

## Chapter 398

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

---

### Introduction

This module calculates the power for testing the difference between two slopes from binary, correlated data that are analyzed using the GEE method. Such data occur in two design types: clustered and longitudinal.

GEE does not require the full specification of the joint distribution of the repeated measurements, as long as the marginal 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

The details of the calculation of sample size and power is given in Ahn, Heo, and Zhang (2015), section 4.5.1, page 119-123. See also Jung and Ahn (2005). We will use their notation here.

Suppose we have  $n_1$  subjects in group 1 (treatment) and  $n_2$  subjects in group 2 (control) 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  $t_m = 1$ . The mean of the binary responses  $y_{kij}$  is modeled by the log-odds model

$$\text{logit}(p_{kij}) = \log\left(\frac{p_{kij}}{1 - p_{kij}}\right) = a_k + \beta_k t_j$$

where

$y_{kij}$  is the  $j^{\text{th}}$  response from subject  $i$  in group  $k$ .

$p_{kij}$  is expectation of  $y_{kij}$ ,

$a_1$  is the regression coefficient giving intercept of the treatment group,

$a_2$  is the regression coefficient giving intercept of the control group,

$\beta_1$  is the regression coefficient giving slope of the treatment group,

$\beta_2$  is the regression coefficient giving slope of the control group.

This model can be expressed as

$$p_{kij} = \frac{\exp(a_k + \beta_k t_j)}{1 + \exp(a_k + \beta_k t_j)}$$

GEE is used to estimate and test hypotheses about the equality of the slopes  $\beta_1$  and  $\beta_2$ .

### Odds Ratio of P1(1) and P2(1)

Notice that when  $a_1 = a_2$  the odds ratio of P1(1) and P2(1) is related to the difference in the slopes as follows.

$$\begin{aligned} \log[OR(P1(1), P2(1))] &= \text{logit}(P1(1)) - \text{logit}(P2(1)) \\ &= a_1 + \beta_1(1) - a_2 + \beta_2(1) \\ &= \beta_1 - \beta_2 \end{aligned}$$

Hence a hypothesis about the difference in slopes is a hypothesis about the odds ratio of P1(1) and P2(1).

## 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 Two Groups in a Repeated Measures Design (Binary 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 Two Groups in a Repeated Measures Design (Binary 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 Two Groups in a Repeated Measures Design (Binary 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. These are summarized here.

GEE is used to estimate the regression coefficients  $\beta_1$  and  $\beta_2$ . The significance of  $\beta_1 - \beta_2$ , the coefficient associated with the difference between the treatment and control group slopes, is tested using a Wald statistic for which the following sample size formula is derived

$$n = \frac{\left(\frac{v_1^2}{r_1} + \frac{v_2^2}{r_2}\right) \left(z_{1-\frac{\alpha}{h}} + z_{1-\gamma}\right)^2}{\delta^2}$$

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

where

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

$\gamma = 1 - \text{power}$

$\alpha = \text{significance level}$

$$v_k^2 = \frac{s_k^2 + c_k^2}{s_k^4}$$

$$\delta = \beta_1 - \beta_2$$

$$s_k^2 = \sum_{j=1}^M \phi_j p_{kj} q_{kj} (t_j - \tau_k)^2$$

$$c_k^2 = \sum_{j \neq j'}^M \phi_{jj'} \rho_{jj'} \sqrt{p_{kj} q_{kj} p_{kj'} q_{kj'}} (t_j - \tau_k) (t_{j'} - \tau_k)$$

$$\tau_k = \frac{\sum_{j=1}^M \phi_j p_{kj} q_{kj} t_j}{\sum_{j=1}^M \phi_j p_{kj} q_{kj}}$$

$$q_{kj} = 1 - p_{kj}$$

$$r_k = n_k / N$$

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

$\rho_{jj'}$  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  $W$ .

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 *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.

---

#### Test

##### Alternative Hypothesis

This option specifies the alternative hypothesis. The null hypothesis is always  $H_0: \beta_1 - \beta_2 = 0$

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  $H_1: \beta_1 - \beta_2 < 0$  or  $H_1: \beta_1 - \beta_2 > 0$ .

- **Two-Sided**

This is the most common selection. Use it for testing the alternative hypotheses  $H_1: 0 \neq \beta_1 - \beta_2$ .

---

#### 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.



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

### 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 1 Allocation Percentage

#### N (Subjects)

Enter a value for the sample size, N, the number of subjects in the study. Each subject is measured  $m$  times, where  $m > 1$ .

You may enter a single value, such as '42'; a range of values, such as '10 to 100 by 10'; or a list of values, such as '10 30 80 90'.

#### R (Group 1 Allocation %)

Enter one or more values for the percentage of the subjects that are (or will be) in group 1, the treatment group. The usual entry is 50 for 50%. The valid range is from 1 to 99. However, a more reasonable range is from 10 to 90. Usually, the treatment group allocation is 50%.

You can enter a single value such as '50' or a series of values such as '40 50 60' or '40 to 70 by 5'.

---

### 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.

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

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'.

$M$ , 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		

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

### 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.

### 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

### P1(0) and P2(0) Input Type

Indicate whether the two event probabilities will be equal or different at the first measurement (time proportion = 0).

The choices are

- **P1(0) = P2(0)**  
Enter one or more values that will be used for both P1(0) and P2(0). In a trial in which a measurement is taken before the treatment is applied, it is quite reasonable to make this assumption.
- **Enter P1(0) and P2(0) Separately**  
Individual entry boxes will appear for P1(0) and P2(0) separately allowing different values. Note that this assumes that the intercepts of the regression equations will be different making the interpretation of the slopes problematic.

### P1(0) and P2(0)

Enter values for the event proportions in group 1 and 2 at the beginning of the study. We assume that group 1 is the 'treatment' group and group 2 is the 'control or reference' group.

Values must be between 0 and 1.

You can enter a single value such as *0.1* or a series of values such as *0.1 0.2 0.3* or *0.1 to 0.5 by 0.1*.

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

### P1(1) Input Type

Indicate what type of values to enter to specify P1(1), the group 1 event probability at the last measurement. Regardless of the entry type chosen, the same test statistic used in the power and sample size calculations. This option is simply given for convenience in specifying the P1(1).

The choices are

- **P1(1)**  
Enter values for P1(1) directly.
- **Difference (P1(1) - P2(1))**  
Enter values for the difference (P1(1) - P2(1)) and values for P2(1). The corresponding value of P1(1) will be computed from these two values.
- **Ratio (P1(1) / P2(1))**  
Enter values for the ratio (P1(1) / P2(1)) and values for P2(1). The corresponding value of P1(1) will be computed from these two values.
- **Odds Ratio (O1(1) / O2(1))**  
Enter values for the odds ratio (O1(1) / O2(1)) and values for P2(1). The corresponding value of P1(1) will be computed from these two values.  
  
Note that  $O1(1) = P1(1)/(1-P1(1))$  and  $O2(1) = P2(1)/(1-P2(1))$

### P1(1)

Enter values for the event probabilities of group 1 at the final measurement of the study. We assume that group 1 is the 'treatment' group and group 2 is the 'control or reference' group.

Values must be between 0 and 1.

You can enter a single value such as *0.1* or a series of values such as *0.1 0.2 0.3* or *0.1 to 0.5 by 0.1*.

#### Note

This value must be different from P2(1) but can be the same as P1(0).

