

Chapter 391

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)

Introduction

This module calculates the power for testing for differences among the time-averaged responses (TAD) of two or more groups from correlated **count** data that are analyzed using the GEE method. Such data can occur in two design types: clustered and longitudinal. This procedure emphasizes longitudinal designs.

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.

Time-averaged response analysis is often used when the outcome to be measured varies with time. For example, suppose that you want to compare three treatment groups based on the means of a certain outcome such as blood pressure. It is known that a person's blood pressure depends on several instantaneous factors such as amount of sleep, excitement level, mood, exercise, etc. If only a single measurement is taken from each patient, then the comparison of group values may be insensitive because of the large degree of variation in blood pressure levels within a patient. The precision of the experiment is increased by taking multiple measurements from each individual and comparing the time-averaged response among the groups. Care must be taken in the analysis because of the correlation that is introduced when several measurements are taken from the same individual. The correlation structure may take on several forms depending on the nature of the experiment and the subjects involved.

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 4.8.3, pages 132-134, and in Zhang and Ahn (2013).

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 $t_M = 1$.

Let y_{kij} be the count response of subject i in group k at time t_j . The count is modeled by the Poisson distribution

$$f(y_{kij}) = \frac{e^{-\mu_{kij}} \mu_{kij}^{y_{kij}}}{y_{kij}!}$$

The mean of y_{kij} is modeled by

$$\log(\mu_{kij}) = \beta_k$$

while y_{kij} is expressed as

$$y_{kij} = \mu_{kij} + \epsilon_{kij}$$

where

y_{kij} is the j^{th} response from subject i in group k ,

μ_{kij} is group-specific treatment effect ($k = 1, \dots, G$),

ϵ_{kij} is a zero-mean error term with variance also given by μ_{kij} .

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

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

$$T = \frac{C'b}{\sqrt{\frac{1}{N} (C'WA_N^{-1}VA_N^{-1}WC)}}$$

where

b_k is the GEE estimate of β_k , $k = 1, \dots, G$

r_k is the proportion of subjects in group k ,

C is the vector of contrast coefficients

N is the total sample size

W is a diagonal matrix of elements $1/\sqrt{r_k}$

A_N is a diagonal matrix of elements $1/n_k \sum_{i=1}^{n_k} \sum_{j=1}^M \exp(b_k)$

V is a diagonal matrix of elements $1/n_k \sum_{i=1}^{n_k} [\sum_{j=1}^M \hat{\epsilon}_{kij}]^2$, where $\hat{\epsilon}_{kij} = y_{kij} - \mu_{kij}(\underline{b})$

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

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 correlations are equal, and all diagonal elements are equal to one. That is

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

where ρ is the baseline correlation.

Banded(1)

A Banded(1) (banded order 1) correlation model assumes that diagonal elements are one, correlations for observations one time period apart are equal to ρ , and correlations 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 ρ is the baseline correlation.

Banded(2)

A Banded(2) (banded order 2) correlation model assumes that diagonal elements are one, correlations for observations one time period or two periods apart are equal to ρ , and correlations 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 ρ is the baseline correlation.

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)

AR1 (Traditional)

This version of AR1 (autoregressive order 1) correlation model assumes that correlations t time periods apart are equal to ρ^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 ρ 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 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 ρ is the baseline correlation. Note that in this pattern, the value of ρ 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 ρ 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 ρ 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 TAD of Multiple Groups in a Repeated Measures Design (Count 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 κ_j . The proportion non-missing at a particular time point is $\phi_j = 1 - \kappa_j$. We will refer to ϕ_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 TAD of Multiple Groups in a Repeated Measures Design (Count 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. The formula for the sample size is summarized here.

$$N = \frac{D \left(z_{1-\frac{\alpha}{2}} + z_{1-\gamma} \right)^2}{E^2}$$

where

γ 1 – power

α significance level

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

D is $C'WA^{-1}VA^{-1}WC$

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)

E	is $C' \underline{\beta}$
$\underline{\beta}$	is a vector of anticipated GEE regression coefficients, β_1, \dots, β_G
C	is the vector of contrast coefficients
N	is the total sample size
W	is a diagonal matrix of elements $1/\sqrt{r_k}$
A	is a diagonal matrix of elements $\bar{M} \exp(\beta_k)$
V	is a diagonal matrix of elements $h \exp(\beta_k)$
\bar{M}	is $\sum_{k=1}^G \phi_k$
h	is $\sum_{j=1}^M \sum_{j'=1}^M \phi_{jj'} \rho_{jj'}$
r_k	is the proportion of subjects in group k .

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.

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 means when in fact the means 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 TAD of Multiple Groups in a Repeated Measures Design (Count 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 means when in fact the means are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Sample Size and Group Allocation

G (Number of Groups)

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

Range

$G \geq 2$.

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 .

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)**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.

 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

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

 n_1, n_2, \dots, n_G (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 n_i '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.

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)**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.

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.

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)

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

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.

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)**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'.

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.

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)**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.

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 - $\mu_1, \mu_2, \dots, \mu_G$ (Mean Time-Averaged Responses) and Contrast
 μ_i 's Input Type

Specify how you want to enter $\mu_1, \mu_2, \dots, \mu_G$, the mean time-averaged response per group for each of the G groups.

Options

- **$\mu_1, \mu_2, \dots, \mu_G$**
Specify the values of the mean time-averaged response for each group. The differences among the logs of these values represent the effect size that you want to detect.
- **Columns Containing Sets of μ 's**
Select one or more columns of the spreadsheet that each contain a set of μ_i 's going down the column. Each column is analyzed separately.

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome) **$\mu_1, \mu_2, \dots, \mu_G$**

Enter the values of the mean time-averaged response for each of the G groups under the alternative hypothesis. The effect size that the study will detect is a function of the differences among the logs of these values.

A time-averaged response for a particular individual is the average of all of the repeated measurements of that individual. This value is the anticipated mean of all subjects in a particular group.

Range

Each μ_i should be numeric and at least one of the values must be different from the rest. Since logs will be taken, all of these values must be greater than zero.

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.

Columns Containing Sets of μ_i 's

Enter one or more spreadsheet columns containing vertical lists of $\mu_1, \mu_2, \dots, \mu_G$.

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 μ_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.

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)**Contrast Input Type**

Specify how you want to enter the contrast coefficients.

Contrast Definition

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

$$-3 \log(\mu_1) - \log(\mu_2) + \log(\mu_3) + 3 \log(\mu_4)$$

The point to realize is that these weights (the coefficients) are used to calculate a specific weighted average of the log means which is to be compared against zero using a standard z (or t) test. Hence, G coefficients (weights) must be defined.

