

## Chapter 392

# GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)

---

### Introduction

This module calculates the power for simultaneously testing the differences among the slopes of two or more groups of continuous, correlated data that are analyzed using the GEE method. Such data can occur in both clustered and longitudinal designs. The slopes are measured across repeated measurements of a specific subject.

GEE is different from mixed models in that it does not require the full specification of the joint distribution of the repeated 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 MM. For clustered designs, GEE often uses a *compound symmetric* (CS) correlation structure. For longitudinal data, an *autoregressive* (AR(1)) correlation structure is often used.

### Missing Values

This procedure allows you to specify various patterns of incomplete (or missing) data. Subjects may miss some appointments, but attend others. This phenomenon of incomplete data can be accounted for in the sample size calculation which can greatly reduce the overall sample size from that calculated by just omitting subjects with incomplete observations.

## Technical Details

### Theory and Notation

Technical details are given in Ahn, Heo, and Zhang (2015), chapter 4, section 4.3.5, pages 100-105 and in Jung and Ahn (2004).

Suppose we have  $n_k$  ( $k = 1, \dots, G$ ) subjects in each of  $G$  groups for a total of  $N$  subjects, each measured on  $M$  occasions at times  $t_j$  ( $j = 1, \dots, M$ ). For convenience, we normalize these time points to the proportion of total time so that  $t_1 = 0$  and  $t_M = 1$ .

For each group  $k$ , the time-specific mean of  $y_{kij}$  is fit using GEE by

$$y_{kij} = \mu_{kij} + \epsilon_{kij} \text{ and } \mu_{kij} = \theta_k + \beta_k t_j$$

where

$y_{kij}$  is the  $j^{\text{th}}$  response from subject  $i$  in group  $k$  at time  $t_j$ , with variance given by  $\sigma^2$ ,

$r_k$  is the proportion of subjects in group  $k$ ,

$\mu_{kij}$  is group-specific treatment effect ( $k = 1, \dots, G$ ) given by  $\theta_k + \beta_k t_j$ ,

$\epsilon_{kij}$  is a zero-mean, random-effect error term with covariance  $\sigma^2 \rho_{jj'}$

$\theta_k$  is the intercept of group  $k$ ,

$\beta_k$  is the slope of group  $k$ ,

$t_j$  is  $j^{\text{th}}$  measurement time, scaled to go from 0 to 1.

Note that at the first scheduled time

$$\mu_{kij} = \theta_k$$

And at the last scheduled time

$$\mu_{kij} = \theta_k + \beta_k.$$

Also, since subjects are assigned randomly to the groups, it is reasonable to assume that the responses will all have the same mean at  $t_1$ . Thus, it is likely that all intercepts will be equal.

The main focus of this design is to compare the average response of each group at the final time point,  $t_M$ . The difference between any two group means will thus reduce to the difference between the corresponding group slopes. That is why this test is popular.

In this procedure, the primary interest is to test  $H_0: \beta_1 = \dots = \beta_G$  against the alternative that at least one slope is different.

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

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

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

## Correlation Patterns

In a longitudinal design with  $N$  subjects, each measured  $M$  times, observations from a single subject are correlated, and a pattern of those correlations through time needs to be specified. Several choices are available.

### Compound Symmetry

A compound symmetry covariance model assumes that all covariances are equal, and all variances on the diagonal are equal. That is

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

where  $\rho$  is the baseline correlation.

### Banded(1)

A Banded(1) (banded order 1) correlation model assumes that all variances on the diagonal are equal, covariances for observations one time period apart are equal to  $\rho$ , and covariances for measurements greater than one time period apart are equal to zero. That is

$$[\rho_{jj'}] = \begin{bmatrix} 1 & \rho & 0 & 0 & \cdots & 0 \\ \rho & 1 & \rho & 0 & \cdots & 0 \\ 0 & \rho & 1 & \rho & \cdots & 0 \\ 0 & 0 & \rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{M \times M}$$

where  $\rho$  is the baseline correlation.

### Banded(2)

A Banded(2) (banded order 2) correlation model assumes that all variances on the diagonal are equal, covariances for observations one time period or two periods apart are equal to  $\rho$ , and covariances for measurements greater than one time period apart are equal to zero. That is

$$[\rho_{jj'}] = \begin{bmatrix} 1 & \rho & \rho & 0 & \cdots & 0 \\ \rho & 1 & \rho & \rho & \cdots & 0 \\ \rho & \rho & 1 & \rho & \cdots & 0 \\ 0 & \rho & \rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{M \times M}$$

where  $\rho$  is the baseline correlation.

## GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)

**AR1 (Traditional)**

This version of AR1 (autoregressive order 1) correlation model assumes that all variances on the diagonal are equal and that correlations  $t$  time periods apart are equal to  $\rho^t$ . That is

$$[\rho_{jj'}] = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{M-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^{M-2} \\ \rho^2 & \rho & 1 & \rho & \dots & \rho^{M-3} \\ \rho^3 & \rho^2 & \rho & 1 & \dots & \rho^{M-4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{M-1} & \rho^{M-2} & \rho^{M-3} & \rho^{M-4} & \dots & 1 \end{bmatrix}_{M \times M}$$

where  $\rho$  is the baseline correlation.

**AR1 (Proportional)**

This version of AR1 (autoregressive order 1) correlation model is described in the book by Ahn et al. (2015). It assumes that all variances on the diagonal are equal and that correlations  $|t_j - t_k|$  time periods apart are equal to  $\rho^{|t_j - t_k|}$ . That is

$$[\rho_{jk}] = [\rho^{|t_j - t_k|}]_{M \times M}$$

where  $\rho$  is the baseline correlation. Note that in this pattern, the value of  $\rho$  is shown in the final column since in this case  $t_j = 0$  and  $t_k = 1$ , so  $|t_j - t_k| = 1$ .

**Dampened Exponential**

A dampened exponential is an extension of the AR(1) correlation model in which the exponents are raised to the power  $Dexp$  ( $\theta = Dexp$  in the diagram below). This causes the resulting correlations to be reduced (dampened). Here is an example

$$[\rho_{jj'}] = \begin{bmatrix} 1 & \rho & \rho^{2\theta} & \rho^{3\theta} & \dots & \rho^{(M-1)\theta} \\ \rho & 1 & \rho & \rho^{2\theta} & \dots & \rho^{(M-2)\theta} \\ \rho^{2\theta} & \rho & 1 & \rho & \dots & \rho^{(M-3)\theta} \\ \rho^{3\theta} & \rho^{2\theta} & \rho & 1 & \dots & \rho^{(M-4)\theta} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{(M-1)\theta} & \rho^{(M-2)\theta} & \rho^{(M-3)\theta} & \rho^{(M-4)\theta} & \dots & 1 \end{bmatrix}_{M \times M}$$

where  $\rho$  is the baseline correlation.

**Dampened Exponential (Proportional)**

This version of the dampened exponential correlation model is described in the book by Ahn et al. (2015). It assumes that all variances on the diagonal are equal and that correlations  $|t_j - t_k|$  time periods apart are equal to  $\rho^{|t_j - t_k|^\theta}$ . That is

$$[\rho_{jk}] = [\rho^{|t_j - t_k|^\theta}]_{M \times M}$$

where  $\rho$  is the baseline correlation. Note that in this pattern, the value of  $\rho^{|t_j - t_k|^\theta}$  turns up in the final column since in this case  $t_j = 0$  and  $t_k = 1$ , so  $|t_j - t_k| = 1$ .

## GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)

### Linear Exponential Decay

A linear exponential decay correlation model is one in which the exponent of the correlation decays according to a linear equation from 1 at the *Base Time Proportion* to a final value,  $E_{max}$ . The resulting pattern looks similar to the dampened exponential. Note that the exponents are applied to the absolute difference between the Measurement Time Proportions. This method allows you to easily construct comparable correlation matrices of different dimensions. Otherwise, differences in the resulting power would be more strongly due to differences in the correlation matrices.

Here is an example. Suppose  $M$  is 6,  $\rho = 0.5$ ,  $E_{max} = 3$ , the *Base Time Proportion* is 0.20, and the Measurement Time Proportions are (0, 0.2, 0.4, 0.6, 0.8, 1). The following correlation matrix would be obtained

$$[\rho_{jj'}] = \begin{bmatrix} 1 & 0.5 & 0.3536 & 0.25 & 0.1768 & 0.125 \\ 0.5 & 1 & 0.5 & 0.3536 & 0.25 & 0.1768 \\ 0.3536 & 0.5 & 1 & 0.5 & 0.3536 & 0.25 \\ 0.25 & 0.3536 & 0.5 & 1 & 0.5 & 0.3536 \\ 0.1768 & 0.25 & 0.3536 & 0.5 & 0.3536 & 0.5 \\ 0.125 & 0.1768 & 0.25 & 0.3536 & 0.5 & 1 \end{bmatrix}_{M \times M}$$

Note that in the top row, the correlation is 0.5 for the second (0.2 - 0) time point and 0.125 ( $0.5^3$ ) at the last (1 - 0) time points. The correlations are obtained by raising 0.5 to the appropriate exponent. The linear equation from 1 to 3 results in the exponents 1, 1.5, 2, 2.5, 3 correspondent to the time proportions 0, 0.2, 0.4, 0.6, 0.8, and 1.

As a further example, note that the correlation for the 0.4 time point is,  $0.5^{1.5} = 0.35355339 \approx 0.3536$ .

This method allows you to compare various values of  $M$  while keeping the correlation matrix similar. To see what we mean, consider what the correlation matrix looks like when  $M$  is reduced to 4 and the measurement time proportions are set to (0, 0.2, 0.6, 1). It becomes

$$[\rho_{jj'}] = \begin{bmatrix} 1 & 0.5 & 0.25 & 0.125 \\ 0.5 & 1 & 0.5 & 0.25 \\ 0.25 & 0.5 & 1 & 0.5 \\ 0.125 & 0.25 & 0.5 & 1 \end{bmatrix}_{M \times M}$$

Note that the correlation at a measurement time difference of 0.6 is equal to 0.25 in both matrices.

### Missing Data Patterns

The problem of missing data occurs for several reasons. In longitudinal studies in which a subject is measured multiple times, missing data becomes more complicated to model because it is possible that a subject is measured only some of the time. In fact, it is probably more common for data to be incomplete than complete. The approach of omitting subjects with incomplete data during the planning phase is very inaccurate. Certainly, subjects with partial measurements are included in the analysis. This procedure provides several missing data patterns to choose from so that your sample size calculations are more realistic.

In the presentation to following, we denote the percent of subjects with a missing response at time point  $t_j$  as  $\kappa_j$ . The proportion non-missing at a particular time point is  $\phi_j = 1 - \kappa_j$ . We will refer to  $\phi_j$  as the *marginal observant probability* at time  $t_j$  and  $\phi_{jj'}$  as a *joint observant probability* at times  $t_j$  and  $t_{j'}$ .

