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, ..., 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

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

This implies that

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

© NCSS, LLC. All Rights Reserved.

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 $C = c_1, \ldots, c_G$ 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

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

H_0 is rejected with a type I error α if $|\boldsymbol{Z}| > z_{1-\alpha/2}$ 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.

© NCSS, LLC. All Rights Reserved.

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

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

where

γ 1 – power

α significance level

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

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

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

$\phi_{jj'}$ 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 intracluster 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.

© NCSS, LLC. All Rights Reserved.

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 Ki 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

If the procedure window is not already open, use the PASS Home window to open it. The parameters 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.

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**

© NCSS, LLC. All Rights Reserved.

Output

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

Numeric Reports

Numeric Results

Solve For: Power
Number of Groups: 3

Power	Number of Subjects N	Number of Clusters K	Clusters Per Group Ki	Average Cluster Size M	Group Proportions Pi	Contrast Coefficients Ci	Linear Combination of Ci and Pi |Ci'Pi|	ICC ρ	Alpha	Missing Proportion
0.3001	300	30	10	10	Pi(1)	C(1)	0.4	0.6	0.05	0
0.2691	300	30	10	10	Pi(1)	C(1)	0.4	0.7	0.05	0
0.2446	300	30	10	10	Pi(1)	C(1)	0.4	0.8	0.05	0
0.7009	900	90	30	10	Pi(1)	C(1)	0.4	0.6	0.05	0
0.6438	900	90	30	10	Pi(1)	C(1)	0.4	0.7	0.05	0
0.5937	900	90	30	10	Pi(1)	C(1)	0.4	0.8	0.05	0
0.8945	1500	150	50	10	Pi(1)	C(1)	0.4	0.6	0.05	0
0.8523	1500	150	50	10	Pi(1)	C(1)	0.4	0.7	0.05	0
0.8096	1500	150	50	10	Pi(1)	C(1)	0.4	0.8	0.05	0
0.9670	2100	210	70	10	Pi(1)	C(1)	0.4	0.6	0.05	0
0.9449	2100	210	70	10	Pi(1)	C(1)	0.4	0.7	0.05	0
0.9186	2100	210	70	10	Pi(1)	C(1)	0.4	0.8	0.05	0

Item	Values
Pi(1)	0.4, 0.2, 0.2
C(1)	-2, 1, 1

Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N	The total number of subjects in the study.
K	The total number of clusters in the study.
Ki	The number of clusters per group.
M	The average cluster size (number of subjects per cluster). This is used for all clusters.
Pi	Group Proportions. Gives the name and number of the set containing the response probabilities for each group.
Ci	Contrast Coefficients. Gives the name of the set containing the contrast coefficients that are combined with the group response probabilities.
|Ci'Pi|	Linear Combination of Ci and Pi. Gives the linear combination of the contrast coefficients and the group response probabilities.
ρ	The intracluster correlation coefficient (ICC) used for all clusters. This is the correlation between any two subjects within a cluster.
Alpha	The probability of rejecting a true null hypothesis.
Missing Proportion	The proportion of each cluster that is assumed to be missing at the end of the study.

Summary Statements

A parallel, 3-group cluster-randomized design will be used to test the difference among the 3 proportions that is defined by the contrast coefficients -2, 1, 1. The comparison will be made using a generalized estimating equation (GEE) logistic model Z-test with a Type I error rate (α) of 0.05. The autocorrelation matrix of the responses within a cluster is assumed to be compound symmetric with an intraclass correlation coefficient (ICC) of 0.6. Missing values are assumed to occur completely at random (MCAR), and the anticipated proportion missing is 0. To detect the group proportions 0.4, 0.2, 0.2, with contrast coefficients -2, 1, 1, with a total of 30 clusters (allocated to the 3 groups as 10, 10, 10), with an average cluster size of 10 subjects per cluster (for a total sample size of 300 subjects), the power is 0.3001.

© NCSS, LLC. All Rights Reserved.

