

Chapter 534

GEE Tests for Multiple Proportions in a Cluster-Randomized Design

Introduction

This module calculates the power for testing for differences among the group proportions from binary, correlated data from a cluster-randomized design that are analyzed using a GEE logistic regression model.

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 at the end of the study.

Technical Details

Theory and Notation

Technical details are given in Wang, Zhang, and Ahn (2018).

Suppose we want to compare the proportions of G groups. Further suppose we have K_g ($g = 1, \dots, G$) clusters in each of the groups, each with an average of M subjects. Let Y_{gki} be the binary response of subject i of cluster k in group g . The response is modeled by a logistic model in which

$$Y_{gki} \sim \text{Bernoulli}(P_g)$$

The average proportion of Y_{gki} is modeled by

$$\text{logit}(P_g) = \log\left(\frac{P_g}{1 - P_g}\right) = \beta_g$$

GEE Tests for Multiple Proportions in a Cluster-Randomized Design

This implies that

$$E(Y_{gki}) = P_g = \frac{e^{\beta_g}}{1 + e^{\beta_g}}$$

The GEE estimator of β_g is b_g , given by

$$b_g = \log \left(\frac{\sum_{k=1}^{K_g} \sum_{i=1}^M Y_{gki} / (K_g M)}{1 - \sum_{k=1}^{K_g} \sum_{i=1}^M Y_{gki} / (K_g M)} \right)$$

In this procedure, the primary interest is to test that a specific contrast based on the coefficients $C = c_1, \dots, c_G$ is zero, that is, that $H_0: \sum_{g=1}^G \beta_g c_g = 0$ against the alternative that it is non-zero.

GEE is used to estimate the β_g 's and test this hypothesis. The test statistic is

$$Z = \frac{C'b}{\sqrt{\text{Var}(C'b)}}$$

H_0 is rejected with a type I error α if $|Z| > z_{1-\alpha/2}$ where $z_{1-\alpha/2}$ is the $100(1 - \alpha/2)$ th percentile of a standard normal distribution.

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).

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 Wang, Zhang, and Ahn (2018). The formula for the cluster count (K) is

$$K = \frac{C'VC \left(z_{1-\frac{\alpha}{2}} + z_{1-\gamma} \right)^2}{(C'\underline{\beta})^2}$$

where

- γ 1 – power
- α significance level

GEE Tests for Multiple Proportions in a Cluster-Randomized Design

$z_{1-\alpha/2}$ is the $100(1 - \alpha/2)$ th percentile of a standard normal distribution.

V is a diagonal matrix of elements $\left\{ h \frac{(1+e^{\beta g})^2}{r_g e^{\beta g}} \right\}$

h is $\frac{\sum_{j=1}^M \sum_{j'=1}^M \phi_{jj'} \rho_{jj'}}{(\sum_{g=1}^G \phi_g)^2}$

$\phi_{jj'}$ is the probability that both subjects j and j' are observed. By definition, this is set to one minus the proportion missing.

$\rho_{jj'}$ is the intraclass correlation coefficient between any two subjects in the same cluster.

r_g is the proportion of subjects in group g .

The above formula is easily rearranged to obtain a formula for power.

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 a zero-contrast value when in fact the contrast value is non-zero.

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.

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) 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.

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.

GEE Tests for Multiple Proportions in a Cluster-Randomized Design

Range

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

Ki (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, ..., KG (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 Ki'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.

Example (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 Proportions 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.

Example (assuming $G = 3$)

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

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.

GEE Tests for Multiple Proportions in a Cluster-Randomized Design

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 in 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 - P1, P2, ..., PG (Time-Averaged Response Probabilities)

Pi's Input Type

Specify how you want to enter P1, P2, ..., PG, the time-averaged response probabilities for each of the G groups.

Options

- **P1, P2, ..., PG**

Specify the values of the time-averaged response probabilities for each group. The differences among these values represent the effect size that you want to detect.

- **Columns Containing Sets of Pi's**

Select one or more columns of the spreadsheet that each contain a set of Pi's going down the column. Each column is analyzed separately.