### Difference (P1(1) - P2(1))

Enter the difference between P1(1) and P2(1) at the final measurement. This difference is used with P2(1) to calculate the value of P1(1) using the formula:  $P1(1) = \text{Diff} + P2(1)$ .

You can enter a single value such as *0.05* or a series of values such as *0.03 0.05 0.10* or *0.01 to 0.09 by 0.02*.

### Range

Differences must be between -1 and 1. They cannot take on the values -1, 0, or 1. The resulting value of P1(1) must be between 0 and 1. If it is not, it is changed so that it is between 0 and 1.

### Ratio (P1(1) / P2(1))

Enter the ratio of the two probabilities P1(1) and P2(1). This ratio is used with P2(1) to calculate the value of P1(1) using the formula:  $P1(1) = \text{Ratio} \times P2(1)$ .

You can enter a single value such as *0.5* or a series of values such as *0.5 0.6 0.7 0.8* or *0.25 to 2.0 by 0.25*.

### Range

Ratios must be greater than zero. They cannot take on the value of one. The resulting value of P1(1) must be between 0 and 1. If it is not, it is changed so that it is between 0 and 1.

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

### Odds Ratio (O1(1) / O2(1))

This option specifies the odds ratio between the two probabilities P1(1) and P2(1). This value is used with P2(1) to calculate the value of P1(1).

You can enter a single value such as *0.5* or a series of values such as *0.5 0.6 0.7 0.8* or *1.25 to 2.0 by 0.25*.

### Range

Odds ratios must be greater than zero. They cannot take on the value of one.

### P2(1)

Enter values for the event probability in group 2 at the final measurement of the study. We assume that group 2 is the 'control or reference' group. The difference between P1(1) and P2(1) is a measure of the effect size.

Values must be between 0 and 1.

You can enter a single value such as *0.1* or a series of values such as *0.1 0.2 0.3* or *0.1 to 0.5 by 0.1*.

### Note

P2(1) must be different from P1(1). That is, you cannot have a zero effect size.

## 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.

1	$\rho$	$\rho$	$\rho$
$\rho$	1	$\rho$	$\rho$
$\rho$	$\rho$	1	$\rho$
$\rho$	$\rho$	$\rho$	1

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.

1	$\rho$	0	0
$\rho$	1	$\rho$	0
0	$\rho$	1	$\rho$
0	0	$\rho$	1

Example first row of correlation matrix

1.0000, 0.5000, 0.0000, 0.0000, 0.0000, 0.0000

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

- **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.

1	$\rho$	$\rho$	0
$\rho$	1	$\rho$	$\rho$
$\rho$	$\rho$	1	$\rho$
0	$\rho$	$\rho$	1

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.

1	$\rho$	$\rho^2$	$\rho^3$
$\rho$	1	$\rho$	$\rho^2$
$\rho^2$	$\rho$	1	$\rho$
$\rho^3$	$\rho^2$	$\rho$	1

Example first row of correlation matrix

1.0000, 0.5000, 0.2500, 0.1250, 0.0625, 0.0313

- **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

$\rho^{0/3}$	$\rho^{1/3}$	$\rho^{2/3}$	$\rho^{3/3}$
$\rho^{1/3}$	$\rho^{0/3}$	$\rho^{1/3}$	$\rho^{2/3}$
$\rho^{2/3}$	$\rho^{1/3}$	$\rho^{0/3}$	$\rho^{1/3}$
$\rho^{3/3}$	$\rho^{2/3}$	$\rho^{1/3}$	$\rho^{0/3}$

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

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

- **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  $\rho_{jk} = \rho^{|t_j - t_k|^{Dexp}}$ , where  $t_j$  and  $t_k$  are time proportions at which measurements are planned. An example with  $Dexp = 1$  is

$$\begin{bmatrix} \rho^{0/3} & \rho^{1/3} & \rho^{2/3} & \rho^{3/3} \\ \rho^{1/3} & \rho^{0/3} & \rho^{1/3} & \rho^{2/3} \\ \rho^{2/3} & \rho^{1/3} & \rho^{0/3} & \rho^{1/3} \\ \rho^{3/3} & \rho^{2/3} & \rho^{1/3} & \rho^{0/3} \end{bmatrix}$$

Example first row of correlation matrix when  $M = 6$ ,  $\rho = 0.1$ , and  $Dexp = 1$ . 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  $\rho_{jk} = \rho^{|t_j/P - t_k/P| * Emax}$ .  $P$  is called the Base Time Proportion. The exponents of  $\rho$  range from 1 at  $P$  to  $Emax$  at  $M$ .

Example first row of correlation matrix with  $Emax = 4$

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.

### **$\rho$ 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

**GEE Tests for the Slope of Two Groups in a Repeated Measures Design (Binary Outcome)** **$\rho$  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.

 **$\rho$  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



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

### 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  $E_{\max}$  at the maximum time.

#### Range

$E_{\max} > 0$ .

#### Typical

Typical values of  $E_{\max}$  are between 1 and 10.

### **$\rho$ 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.

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

- **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 x 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

**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.

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

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

**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:

<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

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

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

**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.

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.

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

### 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.

Miss	Time
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.

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

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.

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

Researchers are planning a study to assess whether a new drug will change the response rate to a particular event. They want to compare subjects who take the new drug with subjects who take a placebo. Their experimental protocol calls for a baseline measurement, followed by administration of the new drug or the placebo, followed by the event of interest, followed by three additional measurements one day apart. They want to detect a difference of 0.2 in the final response probabilities. They also want a sensitivity analysis by considering a range of possible differences from 0.1 to 0.3.

Similar studies have found that 80% of subjects do not show a response before the event. Three hours after the event, the response rate is 50%. Using these results, the basic parameters are  $P1(0) = 0.80$ ,  $P2(0) = 0.80$ , and  $P2(1) = 0.50$ . A difference of 0.2 implies that  $P1(1) = 0.5 + 0.2 = 0.7$ . This difference is equivalent to an odds ratio of 2.333.

These studies also showed an autocorrelation between adjacent measurements on the same individual of 0.7, so they want to try autocorrelations of 0.6, 0.7, and 0.8. The researchers assume that first-order autocorrelation adequately represents the autocorrelation pattern. A two-sided Wald test will be conducted at the 0.05 significance level and at 90% power. The subjects will be divided equally between the treatment and control groups.

The researchers anticipate that the missing data pattern across time will begin at 0% missing and increase steadily to 30% at the fourth measurement. They assume that the pairwise missing probabilities are *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 Two Groups in a Repeated Measures Design (Binary 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>
Alternative Hypothesis .....	<b>Two-Sided</b>
Power .....	<b>0.9</b>
Alpha .....	<b>0.05</b>
R (Group 1 Allocation %) .....	<b>50</b>
Measurement Time Input Type .....	<b>Equally Spaced Measurement Times</b>
M (Number of Measurements) .....	<b>4</b>
P1(0) and P2(0) Input Type .....	<b>P1(0) = P2(0)</b>
P1(0) and P2(0) .....	<b>0.8</b>
P1(1) Input Type .....	<b>Difference (P1(1) - P2(1))</b>
Difference (P1(1) - P2(1)) .....	<b>0.1 0.2 0.3</b>
P2(1) .....	<b>0.5</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>0.3</b>

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

## Annotated Output

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

#### Numeric Results for the Test of Slope-Difference in Binary Data 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  
 Pk(t): Event probability of group k (1=treatment, 2=control) at time proportion t

