

Chapter 531

GEE Tests for Two Means in a Cluster-Randomized Design

Introduction

This module calculates the power for testing the difference between two means obtained from a cluster-randomized design. The mean come from continuous, correlated data that are analyzed using the GEE method. Such data occur in two design types: clustered and longitudinal. This procedure is specific to cluster-randomized designs. A companion procedure power analyzes data from a repeated measures design.

GEE is different from mixed models in that it does not require the full specification of the joint distribution of the clustered measurements, as long as the marginal mean model is correctly specified. Estimation consistency is achieved even if the correlation matrix is incorrect. Also, the correlation matrix of the responses is specified directly, rather than using an intermediate, random effects model as is the case in mixed models. For clustered designs, GEE often uses a *compound symmetric* (CS) correlation structure.

Time-averaged difference analysis is often used when the outcome varies with time. However, in this case, the observations are treated as if they were repeated measurements from a subject (cluster). Each cluster is randomized either in the treatment or the control group.

Making these assumptions, the data may be analyzed using the GEE TAD methodology. This procedure performs a power analysis and sample size calculation for data obtained and analyzed in this manner.

Missing Values

This procedure allows you to specify a proportion of the subjects that are missing from each cluster.

Technical Details

Theory and Notation

Technical details are given in Ahn, Heo, and Zhang (2015), chapter 4, section 4.4.1, pages 110-116.

Suppose we have K_1 clusters in group 1 (treatment) and K_2 clusters in group 2 (control) for a total of K clusters. The average cluster size (number of subjects per cluster) is M . The mean of y_{ij} is modeled by

$$\mu_{ij} = \beta_1 + \beta_2 r_i$$

where

y_{ij} is the response of the j^{th} subject from cluster i , with variance σ^2

μ_{ij} is expectation of y_{ij}

r_i is the cluster treatment indicator with 0 for control and 1 for treatment

β_1 is the regression coefficient giving the intercept of the control group

β_2 is the regression coefficient giving the TAD between the two groups

In this procedure, the primary interest is on β_2 .

This mean model is reparametrized as

$$\mu_{ij} = b_1 + b_2(r_i - \bar{r})$$

where

$$b_1 = \beta_1 + \bar{r}\beta_2$$

$$b_2 = \beta_2$$

The vector of covariates is given by $x_{ij} = (1, r_i - \bar{r})'$.

GEE is used to estimate and test hypotheses about \mathbf{b} with $\hat{\mathbf{b}}$.

Correlation Matrix Structure

In a cluster-randomized design with K clusters, each containing M subjects, observations within a particular cluster are correlated. The pattern of those correlations is assumed to be compound symmetric, with a single correlation, ρ .

A compound symmetry correlation model assumes that all correlations are equal, and all diagonal elements are one. That is

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

Sample Size Calculations

The details of the calculation of sample size and power is given in Ahn, Heo, and Zhang (2015), Chapter 4, page 113. These are summarized here.

As explained above, GEE is used to estimate the regression coefficients \mathbf{b} with $\hat{\mathbf{b}}$. The significance of b_4 , the coefficient associated with the difference between the control and treatment slopes, is tested using a Wald statistic from which the following sample size formula is derived

$$K = \frac{\sigma^2 \eta_0 \left(z_{1-\frac{\alpha}{h}} + z_{1-\gamma} \right)^2}{\delta^2 \mu_0^2 \sigma_r^2}$$

where

K = number of clusters

M = average cluster size

h = 1 (one-sided test) or 2 (two-sided test)

γ = 1 – power

α = significance level

$\delta = \beta_2$ = the mean difference

κ = proportion missing within a cluster

$\mu_0 = M(1 - \kappa)$

$\eta_0 = [M^2 \rho + (1 - \rho)](1 - \kappa)$

$\sigma_r^2 = \bar{r}(1 - \bar{r})$

ρ = the intraclass correlation coefficient, ICC

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 KI , the program searches for the lowest sample size that meets the alpha and power criterion you have specified. The Solve For parameter is displayed on the vertical axis of the plot.