P1, P2, ..., PG

Enter the values of the time-averaged response probabilities, one for each of the G groups. These entries provide values assumed by the alternative hypothesis. The null hypothesis assumes that all of the values are equal. The effect size detected by the study is a function of the differences among these values.

A time-averaged response probability for a particular individual is the average of all of the repeatedly-measured responses of that individual. This value is the anticipated mean response probability of all subjects in a particular group.

The actual analysis uses the logit of each probability where $\text{logit}(P) = \log(P/(1-P))$.

Range

Each Pi can be any number between 0 and 1. Also, at least one of the values must be different from the rest.

GEE Tests for Multiple Proportions in a Cluster-Randomized Design

Example

0.2 0.1 0.1 0.1

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 Pi's

Enter one or more spreadsheet columns containing vertical lists of P1, P2, ..., PG.

Each Pi can be any number between 0 and 1. Also, at least one of the values must be different from the rest.

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 Pi's are entered going down.

Example (assuming G = 3)

C1	C2	C3
0.5	0.5	0.5
0.1	0.2	0.3
0.1	0.3	0.4

Definition of a Single Column

Each column gives one set of probabilities. 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 column is less than G, the missing numbers are set equal to the last entry in the column.

Valid Entries

Numbers between 0 and 1.

Note

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

Effect Size - Contrast

Contrast Input Type

Select how you want to enter the contrast coefficients. Possible options are as follows.

Contrast

A contrast is a weighted average of the logits of the group response probabilities in which the weights sum to zero. For example, suppose you are studying four groups and that the main hypothesis of interest is whether there is a linear trend across the groups. You might use "-3, -1, 1, 3" here. This would form the weighted average of the logits:

$$-3 \text{ logit}(P1) - \text{logit}(P2) + \text{logit}(P3) + 3 \text{ logit}(P4)$$

These weights (the coefficients) are used to calculate a specific weighted average of the logits which is to be tested against zero using a standard z (or t) test. Hence, G coefficients (weights) must be defined.

GEE Tests for Multiple Proportions in a Cluster-Randomized Design

These coefficients must sum to zero. Also, the scale of the coefficients does not matter. That is the powers of "0.5 0.25 0.25", "-2 1 1", and "-200 100 100" are the same.

Logit

The logit of P is defined as $\text{Log}(P/(1-P))$.

Options

- **List of Contrast Coefficients**

Enter a list of coefficients, separated by commas or blanks.

- **Multiple Lists of Contrast Coefficients**

Enter several contrasts, one per column, on the spreadsheet. Each contrast is entered down a column, one coefficient per row.

- **First Group vs Rest**

A contrast is generated appropriate for testing that the logit of the first group is different from the average of the remaining groups. For example, if there were four groups, the generated coefficients would be "-3, 1, 1, 1".

- **Last Group vs Rest**

A contrast is generated appropriate for testing that the logit of the last group is different from the average logit of the remaining groups. For example, if there were four groups, the generated coefficients would be "1, 1, 1, -3".

- **Linear Trend**

A set of coefficients is generated appropriate for testing the alternative hypothesis that there is a linear (straight-line) trend across the group logits. These coefficients assume that the groups are equally spaced across some unspecified, quantitative variable associated with the groups.

- **Maximum Power (if n_i 's equal)**

A set of coefficients is generated which will result in the maximum possible power. This contrast is based on a knowledge of the actual population logits, so in practice it cannot be attained. However, it lets you determine how close your power is compared with the maximum possible.

(List of) Contrast Coefficients

Specify how you want to enter the contrast coefficients.

Enter a list of numbers, separated by blanks or commas, one number for each group. These numbers must sum to zero. The attained power is highly related to the specific contrast that is entered.

Definition

A contrast is a weighted average of the logits of the G group response probabilities in which the weights (coefficients) sum to zero. Each successive coefficient is applied to the corresponding group response probability. For example, suppose $G = 3$ and the first group is a control group. A contrast that might be of interest is "2 -1 -1". This contrast is interpreted as " $2 \text{Logit}(P_1) - \text{Logit}(P_2) - \text{Logit}(P_3)$ ". The coefficients must sum to zero.