Power	Group 1 Alloc Pct		M	P1(0)	P2(0)	P1(1)	P2(1)	Diff	Base Corr $\rho$	First Row of Corr Matrix	Alpha	Missing	Time
	N	R											
0.9000	2518	50	4	0.800	0.800	0.600	0.500	0.100	0.600	$\rho_1(T1)$	0.050	M1(T1)	T1
0.9001	2191	50	4	0.800	0.800	0.600	0.500	0.100	0.700	$\rho_2(T1)$	0.050	M1(T1)	T1
0.9000	1748	50	4	0.800	0.800	0.600	0.500	0.100	0.800	$\rho_3(T1)$	0.050	M1(T1)	T1
0.9000	596	50	4	0.800	0.800	0.700	0.500	0.200	0.600	$\rho_1(T1)$	0.050	M1(T1)	T1
0.9000	518	50	4	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_2(T1)$	0.050	M1(T1)	T1
0.9002	413	50	4	0.800	0.800	0.700	0.500	0.200	0.800	$\rho_3(T1)$	0.050	M1(T1)	T1
0.9007	240	50	4	0.800	0.800	0.800	0.500	0.300	0.600	$\rho_1(T1)$	0.050	M1(T1)	T1
0.9001	208	50	4	0.800	0.800	0.800	0.500	0.300	0.700	$\rho_2(T1)$	0.050	M1(T1)	T1
0.9010	166	50	4	0.800	0.800	0.800	0.500	0.300	0.800	$\rho_3(T1)$	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  
 $\rho_1(T1)$ : 1.000, 0.600, 0.360, 0.216  
 $\rho_2(T1)$ : 1.000, 0.700, 0.490, 0.343  
 $\rho_3(T1)$ : 1.000, 0.800, 0.640, 0.512

#### 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. 2005. 'Sample size for a two-group comparison of repeated binary measurements using GEE'. Statistics in Medicine, Volume 24, pages 2583-2596.

#### Report Definitions

Power is the probability of rejecting a false null hypothesis.  
 N is the total number of subjects in the study.  
 R is the treatment group allocation proportion. It is the proportion of subjects that are in the treatment group.  
 M is the number of time points at which each subject is measured.  
 Diff = (P1(1)-P2(1)) is the difference between the treatment and control event probabilities at the final measurement time.  
 Pk(t) is the event probability of group k (1=treatment, 2=control) at the measurement time proportion t.  
 $\rho$  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.

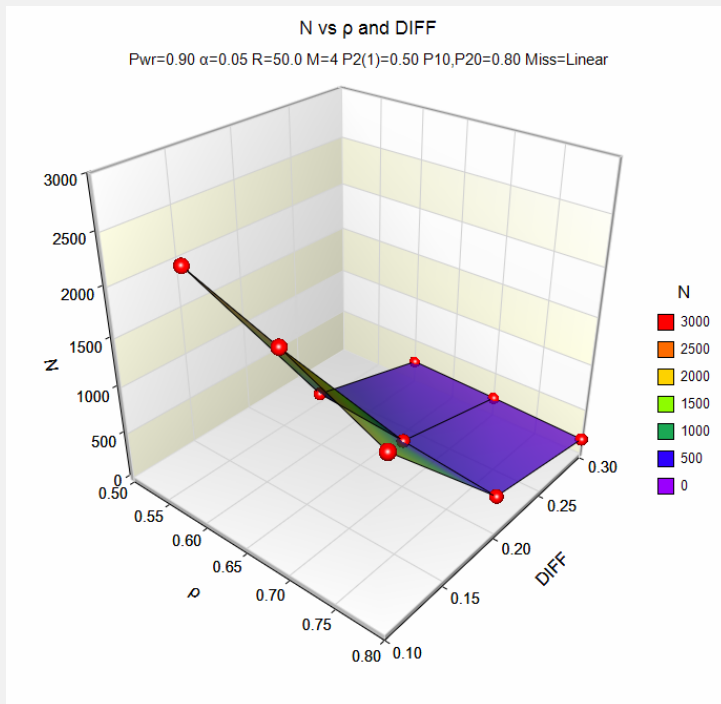
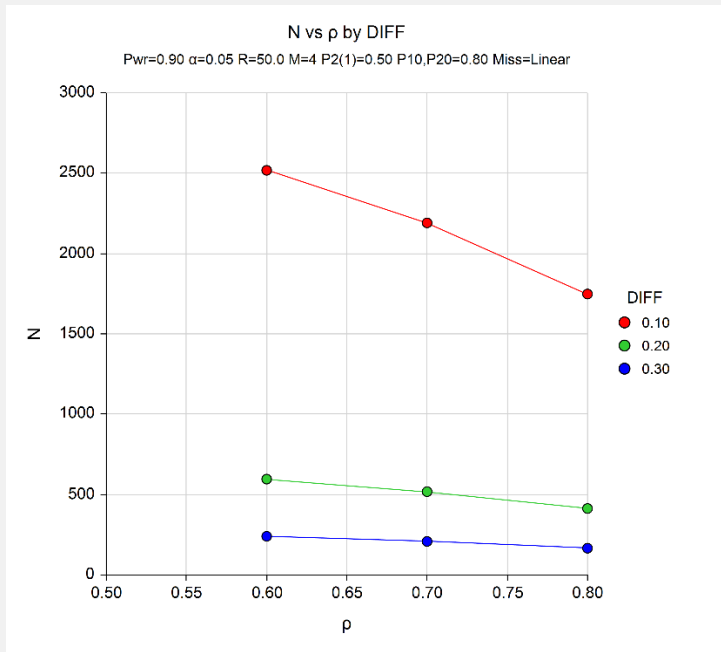
#### Summary Statements

A sample of 2518 subjects, each scheduled to be measured 4 times, achieve a power of 0.9000 when using a two-sided Wald test from a GEE analysis to test whether a P1(1) of 0.600 differs from a P2(1) of 0.500 at a significance level of 0.050. The value of P1(1) was calculated from a difference (P1(1) - P2(1)) of 0.100. P1(0) and P2(0) are assumed to be 0.800. The subjects are randomly split between a treatment group and a control group, with 50% of the subjects assigned to the treatment group. 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, 0.10, 0.20, 0.30}. The first row of the autocorrelation matrix of the responses within a subject is assumed to be {1.000, 0.600, 0.360, 0.216}. Other rows follow the same pattern.

This report gives the sample size for each value of the other parameters. The definitions of each of the items is given in the Reports Definitions section at the end of the report.



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



These charts show the relationship between sample size, difference, and  $\rho$  when the other parameters in the design are held constant.

