{W}^{-1}\mathbf{B} > \chi_{G-1, 1-\alpha}^2$ where $\chi_{G-1, 1-\alpha}^2$ is the 100(1 - α)th percentile of a chi-square distribution with $G-1$ degrees of freedom. The mean deviation vector \mathbf{B} is given by

$$\mathbf{B} = \sqrt{K}(\hat{\mu}_1 - \hat{\mu}, \dots, \hat{\mu}_{G-1} - \hat{\mu})'$$

and \mathbf{W} is a consistent estimate of the variance matrix based on the residuals.

Correlation Patterns

In a cluster-randomized design with K clusters, each consisting of M subjects, observations from a single cluster are correlated. The resulting correlation matrix is assumed to have a *compound symmetric* pattern with a common correlation coefficient ρ . That is, the correlation matrix within a cluster is

$$[\rho_{jj'}] = \begin{bmatrix} 1 & \rho & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \rho & \cdots & \rho \\ \rho & \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \rho & \cdots & 1 \end{bmatrix}_{M \times M}$$

Missing Data

The problem of missing data occurs for several reasons. In these designs, it is assumed that the responses of some proportion, P , of the subjects will be missing.

Sample Size Calculations

The details of the calculation of sample size and power is given in Ahn, Heo, and Zhang (2015), Chapter 4. These are summarized here.

As explained above, GEE is used to estimate the regression coefficients. Under the alternative hypothesis, as K approaches infinity, the test statistic $U'\widehat{W}^{-1}U$ is approximately distributed as a noncentral chi-square random variable with $G - 1$ degrees of freedom and noncentrality parameter $U = K\boldsymbol{\eta}'\mathbf{W}^{-1}\boldsymbol{\eta}$

where

$$\boldsymbol{\eta}' = (\mu_1 - \bar{\mu}, \dots, \mu_{G-1} - \bar{\mu})$$

$$\bar{\mu} = \sum_{g=1}^G r_g \mu_g$$

$$\boldsymbol{\eta}'\mathbf{W}^{-1}\boldsymbol{\eta} = \frac{\bar{m}^2}{s} \left[\sum_{g=1}^{G-1} r_g \eta_g^2 + \frac{1}{r_G} \left(\sum_{g=1}^{G-1} r_g \eta_g \right)^2 \right]$$

$$\bar{m} = (1 - P)M$$

$$s = \sigma^2(M^2\rho + M(1 - \rho))(1 - P)$$

Here, P is the proportion missing.

GEE Tests for Multiple Means in a Cluster-Randomized Design

This noncentral distribution can be used to calculate the power using the following

$$\text{Power} = \Pr(\chi_{G-1}^{\prime 2}(U) > \chi_{G-1, 1-\alpha}^2)$$

Using this power function, the sample size can be determined using a simple binary search.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be solved for. When you choose to solve for *Number of Clusters*, the program searches for the lowest number of clusters that meets the alpha and power criterion you have specified. The *Solve For* parameter is displayed on the vertical axis of the plot.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal slopes when in fact the slopes are different.

Values must be between zero and one. Historically, the value of 0.80 (beta = 0.20) was used for power. Now, 0.90 (beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Cluster Count and Group Allocation

G (Number of Groups)

This is the number of groups (arms) whose means are being compared. The number of items used in the Group Allocation boxes is controlled by this number.

This value must be an integer greater than or equal to two.

GEE Tests for Multiple Means in a Cluster-Randomized Design

Group Allocation Input Type (when Solve For = Power)

Specify how you want to enter the information about how the subjects are allocated to each of the G groups.

Possible options are:

- **Equal ($K_1 = K_2 = \dots = K_G$)**
The cluster count of every group is K_i . All group sample sizes are equal. Enter one or more values for the common cluster count.
- **Enter group multipliers**
Enter a list of group multipliers (r_1, r_2, \dots, r_G) and one or more values of n . The individual group sample sizes are found by multiplying the multipliers by n . For example, $K_1 = r_1 \times K_i$.
- **Enter K_1, K_2, \dots, K_G**
Enter a list of group cluster counts, one for each group.
- **Enter columns of K_i 's**
Select one or more columns of the spreadsheet that each contain a set of group cluster counts going down the column. Each column is analyzed separately.

K_i (Clusters Per Group)

Enter K_i , the number of clusters in each group. The total number of clusters, K , is equal to $K_i \times$ (number of groups).