Incomplete List

If the list is not long enough, the list is padded with zeros.

GEE Tests for Multiple Proportions in a Cluster-Randomized Design

Examples

-1 -1 2

-2 -1 0 1 2

-3 -1 1 3

-3 1 1 1

Multiple Lists of (Contrast) Coefficients

Select a set of columns on the spreadsheet which each contain a contrast defined vertically.

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.

Definition

A contrast is a weighted average of the logits of the G group response probabilities in which the weights (coefficients) sum to zero. Each successive coefficient is applied to the corresponding group mean. For example, suppose $G = 3$ and the first group is a control group. A contrast that might be of interest is "2 -1 -1". This contrast is interpreted as "2 Logit(P1) - Logit(P2) - Logit(P3)". The coefficients must sum to zero. The attained power is highly related to the specific contrast that is entered.

Incomplete List

If the list is not long enough, it is padded with zeros.

Examples

C1 - C3

C1, C2, C3

Con1, Con2, Con3

Spreadsheet Example

If $G = 4$ and there are three columns specified, the spreadsheet might look as follows.

C1	C2	C3
1	-3	1
1	-1	-1
1	1	-1
-3	3	1

Effect Size - ρ (Intracluster Correlation)

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.

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 between 10 and 70. They will assume an average cluster size of 10.

The response is whether the heart rate is above 60 bpm. With the standard drug, the percentage is 40%. The researchers want a sample size large enough to detect a decrease of 20 percentage points in the response. They will use the response percentages of 40, 20, and 20 to represent the magnitude of the difference between proportions that they want to detect. The contrast coefficients that they will use are -2, 1, 1.

Similar studies had an autocorrelation between subjects within a cluster of between 0.6 and 0.8, so they want to try values in that range. 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)	10 30 50 70
M (Average Cluster Size)	10
Pi's Input Type	P1, P2, ..., PG
P1, P2, ..., PG.....	0.4 0.2 0.2
Contrast Input Type	List of Contrast Coefficients
Contrast Coefficients	-2 1 1
ρ (Intracluster Correlation, ICC).....	0.6 0.7 0.8
Missing Input Type.....	Constant = 0

GEE Tests for Multiple Proportions in a Cluster-Randomized Design

Annotated Output

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

Numeric Results

Number of Groups: 3

	Total Subj Cnt N	Total Clus Cnt K	Clus Per Group Ki	Ave Clus Size M	Group Prop Set Pi	Cont Coef Set Ci	Lin Comb of Ci and Pi Ci'Pi	Intra- Clus Corr p	Alpha	Miss Prop
0.3001	300	30	10	10	Pi(1)	C(1)	0.40	0.6000	0.050	0.00
0.2691	300	30	10	10	Pi(1)	C(1)	0.40	0.7000	0.050	0.00
0.2446	300	30	10	10	Pi(1)	C(1)	0.40	0.8000	0.050	0.00
0.7009	900	90	30	10	Pi(1)	C(1)	0.40	0.6000	0.050	0.00
0.6438	900	90	30	10	Pi(1)	C(1)	0.40	0.7000	0.050	0.00
0.5937	900	90	30	10	Pi(1)	C(1)	0.40	0.8000	0.050	0.00
0.8945	1500	150	50	10	Pi(1)	C(1)	0.40	0.6000	0.050	0.00
0.8523	1500	150	50	10	Pi(1)	C(1)	0.40	0.7000	0.050	0.00
0.8096	1500	150	50	10	Pi(1)	C(1)	0.40	0.8000	0.050	0.00
0.9670	2100	210	70	10	Pi(1)	C(1)	0.40	0.6000	0.050	0.00
0.9449	2100	210	70	10	Pi(1)	C(1)	0.40	0.7000	0.050	0.00
0.9186	2100	210	70	10	Pi(1)	C(1)	0.40	0.8000	0.050	0.00

Set(Set Number): Values

Pi(1): 0.40, 0.20, 0.20

C(1): -2.000, 1.000, 1.000

References

Wang, J., Zhang, S., and Ahn, C. 2018. Sample Size Calculations for Comparing Time-Averaged Responses in K-group Repeated Binary Outcomes. (To appear in) Communications for Statistical Applications and Methods.