#### Pairwise Missing Pattern

The program provides three options for how the pairwise (joint) observant probabilities  $\phi_{jj'}$  are calculated. These are

**Independent (Ind):**  $\phi_{jj'} = \phi_j \phi_{j'}$ ,  $\phi_{jj} = \phi_j$

**Monotonic (Mon):**  $\phi_{jj'} = \phi_k$  where  $k = \max(j, j')$

**Mixture:**  $\phi_{jj'} = W(\text{Ind}) + (1 - W)(\text{Mon})$  for weighting factor  $W$ .

**GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)****Missing Input Type**

There are several ways in which the missing value pattern can be specified. Each missing value pattern is a list of missing proportions at each of the  $M$  time points. Each value in the list must be non-negative and less than 1. 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.

- **Piecewise Constant on Spreadsheet**

A set of missing proportions are defined for several time intervals using the spreadsheet. One column contains the missing proportions for the interval, going down the rows. Another column defines the corresponding upper limit of time proportion of the interval. The lower limit is implied by the limit given immediately above. The program assumes that the first time interval starts at 0 percent.

- **Linear (Steady Change)**

The missing proportions fall along a straight-line between 0 and 1 elapsed time. Only the first and last proportions are entered.

- **Piecewise Linear on Spreadsheet**

The missing proportions fall along a set of connected straight-lines that are defined by two columns on the spreadsheet.

- **List**

Enter a list of  $M$  missing proportions, one for each time point.

- **Multiple Lists on Spreadsheet**

Select multiple columns containing vertical lists of missing proportions. Each column contains a set of missing proportions in rows, one for each time point.

- **Pairwise Observed Proportions on Spreadsheet**

Enter an  $M \times M$  matrix of observed probabilities by selecting  $M$  columns. These observed probabilities are the proportion of the responses for both the row and column time points that are observed.

## Sample Size Calculations

The details of the calculation of sample size and power is given in Ahn, Heo, and Zhang (2015), Chapter 4 and Jung and Ahn (2004). These are summarized here.

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

where

$$\boldsymbol{\eta}' = (\beta_1 - \bar{\beta}, \dots, \beta_{G-1} - \bar{\beta})$$

$$\bar{\beta} = \sum_{k=1}^G r_k \beta_k$$

$$\boldsymbol{\eta}'\mathbf{V}^{-1}\boldsymbol{\eta} = \frac{\bar{m}^2 \sigma_t^4}{\sigma^2 S_t^2} \left[ \sum_{k=1}^{G-1} r_k \eta_k^2 + \frac{1}{r_G} \left( \sum_{k=1}^{G-1} r_k \eta_k \right)^2 \right]$$

$$\bar{m} = \sum_{j=1}^m \phi_j$$

$$\sigma_t^2 = \frac{1}{\bar{m}} \sum_{j=1}^m \phi_j (t_j - \bar{t})^2$$

$$\bar{t} = \frac{1}{\bar{m}} \sum_{j=1}^m \phi_j t_j$$

$$s_t^2 = \sum_{j=1}^m \sum_{j'=1}^m \phi_{jj'} \rho_{jj'} (t_j - \bar{t})(t_{j'} - \bar{t})$$

$\phi_j = 1 - \kappa_j$ , where  $\kappa_j$  = proportion missing at the  $j^{\text{th}}$  time point

$\rho_{jj'}$  is the corresponding element from within-subject correlation matrix

$\phi_{jj'}$  is the joint observant probability of observing both  $y_{ij}$  and  $y_{ij'}$  for every subject  $i$

Three possible choices are available to calculate  $\phi_{jj'}$ . These are

Independent:  $\phi_{jj'} = \phi_j \phi_{j'}$ ,  $\phi_{jj} = \phi_j$

Monotonic:  $\phi_{jj'} = \phi_k$  where  $k = \max(j, j')$

Mixture:  $\phi_{jj'} = w(\text{Independent}) + (1 - w)(\text{Monotonic})$  for weighting factor  $w$ .

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 *Sample Size*, 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.

---

#### 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 slopes when in fact the slopes 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 and Group Allocation

##### G (Number of Groups)

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

##### Range

$G \geq 2$ .



**GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)****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 ( $n_1 = n_2 = \dots = n_G = n$ )**  
The sample size of every group is  $n$ . All group sample sizes are equal. Enter one or more values for  $n$ .
- **Enter group multipliers and  $n$**   
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,  $n_1 = r_1 \times n$ .
- **Enter  $n_1, n_2, \dots, n_G$**   
Enter a list of group sample sizes, one for each group.
- **Enter columns of  $n_i$ 's**  
Select one or more columns of the spreadsheet that each contain a set of group sample sizes going down the column. Each column is analyzed separately.

 **$n$  (Sample Size Per Group)**

Enter  $n$ , the number of subjects in each group. The total number of subjects,  $N$ , is equal to  $n \times$  (number of groups).

You can specify a single value or a list.

**Single Value**

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

**List of Values**

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

**Range of  $n$** 

$n > 1$

**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  $n_i = \text{ceiling}[r_i \times n]$ , where  $\text{ceiling}[y]$  is the first integer greater than or equal to  $y$ . For example, the multipliers  $\{1, 1, 2, 2.95\}$  and  $base\ n$  of 10 would result in the sample sizes  $\{10, 10, 20, 30\}$ .

**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 sample sizes must be at least 1.

**GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)****n (Base Sample Size Per Group)**

Enter  $n$ , the base number of subjects in each group. The number of subjects 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 sample size.

**List of Values**

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

**Range**

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

**n1, n2, ..., nG (List)**

Enter a list of  $G$  sample sizes, one for each group.

**Incomplete List**

If the number of items in the list is less than  $G$ , the missing sample sizes 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 ni's**

Enter one or more spreadsheet columns containing vertical lists of group sample sizes.

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 sample sizes 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 sample size 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 the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)****Note**

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

**Group Allocation Input Type (when Solve For = Sample Size)**

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

## Options

- **Equal ( $n_1 = n_2 = \dots = n_G = n$ )**  
All group sample sizes are equal to  $n$ . The value of  $n$  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 sample sizes are found by multiplying these proportions by  $N$  (the total sample size) 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 sample size by dividing each item by the sum of all items. The individual group sample sizes are found by multiplying these proportions by  $N$  (the total sample size) and rounding up.

For example, the pattern  $\{1, 3, 4\}$  will be rescaled to  $\{0.125, 0.375, 0.5\}$ . The group sample sizes will be constrained to these proportions (within rounding) during the search for the sample configuration that just 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 sample sizes 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 the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)

### 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 clicking on them in the spreadsheet

---

## Measurement Times Schedule

### Measurement Time Input Type

Specify how you want to enter the information about the measurement proportion times.

- **Equally Spaced Measurement Times**

Enter one or more values for  $M$ , the number of measurements per subject. Equally-spaced measurement time proportions will automatically be generated.

- **List of Measurement Time Proportions**

Enter a list of measurement time proportions in the box below, one for each time point. The number of measurement times,  $M$ , is set equal to the number of items in this list.

- **Columns of Measurement Time Proportions**

Select one or more columns of the spreadsheet that contain a set of measurement time proportions. Each column is analyzed separately. The number of measurement times,  $M$ , is equal to the number of rows in the active column.

### **$M$ (Measurements)**

Specify one or more values for  $M$ , the total number of measurements scheduled for each subject. It is assumed that these measurements are made at identical time points or under the same conditions.

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 50.

Some examples are

3

3 4 5

3 to 10 by 1

### Measurement Time Proportions

Enter a set of measurement time proportions, one for each time point. The number of measurements,  $M$ , is the number of items in the list. Each entry in the list is the proportion that that time is of the maximum time.

For example, suppose a 24-month study is planned with anticipated measurements at the beginning and then after each six-month period. Hence, the five measurement times are 0, 6, 12, 18, and 24 months. The entry must be proportions of 24, so the actual entry would be '0, 0.25, 0.50, 0.75, 1.0'.

## GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)

M, the number of repeated measurements, is set to the number of items in the list.

Once M is set here, it must match its implicit value in the Correlations and Missing Proportions.

### Auto Rescaling

You can enter any increasing series that is convenient and it will be rescaled so that each entry is a proportion of the maximum. In the example above, you could have entered '0, 6, 12, 18, 24' or even '1, 2, 3, 4, 5' and the program would have rescaled the values into proportions beginning at 0 and ending at 1.

### Valid Entries

Since the numbers are automatically rescaled, the only requirement is that the series be strictly increasing.

### Examples

0 0.25 0.50 0.75 1.0

0 10 50 90 100

0 0.20 0.40 0.90 0.95 1.0

1 2 3 4 5

0 10 20 40

### Column(s) of Time Proportions

Enter one or more spreadsheet columns containing lists of measurement time proportions.

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 measurement times are entered going down.

### Examples

C1	C2	C3
0.0	1	0.1
0.2	2	0.4
0.4	3	0.5
0.6	4	
0.8		
1.0		

### Definition of a Single Column

Each column gives one list of time proportions. Each column results in a new scenario. Different columns are not connected and may be of different lengths.

Each entry in the list is the proportion that that time is of the maximum time. For example, suppose a 24-month study is planned with anticipated measurements at the beginning and then after each six-month period. Hence, the five measurement times are 0, 6, 12, 18, and 24 months. The entry must be proportions of 24, so the actual entry would be '0, 0.25, 0.50, 0.75, 1.0' (entered vertically in a particular column).

### Auto Rescaling

You can enter any increasing series that is convenient and it will be rescaled so that each entry is a proportion of the maximum. In the example above, you could have entered '0 6 12 18 24', or even '1 2 3 4 5', going down the column and the program would have rescaled the values into proportions beginning at 0 and ending at 1.

### Valid Entries

Since the numbers are automatically rescaled, the only requirement is that the series be strictly increasing.

**GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)****Notes**

M, the number of repeated measurements, is set to the number of items in the active column.

Once M is set here, it must match its implicit value in the Correlations and Missing Proportions.

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

---

**Effect Size -  $\beta_1, \beta_2, \dots, \beta_G$  (Group Slopes) and  $\sigma$  (Standard Deviation)**
 **$\beta_i$ 's Input Type**

Specify how you want to enter  $\beta_1, \beta_2, \dots, \beta_G$ , the group slopes for each of the G groups.

Options

- **$\beta_1, \beta_2, \dots, \beta_G$  (Slopes, i.e., Means)**  
Specify the values of the group slopes. The differences among these values represent the effect size that you want to detect.
- **Columns Containing Sets of  $\beta_i$ 's**  
Select one or more columns of the spreadsheet that each contain a set of  $\beta_i$ 's going down the column. Each column is analyzed separately.

 **$\beta_1, \beta_2, \dots, \beta_G$  (Slopes, i.e., Means)**

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

The prediction equation for the mean of the  $t_j$  time point within the  $i$ th group is  $\mu_i(t_j) = \theta_i + \beta_i(t_j)$ .

Often, you assume that all  $\theta_i$  are equal. This assumption is reasonable because subjects are randomly assigned to groups, so the group means at time  $t_0 = 0$  are equal. Hence  $\mu_i(t_0) = \theta_i + \beta_i(t_0) = \theta_i$ .

