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, ..., G) subjects in each of G groups for a total of N subjects, each measured on M occasions at times t_j (j = 1, ..., 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

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

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, ...,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 $C = c_1, \ldots, c_G$ 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 $\beta_k, k = 1, \ldots, G$

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 $1/n_k \sum_{i=1}^{n_k} \left[\sum_{j=1}^{M} \hat{\epsilon}_{kij}\right]^2$, where $\hat{\epsilon}_{kij} = y_{kij} - \mu_{kij}(\underline{b})$

H_0 is rejected with a type I error α if $|\boldsymbol{T}| > z_{1-\alpha/2}$ where $z_{1-\alpha/2}$ is the 100(1 – α/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.

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

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

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

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

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 $|t_j - t_k| = 1$.

Damped Exponential

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

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

where ρ is the baseline correlation.

Damped Exponential (Proportional)

This version of the damped 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

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

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.

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, *Emax*. The resulting pattern looks similar to the damped 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, ρ = 0.5, *Emax* = 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): $\phi_{jj'} = \phi_k \text{ where } k = \max(j, j')$

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

Missing Input Type

The 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 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.

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 – α/2)th percentile of a standard normal distribution.
D	is $C'WA^{-1}VA^{-1}WC$
E	is $C'\underline{\beta}$
$\underline{\beta}$	is a vector of anticipated GEE regression coefficients, $\beta_1, \dots, \beta_G$
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.

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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For	**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**
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)	**0.6 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**

Output

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

Numeric Reports

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

Solve For:	Sample Size
Contrast:	List of Coefficients
Measurement Times:	Equally spaced
Correlation:	AR1: ρ(j,k) = ρ^|j-k|
Missing Pattern:	Range of missing proportions
Observant Proportions:	Assume independence
Number of Groups:	3

Power	Total Sample Size N	Group Allocation Proportions ri	Number of Measurement Times M	Time-Averaged Responses μi	Contrast of μi Coefficients Ci	Contrast of μi Value |Ci'μi|	Base Correlation ρ	First Row of Correlation Matrix	Missing Data Proportions	Measurement Times	Alpha
0.9063	78	ri(1)	4	μi(1)	Con(1)	10	0.6	ρ1(1)	Ms1(1)	T(1)	0.05
0.9028	87	ri(1)	4	μi(1)	Con(1)	10	0.7	ρ2(1)	Ms1(1)	T(1)	0.05
0.9052	99	ri(1)	4	μi(1)	Con(1)	10	0.8	ρ3(1)	Ms1(1)	T(1)	0.05

Item	Values
ri(1)	0.333, 0.333, 0.333
μi(1)	65, 60, 60
Con(1)	-2, 1, 1
ρ1(1)	1, 0.6, 0.36, 0.216
ρ2(1)	1, 0.7, 0.49, 0.343
ρ3(1)	1, 0.8, 0.64, 0.512
Ms1(1)	0, 0.13, 0.27, 0.4
T(1)	0, 0.33, 0.67, 1

Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N	The total number of subjects in the study.
ri	The Group Allocation Proportions Set gives the name of the set containing group allocation proportions.
M	The Measurement Times is the number of time points at which each subject is scheduled to be measured.
μi	The Time-Averaged Response gives the name of the set containing the time-averaged responses for each group.
Ci	The Contrast Coefficient gives the name of the set containing the contrast coefficients that are combined with the group responses.
|Ci'μi|	The contrast value of Ci and μi gives the linear combination of the contrast coefficients and the group responses.
ρ	The base correlation between two responses on the same subject. It may be transformed based on the correlation pattern.
First Row	Presents the top row of the correlation matrix.
Missing Data Proportions	Gives the name of the set containing the individual missing value proportions for each time period.
Measurement Times	Gives the name of the set containing the time proportions. The values represent the proportion of the total study time that has elapsed just before the measurement.
Alpha	The probability of rejecting a true null hypothesis.