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

Options

- **List of Contrast Coefficients**

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

- **Multiple Lists of Contrast Coefficients**

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

- **First Group vs Rest**

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

- **Last Group vs Rest**

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

- **Linear Trend**

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

- **Maximum Power**

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

(List of) Contrast Coefficients

Specify how you want to enter the contrast coefficients.

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

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)**Definition**

A contrast is a weighted average of the logs of the G group means in which the weights (coefficients) sum to zero. Each successive coefficient is applied to the corresponding group mean. For example, suppose $G = 3$ and the first group is a control group. A contrast that might be of interest is "2 -1 -1". This contrast is interpreted as " $2 \text{Log}(\mu_1) - \text{Log}(\mu_2) - \text{Log}(\mu_3)$ ". The coefficients must sum to zero.

Incomplete List

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

Examples

-1 -1 2

-2 -1 0 1 2

-3 -1 1 3

-3 1 1 1

Multiple Lists of (Contrast) Coefficients

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

Press the Spreadsheet icon (directly to the right) to select the columns and then enter the values.

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

Definition

A contrast is a weighted average of the logs of the G group means in which the weights (coefficients) sum to zero. Each successive coefficient is applied to the corresponding group mean. For example, suppose $G = 3$ and the first group is a control group. A contrast that might be of interest is "2 -1 -1". This contrast is interpreted as " $2 \text{Log}(\mu_1) - \text{Log}(\mu_2) - \text{Log}(\mu_3)$ ". The coefficients must sum to zero.

The attained power is highly related to the specific contrast that is entered.

Incomplete List

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

Examples

C1 - C3

C1, C2, C3

Con1, Con2, Con3

Spreadsheet Example

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

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

Effect Size - ρ (Correlation Between Measurements)

Pattern of ρ 's Across Time (or Space)

Specify the type of pattern for the ρ '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 ρ 's Equal)**

A single value of ρ is used for all correlations. This pattern is commonly used for cluster-randomized trials in which case m is the cluster size and ρ 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 ρ '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 ρ '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 ρ 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 ρ . The basic formula is $\rho_{jk} = \rho^{|j-k|}$. Thus the pattern is ρ, ρ^2, ρ^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 TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)

- **AR1 (Proportional)**

A single value of ρ 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 ρ . 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 ρ 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 ρ 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 ρ . 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 ρ 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 ρ 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 TAD of Multiple Groups in a Repeated Measures Design (Count Outcome) **ρ Pattern = Compound Symmetry, Banded (1 Time Period), Banded (2 Time Periods), AR(1)** **ρ (Base Correlation)**

This is the base correlation, ρ , 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

 ρ Pattern = Damped Exponential **ρ (Base Correlation)**

This is the base correlation, ρ , 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 $\rho_{jk} = \rho^{|j-k| \wedge \text{Dexp}}$.

Range

The valid range is $\text{Dexp} > 0$. The typical range is 1 to 3.

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome) **ρ Pattern = Linear Exponential Decay** **ρ (Base Correlation)**

This is the base correlation, ρ , 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 TAD of Multiple Groups in a Repeated Measures Design (Count Outcome) **ρ Pattern = Matrix on Spreadsheet****Columns Containing the ρ '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 TAD of Multiple Groups in a Repeated Measures Design (Count 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:

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

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count 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 TAD of Multiple Groups in a Repeated Measures Design (Count 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.

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

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count 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

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

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)**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 comparing three heart-rate medications: a standard drug and two experimental drugs. The experimental drugs are expected to have about the same impact on heart rate. Each subject will receive four applications of only one drug, two days apart. The researchers want a sample size large enough to detect a time-averaged response difference of 5 between the highest and lowest time-averaged heart rates. They will use the means of 65, 60, and 60 to represent this difference. The contrast coefficients that they will use are -2, 1, 1.

Similar studies have shown an autocorrelation between adjacent measurements on the same individual of 0.7, so the researchers 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 split 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 procedure window. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Power.....	0.90
Alpha.....	0.05
G (Number of Groups).....	3
Group Allocation Input Type	Equal (n1 = n2 = ... = nG = n)
Measurement Time Input Type.....	Equally Spaced Measurement Times
M (Number of Measurement Times).....	4
μ 's Input Type	$\mu_1, \mu_2, \dots, \mu_G$
$\mu_1, \mu_2, \dots, \mu_G$	65 60 60
Contrast Input Type	List of Contrast Coefficients
Contrast Coefficients	-2 1 1
Pattern of ρ 's Across Time.....	AR1 (Traditional)
ρ (Base Correlation)	0.6 0.7 0.8
Missing Input Type.....	Linear (Steady Change)
Pairwise Missing Pattern	Independent (Ind)
First Missing Proportion (Ind)	0
Last Missing Proportion (Ind).....	0.4

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)

Annotated Output

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

Numeric Results for a Time-Averaged Response Test of Count Data using GEE

Contrast: List of Coefficients
 Measurement Times: Equally spaced
 Correlation: AR(1): $\rho(j,k) = \rho^{|j-k|}$
 Missing Pattern: Range of missing proportions
 Observant Proportions: Assume independence
 Number of Groups: 3

	Group Alloc Prop Set	Meas Times	Time Set	Time-Ave Resp Set	Cont Coef Set	Lin Comb of Ci and μ_i	Base Corr ρ	First Row of Corr Matrix	Alpha	Miss Set
Power	N	ri	M	μ_i	Ci	$ Ci'\mu_i $	ρ			
0.9063	78	ri(1)	4	T(1)	$\mu_i(1)$	Con(1)	10.00	0.6000	0.050	Ms1(1)
0.9028	87	ri(1)	4	T(1)	$\mu_i(1)$	Con(1)	10.00	0.7000	0.050	Ms1(1)
0.9052	99	ri(1)	4	T(1)	$\mu_i(1)$	Con(1)	10.00	0.8000	0.050	Ms1(1)

Set(Set Number): Values

ri(1): 0.333, 0.333, 0.333
 T(1): 0.00, 0.33, 0.67, 1.00
 $\mu_i(1)$: 65.00, 60.00, 60.00
 Con(1): -2.00, 1.00, 1.00
 $\rho_1(1)$: 1.0000, 0.6000, 0.3600, 0.2160
 $\rho_2(1)$: 1.0000, 0.7000, 0.4900, 0.3430
 $\rho_3(1)$: 1.0000, 0.8000, 0.6400, 0.5120
 Ms1(1): 0.00, 0.13, 0.27, 0.40

References

Ahn, C., Heo, M., and Zhang, S. 2015. Sample Size Calculations for Clustered and Longitudinal Outcomes in Clinical Research. CRC Press. New York.