Similarly, the scaled value of the last time point is one. Thus, the predicted mean at the last time is given by

$$\mu_i(1) = \theta_i + \beta_i(1) = \theta_i + \beta_i.$$

Since all  $\theta_i$  are equal, the difference between the  $i^{\text{th}}$  and  $j^{\text{th}}$  group means at the final measurement time is  $\beta_i - \beta_j$  (which does not involve the intercept).

Hence, the difference among the slopes represent the difference among the means at the final time point. That is, you can enter the group mean at the final time point as the value of the group slope.

**Range**

Each  $\beta_i$  can be any number and at least one slope 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.

**GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)****Columns Containing Sets of  $\beta_i$ 's**

Enter one or more spreadsheet columns containing vertical lists of  $\beta_1, \beta_2, \dots, \beta_G$ . (See the section above for the interpretation of these slopes).

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  $\mu_i$ '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.

 **$\sigma$  (Standard Deviation)**

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

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'.

## Effect Size - $\rho$ (Correlation Between Measurements)

### Pattern of $\rho$ 's Across Time (or Space)

Specify the type of pattern for the  $\rho$ 's, the correlations between responses of an individual at two points in time or between subjects within a cluster. This pattern often depends on the measurement time proportions. Usually, the correlation shrinks as two points are further and further apart in time. The options are

- **Compound Symmetry (All  $\rho$ 's Equal)**

A single value of  $\rho$  is used for all correlations. This pattern is commonly used for cluster-randomized trials in which case  $m$  is the cluster size and  $\rho$  is the intraclass correlation.

$$\begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

Example first row of correlation matrix

1.0000, 0.5000, 0.5000, 0.5000, 0.5000, 0.5000

- **Banded (1 Time Period)**

This is similar to all  $\rho$ 's equal, except that all elements are set to zero after the first off-diagonal element.

$$\begin{bmatrix} 1 & \rho & 0 & 0 \\ \rho & 1 & \rho & 0 \\ 0 & \rho & 1 & \rho \\ 0 & 0 & \rho & 1 \end{bmatrix}$$

Example first row of correlation matrix

1.0000, 0.5000, 0.0000, 0.0000, 0.0000, 0.0000

- **Banded (2 Time Periods)**

This is similar to all  $\rho$ 's equal, except that all elements are set to zero after the second off-diagonal element.

$$\begin{bmatrix} 1 & \rho & \rho & 0 \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ 0 & \rho & \rho & 1 \end{bmatrix}$$

Example first row of correlation matrix

1.0000, 0.5000, 0.5000, 0.0000, 0.0000, 0.0000

- **AR1 (Traditional) 1st Order Autocorrelation**

A single value of  $\rho$  is used to generate a first order autocorrelation pattern in the correlation matrix. This pattern produces the autocorrelation at each successive step by multiplying the value at the last step by  $\rho$ . The basic formula is  $\rho_{jk} = \rho^{|j-k|}$ . Thus the pattern is  $\rho, \rho^2, \rho^3$ , etc.

$$\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

Example first row of correlation matrix

1.0000, 0.5000, 0.2500, 0.1250, 0.0625, 0.0313



**GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)**

- **AR1 (Proportional)**

A single value of  $\rho$  is used to generate a first order autocorrelation pattern in the correlation matrix. This pattern produces the autocorrelation at each successive step by multiplying the value at the last step by  $\rho$ . The basic formula is  $\rho_{jk} = \rho^{|t_j - t_k|}$ , where  $t_j$  and  $t_k$  are time proportions at which measurements are planned. An example is

$$\begin{array}{cccc} |\rho^{0/3} & \rho^{1/3} & \rho^{2/3} & \rho^{3/3}| \\ |\rho^{1/3} & \rho^{0/3} & \rho^{1/3} & \rho^{2/3}| \\ |\rho^{2/3} & \rho^{1/3} & \rho^{0/3} & \rho^{1/3}| \\ |\rho^{3/3} & \rho^{2/3} & \rho^{1/3} & \rho^{0/3}| \end{array}$$

Example first row of correlation matrix when  $M = 6$  and  $\rho = 0.1$ . Note that  $\rho$  shows up at the end of the list in this case.

1.0000, 0.6310, 0.3981, 0.2512, 0.1585, 0.1000

- **Dampened Exponential (Traditional)**

This pattern is similar to the AR(1), except that the exponents are raised to the power of a second parameter which we label as  $D_{exp}$ . The formula is  $\rho_{jk} = \rho^{|j-k|^{D_{exp}}}$ .

Example first row of correlation matrix with  $D_{exp} = 1.1$

1.0000, 0.5000, 0.2012, 0.0729, 0.0245, 0.0078

- **Dampened Exponential (Proportional)**

A single value of  $\rho$  is used to generate a first order autocorrelation pattern in the correlation matrix. This pattern produces the autocorrelation at each successive step by multiplying the value at the last step by  $\rho$ . The basic formula is  $\rho_{jk} = \rho^{|t_j - t_k|^{D_{exp}}}$ , where  $t_j$  and  $t_k$  are time proportions at which measurements are planned. An example with  $D_{exp} = 1$  is

$$\begin{array}{cccc} |\rho^{0/3} & \rho^{1/3} & \rho^{2/3} & \rho^{3/3}| \\ |\rho^{1/3} & \rho^{0/3} & \rho^{1/3} & \rho^{2/3}| \\ |\rho^{2/3} & \rho^{1/3} & \rho^{0/3} & \rho^{1/3}| \\ |\rho^{3/3} & \rho^{2/3} & \rho^{1/3} & \rho^{0/3}| \end{array}$$

Example first row of correlation matrix when  $M = 6$ ,  $\rho = 0.1$ , and  $D_{exp} = 1$ . Note that  $\rho$  shows up at the end of the list in this case.

1.0000, 0.6310, 0.3981, 0.2512, 0.1585, 0.1000

- **Linear Exponential Decay**

This pattern is similar to AR(1). The formula is  $\rho_{jk} = \rho^{|t_j - t_k|/P} * E_{max}$ .  $P$  is called the Base Time Proportion. The exponents of  $\rho$  range from 1 at  $P$  to  $E_{max}$  at  $M$ .

Example first row of correlation matrix with  $E_{max} = 4$

1.0000, 0.5000, 0.2973, 0.1768, 0.1051, 0.0625

This pattern is particularly useful when you want to compare different values of  $M$  since the exponent only depends on the difference. That is, a difference of 0.10 in the time proportions will always result in the same exponent, no matter what the value of  $M$ . This is NOT the case for AR(1).

- **Matrix on Spreadsheet**

A symmetric matrix is entered in the spreadsheet. The number of columns selected must match  $M$ , the number of time periods.

**GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)** **$\rho$  Pattern = Compound Symmetry, Banded (1 Time Period), Banded (2 Time Periods), AR(1)** **$\rho$  (Base Correlation)**

This is the base correlation,  $\rho$ , between measurements on the same subject taken at two time points. In several cases, this value is modified according to a special formula.

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.

**Recommended**

The value of this parameter depends on time or location pattern at which measurements are taken. In their book on sample size, Machin and Campbell comment that values between 0.60 and 0.75 are typical. Some authors recommend using 0.2 when nothing is known about the actual value.

**Examples**

0.5

0.5 0.6 0.7

0 to 0.9 by 0.1

 **$\rho$  Pattern = Damped Exponential** **$\rho$  (Base Correlation)**

This is the base correlation,  $\rho$ , between measurements on the same subject taken at two time points. In several cases, this value is modified according to a special formula.

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.

**Recommended**

The value of this parameter depends on time or location pattern at which measurements are taken. In their book on sample size, Machin and Campbell comment that values between 0.60 and 0.75 are typical. Some authors recommend using 0.2 when nothing is known about the actual value.

**Examples**

0.5

0.5 0.6 0.7

0 to 0.9 by 0.1

**Dexp (Dampening Exponent)**

This exponent is applied to the AR(1) formula so that the resulting correlations are reduced faster. The formula become  $\rho_{jk} = \rho^{|j-k|^{Dexp}}$ .

Range

The valid range is  $Dexp > 0$ . The typical range is 1 to 3.

**GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)** **$\rho$  Pattern = Linear Exponential Decay** **$\rho$  (Base Correlation)**

This is the base correlation,  $\rho$ , between measurements on the same subject taken at two time points. In several cases, this value is modified according to a special formula.

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.

**Recommended**

The value of this parameter depends on time or location pattern at which measurements are taken. In their book on sample size, Machin and Campbell comment that values between 0.60 and 0.75 are typical. Some authors recommend using 0.2 when nothing is known about the actual value.

**Examples**

0.5

0.5 0.6 0.7

0 to 0.9 by 0.1

**Base Time Proportion**

The base time proportion is used by the Linear Exponential Decay formula to specify the time proportion at which the correlation exponent is one. Thus, at this proportion the base correlation is reproduced.

**Range**

$0 < \text{Time} < 0.5$ .

**Typical Value**

$1/M$

**Note**

When you are comparing values of  $M$ , choose a value for  $M$  near the smallest value of  $1/M$ .

**Emax (Max Decay Exponent)**

This specifies the exponent of the base correlation when the time proportion is 1. Thus, the exponents go linearly from 1 at the base time to Emax at the maximum time.

**Range**

$\text{Emax} > 0$ .

**Typical**

Typical values of Emax are between 1 and 10.

**GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)** **$\rho$  Pattern = Matrix on Spreadsheet****Columns Containing the  $\rho$ 's**

Select a set of  $M$  columns from the spreadsheet that contain a correlation matrix. Each column represents a column of the correlation matrix so it must contain  $M$  rows.

Since this is a correlation matrix, the diagonal elements must be set to one and the absolute values of the off-diagonal elements must be less than one. Also, the matrix must be symmetric.

---

**Missing Proportions Across Measurement Times****Missing Input Type**

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

Each missing value pattern is a list of missing proportions at each of the  $M$  time points. Each value in the list must be non-negative and less than 1.

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.
- **Piecewise Constant on Spreadsheet**  
A set of missing proportions are defined for several time intervals using the spreadsheet. One column contains the missing proportions for the interval, going down the rows. Another column defines the corresponding upper limit of time proportion of the interval. The lower limit is implied by the limit given immediately above. The program assumes that the first time interval starts at 0 percent.
- **Linear (Steady Change)**  
The missing proportions fall along a straight-line between 0 and 1 elapsed time. Only the first and last proportions are entered.
- **Piecewise Linear on Spreadsheet**  
The missing proportions fall along a set of connected straight-lines that are defined by two columns on the spreadsheet.
- **List**  
Enter a list of  $M$  missing proportions, one for each time point.
- **Multiple Lists on Spreadsheet**  
Select multiple columns containing vertical lists of missing proportions. Each column contains a set of missing proportions in rows, one for each time point.
- **Pairwise Observed Proportions on Spreadsheet**  
Enter an  $M \times M$  matrix of observant probabilities by selecting  $M$  columns. These observant probabilities are the proportion of the responses for both the row and column time points that are observed

**GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)****Missing Input Type = Constant Additional Options****Constant Missing Proportion**

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

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

**Missing Input Type = Piecewise Constant on Spreadsheet Additional Options****Pairwise Missing Pattern**

See the Pairwise Missing Pattern section of above.