Summary Statements

A 3-group repeated measures design (with a count response and with 4 measurements for each subject) will be used to test whether there is a group difference in time-averaged response based on a contrast with group contrast coefficients -2, 1, 1. The test will be performed by comparing the contrast of the GEE estimates of the log of the Poisson means to the null value zero, with a Type I error rate (α) of 0.05. The (repeated) measurements of each subject will be made at the following 4 times, expressed as proportions of the total study time: 0, 0.33, 0.67, 1. Missing values are assumed to occur completely at random (MCAR). The missing value proportions will be combined to form the pairwise observant probabilities using an independent pairwise missing pattern. The anticipated proportions missing at each measurement time are 0, 0.13, 0.27, 0.4. The first row of the autocorrelation matrix of the responses within a subject is assumed to be 1, 0.6, 0.36, 0.216, with subsequent rows following the same pattern (AR1: ρ(j,k) = ρ^|j-k|). To detect time-averaged group (Poisson) count means of 65, 60, 60 with 90% power, the total number of needed subjects is 78 (divided into the 3 groups according to the proportions: 0.333, 0.333, 0.333).

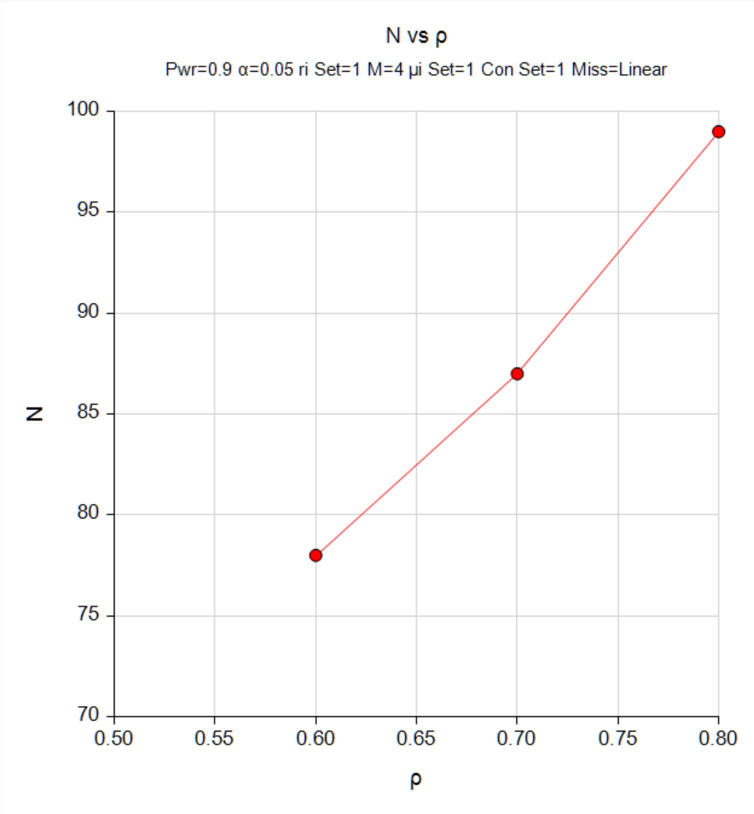
References

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

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

Plots Section

Plots



This shows the relationship among the design parameters.

Autocorrelation Matrices

Autocorrelation Matrix for Report Row 1

Time	T(0)	T(0.33)	T(0.67)	T(1)
T(0)	1.000	0.60	0.36	0.216
T(0.33)	0.600	1.00	0.60	0.360
T(0.67)	0.360	0.60	1.00	0.600
T(1)	0.216	0.36	0.60	1.000

Autocorrelation Matrix for Report Row 2

Time	T(0)	T(0.33)	T(0.67)	T(1)
T(0)	1.000	0.70	0.49	0.343
T(0.33)	0.700	1.00	0.70	0.490
T(0.67)	0.490	0.70	1.00	0.700
T(1)	0.343	0.49	0.70	1.000

Autocorrelation Matrix for Report Row 3

Time	T(0)	T(0.33)	T(0.67)	T(1)
T(0)	1.000	0.80	0.64	0.512
T(0.33)	0.800	1.00	0.80	0.640
T(0.67)	0.640	0.80	1.00	0.800
T(1)	0.512	0.64	0.80	1.000

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

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

Design Tab

Solve For .... **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.