You can specify a single value or a list.

Single Value

Enter a value for the cluster counts of all groups. If you enter '10' here and there are five groups, then each group will be assigned 10 clusters and the total cluster count will be 50.

List of Values

A separate power analysis is calculated for each value of K_i in the list. All analyses assume that the common, group sample size is K_i .

Range of K_i

$n > 1$

Group Multipliers (r_1, r_2, \dots, r_G)

Enter a set of G multipliers, one for each group.

The individual group sample sizes is computed as $K_g = \text{ceiling}[r_g \times K_i]$, where $\text{ceiling}[y]$ is the first integer greater than or equal to y . For example, the multipliers $\{1, 1, 2, 2.95\}$ and *base* K_i of 10 would result in the cluster counts $\{10, 10, 20, 30\}$.

Incomplete List

If the number of items in the list is less than G , the missing multipliers are set equal to the last entry in the list.

Range

The items in the list must be positive. The resulting cluster counts must be at least 1.

GEE Tests for Multiple Means in a Cluster-Randomized Design

K_i (Base Sample Size Per Group)

Enter K_i , the base number of clusters in each group. The number of clusters in the group is found by multiplying this number by the corresponding group multiplier, $\{r_1, r_2, \dots, r_G\}$, and rounding up to the next integer.

You can specify a single value or a list.

Single Value

Enter a value for the base group cluster count.

List of Values

A separate power analysis is calculated for each value of K_i in the list.

Range

$\text{Ceiling}[K_i \times r_i] \geq 1$.

K₁, K₂, ..., K_G (List)

Enter a list of G cluster counts, one for each group.

Incomplete List

If the number of items in the list is less than G , the missing cluster counts are set equal to the last entry in the list.

Range

The items in the list must be positive. At least one item in the list must be greater than 1.

Columns of K_i's

Enter one or more spreadsheet columns containing vertical lists of group cluster counts.

Press the Spreadsheet icon (directly to the right) to select the columns and then enter the values.

Press the Input Spreadsheet icon (to the right and slightly up) to view/edit the spreadsheet. Also note that you can obtain the spreadsheet by selecting "Tools", then "Input Spreadsheet", from the menus.

On the spreadsheet, the group cluster counts are entered going down.

Examples (assuming $G = 3$)

C1	C2	C3
111	1	28
115	20	68
100	30	46

Definition of a Single Column

Each column gives one list. Each column results in a new scenario. The columns are not connected, but all should have exactly G rows.

Each entry in the list is the cluster count of that group.

Incomplete List

If the number of items in the list is less than G , the missing entries are set equal to the last entry in the list.

Valid Entries

All values should be positive integers. At least one value must be greater than one.

GEE Tests for Multiple Means in a Cluster-Randomized Design

Note

The column names (C1, C2, ...) can be changed by right-clicking on them in the spreadsheet.

Group Allocation Input Type (when Solve For = Number of Clusters)

Specify how you want to enter the information about how the clusters are allocated to each of the G groups.

Options

- **Equal ($K_1 = K_2 = \dots = K_G$)**

All group cluster counts are equal to K_i . The value of K_i will be found by conducting a search.

- **Enter group allocation pattern**

Enter an allocation pattern (r_1, r_2, \dots, r_G). The pattern consists of a set of G numbers. These numbers will be rescaled into proportions by dividing each item by the sum of all items. The individual group cluster counts are found by multiplying these proportions by K (the total cluster count) and rounding up.

- **Enter columns of allocation patterns**

Select one or more columns of the spreadsheet that each contain a group allocation pattern going down the column. Each column is analyzed separately.

Group Allocation Pattern (r_1, r_2, \dots, r_G)

Enter an allocation pattern (r_1, r_2, \dots, r_G). The pattern consists of a set of G numbers. These numbers will be rescaled into proportions of total number of clusters by dividing each item by the sum of all items. The individual group cluster counts are found by multiplying these proportions by K (the total cluster count) and rounding up.

For example, the pattern $\{1, 3, 4\}$ will be rescaled to $\{0.125, 0.375, 0.5\}$. The group cluster counts will be constrained to these proportions (within rounding) during the search for the cluster count configuration that meets the power requirement.

Incomplete List

If the number of items in the list is less than G , the missing numbers are set equal to the last entry in the list.

Range

The items in the list must be positive. The resulting cluster counts must be at least 1.