**Column of Missing Proportions (Pairwise)**

Enter the column name on the spreadsheet containing a column of missing proportions, one for each segment. The program assigns this missing proportion to all time points in this time segment. The number of segments is arbitrary and determined by the number of rows.

Note that the usual pattern is to have the missing proportions increase over time.

Press the Spreadsheet icon (directly to the right) to select the column and then enter the missing proportion values.

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

**(Pairwise)**

This refers to the two types of pairwise missing patterns: Ind(ependent) and Mon(otonic).

**Example**

Suppose that 'C1' is selected for the Missing column and 'C2' is selected for the Time column. Suppose that these columns contain the following values:

<b>C1</b>	<b>C2</b>
0.1	0.2
0.3	0.5
0.35	0.75
0.4	0.9
0.6	1.0

This would be interpreted by the program as

<b>Time Interval</b>	<b>Missing</b>
$0 \leq t \leq 0.2$	0.1
$0.2 < t \leq 0.5$	0.3
$0.5 < t \leq 0.75$	0.35
$0.75 < t \leq 0.9$	0.4
$0.9 < t \leq 1.0$	0.6

**GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)****Column of Time Proportions (Pairwise)**

Enter the column name on the spreadsheet containing a column of Upper Limit Time Proportions, one for each segment. The program assigns the missing proportion to all time points in the time segment defined by this upper limit. The number of segments is arbitrary and determined by the number of rows. Note that the final upper limit must be 1.

Press the Spreadsheet icon (directly to the right) to select the column and then enter the missing proportion values.

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

**(Pairwise)**

This refers to the two types of pairwise missing patterns: Ind(ependent) and Mon(otonic).

**Example**

Suppose that 'C1' is selected for the Missing column and 'C2' is selected for the Time column. Suppose that these columns contain the following values:

C1	C2
0.1	0.2
0.3	0.5
0.35	0.75
0.4	0.9
0.6	1.0

This would be interpreted by the program as

Time Interval	Missing
$0 \leq t \leq 0.2$	0.1
$0.2 < t \leq 0.5$	0.3
$0.5 < t \leq 0.75$	0.35
$0.75 < t \leq 0.9$	0.4
$0.9 < t \leq 1.0$	0.6

**Missing Input Type = Linear (Steady Change) Additional Options****Pairwise Missing Pattern**

See the Pairwise Missing Pattern section of above.

**First (or Last) Missing Proportion (Pairwise)**

Specify a set of  $M$  missing proportions by entering only the first and last values of the list. A straight-line formula will be used to calculate the missing proportions for each time point in the interval.

**Range**

$$0 \leq \text{First} \leq \text{Last}$$

$$\text{First} \leq \text{Last} < 1$$

**Example**

Suppose a study is being planned with 5, evenly spaced, measurements. Further suppose that the missing data proportions are anticipated to be 0.1 initially and increase to 0.6 at the end of the study, with the missing proportions in between arranged linearly. If  $M = 5$ , First = 0.1, and Last = 0.6, then the resulting Missing Proportions will be 0.1, 0.225, 0.35, 0.475, and 0.60. Note that these proportions are evenly spaced between the two boundaries.

**GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)**

Note that using this option allows the value of  $M$  to be varied and reasonable comparisons in power and sample size to be made.

**Missing Input Type = Piecewise Linear on Spreadsheet Additional Options****Pairwise Missing Pairwise**

See the Pairwise Missing Pairwise section of above.

**Column of Missing (or Time) Proportions (Pairwise)**

These options let you easily define separate missing proportions for various time intervals of a study.

Select the column name on the spreadsheet of a column containing missing proportions. Also select the column name of the corresponding measurement time proportions. The program calculates a straight-line defined by each successive pair of points (see example below). This set of connected line segments becomes a prediction equation for the missing proportion based on the corresponding time proportion.

The number of line segments is arbitrary and determined from the number of rows in the selected columns. The number of rows does not have to match  $M$ , the number of time points.

**(Pairwise)**

This designates which of the two types of pairwise missing patterns, Ind(ependent) or Mon(otonic), is being entered in this option.

**Notes**

The usual pattern is to for the missing proportions to increase over time. The time value of the **first row of the time column must be 0**. The time value of the **last row of the time column must be 1**.

**Shortcuts**

Press the Spreadsheet icon (directly to the right) to select the column and then enter the missing proportion values. Press the Input Spreadsheet icon (to the right and slightly up) to view/edit the spreadsheet. Note that you can obtain also the spreadsheet window by selecting "Tools", then "Input Spreadsheet", from the menus.

**Example**

Suppose that you have enter data in two columns on the spreadsheet labeled *Miss* and *Time* as follows.

<b>Miss</b>	<b>Time</b>
0.5	0
0.1	0.2
0.3	0.5
0.35	0.75
0.4	0.9
0.6	1

Now suppose that *Miss* is selected for the Missing column and *Time* is selected for the Time column. This would be interpreted by the program as the following segmented-prediction equation.

<b>Time Interval</b>	<b>Slope</b>	<b>Intercept</b>
$0 \leq T \leq 0.2$	$0.25 = (0.1 - 0.05) / (0.2 - 0)$	$0.1 - 0.2 * 0.25 = 0.05$
$0.2 < T \leq 0.5$	$0.67 = (0.3 - 0.1) / (0.5 - 0.2)$	$0.3 - 0.5 * 0.67 = -0.0333$
$0.5 < T \leq 0.75$	$0.2 = (0.35 - 0.3) / (0.75 - 0.5)$	$0.35 - 0.75 * .20 = 0.2$
$0.75 < T \leq 0.9$	$0.33 = (0.4 - 0.35) / (0.9 - 0.75)$	$0.4 - 0.9 * .33 = 0.1$
$0.9 < T \leq 1$	$2 = (0.6 - 0.4) / (1 - 0.9)$	$0.6 - 1 * 2 = -1.4$

**GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)**

A Missing Proportion would then be calculated using the formula

$$\text{Missing} = \text{Intercept} + \text{Slope} * \text{Time}$$

For example, using the above values, the missing proportions at time points 0, 0.1, 0.3, 0.8, and 1.0 would be found as follows:

Time	Equation	Resulting Missing Proportion
0	$0.05 + 0.25(0)$	0.05
0.1	$0.05 + 0.25(0.1)$	0.075
0.3	$-0.0333 + 0.67(0.3)$	0.1667
0.8	$0.1 + 0.33(0.8)$	0.3667
1.0	$-1.4 + 2(1)$	0.6

**Missing Input Type = List Additional Options****Pairwise Missing Pairwise**

See the Pairwise Missing Pairwise section of above.

**List of Missing Proportions (Pairwise)**

These options let you easily define separate missing proportions for various time intervals of a study.

Select the column name on the spreadsheet of a column containing missing proportions. Also select the column name of the corresponding measurement time proportions. The program calculates a straight-line defined by each successive pair of points (see example below). This set of connected line segments becomes a prediction equation for the missing proportion based on the corresponding time proportion.

**Missing Input Type = Multiple Lists on Spreadsheet Additional Options****Pairwise Missing Pairwise**

See the Pairwise Missing Pairwise section of above.

**Columns of Missing Proportions (Pairwise)**

Enter lists of missing value proportions by first entering them on the spreadsheet and then selecting those columns here. The missing value proportions are entered down the rows of a column, one for each time point. Each value is the anticipated proportion of missing responses at the corresponding time point.

Although you can enter any pattern you want, the usual pattern is that the proportion missing increases over time.

When you specify more than  $M$  values in a list (where  $M$  is specified in the Measurement Times section), the extra values are ignored. When you specify too few values, the last value you specify is copied forward.

Press the Spreadsheet icon (directly to the right) to select the column and then enter the missing proportion values.

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

**(Pairwise)**

This designates which of the two types of pairwise missing patterns, Ind(ependent) or Mon(otonic), is being entered in this option.

**Range**

$0 \leq \text{missing proportion} < 1$ .



**Missing Input Type = Pairwise Observed Proportions on Spreadsheet Additional Options****Columns of Pairwise Observed**

Enter an M-by-M matrix of pairwise observed proportions on the spreadsheet and then select it by selecting those M columns here. Each entry gives the proportion (or probability) that the corresponding row and column time point values are both observed (non-missing). The diagonal elements of this matrix contain the probability that an observation is observed at that point (marginal probability).

**Range**

$0 < \text{each value} \leq 1$ .

**Note**

In other missing entries, you enter the proportion missing. This is the only case that requires the proportion observed.

## Example 1 – Determining Sample Size for a Three-Arm Trial

Researchers are planning a study comparing three heart-rate medications: a standard drug and two experimental drugs. The experimental drugs appear to have about the same impact on heart rate. Each subject will receive four applications of one drug, two days apart. The researchers want a sample size large enough to detect a slope difference of 5 between the highest and lowest slopes. They will use the slopes of 65, 60, and 60 to represent this difference.

Similar studies have had a standard deviation of the residuals within a subject across time of between 5 and 7. These studies also showed an autocorrelation between adjacent measurements on the same individual of 0.7, so they want to try values of 0.6, 0.7, and 0.8. The researchers assume that first-order autocorrelation adequately represents the autocorrelation pattern. The test will be conducted at the 0.05 significance level and powered at 90%. The subjects will be randomly assigned equally among the three groups.

The researchers anticipate that the missing pattern across time will begin at 0% missing and increase steadily to 40% at the fourth measurement. They assume that the pairwise missing is *independent*.

What are the sample size requirements 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 the **GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)** 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>
<b>Design Tab</b>	
Solve For .....	<b>Sample Size</b>
Power.....	<b>0.90</b>
Alpha.....	<b>0.05</b>
G (Number of Groups).....	<b>3</b>
Group Allocation Input Type .....	<b>Equal (n1 = n2 = ... = nG = n)</b>
Measurement Time Input Type.....	<b>Equally Spaced Measurement Times</b>
M (Number of Measurement Times).....	<b>4</b>
$\beta_i$ 's Input Type .....	<b><math>\beta_1, \beta_2, \dots, \beta_G</math> (Slopes, i.e., Means)</b>
$\beta_1, \beta_2, \dots, \beta_G$ (Slopes, i.e., Means).....	<b>65 60 60</b>
$\sigma$ (Standard Deviation) .....	<b>5 6 7</b>
Pattern of $\rho$ 's Across Time.....	<b>AR1 (Traditional)</b>
$\rho$ (Base Correlation) .....	<b>0.6 0.7 0.8</b>
Missing Input Type.....	<b>Linear (Steady Change)</b>
Pairwise Missing Pattern .....	<b>Independent (Ind)</b>
First Missing Proportion (Ind) .....	<b>0</b>
Last Missing Proportion (Ind).....	<b>0.4</b>

## Annotated Output

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

### Numeric Results for a Multi-Group Slope Comparison Test using GEE

Measurement Times: Equally spaced  
 Correlation: AR1:  $\rho(j,k) = \rho^{|j-k|}$   
 Missing Pattern: Range of missing proportions  
 Observant Proportions: Assume independence