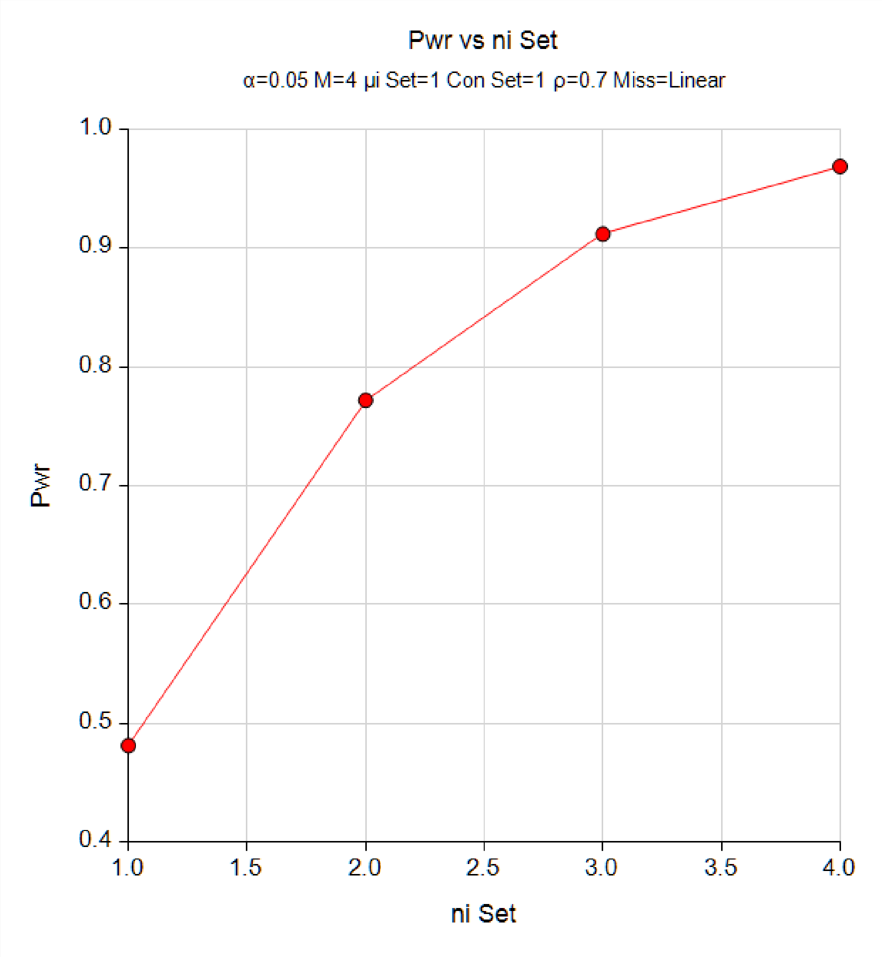
Numeric Results for a Time-Averaged Response Test of Count Data using GEE

Solve For:	Power
Contrast:	List of Coefficients
Measurement Times:	Equally spaced
Correlation:	AR1: ρ(j,k) = ρ^|j-k|
Missing Pattern:	Range of missing proportions
Observant Proportions:	Assume independence
Number of Groups:	3

					Contrast of μi						
Power	Total Sample Size N	Group Sample Sizes ni	Number of Measurement Times M	Time-Averaged Responses μi	Coefficients Ci	Value |Ci'μi|	Base Correlation ρ	First Row of Correlation Matrix	Missing Data Proportions	Measurement Times	Alpha
0.4812	30	ni(1)	4	μi(1)	Con(1)	10	0.7	ρ1(1)	Ms1(1)	T(1)	0.05
0.7720	60	ni(2)	4	μi(1)	Con(1)	10	0.7	ρ1(1)	Ms1(1)	T(1)	0.05
0.9120	90	ni(3)	4	μi(1)	Con(1)	10	0.7	ρ1(1)	Ms1(1)	T(1)	0.05
0.9690	120	ni(4)	4	μi(1)	Con(1)	10	0.7	ρ1(1)	Ms1(1)	T(1)	0.05

