

Chapter 431

Tests for Two Means in a Repeated Measures Design

Introduction

This module calculates the power for testing the time-averaged difference (TAD) between two means in a *repeated measures* design. A repeated measures design is one in which subjects are observed repeatedly over time. Measurements may be taken at pre-determined intervals (e.g. weekly or at specified time points following the administration of a particular treatment), or at random times so there are variable intervals between repeated measurements.

Time-averaged difference analysis is often used when the outcome to be measured varies with time. For example, suppose that you want to compare two 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 mean values from the two groups may be invalid because of the large degree of variation in blood pressure levels among patients. The precision of the experiment is increased by taking multiple measurements from each individual and comparing the time-averaged difference between the two 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 covariance structure may take on several forms depending on the nature of the experiment and the subjects involved. This procedure allows you to calculate sample sizes using four different covariance patterns: Compound Symmetry, AR(1), Banded(1), and Simple.

This procedure can be used to calculate sample size and power for tests of pairwise contrasts in a mixed models analysis of repeated measures data. Mixed models analysis of repeated measures data is also employed to provide more flexibility in covariance specification and a greater degree of robustness in the presence of missing data, provided that the data can be assumed to be missing at random.

Technical Details

Theory and Notation

For a study with n_1 subjects in group 1 and n_2 subjects in group 2 (for a total of N subjects), each measured m times, the time-averaged difference (d) of a continuous response between two groups can be estimated using the following model:

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij}, \quad i = 1, \dots, N; j = 1, \dots, m,$$

where

y_{ij} is the j^{th} response from subject i ,

β_0 is the model intercept,

β_1 is the treatment effect or the time-averaged difference between groups 1 and 2 (i.e. $\beta_1 = d$),

x_i is a binary group assignment variable, which is equal to 1 if the i^{th} subject is in group 1 and equal to 0 if the i^{th} subject is in group 2, and

ε_{ij} is the normal, random error associated with the observation y_{ij} .

Accounting for the relationship between repeated measurements, the model presented above can be written in matrix form as

$$\mathbf{y}_i = \mathbf{X}_i' \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i,$$

where

$\mathbf{y}_i = (y_{i1} \ y_{i2} \ \dots \ y_{im})'$ is an $m \times 1$ vector of responses from subject i ,

$$\mathbf{X}_i = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix}_{m \times 2} \quad \text{if the } i^{\text{th}} \text{ subject is in group 1,}$$

$$\mathbf{X}_i = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{pmatrix}_{m \times 2} \quad \text{if the } i^{\text{th}} \text{ subject is in group 2,}$$

$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ is the vector of model parameters, and

$\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{R}_i)$ is the vector of correlated random errors for the observations from subject i ,

where

$\text{var}(y_{ij}) = \sigma^2$ is the residual variance for a single observation, and \mathbf{R}_i is the $m \times m$ common correlation matrix for all subjects. The contents of \mathbf{R}_i depend on the assumed within-subject correlation structure.

Tests for Two Means in a Repeated Measures Design

We can stack the data in a single vector and matrix form as follows:

$$\begin{aligned}\mathbf{y} &= (\mathbf{y}_1', \mathbf{y}_2', \dots, \mathbf{y}_N')' \\ \mathbf{X} &= (\mathbf{X}_1', \mathbf{X}_2', \dots, \mathbf{X}_N')' \\ \boldsymbol{\varepsilon} &= (\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_2', \dots, \boldsymbol{\varepsilon}_N')'\end{aligned}$$

and the model for the N equations can be compressed into one as

$$\mathbf{y} = \mathbf{X}'\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

with

$$\begin{aligned}\mathbf{V} &= \text{var}(\mathbf{y}) \\ &= \sigma^2 \begin{pmatrix} \mathbf{R}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_N \end{pmatrix} \\ &= \sigma^2 \mathbf{R}\end{aligned}$$

as the covariance (or variance - covariance) matrix.

Covariance Pattern

In a repeated measures design with N subjects, each measured m times, observations from a single subject may be correlated, and a pattern for their covariance must be specified. In this case, \mathbf{V} will have a block-diagonal form:

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V}_3 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{V}_N \end{pmatrix},$$

where \mathbf{V}_i are $m \times m$ covariance matrices corresponding to the i^{th} subject. The $\mathbf{0}$'s represent $m \times m$ matrices of zeros giving zero covariances for observations on different subjects. This routine allows the specification of four different covariance matrix types: Compound Symmetry, AR(1), Banded(1), and Simple.

Compound Symmetry

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

$$\mathbf{V}_i = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \rho & \cdots & \rho \\ \rho & \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \rho & \cdots & 1 \end{pmatrix}_{m \times m}$$

where σ^2 is the residual variance and ρ is the correlation between observations on the same subject.