Power	N	Group Sample Sizes ni's	Group Slopes GJM $\beta_i$ 's	Std Dev $\sigma$	Base Corr $\rho$	First Row of Corr Matrix	Alpha	Missing	Time	
0.9072	123	ni1	3 4	$\beta_{i1}$	5.00	0.6000	$\rho_1(T_1)$	0.050	M1(T1)	T1
0.9078	108	ni2	3 4	$\beta_{i1}$	5.00	0.7000	$\rho_2(T_1)$	0.050	M1(T1)	T1
0.9062	87	ni3	3 4	$\beta_{i1}$	5.00	0.8000	$\rho_3(T_1)$	0.050	M1(T1)	T1
0.9019	174	ni4	3 4	$\beta_{i1}$	6.00	0.6000	$\rho_1(T_1)$	0.050	M1(T1)	T1
0.9030	153	ni5	3 4	$\beta_{i1}$	6.00	0.7000	$\rho_2(T_1)$	0.050	M1(T1)	T1
0.9007	123	ni6	3 4	$\beta_{i1}$	6.00	0.8000	$\rho_3(T_1)$	0.050	M1(T1)	T1
0.9021	237	ni7	3 4	$\beta_{i1}$	7.00	0.6000	$\rho_1(T_1)$	0.050	M1(T1)	T1
0.9012	207	ni8	3 4	$\beta_{i1}$	7.00	0.7000	$\rho_2(T_1)$	0.050	M1(T1)	T1
0.9017	168	ni9	3 4	$\beta_{i1}$	7.00	0.8000	$\rho_3(T_1)$	0.050	M1(T1)	T1

### Footnotes

T1: 0.00, 0.33, 0.67, 1.00  
 M1(T1): 0.00, 0.13, 0.27, 0.40  
 ni1: 41, 41, 41  
 ni2: 36, 36, 36  
 ni3: 29, 29, 29  
 ni4: 58, 58, 58  
 ni5: 51, 51, 51  
 ni6: 41, 41, 41  
 ni7: 79, 79, 79  
 ni8: 69, 69, 69  
 ni9: 56, 56, 56  
 $\beta_{i1}$ : 65.00, 60.00, 60.00  
 $\rho_1(T_1)$ : 1.0000, 0.6000, 0.3600, 0.2160  
 $\rho_2(T_1)$ : 1.0000, 0.7000, 0.4900, 0.3430  
 $\rho_3(T_1)$ : 1.0000, 0.8000, 0.6400, 0.5120

### References

Ahn, C., Heo, M., and Zhang, S. 2015. Sample Size Calculations for Clustered and Longitudinal Outcomes in Clinical Research. CRC Press, New York.  
 Jung, S.H. and Ahn, C. 2004. K-Sample Test and Sample Size Calculation for Comparing Slopes in Data with Repeated Measurements. Biometrical Journal. Vol 46(5). Pages 554-564.

### Report Definitions

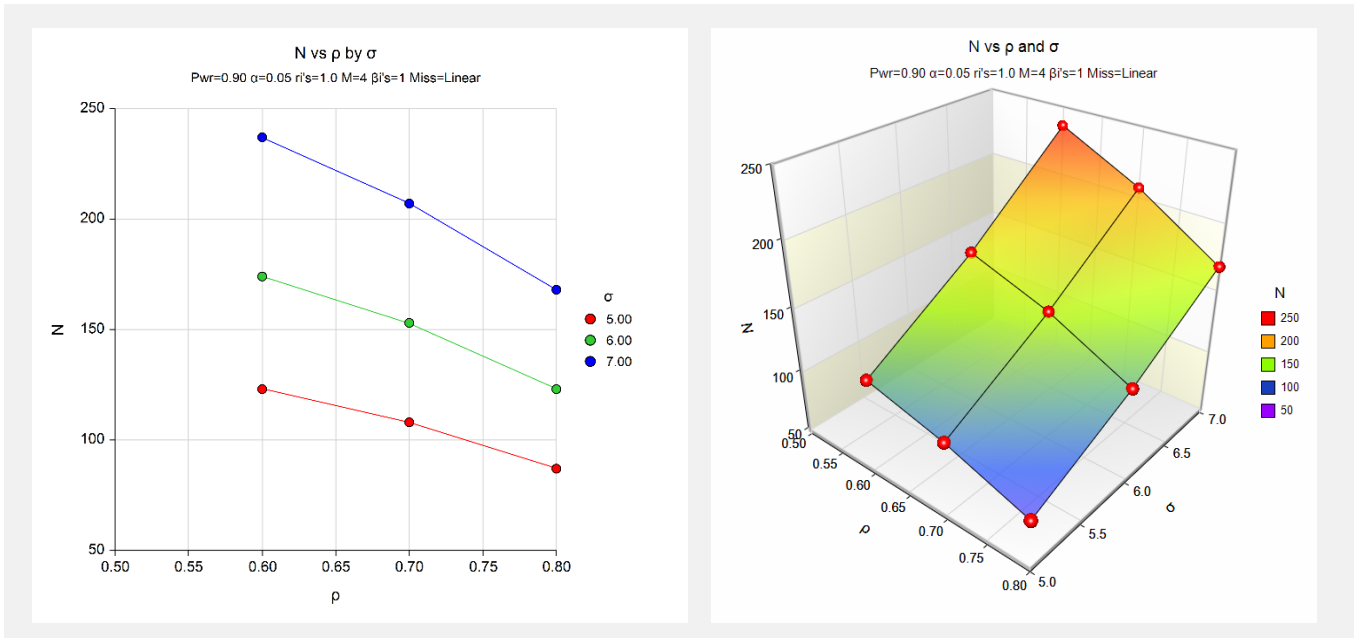
Power is the probability of rejecting a false null hypothesis.  
 N is the total number of subjects in the study.  
 ni's gives the name of the footnote containing the individual group sample sizes.  
 G is the number of treatment groups.  
 M is the number of time points at which each subject is scheduled to be measured.  
 $\beta_i$ 's gives the name of the footnote containing the slopes for each group.  
 $\sigma$  is the standard deviation of a response.  
 $\rho$  is the base correlation between two responses on the same subject. It may be transformed based on the correlation pattern.  
 Alpha is the significance level of the test. The probability of rejecting the null hypothesis when the null hypothesis is true.  
 Missing gives the name of the footnote containing the missing value proportions across time.  
 Time gives the name of the footnote containing the time proportions. The time values represent the proportion of the total study time that has elapsed just before the measurement.

GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)

Summary Statements

A total sample of 123 subjects, divided among 3 groups and each measured 4 times, achieve a power of 0.9072 when using a Chi-squared test with 2 degrees of freedom from a GEE analysis to determine whether the group slopes differ significantly at a significance level of 0.050. The group sample sizes are 41, 41, 41. The group slopes under the alternative hypothesis are 65.00, 60.00, 60.00. The residual standard deviation is anticipated to be 5.00. The measurements of each subject will be made at the following times expressed as proportions of the total study time: 0.00, 0.33, 0.67, 1.00. Missing values are assumed to occur completely at random (MCAR). These missing value proportions will be combined to form the pairwise observant probabilities using the Independent method. The anticipated proportions missing at each measurement time are 0.00, 0.13, 0.27, 0.40. The first row of the autocorrelation matrix of the responses within a subject is assumed to be 1.0000, 0.6000, 0.3600, 0.2160. Other rows follow the same pattern.

This report gives the sample size for each value of the other parameters.



These charts show the relationship among the design parameters.

Autocorrelation Matrix for Report Row 1

Time	T(0.00)	T(0.33)	T(0.67)	T(1.00)
T(0.00)	1.0000	0.6000	0.3600	0.2160
T(0.33)	0.6000	1.0000	0.6000	0.3600
T(0.67)	0.3600	0.6000	1.0000	0.6000
T(1.00)	0.2160	0.3600	0.6000	1.0000

These reports show the autocorrelation matrix for the indicated row of the report.

## Example 2 – Finding the Power

Continuing with Example 1, the researchers want to determine the power corresponding to group sample sizes ranging from 20 to 80 for the middle values of the other parameters.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)** 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>
<b>Design Tab</b>	
Solve For .....	<b>Power</b>
Alpha.....	<b>0.05</b>
G (Number of Groups).....	<b>3</b>
Group Allocation Input Type .....	<b>Equal (n1 = n2 = ... = nG = n)</b>
n (Sample Size Per Group) .....	<b>20 to 80 by 10</b>
Measurement Time Input Type.....	<b>Equally Spaced Measurement Times</b>
M (Number of Measurement Times).....	<b>4</b>
$\beta$ 's Input Type .....	<b><math>\beta_1, \beta_2, \dots, \beta_G</math> (Slopes, i.e., Means)</b>
$\beta_1, \beta_2, \dots, \beta_G$ (Slopes, i.e., Means).....	<b>65 60 60</b>
$\sigma$ (Standard Deviation) .....	<b>6</b>
Pattern of $\rho$ 's Across Time.....	<b>AR1 (Traditional)</b>
$\rho$ (Base Correlation) .....	<b>0.7</b>
Missing Input Type.....	<b>Linear (Steady Change)</b>
Pairwise Missing Pattern .....	<b>Independent (Ind)</b>
First Missing Proportion (Ind) .....	<b>0</b>
Last Missing Proportion (Ind).....	<b>0.4</b>

## Output

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

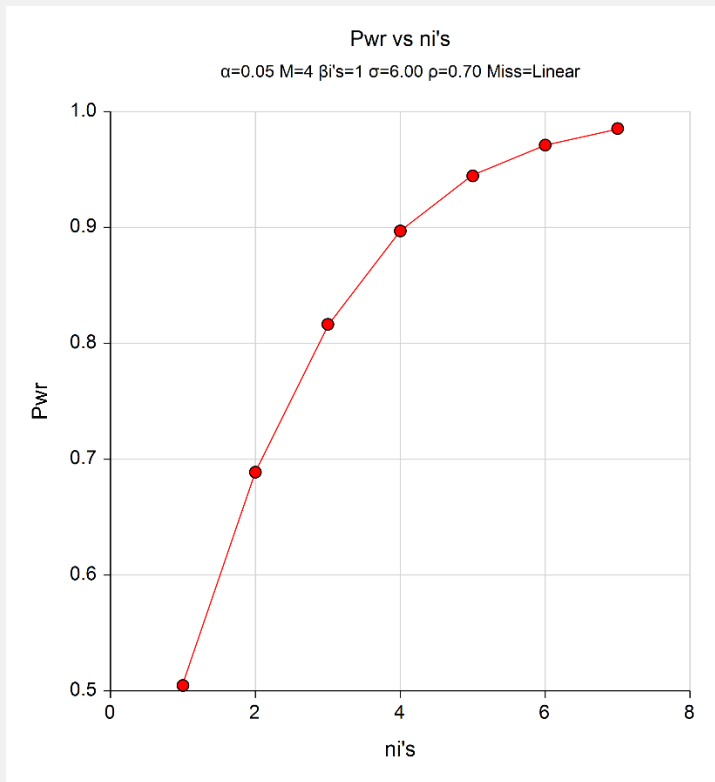
### Numeric Results for a Multi-Group Slope Comparison Test using GEE

Measurement Times: Equally spaced  
 Correlation: AR1:  $\rho(j,k) = \rho^{|j-k|}$   
 Missing Pattern: Range of missing proportions  
 Observant Proportions: Assume independence