Item	Values
ni(1)	10, 10, 10
ni(2)	20, 20, 20
ni(3)	30, 30, 30
ni(4)	40, 40, 40
μi(1)	65, 60, 60
Con(1)	-2, 1, 1
ρ1(1)	1, 0.7, 0.49, 0.343
Ms1(1)	0, 0.13, 0.27, 0.4
T(1)	0, 0.33, 0.67, 1

Plots



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 *μi'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

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

Design Tab

Solve For ....**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**
μi's Input Type ....**Columns containing sets of μi's**
Columns Containing Sets of μi's **C1-C4**
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**

Input Spreadsheet Data

Row	C1	C2	C3	C4
1	65	65	65	65
2	60	61	62	63
3	60	61	62	63

Output

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

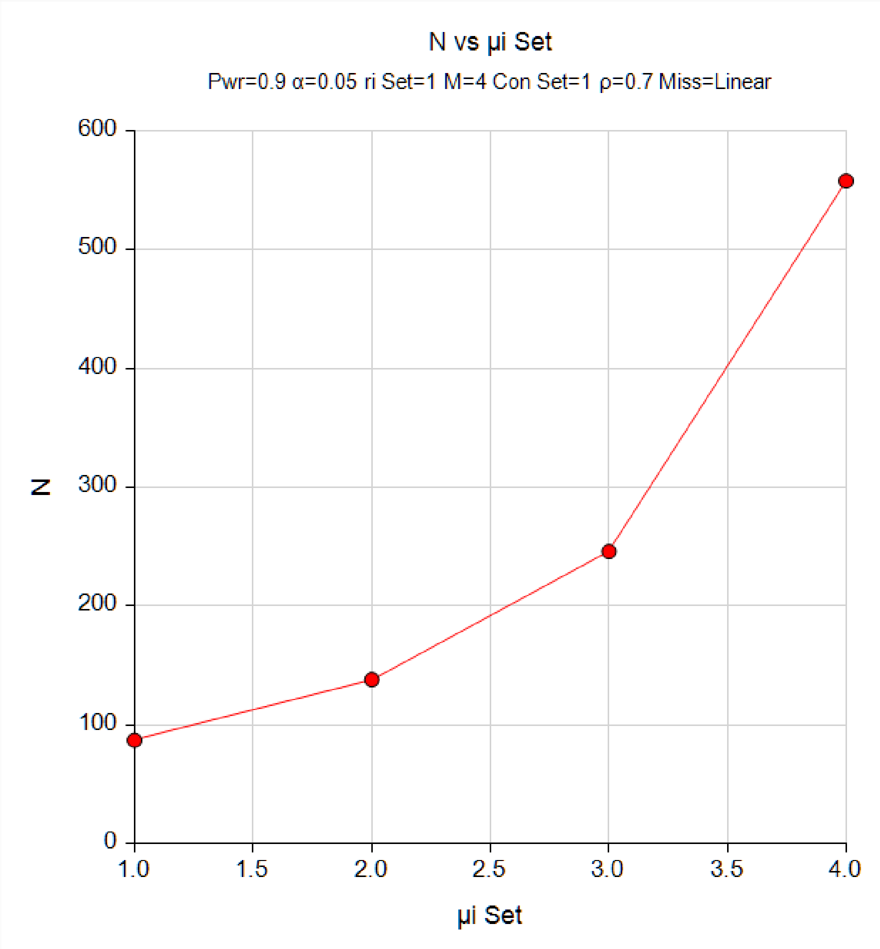
Numeric Results for a Time-Averaged Response Test of Count Data using GEE

Solve For:	Sample Size
Contrast:	List of Coefficients
Measurement Times:	Equally spaced
Correlation:	AR1: ρ(j,k) = ρ^|j-k|
Missing Pattern:	Range of missing proportions
Observant Proportions:	Assume independence
Number of Groups:	3