Columns of Group Allocation Patterns

Enter one or more spreadsheet columns containing vertical lists of group allocation patterns.

Press the Spreadsheet icon (directly to the right) to select the columns and then enter the values.

Press the Input Spreadsheet icon (to the right and slightly up) to view/edit the spreadsheet. Also note that you can obtain the spreadsheet by selecting "Tools", then "Input Spreadsheet", from the menus.

On the spreadsheet, the group allocation patterns are entered going down.

Examples (assuming $G = 3$)

C1	C2	C3
1	1	3
1	2	3
3	2	1

GEE Tests for Multiple Means in a Cluster-Randomized Design

Definition of a Single Column

Each column gives one allocation pattern. Each column results in a new scenario. The columns are not connected, but all should have exactly G rows.

Incomplete List

If the number of items in a list is less than G , the missing numbers are set equal to the last entry in the list before they are rescaled.

Valid Entries

All values should be positive numbers. You can enter decimal values.

Note

The column names (C1, C2, ...) can be changed by right-clicking on them in the spreadsheet

Cluster Size

M (Average Cluster Size)

Specify one or more values for M , the average number of subjects for each cluster.

At least one value must be entered. If multiple values are entered, a separate analysis is run for each value.

The value must be an integer greater than 1. A reasonable range is from 2 to 500.

Some examples are

3

3 4 5

3 to 10 by 1

Effect Size

μ_i 's Input Type

Specify how you want to enter the G group means $\mu_1, \mu_2, \dots, \mu_G$ assumed by the alternative hypothesis. The power is calculated for these values.

Options

- **$\mu_1, \mu_2, \dots, \mu_G$**
Specify values of the group means. The differences among these values are proportional to the effect size that you want to detect.
- **Columns Containing Sets of μ 's**
Select one or more columns of the spreadsheet that each contain a set of μ_i 's going down the column. Each column is analyzed separately.

$\mu_1, \mu_2, \dots, \mu_G$

Enter the values of the G group means under the alternative hypothesis. The effect size that the study will detect is a function of the differences among these values.

The mean for a particular group is the average response of all clusters in that group.

GEE Tests for Multiple Means in a Cluster-Randomized Design

Range

Each μ_i should be numeric and at least one of the values must be different from the rest.

Example

10 10 10 40

Incomplete List

If the number of items in a list is less than G , the missing numbers are set equal to the last entry.

Columns Containing Sets of μ_i 's

Enter one or more spreadsheet columns containing vertical lists of $\mu_1, \mu_2, \dots, \mu_G$.

Press the Spreadsheet icon (directly to the right) to select the columns and then enter the values.

Press the Input Spreadsheet icon (to the right and slightly up) to view/edit the spreadsheet. Also note that you can obtain the spreadsheet by selecting "Tools", then "Input Spreadsheet", from the menus.

On the spreadsheet, the μ_i 's are entered going down.

Examples (assuming $G = 3$)

C1	C2	C3
10	10	30
10	20	30
30	20	10

Definition of a Single Column

Each column gives one set of means. Each column results in a new scenario. The columns are not connected, but all should have exactly G rows.

Incomplete List

If the number of items in a list is less than G , the missing numbers are set equal to the last entry in the list.

Valid Entries

You can enter any numeric value.

Note

The column names (C1, C2, ...) can be changed by clicking on them in the spreadsheet.

σ (Standard Deviation)

Enter a value (or range of values) for the standard deviation of the responses within a cluster, Y_{gki} , where g is the group, k is the cluster, and i is the subject. The analysis assumes that this standard deviation is constant across all clusters.

You can enter a single value such as '10' or a series of values such as '10 20 30 40 50' or '1 to 5 by 0.5'.

ρ (Intracluster Correlation, ICC)

Enter one or more values for ρ , the intracluster correlation coefficient.

At least one value must be entered. If multiple values are entered, a separate analysis is performed for each value.

Range

$0 < \rho < 1$ (negative values are not used). A value near 0 indicates low correlation. A value near 1 indicates high correlation.

GEE Tests for Multiple Means in a Cluster-Randomized Design

Examples

0.5

0.5 0.6 0.7

0 to 0.9 by 0.1

Examples

0.5

0.5 0.6 0.7

0 to 0.9 by 0.1

Missing Data Proportions Within Clusters

Missing Input Type

Specify the way in which you want to specify the proportion of missing values.

Possible input choices are

- **Constant = 0**

All missing proportions are set to 0. That is, there are no missing values.