**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 sample sizes ranging from 100 to 1000 for the main cases 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 Two Groups in a Repeated Measures Design (Binary 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>
Alternative Hypothesis .....	<b>Two-Sided</b>
Alpha.....	<b>0.05</b>
N (Subjects) .....	<b>100 to 1000 by 100</b>
R (Group 1 Allocation %) .....	<b>50</b>
Measurement Time Input Type.....	<b>Equally Spaced Measurement Times</b>
M (Number of Measurements).....	<b>4</b>
P1(0) and P2(0) Input Type .....	<b>P1(0) = P2(0)</b>
P1(0) and P2(0) .....	<b>0.80</b>
P1(1) Input Type .....	<b>Difference (P1(1) - P2(1))</b>
Difference (P1(1) - P2(1)) .....	<b>0.2</b>
P2(1) .....	<b>0.5</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>0.3</b>

GEE Tests for the Slope of Two Groups in a Repeated Measures Design (Binary Outcome)

## Output

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

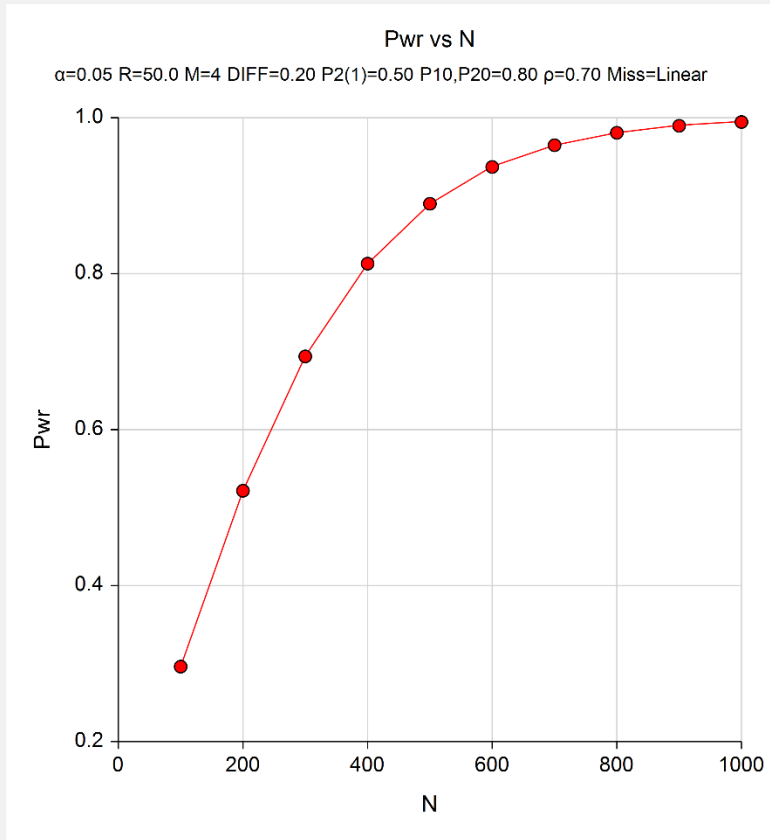
**Numeric Results for the Test of Slope-Difference in Binary Data using GEE**

Measurement Times: Equally spaced  
 Correlation: AR(1):  $\rho(j,k) = \rho^{|j-k|}$   
 Missing Pattern: Range of missing proportions  
 Observant Proportions: Assume independence  
 Pk(t): Event probability of group k (1=treatment, 2=control) at time proportion t

Power	Group 1			M	P1(0)	P2(0)	P1(1)	P2(1)	Diff	Base Corr $\rho$	First Row of Corr Matrix	Alpha	Missing	Time
	N	Alloc Pct R	R											
0.2961	100	50	4	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T_1)$	0.050	M1(T1)	T1	
0.5216	200	50	4	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T_1)$	0.050	M1(T1)	T1	
0.6939	300	50	4	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T_1)$	0.050	M1(T1)	T1	
0.8129	400	50	4	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T_1)$	0.050	M1(T1)	T1	
0.8897	500	50	4	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T_1)$	0.050	M1(T1)	T1	
0.9368	600	50	4	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T_1)$	0.050	M1(T1)	T1	
0.9647	700	50	4	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T_1)$	0.050	M1(T1)	T1	
0.9807	800	50	4	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T_1)$	0.050	M1(T1)	T1	
0.9896	900	50	4	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T_1)$	0.050	M1(T1)	T1	
0.9945	1000	50	4	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T_1)$	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  
 $\rho_1(T_1)$ : 1.000, 0.700, 0.490, 0.343



The reports and plot indicate the power for each value of N.

## Example 3 – Impact of the Number of Repeated Measurements

Continuing with Examples 1 and 2, the researchers want to study the impact on the sample size of changing the number of measurements made on each individual. Their experimental protocol calls for 4 measurements. They want to see the impact of increasing the number of measurements to 7.

### 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 Two Groups in a Repeated Measures Design (Binary 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>Power</b>
Alternative Hypothesis .....	<b>Two-Sided</b>
Alpha.....	<b>0.05</b>
N (Subjects) .....	<b>100 to 1000 by 100</b>
R (Group 1 Allocation %) .....	<b>50</b>
Measurement Time Input Type.....	<b>Equally Spaced Measurement Times</b>
M (Number of Measurements).....	<b>4 7</b>
P1(0) and P2(0) Input Type .....	<b>P1(0) = P2(0)</b>
P1(0) and P2(0) .....	<b>0.8</b>
P1(1) Input Type .....	<b>Difference (P1(1) - P2(1))</b>
Difference (P1(1) - P2(1)) .....	<b>0.2</b>
P2(1) .....	<b>0.5</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>0.3</b>

GEE Tests for the Slope of Two Groups in a Repeated Measures Design (Binary Outcome)

## Output

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

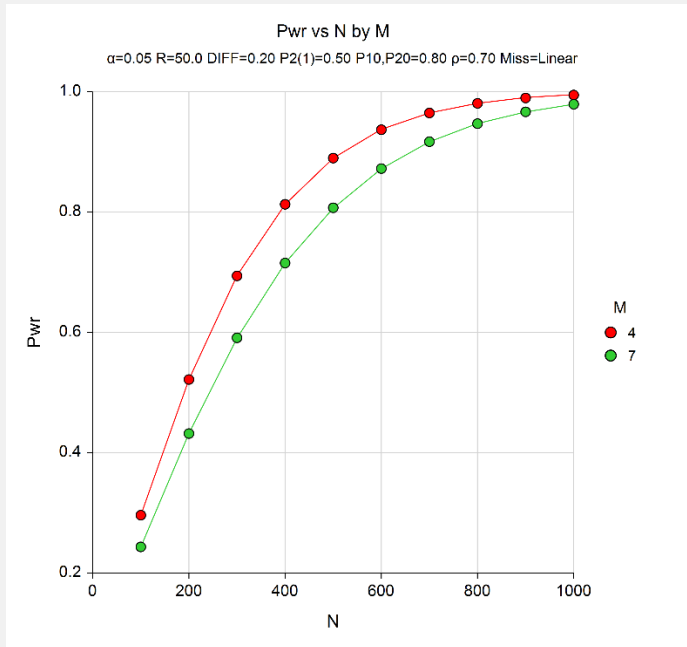
### Numeric Results for the Test of Slope-Difference in Binary Data using GEE

Measurement Times: Equally spaced  
 Correlation: AR(1):  $\rho(j,k) = \rho^{|j-k|}$   
 Missing Pattern: Range of missing proportions  
 Observant Proportions: Assume independence  
 Pk(t): Event probability of group k (1=treatment, 2=control) at time proportion t

Power	Group 1 Alloc Pct			M	P1(0)	P2(0)	P1(1)	P2(1)	Diff	Base Corr $\rho$	First Row of Corr Matrix	Alpha	Missing	Time
	N	R	R											
0.2961	100	50	4	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T1)$	0.050	M1(T1)	T1	
0.2434	100	50	7	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T2)$	0.050	M1(T2)	T2	
0.5216	200	50	4	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T1)$	0.050	M1(T1)	T1	
0.4318	200	50	7	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T2)$	0.050	M1(T2)	T2	
0.6939	300	50	4	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T1)$	0.050	M1(T1)	T1	
0.5910	300	50	7	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T2)$	0.050	M1(T2)	T2	
0.8129	400	50	4	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T1)$	0.050	M1(T1)	T1	
0.7153	400	50	7	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T2)$	0.050	M1(T2)	T2	
0.8897	500	50	4	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T1)$	0.050	M1(T1)	T1	
0.8071	500	50	7	0.800	0.800	0.700	0.500	0.200	0.700	$\rho_1(T2)$	0.050	M1(T2)	T2	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