					Contrast of μi						
Power	Total Sample Size N	Group Allocation Proportions ri	Number of Measurement Times M	Time-Averaged Responses μi	Coefficients Ci	Value |Ci'μi|	Base Correlation ρ	First Row of Correlation Matrix	Missing Data Proportions	Measurement Times	Alpha
0.9028	87	ri(1)	4	C1(1)	Con(1)	10	0.7	ρ1(1)	Ms1(1)	T(1)	0.05
0.9041	138	ri(1)	4	C2(2)	Con(1)	8	0.7	ρ1(1)	Ms1(1)	T(1)	0.05
0.9019	246	ri(1)	4	C3(3)	Con(1)	6	0.7	ρ1(1)	Ms1(1)	T(1)	0.05
0.9012	558	ri(1)	4	C4(4)	Con(1)	4	0.7	ρ1(1)	Ms1(1)	T(1)	0.05

Item	Values
ri(1)	0.333, 0.333, 0.333
C1(1)	65, 60, 60
C2(2)	65, 61, 61
C3(3)	65, 62, 62
C4(4)	65, 63, 63
Con(1)	-2, 1, 1
ρ1(1)	1, 0.7, 0.49, 0.343
Ms1(1)	0, 0.13, 0.27, 0.4
T(1)	0, 0.33, 0.67, 1

Plots



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, μ1 = 2, μ2 = 1, compound symmetry, ρ = 0.6, constant missing = 0.1, α = 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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 4** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For ....**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

Solve For:	Sample Size
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

Power	Total Sample Size N	Group Allocation Proportions ri	Number of Measurement Times M	Time-Averaged Responses μi	Contrast of μi: Coefficients Ci	Contrast of μi: Value |Ci'μi|	Base Correlation ρ	First Row of Correlation Matrix	Missing Data Proportion	Measurement Times	Alpha
0.9028	54	ri(1)	3	μi(1)	Con(1)	1	0.6	ρ1(1)	0.1	T(1)	0.05

Item	Values
ri(1)	0.5, 0.5
μi(1)	2, 1
Con(1)	-1, 1
ρ1(1)	1, 0.6, 0.6
T(1)	0, 0.5, 1

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 α = 0.05. The correlation pattern is Linear Exponential Decay with a base correlation of 0.5, Base Time Proportion of 0.20, and Emax set to 4. The missing input type is set to Linear from 0 to 30% and the pairwise missing assumption is independent. 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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 5** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For .... **Power**
Alpha.. **0.05**
G (Number of Groups) **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**
μi's Input Type .... **μ1, μ2, ..., μG**
μ1, μ2, ..., μG.. **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**
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**

Input Spreadsheet Data

Row	Tm1	Tm2	Tm3	Tm4	Tm5
1	0.0	0.0	0.0	0.0	0.00
2	0.2	0.6	0.1	0.1	0.45
3	0.4	0.7	0.2	0.2	0.50
4	0.6	0.8	0.3	0.8	0.55
5	0.8	0.9	0.4	0.9	0.60
6	1.0	1.0	1.0	1.0	1.00