Report Definitions

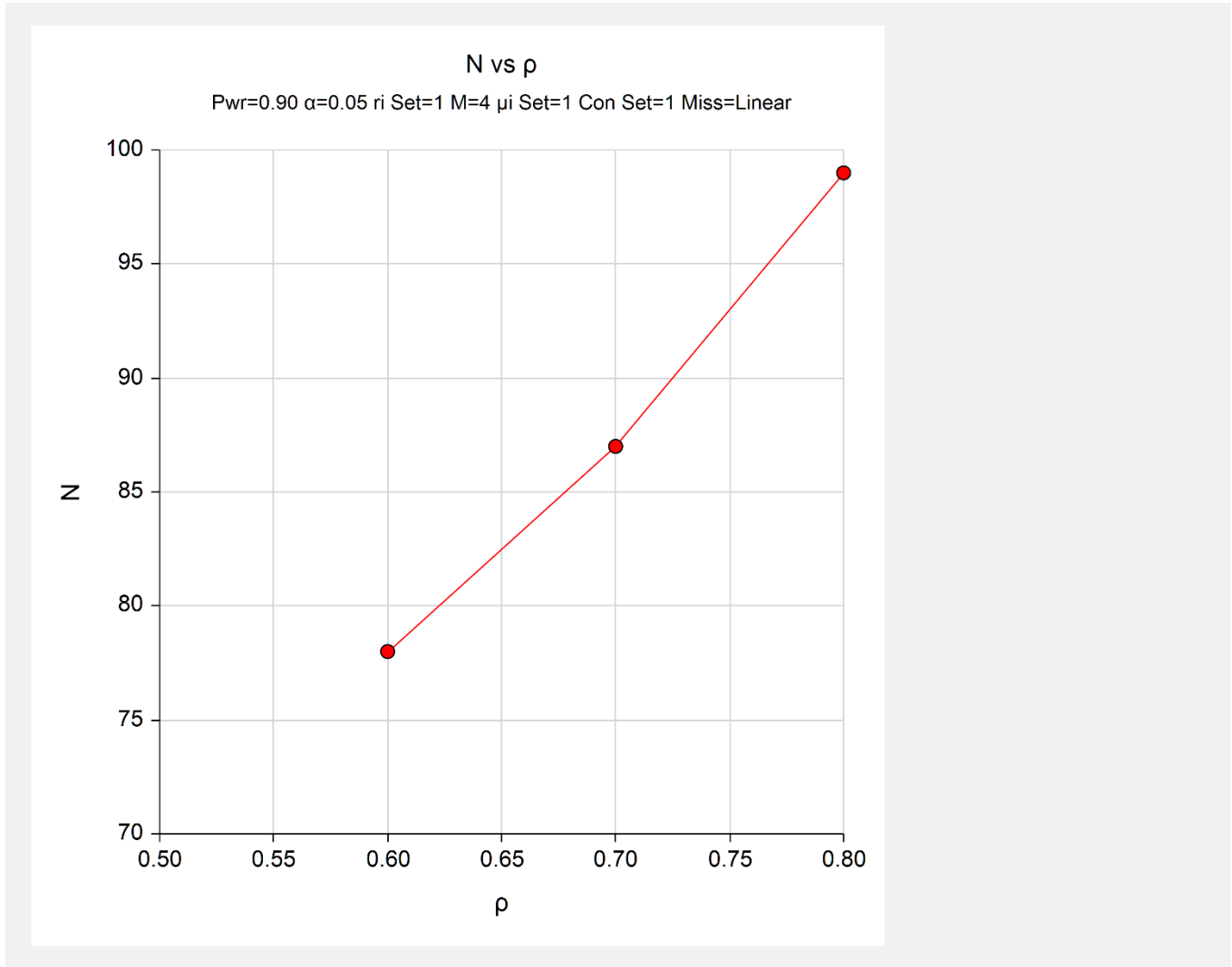
Power is the probability of rejecting a false null hypothesis.
 N is the total number of subjects in the study.
 'Group Alloc Prop Set ri' gives the name of the set containing group allocation proportions.
 'Meas Times M' is the number of time points at which each subject is scheduled to be measured.
 'Time Set' gives the name of the set containing the time proportions. The time values represent the proportion of the total study time that has elapsed just before the measurement.
 'Time-Ave Resp Set μ_i ' gives the name of the set containing the time-averaged responses for each group.
 'Cont Coef Set Ci' gives the name of the set containing the contrast coefficients that are combined with the group response probabilities.
 'Lin Comb of Ci and μ_i $|Ci'\mu_i|$ ' linear combination of the contrast coefficients and the group response probabilities.
 ρ is the base correlation between two responses on the same subject. It may be transformed based on the correlation pattern.
 'First Row of Corr Matrix' presents the top row of the correlation matrix.
 Alpha is the significance level of the test. The probability of rejecting the null hypothesis when the null hypothesis is true.
 'Miss Set' gives the name of the set containing the missing value proportions across time.

Summary Statements

A total sample of 78 subjects, divided among 3 groups and each measured 4 times, achieve a power of 0.9063 when using a normal test from a GEE analysis to determine whether a contrast of the group time-averaged responses differs significantly from zero at a significance level of 0.050. The group sample sizes are 26, 26, 26. The group averages under the alternative hypothesis are 65.00, 60.00, 60.00. The contrast coefficients are -2.00, 1.00, 1.00. The residual standard deviation is anticipated to be Con1. 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.

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)



This shows 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 10 to 40 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 procedure window. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alpha.....	0.05
G (Number of Groups).....	3
Group Allocation Input Type	Equal (n1 = n2 = ... = nG = n)
n (Sample Size Per Group)	10 to 40 by 10
Measurement Time Input Type.....	Equally Spaced Measurement Times
M (Number of Measurement Times).....	4
μi's Input Type	μ1, μ2, ..., μG
μ1, μ2, ..., μG.....	65 60 60
Contrast Input Type	List of Contrast Coefficients
Contrast Coefficients	-2 1 1
Pattern of ρ's Across Time.....	AR1 (Traditional)
ρ (Base Correlation)	0.7
Missing Input Type.....	Linear (Steady Change)
Pairwise Missing Pattern	Independent (Ind)
First Missing Proportion (Ind)	0
Last Missing Proportion (Ind).....	0.4