- **Constant**

All missing proportions are set to constant value.

Missing Proportion

Enter one or more values of the missing value proportion that is to be used for all clusters in all groups.

If multiple values are entered, a separate analysis is performed for each value.

Range

$0 \leq \text{Missing Proportion} < 1$.

Examples

0.2

0.1 0.2 0.4

0 to 0.4 by 0.1

Example 1 – Determining the Power

Researchers are planning a study comparing medications: a standard drug and two experimental drugs. The experimental drugs appear to have about the same impact. All patients within a cluster will receive the same drug. The clusters available for study will be randomly assigned to one of the three groups.

To begin, the researchers want to determine the power for K_i equal to 5, 10, and 15. They will assume an average cluster size of 10.

The researchers want to detect a difference represented by the group means 1, 2, and 3.

Similar studies have had a within-cluster standard deviation of 3. These studies also showed an autocorrelation between subjects within a cluster of between 0.2 and 0.5, so they want to try values those values. The test will be conducted at the 0.05 significance level. The subjects will be divided equally among the three groups.

At this stage of planning, the researchers want to ignore the possibility that some subjects will drop out.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alpha.....	0.05
G (Number of Groups).....	3
Group Allocation Input Type	Equal (K1 = K2 = ... = KG)
Ki (Clusters Per Group)	5 10 15
M (Average Cluster Size)	10
μ 's Input Type	$\mu_1, \mu_2, \dots, \mu_G$
$\mu_1, \mu_2, \dots, \mu_G$	1 2 3
σ (Standard Deviation)	3
ρ (Intracluster Correlation, ICC).....	0.2 0.5
Missing Input Type.....	Constant = 0

Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results											
Number of Groups: 3											
	Total Subj Cnt	Total Clus Cnt	Clus Per Group	Ave Clus Size	Group Means Set	SD of Group Means	Std Dev	Effect Size	Intra-Clus Corr	Alpha	Miss Prop
Power	N	K	Ki	M	μ_i	σ_m	σ	σ_m/σ	ρ		
0.4125	150	15	5	10	$\mu_i(1)$	0.82	3.00	0.272	0.2000	0.050	0.00
0.2275	150	15	5	10	$\mu_i(1)$	0.82	3.00	0.272	0.5000	0.050	0.00
0.7139	300	30	10	10	$\mu_i(1)$	0.82	3.00	0.272	0.2000	0.050	0.00
0.4191	300	30	10	10	$\mu_i(1)$	0.82	3.00	0.272	0.5000	0.050	0.00
0.8805	450	45	15	10	$\mu_i(1)$	0.82	3.00	0.272	0.2000	0.050	0.00
0.5886	450	45	15	10	$\mu_i(1)$	0.82	3.00	0.272	0.5000	0.050	0.00

Set(Set Number): Values
 $\mu_i(1)$: 1.00, 2.00, 3.00

GEE Tests for Multiple Means in a Cluster-Randomized Design

References

- Ahn, C., Heo, M., and Zhang, S. 2015. Sample Size Calculations for Clustered and Longitudinal Outcomes in Clinical Research. CRC Press. New York. See pages 116-119.
 Zhang, S. and Ahn, C. 2013. Sample Size Calculations for Comparing Time-Averaged Responses in K-Group Repeated-Measurement Studies. Computational Statistics and Data Analysis. Vol 58(1). Pages 283-291.