### Footnotes

T1: 0.00, 0.33, 0.67, 1.00  
 T2: 0.00, 0.17, 0.33, 0.50, 0.67, 0.83, 1.00  
 M1(T1): 0.00, 0.10, 0.20, 0.30  
 M1(T2): 0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30  
 $\rho_1(T1)$ : 1.000, 0.700, 0.490, 0.343  
 $\rho_1(T2)$ : 1.000, 0.700, 0.490, 0.343, 0.240, 0.168, 0.118



Note that increasing the number of measurements has had the surprising result of decreasing the power, probably because the assumption of the AR(1) model for the autocorrelation has changed the way in which the correlations are formed. Note from the footnotes that the final autocorrelation drops from 0.343 when M = 4 to 0.118 when M = 7. Look at the next example to see how the autocorrelations can be put on a more equal footing.

## Example 4 – Impact of Changing M with Linear Exponential Decay

We saw in Example 3 that the increasing the number of measurements from 4 to 7 had the counter-intuitive result of reducing the power when the sample size was held constant. We surmised that this was partially due to the differing autocorrelation matrices that were used when the AR(1) model as assumed. In this example, we will leave all parameters the same, except that we will use a Linear Exponential Decay model for the autocorrelation. This will keep the autocorrelation matrices more comparable.

### 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 Two Groups in a Repeated Measures Design (Binary 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>Power</b>
Alternative Hypothesis .....	<b>Two-Sided</b>
Alpha.....	<b>0.05</b>
N (Subjects) .....	<b>100 to 1000 by 100</b>
R (Group 1 Allocation %) .....	<b>50</b>
Measurement Time Input Type.....	<b>Equally Spaced Measurement Times</b>
M (Number of Measurements).....	<b>4 7</b>
P1(0) and P2(0) Input Type .....	<b>P1(0) = P2(0)</b>
P1(0) and P2(0) .....	<b>0.8</b>
P1(1) Input Type .....	<b>Difference (P1(1) - P2(1))</b>
Difference (P1(1) - P2(1)) .....	<b>0.2</b>
P2(1) .....	<b>0.5</b>
Pattern of $\rho$ 's Across Time.....	<b>Linear Exponential Decay</b>
$\rho$ (Base Correlation) .....	<b>0.7</b>
Base Time Proportion .....	<b>0.1666666</b>
E <sub>max</sub> (Max Decay Exponent) .....	<b>3</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>0.3</b>

GEE Tests for the Slope of Two Groups in a Repeated Measures Design (Binary Outcome)

# Output

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

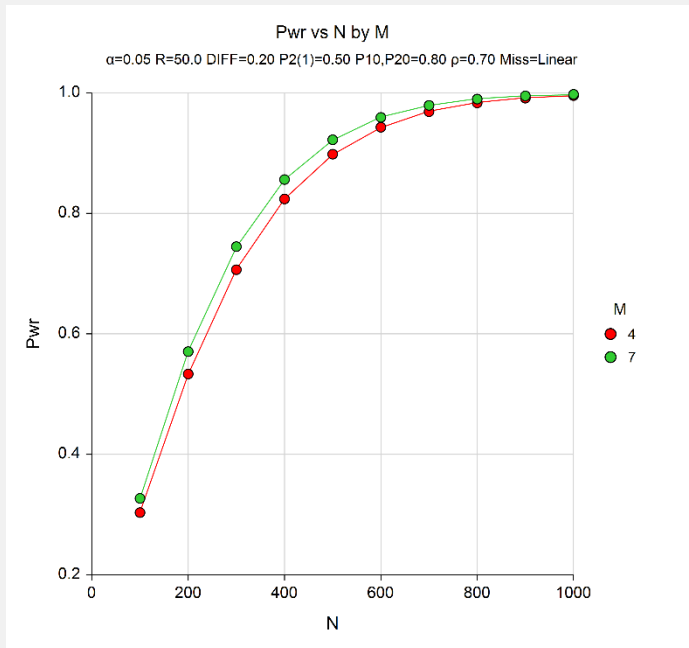
**Numeric Results for the Test of Slope-Difference in Binary Data using GEE**

Measurement Times: Equally spaced  
 Correlation: Linear exponential decay, with Emax = 3 and Base Time Prop = 0.1666666  
 Missing Pattern: Range of missing proportions  
 Observant Proportions: Assume independence  
 Pk(t): Event probability of group k (1=treatment, 2=control) at time proportion t

Power	Group 1 Alloc Pct			M	P1(0)	P2(0)	P1(1)	P2(1)	Diff	Base Corr ρ	First Row of Corr Matrix	Alpha	Missing	Time
	N	R	Pct											
0.3034	100	50	4	0.800	0.800	0.700	0.500	0.200	0.700	ρ1(T1)	0.050	M1(T1)	T1	
0.3269	100	50	7	0.800	0.800	0.700	0.500	0.200	0.700	ρ1(T2)	0.050	M1(T2)	T2	
0.5334	200	50	4	0.800	0.800	0.700	0.500	0.200	0.700	ρ1(T1)	0.050	M1(T1)	T1	
0.5705	200	50	7	0.800	0.800	0.700	0.500	0.200	0.700	ρ1(T2)	0.050	M1(T2)	T2	
0.7065	300	50	4	0.800	0.800	0.700	0.500	0.200	0.700	ρ1(T1)	0.050	M1(T1)	T1	
0.7448	300	50	7	0.800	0.800	0.700	0.500	0.200	0.700	ρ1(T2)	0.050	M1(T2)	T2	
0.8239	400	50	4	0.800	0.800	0.700	0.500	0.200	0.700	ρ1(T1)	0.050	M1(T1)	T1	
0.8562	400	50	7	0.800	0.800	0.700	0.500	0.200	0.700	ρ1(T2)	0.050	M1(T2)	T2	
0.8982	500	50	4	0.800	0.800	0.700	0.500	0.200	0.700	ρ1(T1)	0.050	M1(T1)	T1	
0.9222	500	50	7	0.800	0.800	0.700	0.500	0.200	0.700	ρ1(T2)	0.050	M1(T2)	T2	
0.9429	600	50	4	0.800	0.800	0.700	0.500	0.200	0.700	ρ1(T1)	0.050	M1(T1)	T1	
0.9593	600	50	7	0.800	0.800	0.700	0.500	0.200	0.700	ρ1(T2)	0.050	M1(T2)	T2	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

**Footnotes**

T1: 0.00, 0.33, 0.67, 1.00  
 T2: 0.00, 0.17, 0.33, 0.50, 0.67, 0.83, 1.00  
 M1(T1): 0.00, 0.10, 0.20, 0.30  
 M1(T2): 0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30  
 ρ1(T1): 1.000, 0.607, 0.456, 0.343  
 ρ1(T2): 1.000, 0.700, 0.607, 0.526, 0.456, 0.396, 0.343



Note from the footnotes that the final autocorrelation between the two models is now identical at 0.3430 (= 0.7<sup>3</sup>) when M is increased from 4 to 7. Now that the autocorrelation matrices are more comparable, the power values have increased in all cases, although only slightly. We see that increasing M has not had a huge impact on power.