Output

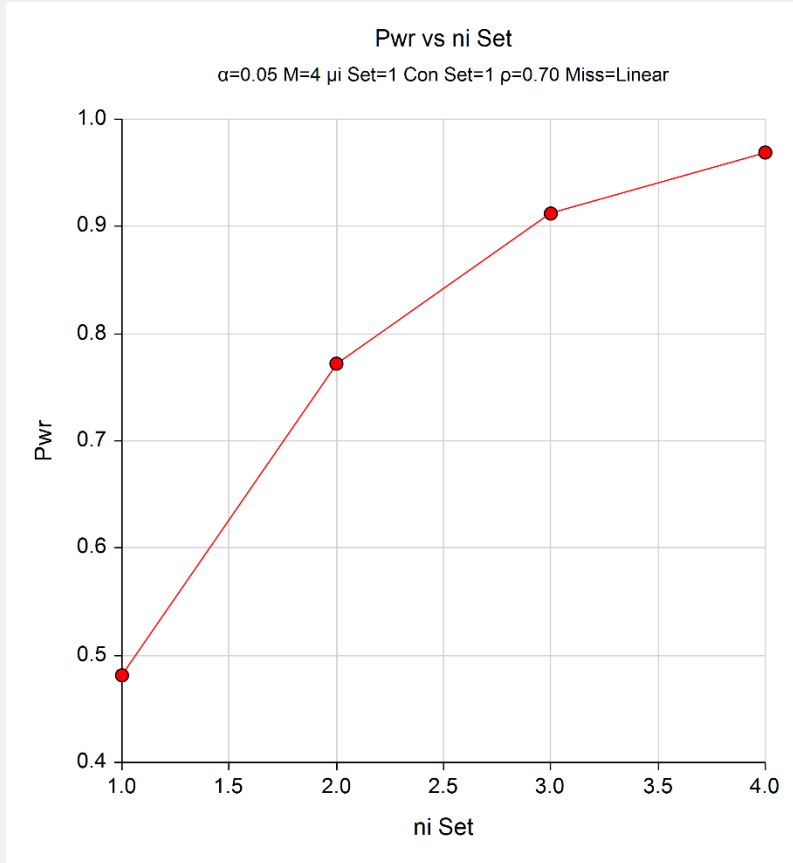
Click the Calculate button to perform the calculations and generate the following output.

Numeric Results for a Time-Averaged Response Test of Count Data using GEE											
Contrast:	List of Coefficients										
Measurement Times:	Equally spaced										
Correlation:	AR(1): $\rho(j,k) = \rho^{ j-k }$										
Missing Pattern:	Range of missing proportions										
Observant Proportions:	Assume independence										
Number of Groups:	3										
Power	Group Sample Sizes N	Meas Set ni	Time M	Time Set	Time-Ave Resp Set μi	Cont Coef Set Ci	Lin Comb of Ci and μi Ci'μi 	Base Corr ρ	First Row of Corr Matrix	Alpha	Miss Set
0.4812	30	ni(1)	4	T(1)	μi(1)	Con(1)	10.00	0.7000	ρ1(1)	0.050	Ms1(1)
0.7720	60	ni(2)	4	T(1)	μi(1)	Con(1)	10.00	0.7000	ρ1(1)	0.050	Ms1(1)
0.9120	90	ni(3)	4	T(1)	μi(1)	Con(1)	10.00	0.7000	ρ1(1)	0.050	Ms1(1)
0.9690	120	ni(4)	4	T(1)	μi(1)	Con(1)	10.00	0.7000	ρ1(1)	0.050	Ms1(1)

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)

Set(Set Number): Values

ni(1): 10, 10, 10
 ni(2): 20, 20, 20
 ni(3): 30, 30, 30
 ni(4): 40, 40, 40
 T(1): 0.00, 0.33, 0.67, 1.00
 μ i(1): 65.00, 60.00, 60.00
 Con(1): -2.00, 1.00, 1.00
 ρ 1(1): 1.0000, 0.7000, 0.4900, 0.3430
 Ms1(1): 0.00, 0.13, 0.27, 0.40



Note that the horizontal axis is scaled according to the *ni* set index. Hence, 1 is for *ni*(1), 2 is for *ni*(2), and so on. The same is true for the μ 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 various sets of mean time-averaged responses on the sample size. To do this, they will compare the four sets of means 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 procedure window. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Power.....	0.90
Alpha.....	0.05
G (Number of Groups).....	3
Group Allocation Input Type	Equal (n1 = n2 = ... = nG = n)
Measurement Time Input Type.....	Equally Spaced Measurement Times
M (Number of Measurement Times).....	4
μ 's Input Type	Columns containing sets of μ's
Columns Containing Sets of μ 's.....	C1-C4
<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>	
Contrast Input Type	List of Contrast Coefficients
Contrast Coefficients	-2 1 1
Pattern of ρ 's Across Time.....	AR1 (Traditional)
ρ (Base Correlation)	0.7
Missing Input Type.....	Linear (Steady Change)
Pairwise Missing Pattern	Independent (Ind)
First Missing Proportion (Ind)	0
Last Missing Proportion (Ind).....	0.4

Output

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

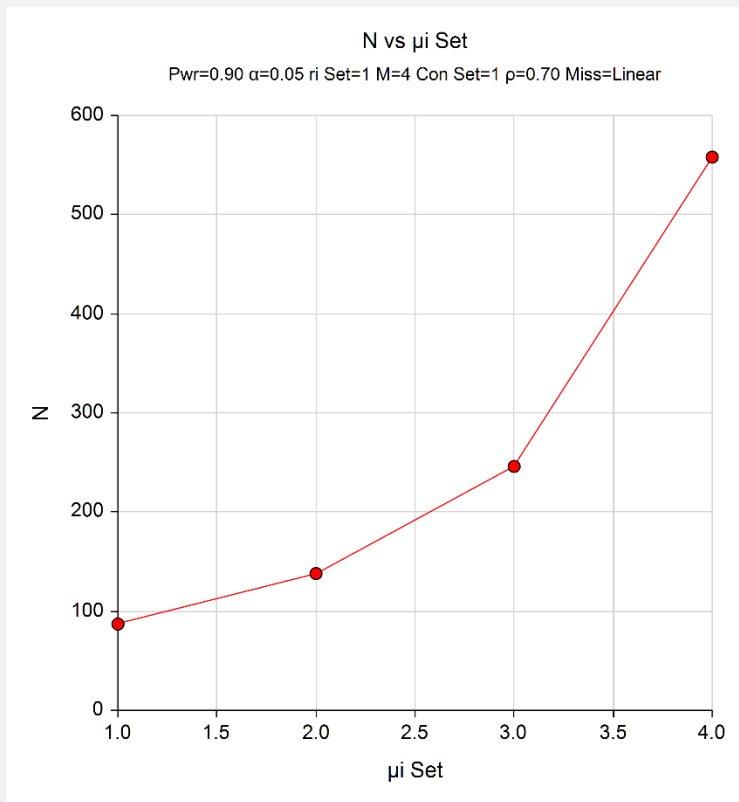
Numeric Results for a Time-Averaged Response Test of Count Data using GEE