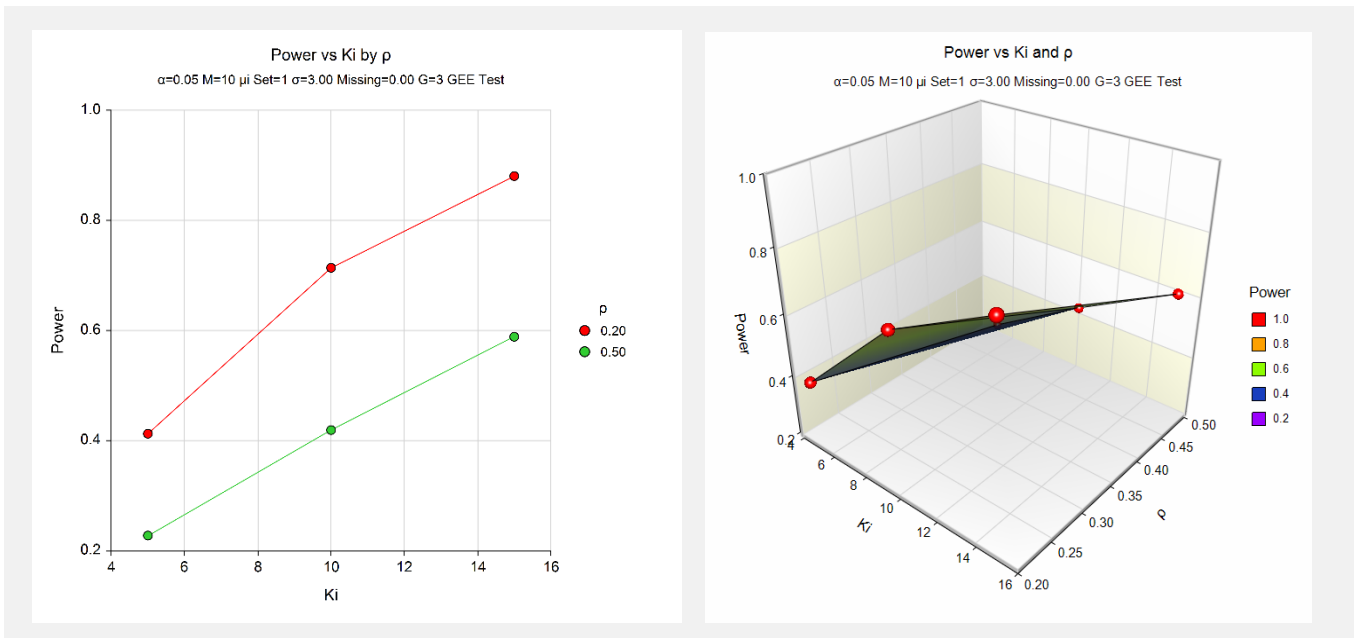
Report Definitions

- Power is the probability of rejecting a false null hypothesis.
- Total Subj Cnt N is the total number of subjects in the study.
- Total Clus Cnt K is the total number of clusters in the study.
- Clusters Per Group K_i is the number of clusters per group.
- Ave Clus Size M is the average cluster size (number of subjects per cluster) used for all clusters.
- Group Means Set μ_i gives the name and number of the set containing the mean responses for each group.
- SD of Group Means σ_m is the population standard deviation of the group means.
- Std Dev σ is the standard deviation of the responses within a cluster.
- Effect Size σ_m/σ is a measure of the effect size. It is the ratio of σ_m and σ .
- Intra-Clus Corr ρ is the intraclass correlation coefficient (ICC) used for all clusters. This is the correlation between any two subjects within a cluster.
- Alpha is the significance level of the test: the probability of rejecting the null hypothesis of equal means when it is true.
- Miss Prop is the proportion of each cluster that is assumed to be missing at the end of the study.

Summary Statements

A sample of 15 clusters, divided among 3 groups, each contains 10 subjects, for a total of 150 subjects, achieves a power of 0.4125. This power assumes that a chi-squared test from a GEE analysis with 2 degrees of freedom is used at a significance level of 0.050. The group cluster counts are 5, 5, 5. The group means under the alternative hypothesis are 1.00, 2.00, 3.00. The standard deviation of these means is 0.82. The standard deviation of the responses is 3.00. The effect size is 0.272. Missing values are assumed to occur completely at random (MCAR). The anticipated proportion missing in each cluster is 0.00. The autocorrelation matrix of the responses within a cluster is assumed to be compound symmetric with an intraclass correlation coefficient of 0.2000.

This report gives the power for various values of the other parameters. The definitions of each of the columns in the report are shown in the Report Definitions section. Note that the details of the μ_i set are shown below the table.



These charts show the relationship among the design parameters.

Example 2 – Determining the Power with Varying Cluster Counts

Continuing with Example 1, we want to give an example that stores the cluster counts on a spreadsheet. This allows us to compare various cluster count patterns.

This example will use all the settings of Example 1, except that various patterns of cluster counts across the three groups will be used. The following cluster count patterns are entered on the spreadsheet.

<u>Eq</u>	<u>Add1</u>	<u>Add5</u>
10	9	5
10	10	10
10	11	15

Note that we renamed column 1 from *C1* to *Eq*, column 2 from *C2* to *Add1*, and column 3 from *C3* to *Add5* by right clicking the column names and entering the new values.