References

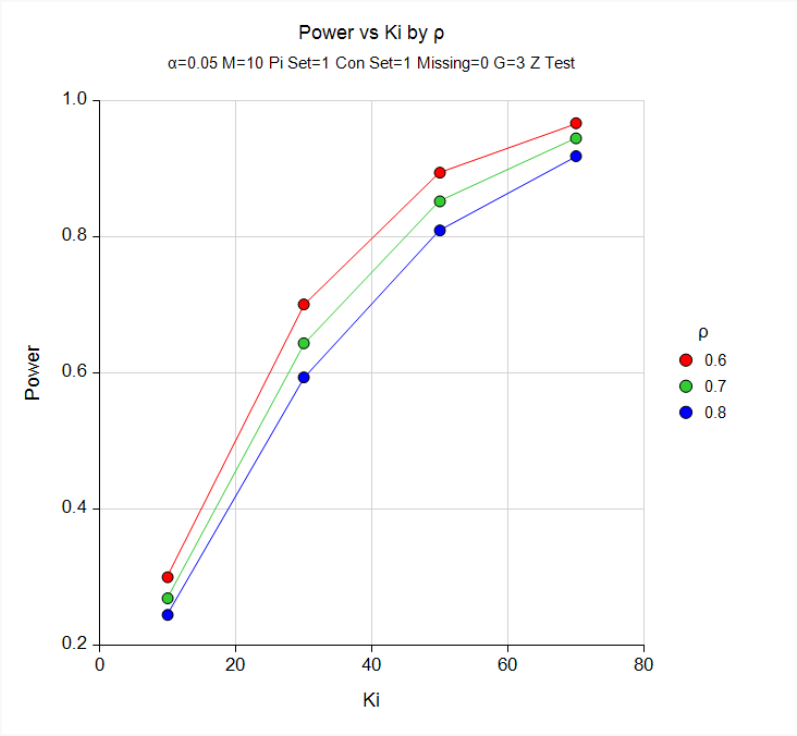
Wang, J., Zhang, S., and Ahn, C. 2018. Sample Size Calculations for Comparing Time-Averaged Responses in K-group Repeated Binary Outcomes. (To appear in) Communications for Statistical Applications and Methods.

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.

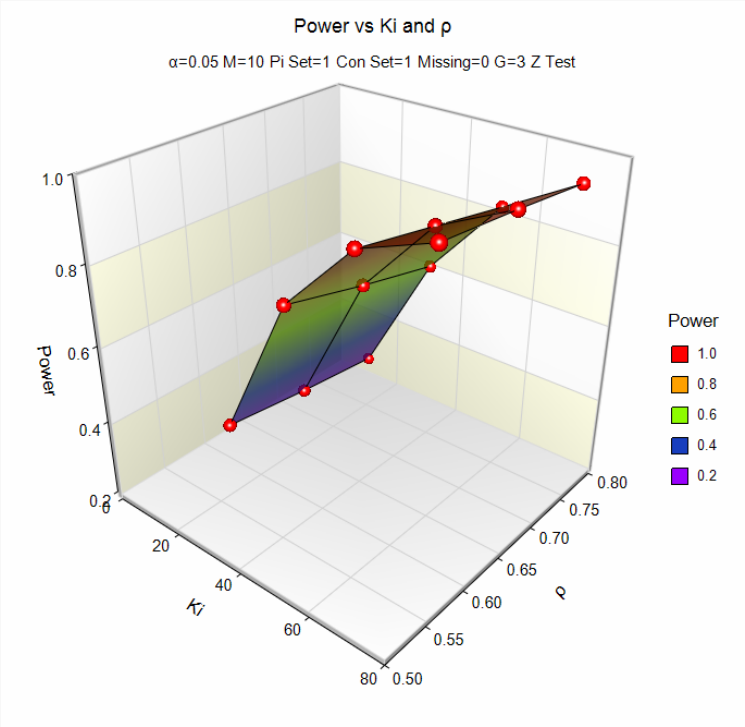
Note that the details of the Pi and the Ci sets are shown below the table.

Plots Section

Plots



© NCSS, LLC. All Rights Reserved.



These plots show the relationship among the design parameters.

© NCSS, LLC. All Rights Reserved.

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.

C1	C2	C3
2	1	1
2	1	2
2	4	3

Also note that each column sums to 6.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

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**

Input Spreadsheet Data

Row	C1	C2	C3
1	2	1	1
2	2	1	2
3	2	4	3

© NCSS, LLC. All Rights Reserved.

Output

Numeric Reports

Numeric Results

Solve For: K (Number of Clusters)
Number of Groups: 3

Power	Number of Subjects N	Number of Clusters K	Group Cluster Allocation ri	Average Cluster Size M	Group Proportions Pi	Contrast Coefficients Ci	Linear Combination of Ci and Pi |Ci'Pi|	ICC ρ	Alpha	Missing Proportion
0.9002	1530	153	C1(1)	10	Pi(1)	C(1)	0.4	0.6	0.05	0
0.9041	1770	177	C1(1)	10	Pi(1)	C(1)	0.4	0.7	0.05	0
0.9030	1980	198	C1(1)	10	Pi(1)	C(1)	0.4	0.8	0.05	0
0.9015	2580	258	C2(2)	10	Pi(1)	C(1)	0.4	0.6	0.05	0
0.9012	2940	294	C2(2)	10	Pi(1)	C(1)	0.4	0.7	0.05	0
0.9010	3300	330	C2(2)	10	Pi(1)	C(1)	0.4	0.8	0.05	0
0.9056	2340	234	C3(3)	10	Pi(1)	C(1)	0.4	0.6	0.05	0
0.9026	2640	264	C3(3)	10	Pi(1)	C(1)	0.4	0.7	0.05	0
0.9001	2940	294	C3(3)	10	Pi(1)	C(1)	0.4	0.8	0.05	0