Power	N	Group Sample Sizes ni's	Group Slopes GJM $\beta_i$ 's	Std Dev $\sigma$	Base Corr $\rho$	First Row of Corr Matrix	Alpha	Missing	Time
0.5047	60	ni1	3 4 $\beta_1$	6.00	0.7000	$\rho_1(T_1)$	0.050	M1(T1)	T1
0.6888	90	ni2	3 4 $\beta_1$	6.00	0.7000	$\rho_1(T_1)$	0.050	M1(T1)	T1
0.8164	120	ni3	3 4 $\beta_1$	6.00	0.7000	$\rho_1(T_1)$	0.050	M1(T1)	T1
0.8970	150	ni4	3 4 $\beta_1$	6.00	0.7000	$\rho_1(T_1)$	0.050	M1(T1)	T1
0.9445	180	ni5	3 4 $\beta_1$	6.00	0.7000	$\rho_1(T_1)$	0.050	M1(T1)	T1
0.9711	210	ni6	3 4 $\beta_1$	6.00	0.7000	$\rho_1(T_1)$	0.050	M1(T1)	T1
0.9854	240	ni7	3 4 $\beta_1$	6.00	0.7000	$\rho_1(T_1)$	0.050	M1(T1)	T1

### Footnotes

T1: 0.00, 0.33, 0.67, 1.00  
 M1(T1): 0.00, 0.13, 0.27, 0.40  
 ni1: 20, 20, 20  
 ni2: 30, 30, 30  
 ni3: 40, 40, 40  
 ni4: 50, 50, 50  
 ni5: 60, 60, 60  
 ni6: 70, 70, 70  
 ni7: 80, 80, 80  
 $\beta_1$ : 65.00, 60.00, 60.00  
 $\rho_1(T_1)$ : 1.0000, 0.7000, 0.4900, 0.3430



Note that the horizontal axis is scaled according to the  $ni$  footnote index. Hence, 1 is for  $ni1$ , 2 is for  $ni2$ , and so on. The same is true for the  $\beta_i$ 's and the *Linear Miss* in the subtitle.

## Example 3 – Comparing Various Effect Sizes

Continuing with Examples 1 and 2, the researchers want to compare the impact of various sets of slopes on sample size. To do this, they will compare the four sets of slopes shown in the following table. The first row is the standard medication. The second and third rows give the anticipated response to the experimental medications.

The values in this table must be loaded into the spreadsheet.

**Table of Four Sets of Time-Averaged Responses**

C1	C2	C3	C4
65	65	65	65
60	61	62	63
60	61	62	63

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)** 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>
<b>Design Tab</b>	
Solve For .....	<b>Sample Size</b>
Power.....	<b>0.90</b>
Alpha.....	<b>0.05</b>
G (Number of Groups).....	<b>3</b>
Group Allocation Input Type .....	<b>Equal (n1 = n2 = ... = nG = n)</b>
Measurement Time Input Type.....	<b>Equally Spaced Measurement Times</b>
M (Number of Measurement Times).....	<b>4</b>
$\beta$ 's Input Type .....	<b>Columns containing sets of <math>\beta</math>'s</b>
Columns Containing Sets of $\beta$ 's.....	<b>C1-C4</b>
<i>The above table must be entered into the spreadsheet and the column names changed to C1 – C4.</i>	
<i>Press the Input Spreadsheet icon to the right to view the spreadsheet.</i>	
$\sigma$ (Standard Deviation) .....	<b>6</b>
Pattern of $\rho$ 's Across Time.....	<b>AR1 (Traditional)</b>
$\rho$ (Base Correlation) .....	<b>0.7</b>
Missing Input Type.....	<b>Linear (Steady Change)</b>
Pairwise Missing Pattern .....	<b>Independent (Ind)</b>
First Missing Proportion (Ind) .....	<b>0</b>
Last Missing Proportion (Ind).....	<b>0.4</b>

## Output

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

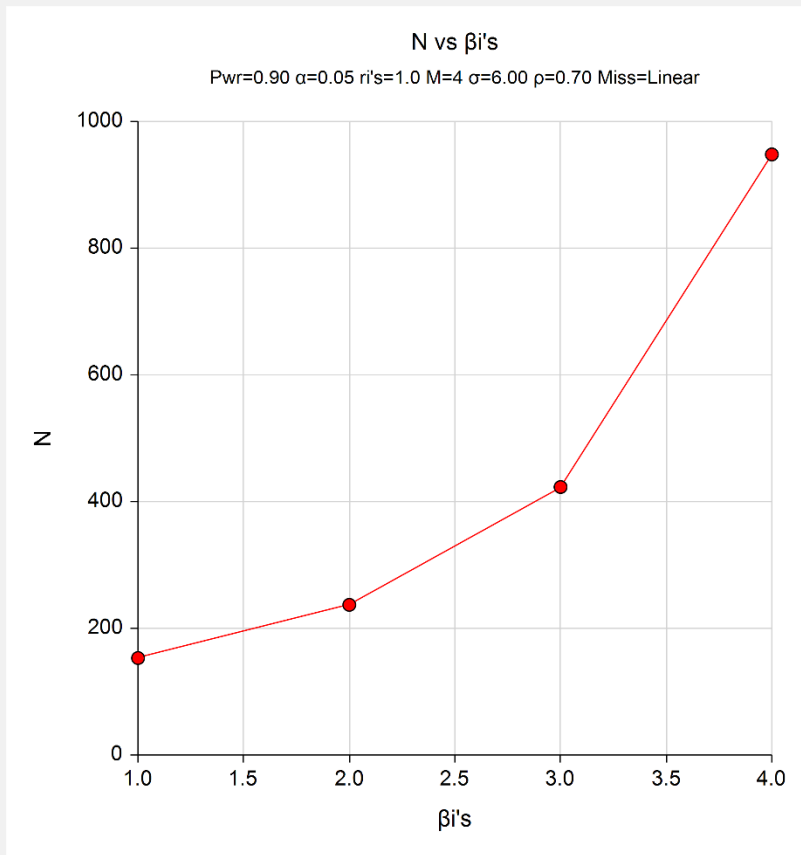
### Numeric Results for a Multi-Group Slope Comparison Test using GEE

Measurement Times: Equally spaced  
 Correlation: AR(1):  $\rho(j,k) = \rho^{|j-k|}$   
 Missing Pattern: Range of missing proportions  
 Observant Proportions: Assume independence

Power	N	Group Sample Sizes ni's	Group Slopes GJM $\beta$ 's	Std Dev $\sigma$	Base Corr $\rho$	First Row of Corr Matrix	Alpha	Missing	Time	
0.9030	153	ni1	3 4	C1	6.00	0.7000	$\rho_1(T_1)$	0.050	M1(T1)	T1
0.9004	237	ni2	3 4	C2	6.00	0.7000	$\rho_1(T_1)$	0.050	M1(T1)	T1
0.9016	423	ni3	3 4	C3	6.00	0.7000	$\rho_1(T_1)$	0.050	M1(T1)	T1
0.9004	948	ni4	3 4	C4	6.00	0.7000	$\rho_1(T_1)$	0.050	M1(T1)	T1

### Footnotes

T1: 0.00, 0.33, 0.67, 1.00  
 M1(T1): 0.00, 0.13, 0.27, 0.40  
 ni1: 51, 51, 51  
 ni2: 79, 79, 79  
 ni3: 141, 141, 141  
 ni4: 316, 316, 316  
 C1: 65.00, 60.00, 60.00  
 C2: 65.00, 61.00, 61.00  
 C3: 65.00, 62.00, 62.00  
 C4: 65.00, 63.00, 63.00  
 $\rho_1(T_1)$ : 1.0000, 0.7000, 0.4900, 0.3430



Note that the horizontal axis gives the index of the slopes footnotes. Thus, '1.0' is column C1, '2.0' is column C2, and so on. Obviously, as the slopes get closer together, the required sample size gets larger.



## Example 4 – Validation of Sample Size Calculation using Ahn, Heo, and Zhang (2015)

Ahn, Heo, and Zhang (2015) page 110 present sample size results for several scenarios. We will validate this procedure using a few results from the table found there. In this example, the standard deviation is 28.56, the slope difference is 28.6, the significance level is 0.05, the power is 0.90, group sizes are equal, and the number of repeated measurements per subject is 6. We will use the compound symmetry option for the correlation matrices with  $\rho = 0.1, 0.25, 0.4$ . The proportion missing at each time point is equal to 0, 0.1, 0.22, 0.33, 0.46, 0.59. The pairwise observant probabilities assume independence. Using these settings, the sample sizes were reported as 86, 76, and 67

### 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 **GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)** procedure window. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Sample Size</b>
Power.....	<b>0.90</b>
Alpha.....	<b>0.05</b>
G (Number of Groups).....	<b>2</b>
Group Allocation Input Type .....	<b>Equal (n1 = n2 = ... = nG = n)</b>
Measurement Time Input Type.....	<b>Equally Spaced Measurement Times</b>
M (Number of Measurement Times).....	<b>6</b>
$\beta$ 's Input Type .....	<b><math>\beta_1, \beta_2, \dots, \beta_G</math> (Slopes, i.e., Means)</b>
$\beta_1, \beta_2, \dots, \beta_G$ (Slopes, i.e., Means).....	<b>0 28.6</b>
$\sigma$ (Standard Deviation) .....	<b>28.56</b>
Pattern of $\rho$ 's Across Time.....	<b>Compound Symmetry (All <math>\rho</math>'s Equal)</b>
$\rho$ (Base Correlation) .....	<b>0.1 0.25 0.4</b>
Missing Input Type.....	<b>List</b>
Pairwise Missing Pattern .....	<b>Independent (Ind)</b>
List of Missing Proportions (Ind) .....	<b>0 0.1 0.22 0.33 0.46 0.59</b>

## GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)

## Output

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

**Numeric Results for a Multi-Group Slope Comparison Test using GEE**

Measurement Times: Equally spaced  
 Correlation: Compound symmetry (all  $\rho$ 's equal)  
 Missing Pattern: List of missing proportions  
 Observant Proportions: Assume independence

Power	Group Sample Sizes		GJM	Group Slopes	Std Dev $\sigma$	Base Corr $\rho$	First Row of Corr	Alpha	Missing	Time
	N	ni's		$\beta$ 's			Matrix			
0.9022	86	ni1	2 6	$\beta$ i1	28.56	0.1000	$\rho$ 1(T1)	0.050	M1	T1
0.9011	76	ni2	2 6	$\beta$ i1	28.56	0.2500	$\rho$ 2(T1)	0.050	M1	T1
0.9079	68	ni3	2 6	$\beta$ i1	28.56	0.4000	$\rho$ 3(T1)	0.050	M1	T1

**Footnotes**

T1: 0.00, 0.20, 0.40, 0.60, 0.80, 1.00

M1: 0.00, 0.10, 0.22, 0.33, 0.46, 0.59

ni1: 43, 43

ni2: 38, 38

ni3: 34, 34

$\beta$ i1: 0.00, 28.60

$\rho$ 1(T1): 1.0000, 0.1000, 0.1000, 0.1000, 0.1000, 0.1000

$\rho$ 2(T1): 1.0000, 0.2500, 0.2500, 0.2500, 0.2500, 0.2500

$\rho$ 3(T1): 1.0000, 0.4000, 0.4000, 0.4000, 0.4000, 0.4000

The sample sizes of 86 and 76 match Ahn, Heo, and Zhang (2015). The sample size of the third scenario was calculated as 68, which is one more than the reported 67. This difference occurred because **PASS** searches among only balanced designs. These require an even sample size, so **PASS** selected 68, the next balanced design above the reported 67.