Tests for Two Means in a Repeated Measures Design

AR(1)

An AR(1) (autoregressive order 1) covariance model assumes that all variances on the diagonal are equal and that covariances t time periods apart are equal to $\sigma^2 \rho^t$. That is

$$\mathbf{V}_i = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{m-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^{m-2} \\ \rho^2 & \rho & 1 & \rho & \dots & \rho^{m-3} \\ \rho^3 & \rho^2 & \rho & 1 & \dots & \rho^{m-4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{m-1} & \rho^{m-2} & \rho^{m-3} & \rho^{m-4} & \dots & 1 \end{pmatrix}_{m \times m}$$

where σ^2 is the residual variance and ρ is the correlation between observations on the same subject.

Banded(1)

A Banded(1) (banded order 1) covariance model assumes that all variances on the diagonal are equal, covariances for observations one time period apart are equal to $\sigma^2 \rho$, and covariances for measurements greater than one time period apart are equal to zero. That is

$$\mathbf{V}_i = \sigma^2 \begin{pmatrix} 1 & \rho & 0 & 0 & \dots & 0 \\ \rho & 1 & \rho & 0 & \dots & 0 \\ 0 & \rho & 1 & \rho & \dots & 0 \\ 0 & 0 & \rho & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}_{m \times m}$$

where σ^2 is the residual variance and ρ is the correlation between observations on the same subject.

Simple

A simple covariance model assumes that all variances on the diagonal are equal and that all covariances are equal to zero. That is

$$\mathbf{V}_i = \sigma^2 \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}_{m \times m}$$

where σ^2 is the residual variance.

Tests for Two Means in a Repeated Measures Design

Model Estimation

With $\hat{\mathbf{V}} = \hat{\sigma}^2 \hat{\mathbf{R}}$, then estimates of the regression coefficients from the above regression model are given as

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \\ &= (\mathbf{X}' \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\mathbf{V}}^{-1} \mathbf{y}\end{aligned}$$

and the variance of $\hat{\boldsymbol{\beta}}$ is

$$\begin{aligned}\text{var}(\hat{\boldsymbol{\beta}}) &= \begin{pmatrix} \text{var}(\hat{\beta}_0) & \text{cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \text{cov}(\hat{\beta}_0, \hat{\beta}_1) & \text{var}(\hat{\beta}_1) \end{pmatrix}, \\ &= (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1}\end{aligned}$$

which is estimated by substituting $\hat{\mathbf{V}}$ for \mathbf{V} .

Hypothesis Test

A two-sided test that the time-averaged difference between the two groups is equal to zero is equivalent to the test of $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$. Similarly, the upper and lower one-sided tests are $H_0 : \beta_1 \leq 0$ vs. $H_1 : \beta_1 > 0$ and $H_0 : \beta_1 \geq 0$ vs. $H_1 : \beta_1 < 0$, respectively. The test can be carried out using the test statistic

$$z = \frac{\hat{\beta}_1}{\sqrt{\text{var}(\hat{\beta}_1)}} \rightarrow N(0,1).$$

If the standard deviation is unknown and estimated, a t test should be used. In practice, this test is often carried out by calculating the average response for each individual and then using a two-sample t test. If the data are balanced, the test can also be carried out in NCSS using Repeated Measures GLM and specifying a comparison such as "Each with First". In the case where the data are not balanced, the test could be carried out using SAS[®] PROC MIXED or SAS[®] PROC GLM. In both cases a REPEATED statement should be used, along with a statement such as

```
ESTIMATE 'A-B' treat 1 -1; or LSMEANS treat/ PDIF;
```

Power Calculations

Sample sizes for repeated measures studies are often calculated as if a simple trial with no repeated measures was planned, which results in a higher calculated sample size than would be found if the correlation between repeated measures were taken into consideration. With an idea of the correct covariance structure, and an estimate of the within-patient correlation, you can get a better estimate of the power and sample size necessary to achieve your objectives. If you have no indication of the correct covariance structure for the experiment, then the compound symmetry (program default) is likely to be adequate. If you have no previous estimate of the within-patient correlation, then Brown and Prescott (2006) suggest using a conservative prediction of the correlation, i.e. a higher correlation than anticipated.

Tests for Two Means in a Repeated Measures Design

For a two-sided test where it is assumed that $d > 0$ (without loss of generality),

$$\begin{aligned}
 \text{Power} &= 1 - \beta = \Pr(\text{rejecting } H_0 \mid H_1) \\
 &= \Pr\left(\left|\frac{\hat{\beta}_1}{\sqrt{\text{var}(\hat{\beta}_1)}}\right| > z_{1-\alpha/2} \mid H_1\right) \\
 &\approx \Pr\left(\frac{\hat{\beta}_1}{\sqrt{\text{var}(\hat{\beta}_1)}} > z_{1-\alpha/2} \mid H_1\right) \text{ since it is assumed that } d > 0 \\
 &= \Pr\left(\frac{\hat{\beta}_1 - d}{\sqrt{\text{var}(\hat{\beta}_1)}} > z_{1-\alpha/2} - \frac{d}{\sqrt{\text{var}(\hat{\beta}_1)}} \mid H_1\right) \\
 &= 1 - \Phi\left(z_{1-\alpha/2} - \frac{d}{\sqrt{\text{var}(\hat{\beta}_1)}}\right),
 \end{aligned}$$

where $\Phi()$ is the standard normal density function, and α and β are the probabilities of type I and type II error, respectively. For a one-sided test, α is used in place of $\alpha/2$.