GEE Tests for Two Means in a Cluster-Randomized Design

Test

Alternative Hypothesis

This option specifies the alternative hypothesis. The null hypothesis is always $H_0: \delta = 0$

Note that the alternative hypothesis enters into power calculations by specifying the rejection region of the hypothesis test. Its accuracy is critical.

Possible selections are:

- **One-Sided**

This option yields a *one-tailed* test. Use it for testing the alternative hypotheses $H_1: \delta < 0$ or $H_1: \delta > 0$.

- **Two-Sided**

This is the most common selection. Use it for testing the alternative hypotheses $H_1: \delta \neq 0$.

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. This means that about one test in twenty will falsely reject the null hypothesis. 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*.

Sample Size – Number of Clusters and Cluster Size

K1 (Number of Group 1 Clusters)

Enter a value for K1, the number of clusters assigned to group 1 (the treatment group). Each cluster is assumed to have M subjects.

You may enter a single value such as *42*, a list of values such as *10, 20, 30*, or a range of values such as *10 to 100 by 10*.

K2 (Number of Group 2 Clusters)

Enter one or more values for the number of clusters in group 2, which we have designated as the control group. The sample size for this group is equal to the number of clusters times the number of subjects per cluster.

This value must be a positive number.

It can be a decimal (fractional) number such as '12.7'. The resulting value of N2 will be rounded to an integer.

GEE Tests for Two Means in a Cluster-Randomized Design

Using Multiples of K1

If you simply want a multiple of the value of group one, you would enter the multiple followed by "K1", with no blanks. If you want to use K1 directly, you do not have to enter the leading "1". For example, all of the following are valid entries:

10K1 2K1 0.5K1 K1.

You can use a list of values such as "10 20 30" or "K1 2K1 3K1".

M (Average Cluster Size)

Specify one or more values for M , the average number of subjects per 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 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

δ (Mean Difference = $\mu_1 - \mu_2$)

Enter a value (or range of values) for the cluster average difference (TAD) between the two treatments that you want to be able to detect. This value is in the same scale as σ .

These values must be positive.

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

σ (Standard Deviation)

Enter a value (or range of values) for the standard deviation of the response, Y_{ij} where i is the cluster and j is the subject. The analysis assumes that this standard deviation is constant across all subjects.

These values must be positive.

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.

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 pattern of missing values.

Possible input choices are

- **Constant = 0**
The missing proportion is set to 0. That is, there are no missing values.
- **Constant**
All missing proportions are set to particular value.

Missing Proportion

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

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 Sample Size

Researchers are planning a cluster-randomized study to compare a new drug with the standard drug for the treatment of a certain disease. The researchers want to investigate what happens when the cluster size is between 10 and 40 patients. The outcome will be the change in heart rate.

The study will be powered to detect a difference of 4. Similar studies have found a standard deviation between 8.0 and 10.0. These studies had an ICC of 0.2. The two-sided test will be conducted at the 0.05 significance level and at 90% power. They are planning on dividing the subjects equally between the treatment and control groups. The researchers anticipate the proportion missing will be 0.10.

How many clinics will be required for this study?

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load 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	K1 (Number of Group 1 Clusters)
Alternative Hypothesis	Two-Sided
Power	0.90
Alpha	0.05
K2 (Number of Group 2 Clusters)	K1
M (Average Cluster Size)	10 20 30 40
δ (Mean Difference = $\mu_1 - \mu_2$)	4
σ (Standard Deviation)	8 9 10
ρ (Intracluster Correlation, ICC)	0.2
Missing Input Type	Constant
Missing Proportion	0.1