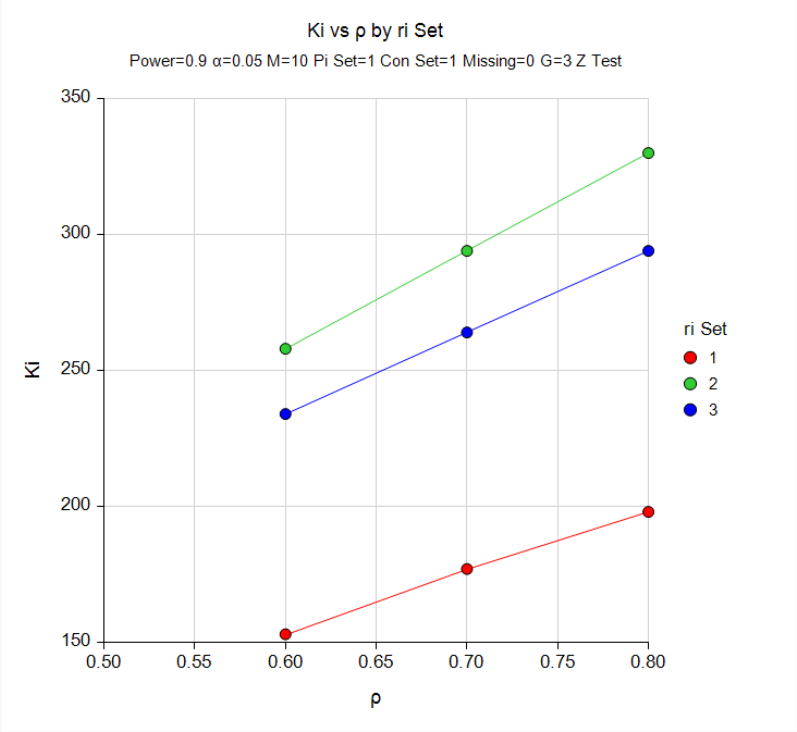
Item	Values
C1(1)	0.333, 0.333, 0.333
C2(2)	0.167, 0.167, 0.667
C3(3)	0.167, 0.333, 0.5
Pi(1)	0.4, 0.2, 0.2
C(1)	-2, 1, 1

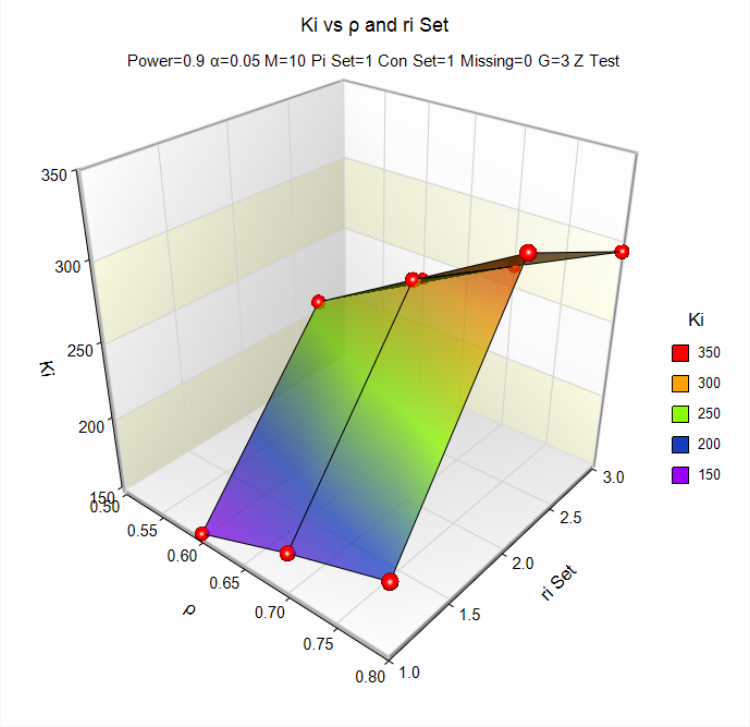
This report gives the number of clusters for various values of the other parameters. Note that the details of the cluster allocation columns and the proportions are shown in the Set footnote below the numeric results.

© NCSS, LLC. All Rights Reserved.

Plots Section

Plots





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.

© NCSS, LLC. All Rights Reserved.

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, ρ = 0.3, and no missing data, the required sample size (N) is calculated as 284.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

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**

© NCSS, LLC. All Rights Reserved.

Output

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

Numeric Results

Solve For: K (Number of Clusters)
Number of Groups: 4

Power	Number of Subjects N	Number of Clusters K	Group Cluster Allocation ri	Average Cluster Size M	Group Proportions Pi	Contrast Coefficients Ci	Linear Combination of Ci and Pi |Ci'Pi|	ICC ρ	Alpha	Missing Proportion
0.8007	1704	284	ri(1)	6	Pi(1)	C(1)	0.37	0.3	0.05	0

Item	Values
ri(1)	0.25, 0.25, 0.25, 0.25
Pi(1)	0.5, 0.62, 0.62, 0.62
C(1)	-3, 1, 1, 1

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

© NCSS, LLC. All Rights Reserved.