Also note that each column sums to 30. That is, each column provides a pattern that results in a total cluster count of 30. However, the individual patterns are different. In the first column, the K_i 's are all equal. In the second column, the K_i increase by 1. In the third column, the K_i increase by 5. This analysis is being run to determine the impact of unequal allocation.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alpha.....	0.05
G (Number of Groups).....	3
Group Allocation Input Type	Enter columns of K_i's
Columns of K_i 's.....	EQ-ADD5
M (Average Cluster Size)	10
μ_i 's Input Type	$\mu_1, \mu_2, \dots, \mu_G$
$\mu_1, \mu_2, \dots, \mu_G$	1 2 3
σ (Standard Deviation)	3
ρ (Intracluster Correlation, ICC).....	0.2 0.5
Missing Input Type.....	Constant = 0

Note that you will have to open the spreadsheet and enter the cluster count values given above.

Output

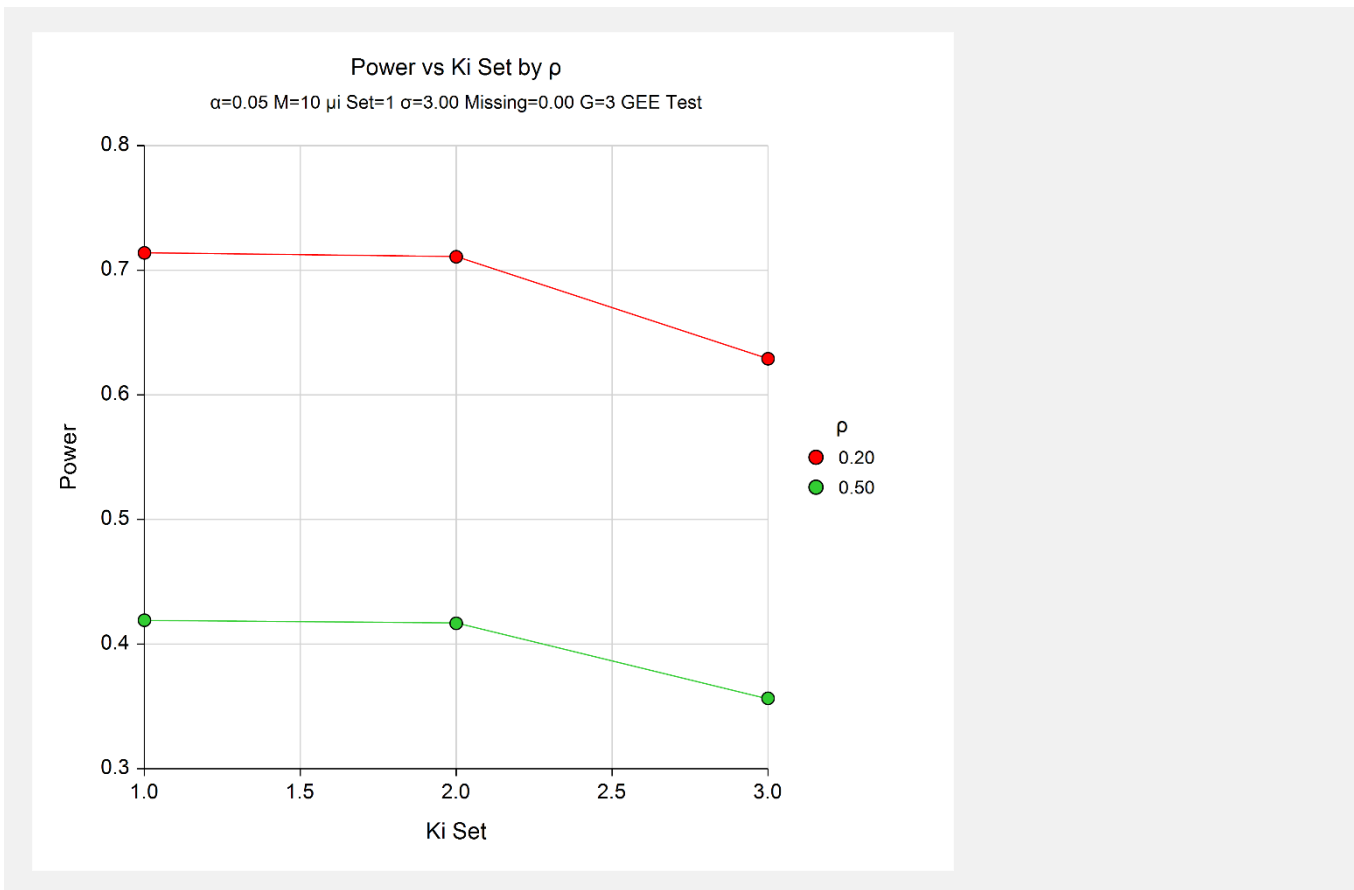
Click the Calculate button to perform the calculations and generate the following output.