## Example 5 – Impact of Measurement Time Distribution

This example will investigate the impact of measurement time distribution on power. It will compare the power of studies in which the measurements are evenly spaced with those that take more measurements at the beginning of the study, near the middle of the study, or at the end of the study.

This example uses  $G = 4$ ,  $m = 6$ ,  $\sigma = 14.3$ ,  $\alpha = 0.05$ , and power = 0.9. The correlation pattern is Linear Exponential Decay with a base correlation of 0.8, Base Time Proportion of 0.20, and Emax set to 4. The missing input type is set to Linear from 0 to 30% and the pairwise missing assumption is independent. Group slopes are 5, 5, 7, 10. The per group sample size is 200.

The measurement times for five scenarios are given in the following table.

**Table of Measurement Times in Proportion of Total Study Time**

Tm1	Tm2	Tm3	Tm4	Tm5
0	0	0	0	0
0.20	0.60	0.10	0.10	0.45
0.40	0.70	0.20	0.20	0.50
0.60	0.80	0.30	0.80	0.55
0.80	0.90	0.40	0.90	0.60
1.00	1.00	1.00	1.00	1.00

Note that the measurements in Tm1 are evenly spaced, those in Tm2 are loaded near the end, those of Tm3 occur at the beginning, those of Tm4 occur only at the beginning and the end, and those of Tm5 occur mostly near the middle.

## Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)** procedure window. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Power</b>
Alpha.....	<b>0.05</b>
G (Number of Groups).....	<b>4</b>
Group Allocation Input Type .....	<b>Equal (n1 = n2 = ... = nG = n)</b>
n (Sample Size Per Group) .....	<b>200</b>
Measurement Time Input Type.....	<b>Columns of Measurement Time Proportions</b>
Columns of Time Proportions .....	<b>TM1-TM5</b>
<i>The above table must be entered into the spreadsheet and the column names changed to Tm1 – Tm5.</i>	
<i>Press the Input Spreadsheet icon to the right to view the spreadsheet.</i>	
β's Input Type .....	<b>β1, β2, ..., βG (Slopes, i.e., Means)</b>
β1, β2, ..., βG (Slopes, i.e., Means).....	<b>5 5 7 10</b>
σ (Standard Deviation) .....	<b>14.3</b>
Pattern of ρ's Across Time.....	<b>Linear Exponential Decay</b>
ρ (Base Correlation) .....	<b>0.8</b>
Base Time Proportion .....	<b>0.2</b>
Emax (Max Decay Exponent).....	<b>4</b>

**GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)**

Missing Input Type..... **Linear (Steady Change)**

Pairwise Missing Pattern ..... **Independent (Ind)**

First Missing Proportion (Ind) ..... **0**

Last Missing Proportion (Ind)..... **0.30**

## Output

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

**Numeric Results for a Multi-Group Slope Comparison Test using GEE**

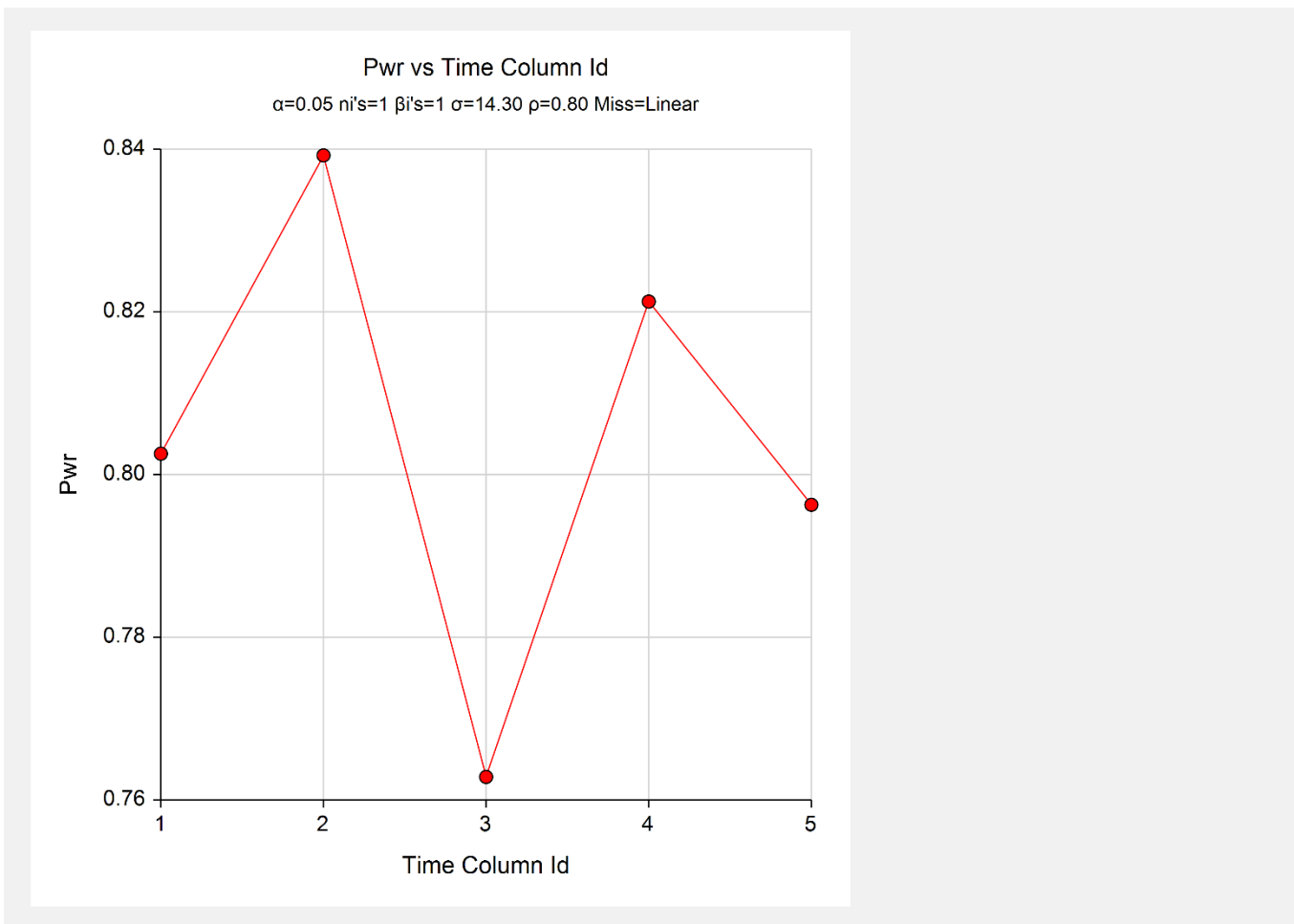
Measurement Times: Lists in spreadsheet columns: {TM1-TM5}  
 Correlation: Linear exponential decay, with Emax = 4 and Base Time Prop = 0.2  
 Missing Pattern: Range of missing proportions  
 Observant Proportions: Assume independence

Power	N	Group Sample Sizes ni's	Group Slopes GJM $\beta_i$ 's	Std Dev $\sigma$	Base Corr $\rho$	First Row of Corr Matrix	Alpha	Missing	Time
0.8026	800	ni1	4 6	$\beta_{i1}$	14.30	$\rho_1(Tm1)$	0.050	M1(Tm1)	Tm1
0.8392	800	ni1	4 6	$\beta_{i1}$	14.30	$\rho_1(Tm2)$	0.050	M1(Tm2)	Tm2
0.7628	800	ni1	4 6	$\beta_{i1}$	14.30	$\rho_1(Tm3)$	0.050	M1(Tm3)	Tm3
0.8213	800	ni1	4 6	$\beta_{i1}$	14.30	$\rho_1(Tm4)$	0.050	M1(Tm4)	Tm4
0.7963	800	ni1	4 6	$\beta_{i1}$	14.30	$\rho_1(Tm5)$	0.050	M1(Tm5)	Tm5

**Footnotes**

Tm1: 0.00, 0.20, 0.40, 0.60, 0.80, 1.00  
 Tm2: 0.00, 0.60, 0.70, 0.80, 0.90, 1.00  
 Tm3: 0.00, 0.10, 0.20, 0.30, 0.40, 1.00  
 Tm4: 0.00, 0.10, 0.20, 0.80, 0.90, 1.00  
 Tm5: 0.00, 0.45, 0.50, 0.55, 0.60, 1.00  
 M1(Tm1): 0.00, 0.06, 0.12, 0.18, 0.24, 0.30  
 M1(Tm2): 0.00, 0.18, 0.21, 0.24, 0.27, 0.30  
 M1(Tm3): 0.00, 0.03, 0.06, 0.09, 0.12, 0.30  
 M1(Tm4): 0.00, 0.03, 0.06, 0.24, 0.27, 0.30  
 M1(Tm5): 0.00, 0.14, 0.15, 0.17, 0.18, 0.30  
 ni1: 200, 200, 200, 200  
 $\beta_{i1}$ : 5.00, 5.00, 7.00, 10.00  
 $\rho_1(Tm1)$ : 1.0000, 0.8000, 0.6767, 0.5724, 0.4842, 0.4096  
 $\rho_1(Tm2)$ : 1.0000, 0.5724, 0.5265, 0.4842, 0.4453, 0.4096  
 $\rho_1(Tm3)$ : 1.0000, 0.8698, 0.8000, 0.7358, 0.6767, 0.4096  
 $\rho_1(Tm4)$ : 1.0000, 0.8698, 0.8000, 0.4842, 0.4453, 0.4096  
 $\rho_1(Tm5)$ : 1.0000, 0.6490, 0.6224, 0.5969, 0.5724, 0.4096

## GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)



The horizontal axis, *Time Column Id*, gives the sequence number of the measurement columns. Thus, 1 is Tm1, 2 is Tm2, and so on.

The pattern Tm2 produces the highest power. Remember that Tm2 put the measurements at the end.

Patterns Tm1, Tm4, and Tm5 are nearly tied for achieving the medium powers. Tm1, the equally-spaced measurement scenario, is near the middle.

## Example 6 – Entering a Correlation Matrix

This example will show how a correlation matrix can be loaded directly.

This example uses  $G = 4$ ,  $M = 4$ ,  $\sigma = 14.3$ ,  $\alpha = 0.05$ , and power = 0.9. The missing input type will be set to Linear from 0 to 30% and the pairwise missing assumption is independent. Group means are 5, 5, 7, 10. The per group sample sizes range from 150 to 300. A correlation matrix (shown below) is available from a previous study.