Contrast: List of Coefficients
 Measurement Times: Equally spaced
 Correlation: AR(1): $\rho(j,k) = \rho^{|j-k|}$
 Missing Pattern: Range of missing proportions
 Observant Proportions: Assume independence
 Number of Groups: 3

Power	Group Alloc Prop Set	Meas Times	Time Set	Time-Ave Resp Set	Cont Coef Set	Lin Comb of Ci and μ_i	Base Corr ρ	First Row of Corr Matrix	Alpha	Miss Set	
	N	ri	M	μ_i	Ci						
0.9028	87	ri(1)	4	T(1)	C1(1)	Con(1)	10.00	0.7000	$\rho_1(1)$	0.050	Ms1(1)
0.9041	138	ri(1)	4	T(1)	C2(2)	Con(1)	8.00	0.7000	$\rho_1(1)$	0.050	Ms1(1)
0.9019	246	ri(1)	4	T(1)	C3(3)	Con(1)	6.00	0.7000	$\rho_1(1)$	0.050	Ms1(1)
0.9012	558	ri(1)	4	T(1)	C4(4)	Con(1)	4.00	0.7000	$\rho_1(1)$	0.050	Ms1(1)

Set(Set Number): Values

ri(1): 0.333, 0.333, 0.333
 T(1): 0.00, 0.33, 0.67, 1.00
 C1(1): 65.00, 60.00, 60.00
 C2(2): 65.00, 61.00, 61.00
 C3(3): 65.00, 62.00, 62.00
 C4(4): 65.00, 63.00, 63.00
 Con(1): -2.00, 1.00, 1.00
 $\rho_1(1)$: 1.0000, 0.7000, 0.4900, 0.3430
 Ms1(1): 0.00, 0.13, 0.27, 0.40



Note that the horizontal axis gives the index of the means footnotes. Thus, '1.0' is column C1, '2.0' is column C2, and so on.

Example 4 – Validation of Sample Size Calculation using Hand Calculations

Ahn, Heo, and Zhang (2015) do not present a numerical example for this procedure. We will calculate a result by hand using their formulas and compare this with the program output to validate this procedure.

The sample size formula for only two groups is

$$n = \frac{\sigma_2^2 \left(z_{1-\frac{\alpha}{h}} + z_{1-\gamma} \right)^2}{\beta_2^2}$$

where

$$\sigma_2^2 = \frac{\bar{\mu} \sum_{j=1}^m \sum_{j'=1}^m \phi_{jj'} \rho_{jj'}}{\left(\sum_{j=1}^m \phi_j \right)^2 \sigma_r^2 \mu_1 \mu_2}$$

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

$$\bar{\mu} = \bar{r}\mu_1 + (1 - \bar{r})\mu_2$$

Using $m = 3$, $\mu_1 = 2$, $\mu_2 = 1$, compound symmetry, $\rho = 0.6$, constant missing = 0.1, $\alpha = 0.05$, and power = 0.9, we obtain

$$\bar{\mu} = 1.5$$

$$\sum_{j=1}^m \phi_j = 2.7$$

$$\sum_{j=1}^m \sum_{j'=1}^m \phi_{jj'} \rho_{jj'} = 5.94$$

$$\sigma_2^2 = \frac{(1.5)(5.94)}{(2.7)^2(0.25)^2} = 2.4444$$

$$n = \text{ceiling} \left(\frac{2.4444(1.96 + 1.28155)^2}{0.69314^2} \right) = \text{ceiling}(53.46) = 54.$$

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Power.....	0.90
Alpha.....	0.05
G (Number of Groups).....	2
Group Allocation Input Type	Equal (n1 = n2 = ... = nG = n)
Measurement Time Input Type.....	Equally Spaced Measurement Times
M (Number of Measurement Times).....	3
μi's Input Type	μ1, μ2, ..., μG
μ1, μ2, ..., μG.....	2 1
Contrast Input Type	List of Contrast Coefficients
Contrast Coefficients	-1 1
Pattern of ρ's Across Time.....	Compound Symmetry (All ρ's Equal)
ρ (Base Correlation)	0.6
Missing Input Type.....	Constant
Constant Missing Proportion	0.10

Output

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

Numeric Results for a Time-Averaged Response Test of Count Data using GEE

Contrast: List of Coefficients
 Measurement Times: Equally Spaced
 Correlation: Compound symmetry (all ρ's equal)
 Missing Pattern: Constant (All missing proportions are equal)
 Number of Groups: 2

	Group Alloc Prop Set	Meas Times M	Time Set	Time-Ave Resp Set μi	Cont Coef Set Ci	Lin Comb of Ci and μi Ci'μi	Base Corr ρ	First Row of Corr Matrix ρ1(1)	Alpha	Prop Miss
Power	N	ri	M	Time Set	μi	Cont Coef Set Ci	Base Corr ρ	First Row of Corr Matrix ρ1(1)	Alpha	Prop Miss
0.9028	54	ri(1)	3	T(1)	μi(1)	Con(1)	1.00	0.6000	0.050	0.10

Set(Set Number): Values
 ri(1): 0.500, 0.500
 T(1): 0.00, 0.50, 1.00
 μi(1): 2.00, 1.00
 Con(1): -1.00, 1.00
 ρ1(1): 1.0000, 0.6000, 0.6000

The sample size of 54 matches our calculated result exactly.

Example 5 – 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.

This example uses $G = 4$, $m = 6$, and $\alpha = 0.05$. The correlation pattern is Linear Exponential Decay with a base correlation of 0.5, Base Time Proportion of 0.20, and E_{\max} set to 4. The missing input type is set to Linear from 0 to 30% and the pairwise missing assumption is independent. The group means are 1, 1, 1.1, 1.5. A linear trend contrast will be used. The per group sample sizes range from 20 to 80.