## Example 5 – Validation of Sample Size Calculation using Ahn, Heo, and Zhang (2015)

Ahn, Heo, and Zhang (2015) page 123 present an example that we will use to validate this procedure. In their example,  $P_1(0) = 0.75$ ,  $P_2(0) = 0.75$ ,  $P_1(1) = 0.75$ ,  $P_2(1) = 0.50$ , the significance level is 0.05, the power is 0.80, R is 50%, M is 6, and the correlation matrix is AR1 with  $\rho = 0.8$ . The proportions missing at each time point are 0.0, 0.05, 0.1, 0.15, 0.2, and 0.25. They calculate the sample size to be 215.

### 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 Two Groups in a Repeated Measures Design (Binary 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>Sample Size</b>
Alternative Hypothesis .....	<b>Two-Sided</b>
Power .....	<b>0.80</b>
Alpha .....	<b>0.05</b>
R (Group 1 Allocation %) .....	<b>50</b>
Measurement Time Input Type .....	<b>Equally Spaced Measurement Times</b>
M (Number of Measurements) .....	<b>6</b>
P1(0) and P2(0) Input Type .....	<b>P1(0) = P2(0)</b>
P1(0) and P2(0) .....	<b>0.75</b>
P1(1) Input Type .....	<b>P1(1)</b>
P1(1) .....	<b>0.75</b>
P2(1) .....	<b>0.50</b>
Pattern of $\rho$ 's Across Time .....	<b>AR1(Traditional)</b>
$\rho$ (Base Correlation) .....	<b>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>0.25</b>



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

**Output**

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

**Numeric Results for the Test of Slope-Difference in Binary Data using GEE**

Measurement Times: Equally spaced  
 Correlation: AR1:  $\rho(j,k) = \rho^{|j-k|}$   
 Missing Pattern: Range of missing proportions  
 Observant Proportions: Assume independence  
 Pk(t): Event probability of group k (1=treatment, 2=control) at time proportion t

Power	Group 1 Alloc Pct		M	P1(0)	P2(0)	P1(1)	P2(1)	Diff	Base Corr $\rho$	First Row of Corr Matrix $\rho_1(T1)$	Alpha	Missing	Time
	N	R											
0.8017	215	50	6	0.750	0.750	0.750	0.500	0.250	0.800	$\rho_1(T1)$	0.050	M1(T1)	T1

**Footnotes**

T1: 0.00, 0.20, 0.40, 0.60, 0.80, 1.00  
 M1(T1): 0.00, 0.05, 0.10, 0.15, 0.20, 0.25  
 $\rho_1(T1)$ : 1.000, 0.800, 0.640, 0.512, 0.410, 0.328

Note that the sample size is 215, which matches that of Ahn, Heo, and Zhang (2015) exactly.

## Example 6 – Impact of Measurement Time Distribution

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

In this example the basic parameters are  $P1(0) = 0.72$ ,  $P2(0) = 0.72$ ,  $P1(1) = 0.75$ , and  $P2(1) = 0.55$ . The significance level is 0.05, the sample size ranges from 100 to 500, and R is 50%. The correlation pattern will be Linear Exponential Decay with a base correlation of 0.6, Base Time Proportion of 0.10, and Emax set to 3. The missing input type will be set to Linear from 0 to 0.30 and the pairwise missing assumption will be independent.

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 Two Groups in a Repeated Measures Design (Binary 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>
Alternative Hypothesis .....	<b>Two-Sided</b>
Alpha .....	<b>0.05</b>
N (Subjects) .....	<b>100 to 500 by 100</b>
R (Group 1 Allocation %) .....	<b>50</b>
Measurement Time Input Type .....	<b>Columns of Measurement Time Proportions</b>
Column(s) of Measurement 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>	
P1(0) and P2(0) Input Type .....	<b>P1(0) = P2(0)</b>
P1(0) and P2(0) .....	<b>0.72</b>
P1(1) Input Type .....	<b>P1(1)</b>
P1(1) .....	<b>0.75</b>
P2(1) .....	<b>0.55</b>
Pattern of $\rho$ 's Across Time .....	<b>Linear Exponential Decay</b>
$\rho$ (Base Correlation) .....	<b>0.6</b>
Base Time Proportion .....	<b>0.10</b>
Emax (Max Decay Exponent) .....	<b>3</b>
Missing Input Type .....	<b>Linear (Steady Change)</b>

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

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

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

Last Missing Proportion (Ind)..... **0.3**

## Output

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

**Numeric Results for the Test of Slope-Difference in Binary Data using GEE**

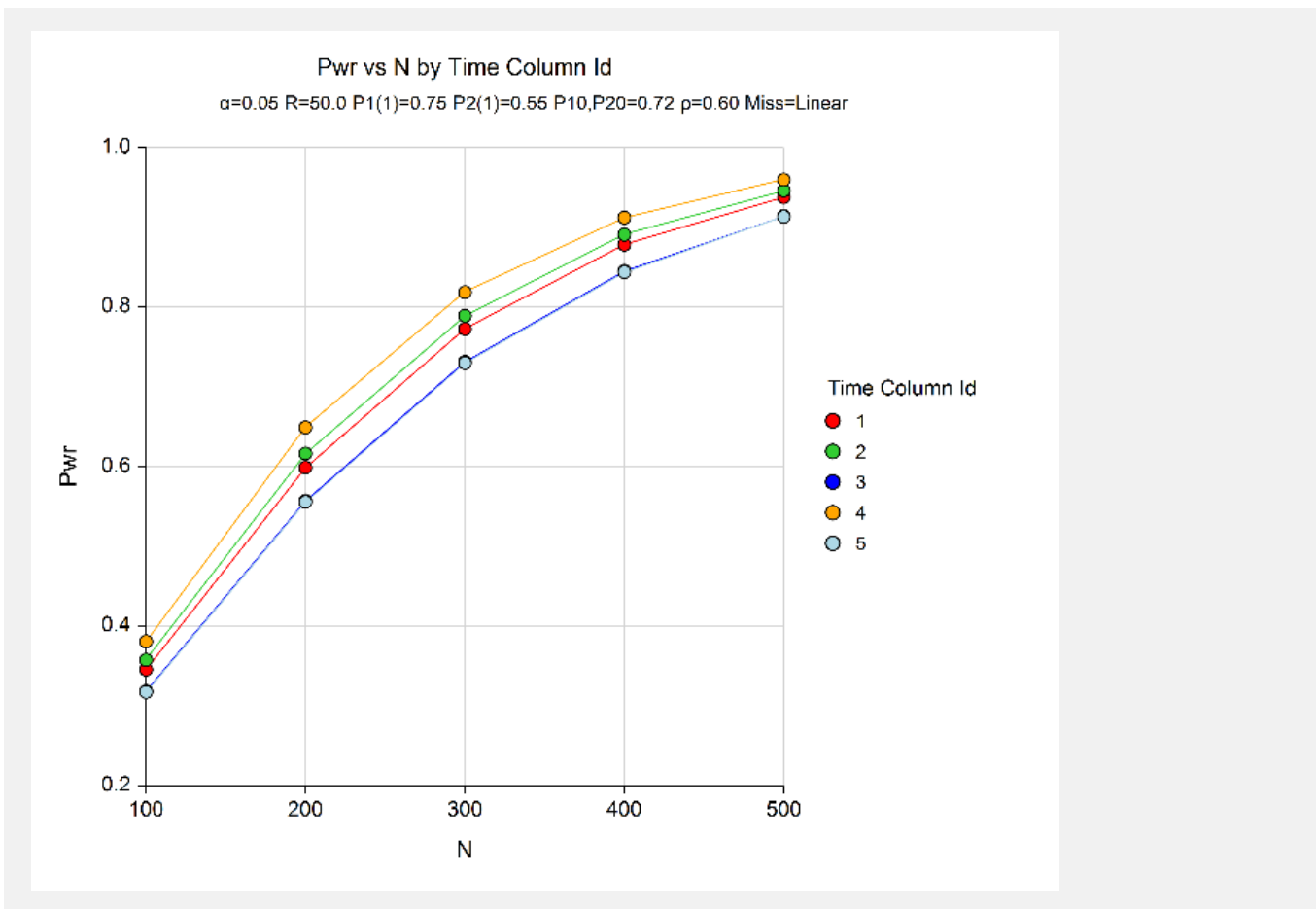
Measurement Times: Equally spaced  
 Correlation: Linear exponential decay, with Emax = 3 and Base Time Prop = 0.10  
 Missing Pattern: Range of missing proportions  
 Observant Proportions: Assume independence  
 Pk(t): Event probability of group k (1=treatment, 2=control) at time proportion t