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 Ki 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 Prop Set Pi gives the name and number of the set containing the response probabilities for each group.

Cont Coef Set Ci gives the name of the set containing the contrast coefficients that are combined with the group response probabilities.

Lin Comb of Ci and Pi |Ci'Pi| gives the linear combination of the contrast coefficients and the group response probabilities.

Intra-Clus Corr p 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 proportions when it is true.

Miss Prop is the proportion of each cluster that is assumed to be missing at the end of the study.

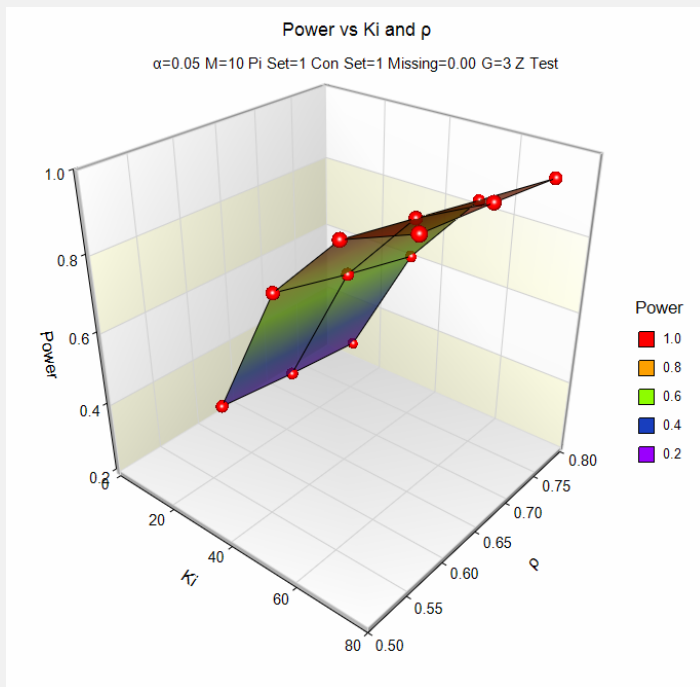
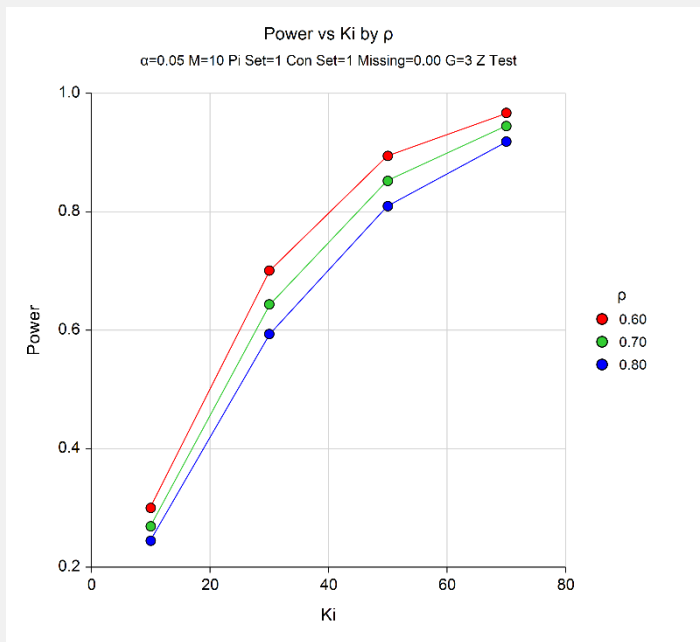
Summary Statements

A sample of 30 clusters, divided among 3 groups, each contains 10 subjects, for a total of 300 subjects, achieves a power of 0.3001. This power assumes that a z-test from a GEE analysis of a logistic model is used at a significance level of 0.050. The group cluster counts are 10, 10, 10. The group proportions under the alternative hypothesis are 0.40, 0.20, 0.20. 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 0.6000.