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 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>
Design Tab	
Solve For	Power
Alpha.....	0.05
G (Number of Groups).....	4
Group Allocation Input Type	Equal (n1 = n2 = ... = nG = n)
n (Sample Size Per Group)	20 40 60 80
Measurement Time Input Type.....	Columns of Measurement Time Proportions
Columns of Time Proportions	Tm1-Tm5
<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>	
μ 's Input Type	$\mu_1, \mu_2, \dots, \mu_G$
$\mu_1, \mu_2, \dots, \mu_G$	1 1 1.1 1.5
Contrast Input Type	Linear Trend
Pattern of ρ 's Across Time.....	Linear Exponential Decay
ρ (Base Correlation)	0.5
Base Time Proportion	0.2
E_{\max} (Max Decay Exponent).....	4

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)

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

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 a Time-Averaged Response Test of Count Data using GEE

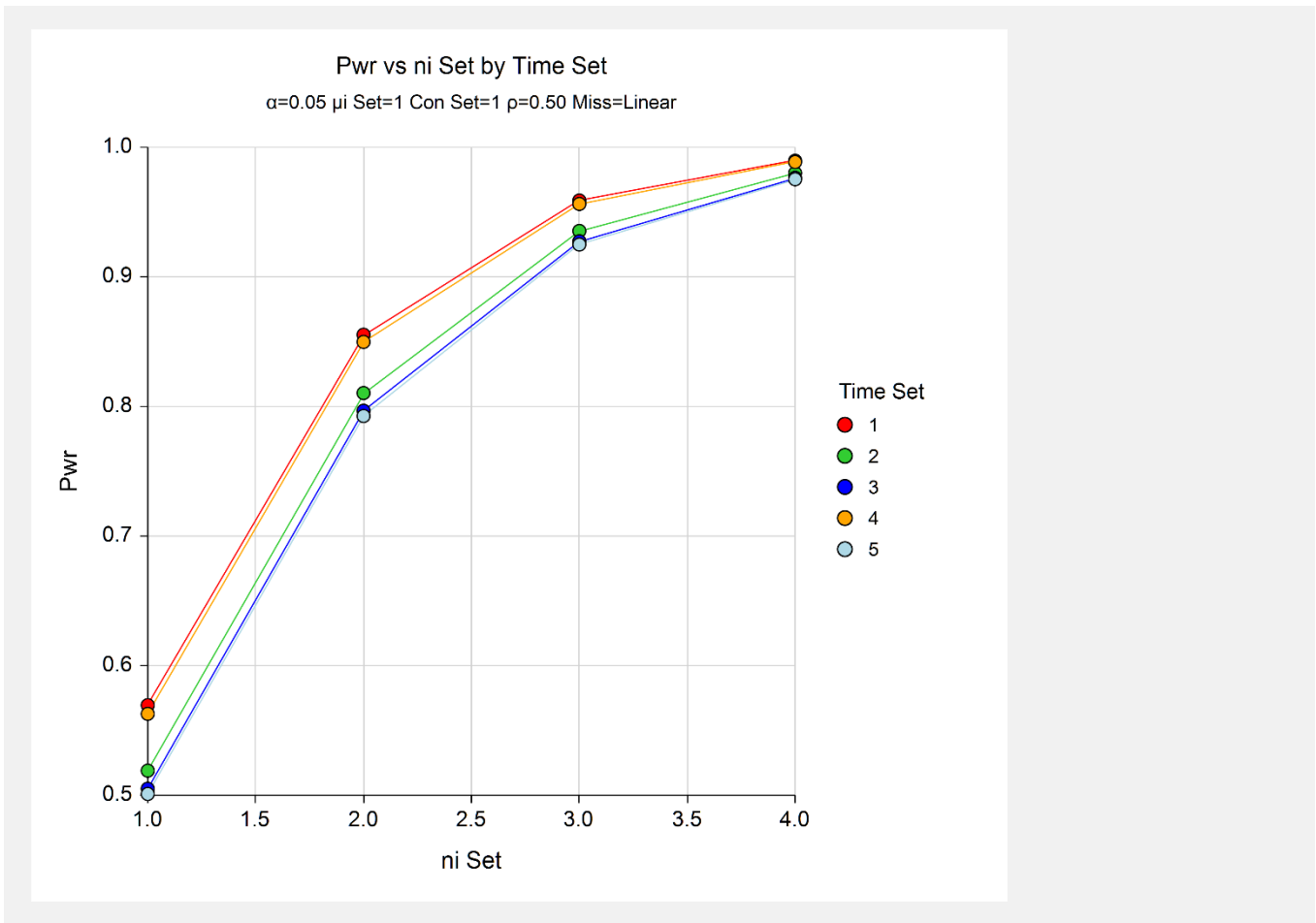
Contrast: Linear Trend
 Measurement Times: Lists in spreadsheet columns: {TM1-TM5}
 Correlation: Linear exponential decay, with Emax = 4 and Base Time Prop = 0.2
 Missing Pattern: Range of missing proportions
 Observant Proportions: Assume independence
 Number of Groups: 4

Power	Group Sample Sizes N ni	Meas Times M	Time Set Time Set	Time-Ave Resp Set μ_i	Cont Coef Set Ci	Lin Comb of Ci and μ_i Ci μ_i	Base Corr ρ	First Row of Corr Matrix Alpha	Miss Set		
0.5696	80	ni(1)	6	Tm1(1)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm1)$	0.050	Ms1(Tm1)
0.5190	80	ni(1)	6	Tm2(2)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm2)$	0.050	Ms1(Tm2)
0.5051	80	ni(1)	6	Tm3(3)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm3)$	0.050	Ms1(Tm3)
0.5628	80	ni(1)	6	Tm4(4)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm4)$	0.050	Ms1(Tm4)
0.5010	80	ni(1)	6	Tm5(5)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm5)$	0.050	Ms1(Tm5)
0.8553	160	ni(2)	6	Tm1(1)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm1)$	0.050	Ms1(Tm1)
0.8104	160	ni(2)	6	Tm2(2)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm2)$	0.050	Ms1(Tm2)
0.7967	160	ni(2)	6	Tm3(3)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm3)$	0.050	Ms1(Tm3)
0.8498	160	ni(2)	6	Tm4(4)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm4)$	0.050	Ms1(Tm4)
0.7926	160	ni(2)	6	Tm5(5)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm5)$	0.050	Ms1(Tm5)
0.9589	240	ni(3)	6	Tm1(1)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm1)$	0.050	Ms1(Tm1)
0.9354	240	ni(3)	6	Tm2(2)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm2)$	0.050	Ms1(Tm2)
0.9274	240	ni(3)	6	Tm3(3)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm3)$	0.050	Ms1(Tm3)
0.9563	240	ni(3)	6	Tm4(4)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm4)$	0.050	Ms1(Tm4)
0.9250	240	ni(3)	6	Tm5(5)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm5)$	0.050	Ms1(Tm5)
0.9896	320	ni(4)	6	Tm1(1)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm1)$	0.050	Ms1(Tm1)
0.9801	320	ni(4)	6	Tm2(2)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm2)$	0.050	Ms1(Tm2)
0.9765	320	ni(4)	6	Tm3(3)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm3)$	0.050	Ms1(Tm3)
0.9886	320	ni(4)	6	Tm4(4)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm4)$	0.050	Ms1(Tm4)
0.9753	320	ni(4)	6	Tm5(5)	$\mu_i(1)$	Con(1)	0.80	0.5000	$\rho_1(Tm5)$	0.050	Ms1(Tm5)