Annotated Output

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

Numeric Results												
Alternative Hypothesis: Two-Sided												
Power	Number of Subjects	Clus Cnt Total	Clus Cnt Grp 1	Clus Cnt Grp 2	Alloc Ratio K1/K2	Clus Size Ave M	Mean Diff $\mu_1 - \mu_2$	Std Dev σ	Intra-Clus Corr ρ	Alpha	Miss Prop	
0.9088	540	54	27	27	1.00	10	4	8	0.2	0.050	0.10	
0.9075	680	68	34	34	1.00	10	4	9	0.2	0.050	0.10	
0.9010	820	82	41	41	1.00	10	4	10	0.2	0.050	0.10	
0.9072	920	46	23	23	1.00	20	4	8	0.2	0.050	0.10	
0.9061	1160	58	29	29	1.00	20	4	9	0.2	0.050	0.10	
0.9076	1440	72	36	36	1.00	20	4	10	0.2	0.050	0.10	
0.9106	1320	44	22	22	1.00	30	4	8	0.2	0.050	0.10	
0.9022	1620	54	27	27	1.00	30	4	9	0.2	0.050	0.10	
0.9076	2040	68	34	34	1.00	30	4	10	0.2	0.050	0.10	
0.9061	1680	42	21	21	1.00	40	4	8	0.2	0.050	0.10	
0.9104	2160	54	27	27	1.00	40	4	9	0.2	0.050	0.10	
0.9076	2640	66	33	33	1.00	40	4	10	0.2	0.050	0.10	

GEE Tests for Two 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 page 113.

Report Definitions

Power is the probability of rejecting a false null hypothesis.

N is the total number of subjects in the study.

K is the total number of clusters. $K = K1 + K2$.

K1 is the number of clusters in group 1, the treatment group.

K2 is the number of clusters in group 2, the control group.

K1/K2 is the allocation ratio. The smallest sample size occurs when it is 1.0.

M is the average cluster size. That is, it is the number of subjects per cluster.

δ is the time-averaged difference (TAD) at which the power is calculated. It is in the same scale as σ .

σ is the standard deviation of a response.

ρ is the intraclass correlation coefficient or ICC. It is the correlation between any two subjects within a cluster.

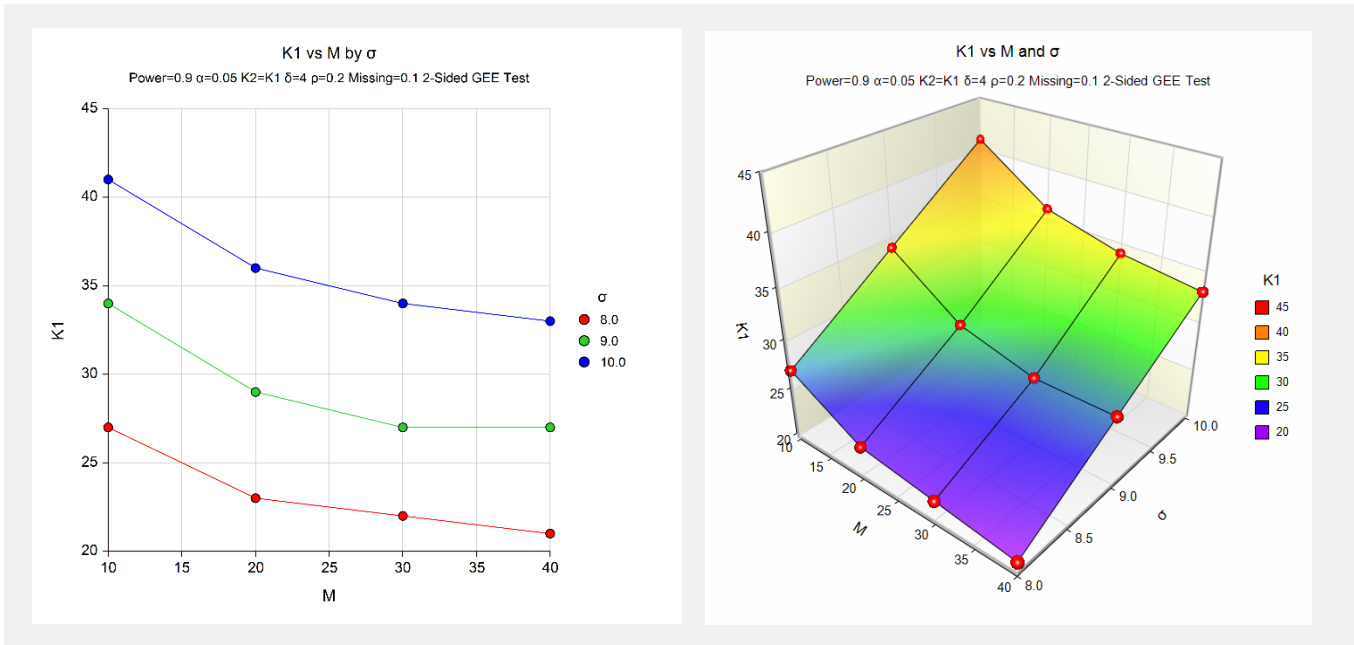
Alpha is the significance level of the test. It is probability of rejecting the null hypothesis when it is true.