Output

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

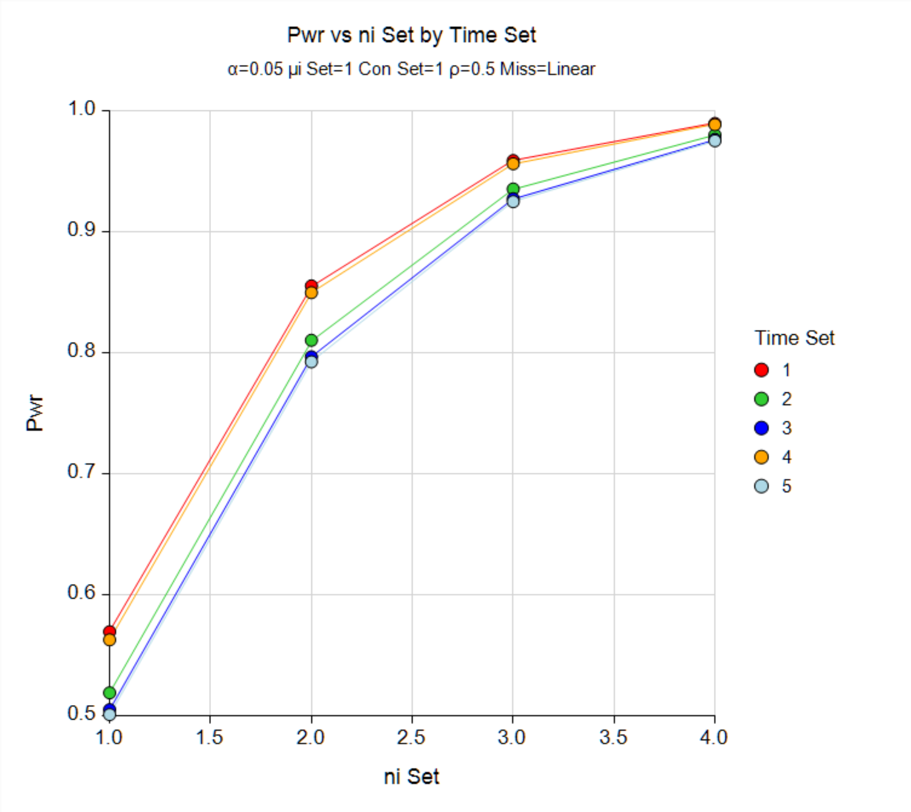
Numeric Results for a Time-Averaged Response Test of Count Data using GEE

Solve For:	Power
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	Total Sample Size N	Group Sample Sizes ni	Number of Measurement Times M	Time-Averaged Responses μi	Contrast of μi Coefficients Ci	Contrast of μi Value |Ci'μi|	Base Correlation ρ	First Row of Correlation Matrix	Missing Data Proportions	Measurement Times	Alpha
0.5696	80	ni(1)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm1)	Ms1(Tm1)	Tm1(1)	0.05
0.5190	80	ni(1)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm2)	Ms1(Tm2)	Tm2(2)	0.05
0.5051	80	ni(1)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm3)	Ms1(Tm3)	Tm3(3)	0.05
0.5628	80	ni(1)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm4)	Ms1(Tm4)	Tm4(4)	0.05
0.5010	80	ni(1)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm5)	Ms1(Tm5)	Tm5(5)	0.05
0.8553	160	ni(2)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm1)	Ms1(Tm1)	Tm1(1)	0.05
0.8104	160	ni(2)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm2)	Ms1(Tm2)	Tm2(2)	0.05
0.7967	160	ni(2)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm3)	Ms1(Tm3)	Tm3(3)	0.05
0.8498	160	ni(2)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm4)	Ms1(Tm4)	Tm4(4)	0.05
0.7926	160	ni(2)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm5)	Ms1(Tm5)	Tm5(5)	0.05
0.9589	240	ni(3)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm1)	Ms1(Tm1)	Tm1(1)	0.05
0.9354	240	ni(3)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm2)	Ms1(Tm2)	Tm2(2)	0.05
0.9274	240	ni(3)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm3)	Ms1(Tm3)	Tm3(3)	0.05
0.9563	240	ni(3)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm4)	Ms1(Tm4)	Tm4(4)	0.05
0.9250	240	ni(3)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm5)	Ms1(Tm5)	Tm5(5)	0.05
0.9896	320	ni(4)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm1)	Ms1(Tm1)	Tm1(1)	0.05
0.9801	320	ni(4)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm2)	Ms1(Tm2)	Tm2(2)	0.05
0.9765	320	ni(4)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm3)	Ms1(Tm3)	Tm3(3)	0.05
0.9886	320	ni(4)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm4)	Ms1(Tm4)	Tm4(4)	0.05
0.9753	320	ni(4)	6	μi(1)	Con(1)	0.8	0.5	ρ1(Tm5)	Ms1(Tm5)	Tm5(5)	0.05