Set(Set Number): Values

ni(1): 20, 20, 20, 20
 ni(2): 40, 40, 40, 40
 ni(3): 60, 60, 60, 60
 ni(4): 80, 80, 80, 80
 Tm1(1): 0.00, 0.20, 0.40, 0.60, 0.80, 1.00
 Tm2(2): 0.00, 0.60, 0.70, 0.80, 0.90, 1.00
 Tm3(3): 0.00, 0.10, 0.20, 0.30, 0.40, 1.00
 Tm4(4): 0.00, 0.10, 0.20, 0.80, 0.90, 1.00
 Tm5(5): 0.00, 0.45, 0.50, 0.55, 0.60, 1.00
 $\mu_i(1)$: 1.00, 1.00, 1.10, 1.50
 Con(1): -1.50, -0.50, 0.50, 1.50
 $\rho_1(Tm1)$: 1.0000, 0.5000, 0.2973, 0.1768, 0.1051, 0.0625
 $\rho_1(Tm2)$: 1.0000, 0.1768, 0.1363, 0.1051, 0.0811, 0.0625
 $\rho_1(Tm3)$: 1.0000, 0.6484, 0.5000, 0.3856, 0.2973, 0.0625
 $\rho_1(Tm4)$: 1.0000, 0.6484, 0.5000, 0.1051, 0.0811, 0.0625
 $\rho_1(Tm5)$: 1.0000, 0.2611, 0.2293, 0.2013, 0.1768, 0.0625
 Ms1(Tm1): 0.00, 0.06, 0.12, 0.18, 0.24, 0.30
 Ms1(Tm2): 0.00, 0.18, 0.21, 0.24, 0.27, 0.30
 Ms1(Tm3): 0.00, 0.03, 0.06, 0.09, 0.12, 0.30
 Ms1(Tm4): 0.00, 0.03, 0.06, 0.24, 0.27, 0.30
 Ms1(Tm5): 0.00, 0.14, 0.15, 0.17, 0.18, 0.30

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)



The legend, *Time Set*, gives the set number of the measurement columns. Thus, 1 is Tm1, 2 is Tm2, and so on. The pattern Tm1 produces the highest power. Remember that Tm1 spreads the measurements evenly through time.

Patterns Tm3 and Tm5 are nearly tied for achieving the low powers, followed closely by Tm2.

Example 6 – Comparing Contrast Coefficients

This example will show how important the contrast coefficients are in achieving a certain power.

This example uses $G = 4$, $M = 6$, and $\alpha = 0.05$. The missing input type will be set to Linear from 0 to 30% and the pairwise missing assumption is independent. Group means are 1, 1, 1.1, 1.5. The per group sample sizes range from 20 to 80. 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 Table of Contrasts shows the contrasts that will be compared.

Table of Contrasts

C1	C2	C3	C4
-3	-3	1	-1
1	-1	1	-2
1	1	1	2
1	3	-3	1

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may then make the appropriate entries as listed below, or open **Example 6** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alpha.....	0.05
G (Number of Groups).....	4
Group Allocation Input Type	Equal (n1 = n2 = ... = nG = n)
n (Sample Size Per Group)	20 40 60 80
Measurement Time Input Type.....	Equally Spaced Measurement Times
M (Number of Measurement Times).....	6
μ 's Input Type	$\mu_1, \mu_2, \dots, \mu_G$
$\mu_1, \mu_2, \dots, \mu_G$	1 1 1.1 1.5
Contrast Input Type	Multiple Lists of Contrast Coefficients
Multiple Lists of Coefficients	C1-C4
Pattern of ρ 's Across Time.....	Linear Exponential Decay
ρ (Base Correlation)	0.5
Base Time Proportion	0.2
Emax (Max Decay Exponent).....	4
Missing Input Type.....	Linear (Steady Change)
Pairwise Missing Pattern	Independent (Ind)
First Missing Proportion (Ind)	0
Last Missing Proportion (Ind).....	0.3

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)