Power	Group 1		M	P1(0)	P2(0)	P1(1)	P2(1)	Diff	First Base Row of Corr Corr	Alpha	Missing	Time
	N	R										
0.3457	100	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm1)	0.050	M1(Tm1)	Tm1
0.3576	100	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm2)	0.050	M1(Tm2)	Tm2
0.3184	100	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm3)	0.050	M1(Tm3)	Tm3
0.3808	100	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm4)	0.050	M1(Tm4)	Tm4
0.3175	100	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm5)	0.050	M1(Tm5)	Tm5
0.5989	200	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm1)	0.050	M1(Tm1)	Tm1
0.6163	200	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm2)	0.050	M1(Tm2)	Tm2
0.5572	200	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm3)	0.050	M1(Tm3)	Tm3
0.6491	200	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm4)	0.050	M1(Tm4)	Tm4
0.5558	200	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm5)	0.050	M1(Tm5)	Tm5
0.7725	300	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm1)	0.050	M1(Tm1)	Tm1
0.7890	300	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm2)	0.050	M1(Tm2)	Tm2
0.7313	300	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm3)	0.050	M1(Tm3)	Tm3
0.8184	300	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm4)	0.050	M1(Tm4)	Tm4
0.7298	300	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm5)	0.050	M1(Tm5)	Tm5
0.8782	400	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm1)	0.050	M1(Tm1)	Tm1
0.8907	400	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm2)	0.050	M1(Tm2)	Tm2
0.8451	400	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm3)	0.050	M1(Tm3)	Tm3
0.9120	400	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm4)	0.050	M1(Tm4)	Tm4
0.8438	400	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm5)	0.050	M1(Tm5)	Tm5
0.9376	500	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm1)	0.050	M1(Tm1)	Tm1
0.9459	500	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm2)	0.050	M1(Tm2)	Tm2
0.9141	500	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm3)	0.050	M1(Tm3)	Tm3
0.9594	500	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(Tm4)	0.050	M1(Tm4)	Tm4
0.9132	500	50	6	0.720	0.720	0.750	0.550	0.200	0.600 ρ1(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  
 ρ1(Tm1): 1.000, 0.536, 0.427, 0.340, 0.271, 0.216  
 ρ1(Tm2): 1.000, 0.340, 0.304, 0.271, 0.242, 0.216  
 ρ1(Tm3): 1.000, 0.600, 0.536, 0.478, 0.427, 0.216  
 ρ1(Tm4): 1.000, 0.600, 0.536, 0.271, 0.242, 0.216  
 ρ1(Tm5): 1.000, 0.403, 0.381, 0.360, 0.340, 0.216

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



The legend, *Time Column Id*, gives the sequence number of the measurement columns. Thus, 1.0 is Tm1, 2.0 is Tm2, and so on.

The pattern Tm4 consistently produces the highest power across all sample sizes. Remember that Tm4 put the measurements at the beginning and the end, but none in the middle.

Patterns Tm3 and Tm5 are nearly tied for achieving the lowest powers. Tm3 put most of the measurements at the beginning of the study. Tm5 put most of the measurements during the middle of the study.

Note that Tm1, the equally-spaced times, is in the middle of the pack.

## Example 7 – Entering a Correlation Matrix

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

In this example the basic parameters are  $P1(0) = 0.72$ ,  $P2(0) = 0.72$ ,  $P1(1) = 0.75$ , and  $P2(1) = 0.55$ . The significance level is 0.05, the sample size ranges from 100 to 500, and R is 50%. There are to be 4 equally spaced time measurements. A correlation matrix (shown below) is available from a previous study. The missing input type will be set to Linear from 0 to 30% and the pairwise missing assumption is independent.

**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 Two Groups in a Repeated Measures Design (Binary 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>
Alternative Hypothesis .....	<b>Two-Sided</b>
Alpha .....	<b>0.05</b>
N (Subjects) .....	<b>100 to 500 by 100</b>
R (Group 1 Allocation %) .....	<b>50</b>
Measurement Time Input Type .....	<b>Equally Spaced Measurement Times</b>
M (Number of Measurements) .....	<b>4</b>
P1(0) and P2(0) Input Type .....	<b>P1(0) = P2(0)</b>
P1(0) and P2(0) .....	<b>0.72</b>
P1(1) Input Type .....	<b>P1(1)</b>
P1(1) .....	<b>0.75</b>
P2(1) .....	<b>0.55</b>
Pattern of p's Across Time .....	<b>Matrix on Spreadsheet</b>
Columns Containing the pjk'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>0.3</b>

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

**Output**

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

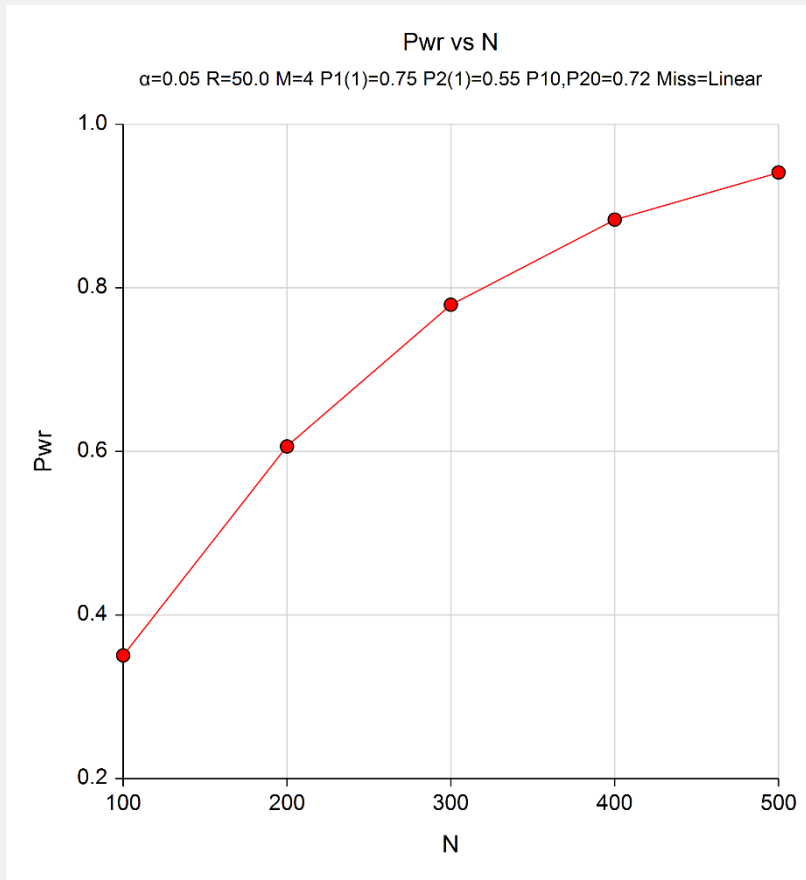
**Numeric Results for the Test of Slope-Difference in Binary Data using GEE**