Numeric Results
Number of Groups: 3

Power	Total Subj Cnt N	Total Clus Cnt K	Group Clus Cnts Set Ki	Ave Clus Size M	Group Means Set μ_i	SD of Group Means σ_m	Std Dev σ	Effect Size σ_m/σ	Intra-Clus Corr ρ	Alpha	Miss Prop
0.7139	300	30	Eq(1)	10	$\mu_i(1)$	0.82	3.00	0.272	0.2000	0.050	0.00
0.4191	300	30	Eq(1)	10	$\mu_i(1)$	0.82	3.00	0.272	0.5000	0.050	0.00
0.7108	300	30	Add1(2)	10	$\mu_i(1)$	0.82	3.00	0.272	0.2000	0.050	0.00
0.4167	300	30	Add1(2)	10	$\mu_i(1)$	0.82	3.00	0.272	0.5000	0.050	0.00
0.6290	300	30	Add5(3)	10	$\mu_i(1)$	0.82	3.00	0.272	0.2000	0.050	0.00
0.3565	300	30	Add5(3)	10	$\mu_i(1)$	0.82	3.00	0.272	0.5000	0.050	0.00

Set(Set Number): Values
 Eq(1): 10, 10, 10
 Add1(2): 9, 10, 11
 Add5(3): 5, 10, 15
 $\mu_i(1)$: 1.00, 2.00, 3.00

This report gives the power for various values of the other parameters. Note that the details of the Ki columns and the μ_i sets are shown in the Set footnote below the numeric results.



Notice that the horizontal axis of this plot is labelled *Ki Set* and that the values of the horizontal variable are 1, 2, and 3. These are the Set Numbers. Hence the Ki Sets are named Eq(1), Add1(2), and Add5(3). The number in parenthesis is the number that is plotted. We note that the pattern can make a big difference in power.

Example 3 – Validation of Sample Size Calculation using Zhang and Ahn (2013)

Zhang and Ahn (2013) presents an example which we will use to validate this procedure. Their example uses $G = 4$, $m = 3$, $\sigma = 1.43178$, $\rho = 0.45$, $\alpha = 0.05$, power = 0.9, compound symmetry correlation pattern, group cluster counts of 25, 25, 25, 26 ($K = 101$), no missing data, and group means of 1.99, 1.99, 1.99, 1.0. The resulting power was 0.90.

Setup

This section presents the values of each of the parameters needed to generate the upper-left-quarter of this example. First, from the PASS Home window, load the procedure window. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alpha.....	0.05
G (Number of Groups).....	4
Group Allocation Input Type	Enter K1, K2, ..., KG
K1, K2, ..., KG (List).....	25 25 25 26
M (Average Cluster Size)	3
μ 's Input Type	$\mu_1, \mu_2, \dots, \mu_G$
$\mu_1, \mu_2, \dots, \mu_G$	1.99 1.99 1.99 1
σ (Standard Deviation)	1.43178
ρ (Intraclass Correlation, ICC).....	0.45
Missing Input Type.....	Constant = 0

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results												
Number of Groups: 3												
	Total Subj Cnt	Total Clus Cnt	Group Clus Cnts Set Ki	Ave Clus Size M	Group Means Set μ_i	SD of Group Means σ_m	Std Dev σ	Effect Size σ_m/σ	Intra- Clus Corr ρ	Alpha	Miss Prop	
Power	N	K	Ki(1)	M	$\mu_i(1)$	σ_m	σ	σ_m/σ	ρ	Alpha	Miss Prop	
0.9086	303	101	Ki(1)	3	$\mu_i(1)$	0.43	1.43	0.299	0.4500	0.050	0.00	
Set(Set Number): Values												
Ki(1): 25, 25, 25, 26												
$\mu_i(1)$: 1.99, 1.99, 1.99, 1.00												

The power of 0.9086 matches Zhang and Ahn (2013). Thus, the procedure is validated.