Miss Prop is the missing proportion. This is the proportion of subjects that are not expected to complete the study.

Summary Statements

A sample of 54 clusters, each with a cluster size of 27, achieves a power of 0.9088 when using a two-sided Wald test from a GEE analysis to test whether the treatment group differs from the control group by more than 4 at a significance level of 0.050. The total number of subjects in the study is 540. The treatment group contains 27 clusters and the control group contains 27 clusters. The residual standard deviation is anticipated to be 8. The intraclass correlation coefficient (ICC) is 0.2. The anticipated proportion missing is 0.10.

This report gives the necessary cluster count for each value of the other parameters.



These charts show the relationship among the parameters varied.

Example 2 – Finding the Power

Continuing with Example 1, the researchers want to determine the power corresponding to K1 values ranging from 20 to 40 and a standard deviation of 9.

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**.

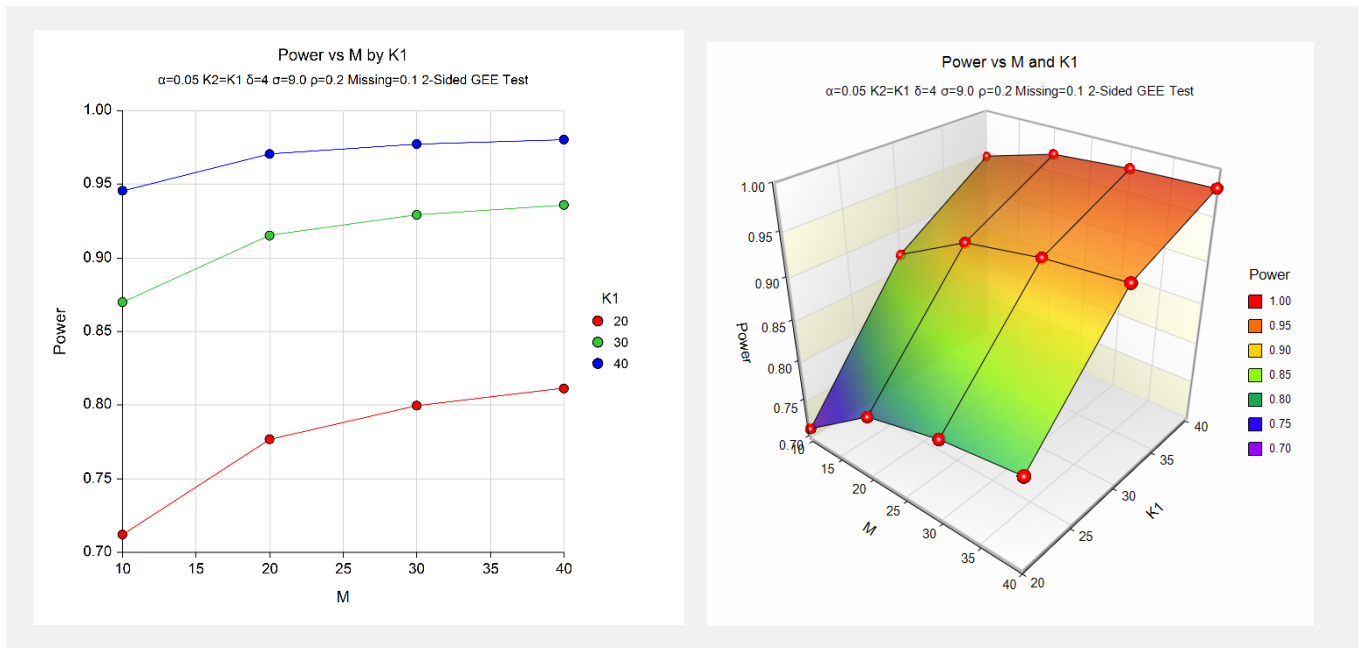
<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alternative Hypothesis	Two-Sided
Alpha	0.05
K1 (Number of Group 1 Clusters)	20 30 40
K2 (Number of Group 2 Clusters)	K1
M (Average Cluster Size)	10 20 30 40
δ (Mean Difference = $\mu_1 - \mu_2$)	4
σ (Standard Deviation)	9
ρ (Intracluster Correlation, ICC)	0.2
Missing Input Type	Constant
Missing Proportion	0.1