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 Pi and the Ci sets are shown below the table.

GEE Tests for Multiple Proportions in a Cluster-Randomized Design



These charts show the relationship among the design parameters.

Example 2 – Finding the Sample Size

Continuing with Example 1, we want to give an example that uses the spreadsheet. This allows us to compare sample size requirements for various cluster allocation patterns.

This example will use all the settings of Example 1, except that three cluster allocation patterns will be compared. The following cluster allocation patterns are entered on the spreadsheet.

<u>C1</u>	<u>C2</u>	<u>C3</u>
2	1	1
2	1	2
2	4	3

Also note that each column sums to 6.

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	K (Number of Clusters)
Power.....	0.90
Alpha.....	0.05
G (Number of Groups).....	3
Group Allocation Input Type	Enter columns of allocation patterns
Columns of Group Allocation Patterns ...	C1-C3
M (Average Cluster Size)	10
Pi's Input Type	P1, P2, ..., PG
P1, P2, ..., PG	0.4 0.2 0.2
Contrast Input Type	List of Contrast Coefficients
Contrast Coefficients	-2 1 1
ρ (Intracluster Correlation, ICC).....	0.6 0.7 0.8
Missing Input Type.....	Constant = 0

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

Output

Numeric Results

Number of Groups: 3

Power	Total Subj Cnt N	Total Clus Cnt K	Group Clus Alloc Set ri	Ave Clus Size M	Group Prop Set Pi	Cont Coef Set Ci	Lin Comb of Ci and Pi Ci*Pi	Intra-Clus Corr ρ	Alpha	Miss Prop
0.9002	1530	153	C1(1)	10	Pi(1)	C(1)	0.40	0.6000	0.050	0.00
0.9041	1770	177	C1(1)	10	Pi(1)	C(1)	0.40	0.7000	0.050	0.00
0.9030	1980	198	C1(1)	10	Pi(1)	C(1)	0.40	0.8000	0.050	0.00
0.9015	2580	258	C2(2)	10	Pi(1)	C(1)	0.40	0.6000	0.050	0.00
0.9012	2940	294	C2(2)	10	Pi(1)	C(1)	0.40	0.7000	0.050	0.00
0.9010	3300	330	C2(2)	10	Pi(1)	C(1)	0.40	0.8000	0.050	0.00
0.9056	2340	234	C3(3)	10	Pi(1)	C(1)	0.40	0.6000	0.050	0.00
0.9026	2640	264	C3(3)	10	Pi(1)	C(1)	0.40	0.7000	0.050	0.00
0.9001	2940	294	C3(3)	10	Pi(1)	C(1)	0.40	0.8000	0.050	0.00