Output

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

Numeric Results for a Time-Averaged Response Test of Count Data using GEE

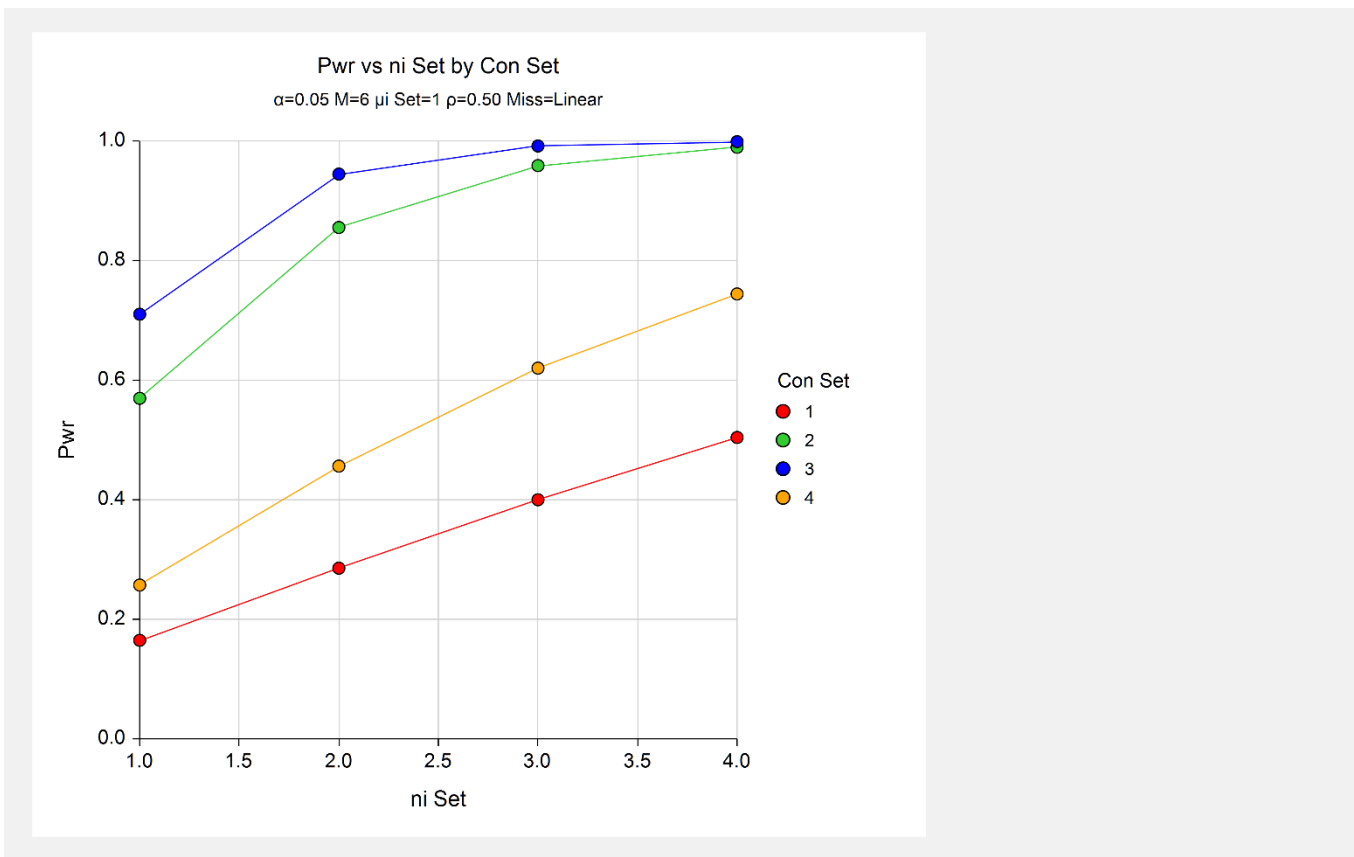
Contrast: Multiple Lists of Coefficients: {C1-C4}
 Measurement Times: Equally spaced
 Correlation: Linear exponential decay , with Emax = 4 and Base Time Prop = 0.2
 Missing Pattern: Range of missing proportions
 Observant Proportions: Assume independence
 Number of Groups: 4

Power	N	Group Sample Sizes Set ni	Meas Times M	Time Set	Time-Ave Resp Set μ_i	Cont Coef Set Ci	Lin Comb of Ci and μ_i Ci μ_i	Base Corr ρ	First Row of Corr Matrix	Alpha	Miss Set
0.1648	80	ni(1)	6	T(1)	$\mu_i(1)$	C1(1)	0.60	0.5000	$\rho_1(1)$	0.050	Ms1(1)
0.5696	80	ni(1)	6	T(1)	$\mu_i(1)$	C2(2)	1.60	0.5000	$\rho_1(1)$	0.050	Ms1(1)
0.7103	80	ni(1)	6	T(1)	$\mu_i(1)$	C3(3)	1.40	0.5000	$\rho_1(1)$	0.050	Ms1(1)
0.2573	80	ni(1)	6	T(1)	$\mu_i(1)$	C4(4)	0.70	0.5000	$\rho_1(1)$	0.050	Ms1(1)
0.2855	160	ni(2)	6	T(1)	$\mu_i(1)$	C1(1)	0.60	0.5000	$\rho_1(1)$	0.050	Ms1(1)
0.8553	160	ni(2)	6	T(1)	$\mu_i(1)$	C2(2)	1.60	0.5000	$\rho_1(1)$	0.050	Ms1(1)
0.9447	160	ni(2)	6	T(1)	$\mu_i(1)$	C3(3)	1.40	0.5000	$\rho_1(1)$	0.050	Ms1(1)
0.4562	160	ni(2)	6	T(1)	$\mu_i(1)$	C4(4)	0.70	0.5000	$\rho_1(1)$	0.050	Ms1(1)
0.3999	240	ni(3)	6	T(1)	$\mu_i(1)$	C1(1)	0.60	0.5000	$\rho_1(1)$	0.050	Ms1(1)
0.9589	240	ni(3)	6	T(1)	$\mu_i(1)$	C2(2)	1.60	0.5000	$\rho_1(1)$	0.050	Ms1(1)
0.9917	240	ni(3)	6	T(1)	$\mu_i(1)$	C3(3)	1.40	0.5000	$\rho_1(1)$	0.050	Ms1(1)
0.6201	240	ni(3)	6	T(1)	$\mu_i(1)$	C4(4)	0.70	0.5000	$\rho_1(1)$	0.050	Ms1(1)
0.5042	320	ni(4)	6	T(1)	$\mu_i(1)$	C1(1)	0.60	0.5000	$\rho_1(1)$	0.050	Ms1(1)
0.9896	320	ni(4)	6	T(1)	$\mu_i(1)$	C2(2)	1.60	0.5000	$\rho_1(1)$	0.050	Ms1(1)
0.9989	320	ni(4)	6	T(1)	$\mu_i(1)$	C3(3)	1.40	0.5000	$\rho_1(1)$	0.050	Ms1(1)
0.7442	320	ni(4)	6	T(1)	$\mu_i(1)$	C4(4)	0.70	0.5000	$\rho_1(1)$	0.050	Ms1(1)

Set(Set Number): Values

ni(1): 20, 20, 20, 20
 ni(2): 40, 40, 40, 40
 ni(3): 60, 60, 60, 60
 ni(4): 80, 80, 80, 80
 T(1): 0.00, 0.20, 0.40, 0.60, 0.80, 1.00
 $\mu_i(1)$: 1.00, 1.00, 1.10, 1.50
 C1(1): -3.00, 1.00, 1.00, 1.00
 C2(2): -3.00, -1.00, 1.00, 3.00
 C3(3): 1.00, 1.00, 1.00, -3.00
 C4(4): -1.00, -2.00, 2.00, 1.00
 $\rho_1(1)$: 1.0000, 0.5000, 0.2973, 0.1768, 0.1051, 0.0625
 Ms1(1): 0.00, 0.06, 0.12, 0.18, 0.24, 0.30

GEE Tests for the TAD of Multiple Groups in a Repeated Measures Design (Count Outcome)



Note that the horizontal axis corresponds to the sequence number of the n_i 's. Hence, 1 is n_{i1} , 2 is n_{i2} , and so on. An interesting exercise will be to change the configuration of the means and watch the corresponding change in the power.

Further Examples of GEE Options

The PASS GEE procedures offer many options that allow you to investigate various designs in detail. These are available at the end of Chapter 399, "GEE Tests for the Slope of Two Groups in a Repeated Measures Design (Continuous Outcome)." We suggest that you take time to look through those examples.