Item	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
μi(1)	1, 1, 1.1, 1.5
Con(1)	-1.5, -0.5, 0.5, 1.5
ρ1(Tm1)	1, 0.5, 0.2973, 0.1768, 0.1051, 0.0625
ρ1(Tm2)	1, 0.1768, 0.1363, 0.1051, 0.0811, 0.0625
ρ1(Tm3)	1, 0.6484, 0.5, 0.3856, 0.2973, 0.0625
ρ1(Tm4)	1, 0.6484, 0.5, 0.1051, 0.0811, 0.0625
ρ1(Tm5)	1, 0.2611, 0.2293, 0.2013, 0.1768, 0.0625
Ms1(Tm1)	0, 0.06, 0.12, 0.18, 0.24, 0.3
Ms1(Tm2)	0, 0.18, 0.21, 0.24, 0.27, 0.3
Ms1(Tm3)	0, 0.03, 0.06, 0.09, 0.12, 0.3
Ms1(Tm4)	0, 0.03, 0.06, 0.24, 0.27, 0.3
Ms1(Tm5)	0, 0.14, 0.15, 0.17, 0.18, 0.3
Tm1(1)	0, 0.2, 0.4, 0.6, 0.8, 1
Tm2(2)	0, 0.6, 0.7, 0.8, 0.9, 1
Tm3(3)	0, 0.1, 0.2, 0.3, 0.4, 1
Tm4(4)	0, 0.1, 0.2, 0.8, 0.9, 1
Tm5(5)	0, 0.45, 0.5, 0.55, 0.6, 1

Plots



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 α = 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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 6** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For .... **Power**
Alpha... **0.05**
G (Number of Groups) **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**
μi's Input Type .... **μ1, μ2, ..., μG**
μ1, μ2, ..., μG.. **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**

Input Spreadsheet Data

Row	C1	C2	C3	C4
1	-3	-3	1	-1
2	1	-1	1	-2
3	1	1	1	2
4	1	3	-3	1

Output

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

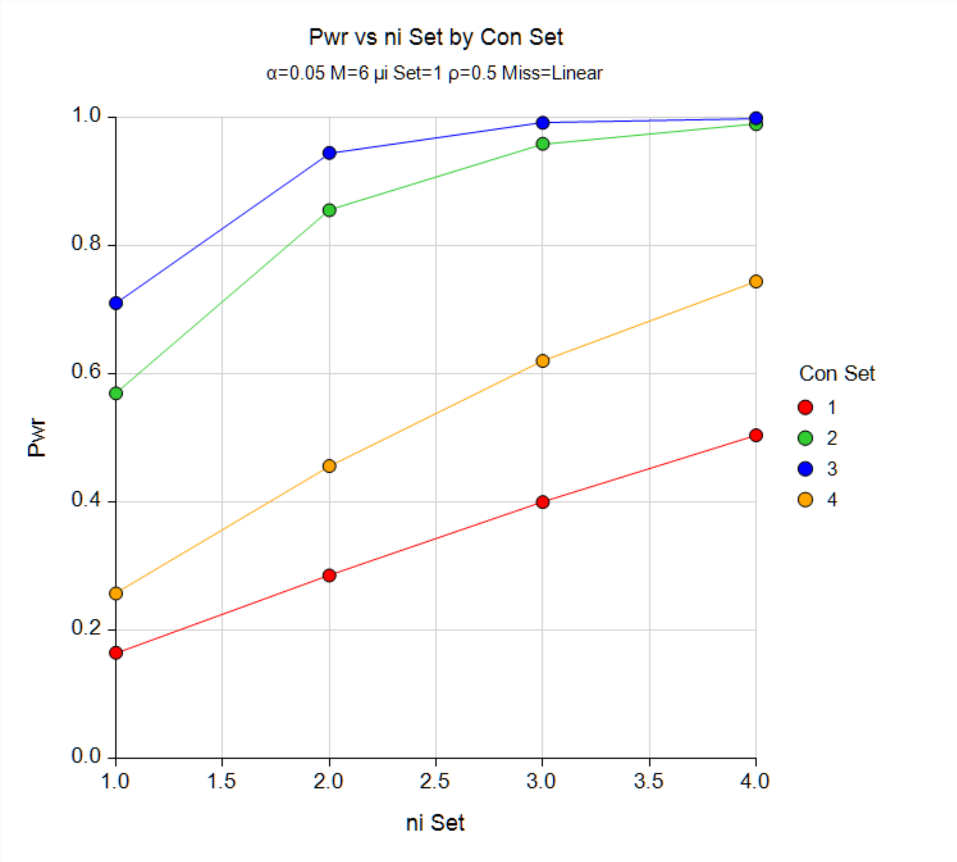
Numeric Results for a Time-Averaged Response Test of Count Data using GEE