Set(Set Number): Values

C1(1): 0.333, 0.333, 0.333

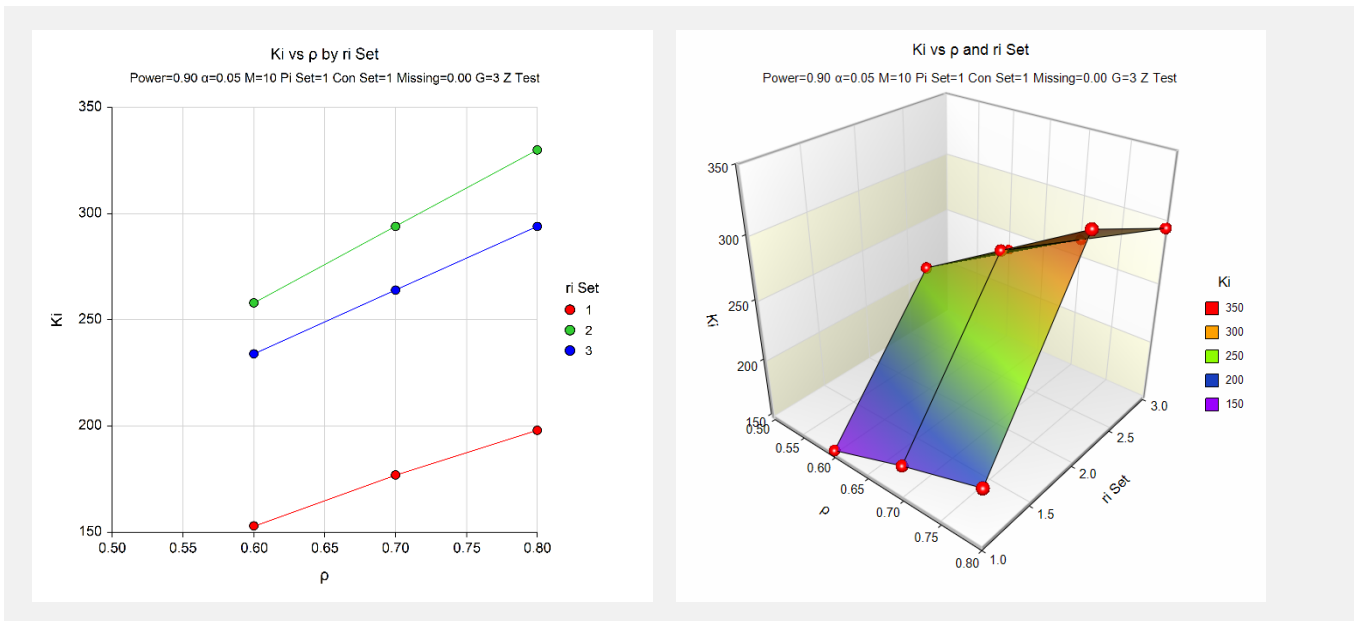
C2(2): 0.167, 0.167, 0.667

C3(3): 0.167, 0.333, 0.500

Pi(1): 0.40, 0.20, 0.20

C(1): -2.000, 1.000, 1.000

This report gives number of clusters for various values of the other parameters. Note that the details of the cluster allocation columns and the proportions are shown in the Set footnote below the numeric results.



The plot shows the sample size requirements for the three cluster allocation patterns that were used. Note that the equal allocation pattern, C1, requires the smallest number of clusters.

Example 3 – Validation of Sample Size Calculation

We could not find a validation example in the literature. However, we will use results from a previously validated procedure to validate this procedure. That procedure is *GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Binary Outcome)* which is procedure 471.

In the other procedure, if we set power = 0.8, alpha = 0.05, G = 4, Group Allocation Input Type = Equally Spaced Measurement Times, M = 6, P1...PG = 0.5 0.62245933, Contrast Coefficients = -3 1 1 1, compound symmetry correlation pattern, $\rho = 0.3$, and no missing data, the required sample size (N) is calculated as 284.

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 3** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	K (Number of Clusters)
Power.....	0.8
Alpha.....	0.05
G (Number of Groups).....	4
Group Allocation Input Type	Equal (K1 = K2 = ... = KG)
M (Average Cluster Size)	6
Pi's Input Type	P1, P2, ..., PG
P1, P2, ..., PG	0.5 0.62245933
Contrast Input Type	List of Contrast Coefficients
Contrast Coefficients	-3 1 1 1
ρ (Intracluster Correlation, ICC).....	0.3
Missing Input Type.....	Constant = 0

Output

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

Numeric Results											
Number of Groups: 4											
	Total Subj Cnt	Total Clus Cnt	Group Clus Alloc Set ri	Ave Clus Size M	Group Prop Set Pi	Cont Coef Set Ci	Lin Comb of Ci and Pi Ci'Pi	Intra- Clus Corr ρ	Alpha	Miss Prop	
Power	N	K	ri(1)	M	Pi(1)	C(1)	0.37	0.3000	0.050	0.00	
0.8007	1704	284		6							
Set(Set Number): Values											
ri(1): 0.250, 0.250, 0.250, 0.250											
Pi(1): 0.50, 0.62, 0.62, 0.62											
C(1): -3.000, 1.000, 1.000, 1.000											

The number of clusters, K, of 284 matches the value found in the other procedure. Thus, the procedure is validated.