Output

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

Numeric Results											
Alternative Hypothesis: Two-Sided											
Power	Number of Subjects	Clus Cnt Total	Clus Cnt Grp 1	Clus Cnt Grp 2	Alloc Ratio K1/K2	Clus Size Ave	Mean Diff $\mu_1 - \mu_2$	Std Dev	Intra-Clus Corr	Alpha	Miss Prop
	N	K	K1	K2		M	δ	σ	ρ		
0.7122	400	40	20	20	1.00	10	4	9	0.2	0.050	0.10
0.7769	800	40	20	20	1.00	20	4	9	0.2	0.050	0.10
0.7997	1200	40	20	20	1.00	30	4	9	0.2	0.050	0.10
0.8113	1600	40	20	20	1.00	40	4	9	0.2	0.050	0.10
0.8699	600	60	30	30	1.00	10	4	9	0.2	0.050	0.10
0.9152	1200	60	30	30	1.00	20	4	9	0.2	0.050	0.10
0.9292	1800	60	30	30	1.00	30	4	9	0.2	0.050	0.10
0.9359	2400	60	30	30	1.00	40	4	9	0.2	0.050	0.10
0.9456	800	80	40	40	1.00	10	4	9	0.2	0.050	0.10
0.9706	1600	80	40	40	1.00	20	4	9	0.2	0.050	0.10
0.9773	2400	80	40	40	1.00	30	4	9	0.2	0.050	0.10
0.9803	3200	80	40	40	1.00	40	4	9	0.2	0.050	0.10

GEE Tests for Two Means in a Cluster-Randomized Design



The report and plot show the power for each scenario.

Example 3 – Validation using Ahn, Heo, and Zhang (2015)

Ahn, Heo, and Zhang (2015) do not present any numerical examples for this procedure. However, on page 115 they give a simplified formula for the compound symmetry case. We will calculate a result by hand using this formula and then calculate with the program to validate this procedure.

The simplified sample size formula is

$$K = \frac{4\sigma^2[1 + (M - 1)\rho] \left(z_{1-\frac{\alpha}{h}} + z_{1-\gamma} \right)^2}{\delta^2 M}$$

Using $M = 5$, $\sigma = 3$, $\rho = 0.5$, $\delta = 1.5$, $\alpha = 0.05$, and power = 0.9, we obtain

$$\begin{aligned} K &= \text{ceiling} \left(\frac{4(3^2)[1 + (5 - 1)0.5](1.96 + 1.28)^2}{1.5^2(5)} \right) \\ &= \text{ceiling}(100.78) \\ &= 101 \end{aligned}$$

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	K1 (Number of Group 1 Clusters)
Alternative Hypothesis	Two-Sided
Power	0.90
Alpha	0.05
K2 (Number of Group 2 Clusters)	K1
M (Average Cluster Size)	5
δ (Mean Difference = $\mu_1 - \mu_2$)	1.5
σ (Standard Deviation)	3
ρ (Intracluster Correlation, ICC)	0.5
Missing Input Type	Constant = 0

Output

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

Numeric Results												
Alternative Hypothesis: Two-Sided												
	Number of Subjects	Clus Cnt Total	Clus Cnt Grp 1	Clus Cnt Grp 2	Alloc Ratio K1/K2	Clus Size Ave M	Mean Diff $\mu_1 - \mu_2$ δ	Std Dev σ	Intra-Clus Corr ρ	Alpha	Miss Prop	
Power	0.9031	510	102	51	51	1.00	5	1.5	3	0.5	0.050	0.00

The number of clusters of 102 matches the above calculations very closely. Actually, the hand calculations of 101 clusters did not take into account that we wanted $K_1 = K_2$, which require an even number. PASS gave us the desired even number by increasing the number of clusters from 101 to 102.