Solve For:	Power
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	Total Sample Size N	Group Sample Sizes ni	Number of Measurement Times M	Time-Averaged Responses μi	Contrast of μi Coefficients Ci	Contrast of μi Value |Ci'μi|	Base Correlation ρ	First Row of Correlation Matrix	Missing Data Proportions	Measurement Times	Alpha
0.1648	80	ni(1)	6	μi(1)	C1(1)	0.6	0.5	ρ1(1)	Ms1(1)	T(1)	0.05
0.5696	80	ni(1)	6	μi(1)	C2(2)	1.6	0.5	ρ1(1)	Ms1(1)	T(1)	0.05
0.7103	80	ni(1)	6	μi(1)	C3(3)	1.4	0.5	ρ1(1)	Ms1(1)	T(1)	0.05
0.2573	80	ni(1)	6	μi(1)	C4(4)	0.7	0.5	ρ1(1)	Ms1(1)	T(1)	0.05
0.2855	160	ni(2)	6	μi(1)	C1(1)	0.6	0.5	ρ1(1)	Ms1(1)	T(1)	0.05
0.8553	160	ni(2)	6	μi(1)	C2(2)	1.6	0.5	ρ1(1)	Ms1(1)	T(1)	0.05
0.9447	160	ni(2)	6	μi(1)	C3(3)	1.4	0.5	ρ1(1)	Ms1(1)	T(1)	0.05
0.4562	160	ni(2)	6	μi(1)	C4(4)	0.7	0.5	ρ1(1)	Ms1(1)	T(1)	0.05
0.3999	240	ni(3)	6	μi(1)	C1(1)	0.6	0.5	ρ1(1)	Ms1(1)	T(1)	0.05
0.9589	240	ni(3)	6	μi(1)	C2(2)	1.6	0.5	ρ1(1)	Ms1(1)	T(1)	0.05
0.9917	240	ni(3)	6	μi(1)	C3(3)	1.4	0.5	ρ1(1)	Ms1(1)	T(1)	0.05
0.6201	240	ni(3)	6	μi(1)	C4(4)	0.7	0.5	ρ1(1)	Ms1(1)	T(1)	0.05
0.5042	320	ni(4)	6	μi(1)	C1(1)	0.6	0.5	ρ1(1)	Ms1(1)	T(1)	0.05
0.9896	320	ni(4)	6	μi(1)	C2(2)	1.6	0.5	ρ1(1)	Ms1(1)	T(1)	0.05
0.9989	320	ni(4)	6	μi(1)	C3(3)	1.4	0.5	ρ1(1)	Ms1(1)	T(1)	0.05
0.7442	320	ni(4)	6	μi(1)	C4(4)	0.7	0.5	ρ1(1)	Ms1(1)	T(1)	0.05

Item	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
μi(1)	1, 1, 1.1, 1.5
C1(1)	-3, 1, 1, 1
C2(2)	-3, -1, 1, 3
C3(3)	1, 1, 1, -3
C4(4)	-1, -2, 2, 1
ρ1(1)	1, 0.5, 0.2973, 0.1768, 0.1051, 0.0625
Ms1(1)	0, 0.06, 0.12, 0.18, 0.24, 0.3
T(1)	0, 0.2, 0.4, 0.6, 0.8, 1

Plots



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