**Table of Correlations**

C1	C2	C3	C4
1.0000	0.7000	0.4900	0.3430
0.7000	1.0000	0.7000	0.4900
0.4900	0.7000	1.0000	0.7000
0.3430	0.4900	0.7000	1.0000

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)** procedure window. You may then make the appropriate entries as listed below, or open **Example 6** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Power</b>
Alpha.....	<b>0.05</b>
G (Number of Groups).....	<b>4</b>
Group Allocation Input Type .....	<b>Equal (n1 = n2 = ... = nG = n)</b>
n (Sample Size Per Group) .....	<b>150 200 250 300</b>
Measurement Time Input Type.....	<b>Equally Spaced Measurement Times</b>
M (Number of Measurement Times).....	<b>4</b>
$\beta$ 's Input Type .....	<b><math>\beta_1, \beta_2, \dots, \beta_G</math> (Slopes, i.e., Means)</b>
$\beta_1, \beta_2, \dots, \beta_G$ (Slopes, i.e., Means).....	<b>5 5 7 10</b>
$\sigma$ (Standard Deviation) .....	<b>14.3</b>
Pattern of $\rho$ 's Across Time.....	<b>Matrix on Spreadsheet</b>
Columns Containing the $\rho_{jk}$ 's .....	<b>C1-C4</b>
<i>You must enter the correlation table using the Spreadsheet icon to the right.</i>	
Missing Input Type.....	<b>Linear (Steady Change)</b>
Pairwise Missing Pattern .....	<b>Independent (Ind)</b>
First Missing Proportion (Ind) .....	<b>0</b>
Last Missing Proportion (Ind).....	<b>0.30</b>

## Output

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

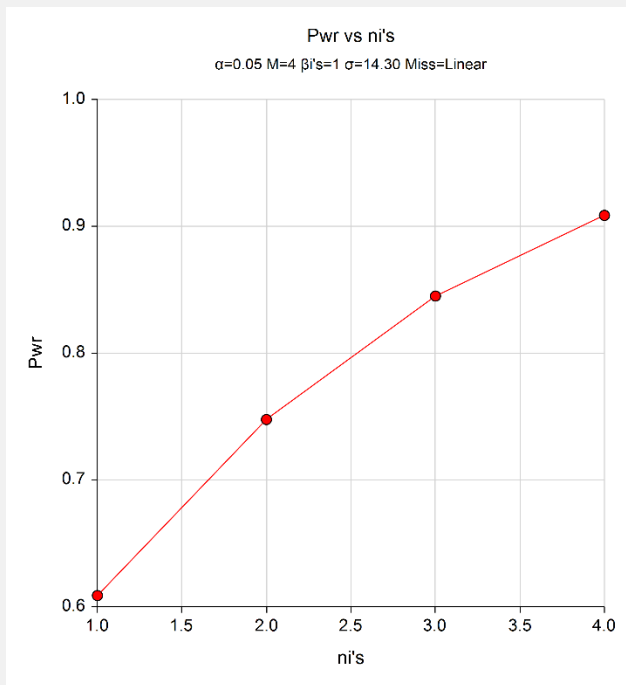
### Numeric Results for a Multi-Group Slope Comparison Test using GEE

Measurement Times: Equally spaced  
 Correlation: Matrix stored on spreadsheet in C1-C4  
 Missing Pattern: Range of missing proportions  
 Observant Proportions: Assume independence

Power	N	Group Sample Sizes ni's	Group Slopes GJM	Group Slopes $\beta$ 's	Std Dev $\sigma$	Base Corr $\rho$	First Row of Corr Matrix	Alpha	Missing	Time
0.6088	600	ni1	4 4	$\beta$ i1	14.30	n/a	$\rho$ 1(T1)	0.050	M1(T1)	T1
0.7476	800	ni2	4 4	$\beta$ i1	14.30	n/a	$\rho$ 1(T1)	0.050	M1(T1)	T1
0.8450	1000	ni3	4 4	$\beta$ i1	14.30	n/a	$\rho$ 1(T1)	0.050	M1(T1)	T1
0.9086	1200	ni4	4 4	$\beta$ i1	14.30	n/a	$\rho$ 1(T1)	0.050	M1(T1)	T1

### Footnotes

T1: 0.00, 0.33, 0.67, 1.00  
 M1(T1): 0.00, 0.10, 0.20, 0.30  
 ni1: 150, 150, 150, 150  
 ni2: 200, 200, 200, 200  
 ni3: 250, 250, 250, 250  
 ni4: 300, 300, 300, 300  
 $\beta$ i1: 5.00, 5.00, 7.00, 10.00  
 $\rho$ 1(T1): 1.0000, 0.7000, 0.4900, 0.3430



### Autocorrelation Matrix for Report Row 1

Time	T(0.00)	T(0.33)	T(0.67)	T(1.00)
T(0.00)	1.0000	0.7000	0.4900	0.3430
T(0.33)	0.7000	1.0000	0.7000	0.4900
T(0.67)	0.4900	0.7000	1.0000	0.7000
T(1.00)	0.3430	0.4900	0.7000	1.0000

Note that the horizontal axis corresponds to the sequence number of the ni's. Hence, 1 is ni1, 2 is ni2, and so on.

## Example 7 – Entering an Observant Probabilities Matrix

This example will show how an observant probabilities matrix can be loaded directly.

This example uses  $G = 4$ ,  $M = 4$ ,  $\sigma = 14.3$ , and  $\alpha = 0.05$ . The per group sample sizes range from 150 to 300. Group means are 5, 5, 7, 10. The correlation pattern will be Linear Exponential Decay with a base correlation of 0.8, Base Time Proportion of 0.1, and Emax set to 4.

The missing input type will be set to Matrix of Pairwise Missing.

**Table of Observant Probabilities**

Row	C1	C2	C3	C4
1	1.00	0.90	0.80	0.70
2	0.90	0.90	0.72	0.63
3	0.80	0.72	0.80	0.56
4	0.70	0.63	0.56	0.70

This table gives the pairwise observant probabilities. That is, each entry gives the probability of obtaining a response for both the row and column time points. For example, 0.63 is the probability of observing both the second response and the fourth response.

## Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **GEE Tests for the Slope of Multiple Groups in a Repeated Measures Design (Continuous Outcome)** procedure window. You may then make the appropriate entries as listed below, or open **Example 7** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Power</b>
Alpha.....	<b>0.05</b>
G (Number of Groups).....	<b>4</b>
Group Allocation Input Type .....	<b>Equal (n1 = n2 = ... = nG = n)</b>
n (Sample Size Per Group) .....	<b>150 200 250 300</b>
Measurement Time Input Type.....	<b>Equally Spaced Measurement Times</b>
M (Number of Measurement Times).....	<b>4</b>
$\beta$ 's Input Type .....	<b><math>\beta_1, \beta_2, \dots, \beta_G</math> (Slopes, i.e., Means)</b>
$\beta_1, \beta_2, \dots, \beta_G$ (Slopes, i.e., Means).....	<b>5 5 7 10</b>
$\sigma$ (Standard Deviation) .....	<b>14.3</b>
Pattern of $\rho$ 's Across Time.....	<b>Linear Exponential Decay</b>
$\rho$ (Base Correlation) .....	<b>0.8</b>
Base Time Proportion .....	<b>0.1</b>
Emax (Max Decay Exponent).....	<b>4</b>
Missing Input Type.....	<b>Pairwise Observed Proportions on Spreadsheet</b>
Columns of Pairwise Observed .....	<b>C1-C4</b>
<i>You must enter the pairwise probabilities table using the Spreadsheet icon to the right.</i>	



## Output

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

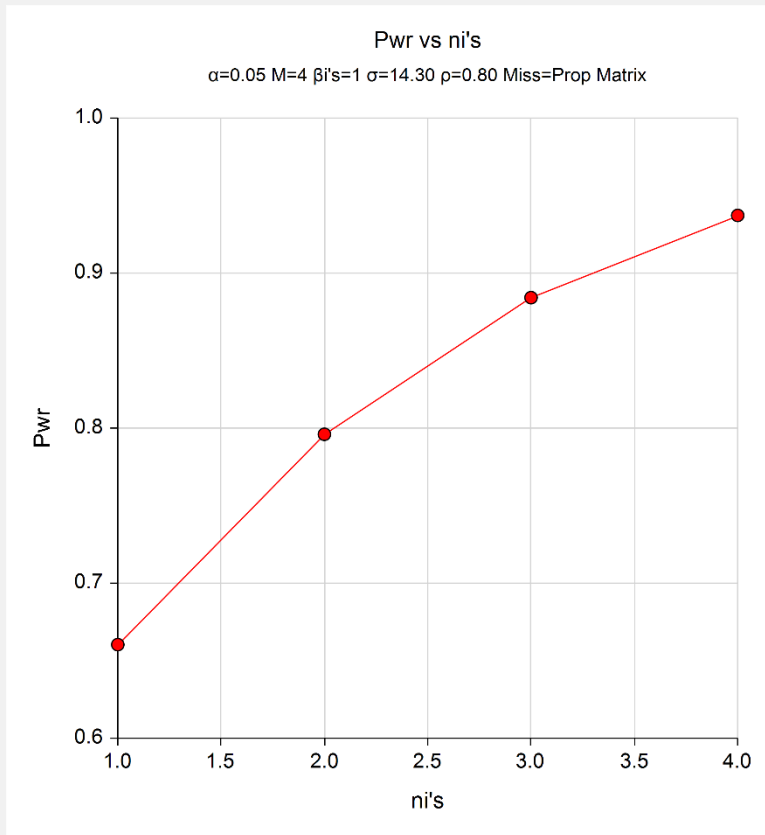
### Numeric Results for a Multi-Group Slope Comparison Test using GEE

Measurement Times: Equally spaced  
 Correlation: Linear exponential decay, with Emax = 4 and Base Time Prop = 0.1  
 Missing Pattern: N/A. Matrix of observant probabilities entered in columns C1-C4.

Power	N	Group Sample Sizes ni's	GJM	Group Slopes $\beta_i$ 's	Std Dev $\sigma$	Base Corr $\rho$	First Row of Corr Matrix	Alpha	Missing	Time
0.6604	600	ni1	4 4	$\beta_{i1}$	14.30	0.8000	$\rho_1(T_1)$	0.050	n/a	T1
0.7960	800	ni2	4 4	$\beta_{i1}$	14.30	0.8000	$\rho_1(T_1)$	0.050	n/a	T1
0.8842	1000	ni3	4 4	$\beta_{i1}$	14.30	0.8000	$\rho_1(T_1)$	0.050	n/a	T1
0.9372	1200	ni4	4 4	$\beta_{i1}$	14.30	0.8000	$\rho_1(T_1)$	0.050	n/a	T1

### Footnotes

T1: 0.00, 0.33, 0.67, 1.00  
 ni1: 150, 150, 150, 150  
 ni2: 200, 200, 200, 200  
 ni3: 250, 250, 250, 250  
 ni4: 300, 300, 300, 300  
 $\beta_{i1}$ : 5.00, 5.00, 7.00, 10.00  
 $\rho_1(T_1)$ : 1.0000, 0.6725, 0.5249, 0.4096



The standard reports are displayed.