Measurement Times: Equally spaced  
 Correlation: Matrix stored on spreadsheet in columns C1-C4  
 Missing Pattern: Range of missing proportions  
 Observant Proportions: Assume independence  
 Pk(t): Event probability of group k (1=treatment, 2=control) at time proportion t

Power	Group 1			M	P1(0)	P2(0)	P1(1)	P2(1)	Diff	Base Corr $\rho$	First Row of Corr Matrix	Alpha	Missing	Time
	N	R	Alloc Pct											
0.3506	100	50	4	0.720	0.720	0.750	0.550	0.200	n/a	$\rho_1(T_1)$	0.050	M1(T1)	T1	
0.6061	200	50	4	0.720	0.720	0.750	0.550	0.200	n/a	$\rho_1(T_1)$	0.050	M1(T1)	T1	
0.7794	300	50	4	0.720	0.720	0.750	0.550	0.200	n/a	$\rho_1(T_1)$	0.050	M1(T1)	T1	
0.8835	400	50	4	0.720	0.720	0.750	0.550	0.200	n/a	$\rho_1(T_1)$	0.050	M1(T1)	T1	
0.9412	500	50	4	0.720	0.720	0.750	0.550	0.200	n/a	$\rho_1(T_1)$	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  
 $\rho_1(T_1)$ : 1.000, 0.700, 0.490, 0.343



The standard reports are displayed.

## Example 8 – Entering an Observant Probabilities Matrix

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

In this example the basic parameters are  $P1(0) = 0.72$ ,  $P2(0) = 0.72$ ,  $P1(1) = 0.75$ , and  $P2(1) = 0.55$ . The significance level is 0.05, the sample size ranges from 100 to 500, and R is 50%. There are to be 4 equally spaced time measurements. The correlation pattern will be Linear Exponential Decay with a base correlation of 0.4, 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 Two Groups in a Repeated Measures Design (Binary Outcome)** procedure window. You may then make the appropriate entries as listed below, or open **Example 8** 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>
Alternative Hypothesis .....	<b>Two-Sided</b>
Alpha .....	<b>0.05</b>
N (Subjects) .....	<b>100 to 500 by 100</b>
R (Group 1 Allocation %) .....	<b>50</b>
Measurement Time Input Type .....	<b>Equally Spaced Measurement Times</b>
M (Number of Measurements) .....	<b>4</b>
P1(0) and P2(0) Input Type .....	<b>P1(0) = P2(0)</b>
P1(0) and P2(0) .....	<b>0.72</b>
P1(1) Input Type .....	<b>P1(1)</b>
P1(1) .....	<b>0.75</b>
P2(1) .....	<b>0.55</b>
Pattern of $\rho$ 's Across Time .....	<b>Linear Exponential Decay</b>
$\rho$ (Base Correlation) .....	<b>0.4</b>
Base Time Proportion .....	<b>0.10</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>	

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

## Output

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

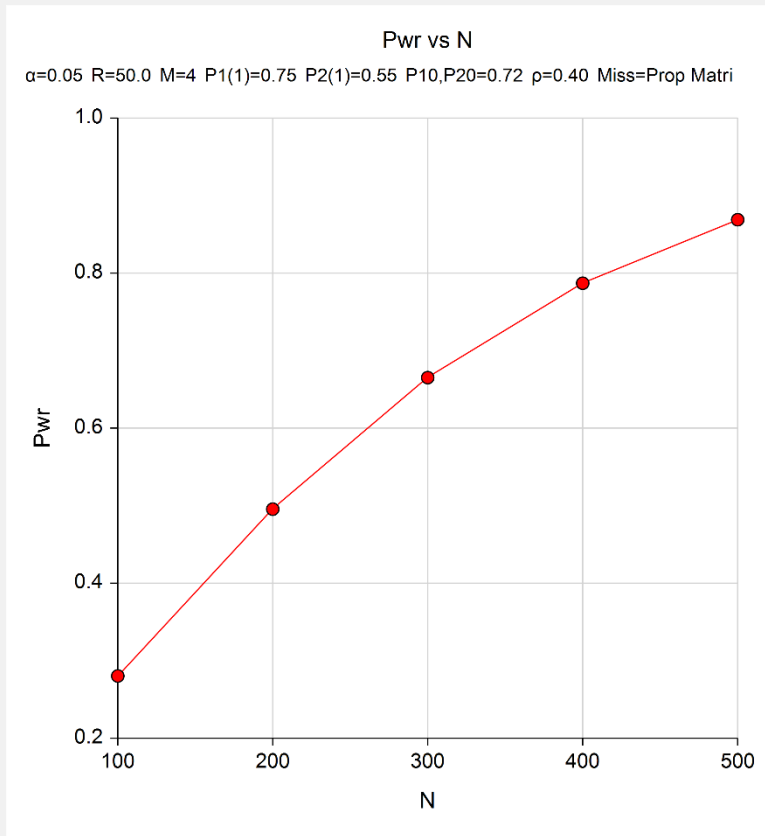
**Numeric Results for the Test of Slope-Difference in Binary Data using GEE**

Measurement Times: Equally spaced  
 Correlation: Linear exponential decay, with  $E_{max} = 4$  and Base Time Prop = 0.10  
 Missing Pattern: Range of missing proportions  
 Observant Proportions: N/A. Matrix of observant probabilities entered in columns C1-C4.  
 Pk(t): Event probability of group k (1=treatment, 2=control) at time proportion t

Power	Group 1			P1(0)	P2(0)	P1(1)	P2(1)	Diff	Base Corr $\rho$	First Row of Corr Matrix	Alpha	Missing	Time
	N	R	M										
0.2802	100	50	4	0.720	0.720	0.750	0.550	0.200	0.400	$\rho_1(T1)$	0.050	n/a	T1
0.4955	200	50	4	0.720	0.720	0.750	0.550	0.200	0.400	$\rho_1(T1)$	0.050	n/a	T1
0.6651	300	50	4	0.720	0.720	0.750	0.550	0.200	0.400	$\rho_1(T1)$	0.050	n/a	T1
0.7869	400	50	4	0.720	0.720	0.750	0.550	0.200	0.400	$\rho_1(T1)$	0.050	n/a	T1
0.8689	500	50	4	0.720	0.720	0.750	0.550	0.200	0.400	$\rho_1(T1)$	0.050	n/a	T1

**Footnotes**

T1: 0.00, 0.33, 0.67, 1.00  
 $\rho_1(T1)$ : 1.000, 0.196, 0.071, 0.026



The standard reports are displayed.