Since a t test is usually used to test for a group difference in a case such as this, we should note here that the power calculation using the standard normal distribution represents an approximation to the actual power achieved by the t test. We feel that it might be more appropriate to use the non-central t distribution; however, since the calculation is based on numerous assumptions about the covariance structure that influence the results, it seems unnecessary to worry about the small gain in precision that may occur by using the non-central t distribution. For this reason, along with the fact that this is the published method, we have elected to follow the methods of Brown and Prescott (2006), Liu and Wu (2005), Diggle et al. (1994) and use the standard normal distribution in power and sample size calculations.

Calculating Power for Testing Pairwise Contrasts of Fixed Effects in Mixed Models

Mixed Model Theory and Notation

A linear mixed model incorporates both fixed and random effects. Fixed effects are those effects in the model whose values are assumed constant, or unchanging. Random effects are those effects in the model that are assumed to have arisen from a distribution, resulting in another source of random variation other than residual variation. Brown and Prescott (2006) demonstrates how this methodology may be used to calculate the sample size and power for testing pairwise contrasts of fixed effects in a mixed models analysis of repeated measures data. For an experiment with N subjects, p fixed effect parameters, and q random effect parameters, the general mixed model can be expressed using matrix notation as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, N$$

where

\mathbf{y}_i is an $n_i \times 1$ vector of responses for subject i ,

\mathbf{X}_i is an $n_i \times p$, full-rank design matrix of fixed effects for subject i ,

$\boldsymbol{\beta}$ is an $p \times 1$ vector of fixed effects parameters,

\mathbf{Z}_i is an $n_i \times q$ design matrix of the random effects for subject i ,

Tests for Two Means in a Repeated Measures Design

\mathbf{u}_i is a $q \times 1$ vector of random effects for subject i which has means of zero and scaled covariance matrix \mathbf{G} ,

$\boldsymbol{\varepsilon}_i$ is an $n_i \times 1$ vector of errors for subject i with zero mean and scaled covariance $\boldsymbol{\Sigma}_i$.

The covariance of \mathbf{y}_i , $\text{var}(\mathbf{y}_i) = \mathbf{V}_i$, can be written as

$$\begin{aligned}\mathbf{V}_i &= \text{var}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i) \\ &= \mathbf{Z} \text{var}(\mathbf{u}_i)\mathbf{Z}' + \text{var}(\boldsymbol{\varepsilon}_i) \\ &= \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i' + \boldsymbol{\Sigma}_i.\end{aligned}$$

We can stack the data in a single vector and matrix form as follows:

$$\begin{aligned}\mathbf{y} &= (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)' \\ \mathbf{X} &= (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N)' \\ \mathbf{Z} &= \begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}_N \end{pmatrix} \\ \mathbf{u} &= (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N)' \\ \boldsymbol{\varepsilon} &= (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_N)'\end{aligned}$$

and the mixed model for the N equations can be compressed into one as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon},$$

with

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V}_N \end{pmatrix} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \boldsymbol{\Sigma}$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_N \end{pmatrix}$$

is the covariance (or variance-covariance) matrix.

Mixed Model Estimation

Estimates of the variance components \mathbf{G} and $\boldsymbol{\Sigma}$ are found using maximum likelihood (ML) or restricted/residual maximum likelihood (REML) methods. From these estimates, $\hat{\mathbf{G}}$ and $\hat{\boldsymbol{\Sigma}}$, an estimate of \mathbf{V} is obtained as $\hat{\mathbf{V}} = \mathbf{Z}\hat{\mathbf{G}}\mathbf{Z}' + \hat{\boldsymbol{\Sigma}}$. The fixed effects are then estimated as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}$$

with the variance of $\hat{\boldsymbol{\beta}}$ estimated as

$$\text{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}.$$

If $\mathbf{Z}\mathbf{G}\mathbf{Z}' = \mathbf{0}$ and $\boldsymbol{\Sigma} = \sigma^2\mathbf{R}$, then these estimation equations are identical to the TAD estimation equations presented earlier, except for the fact that $\boldsymbol{\beta}$ may contain more than two parameters, i.e. a parameter for each fixed effect being modeled. In the TAD model presented above, β_1 represents the difference between two treatment

Tests for Two Means in a Repeated Measures Design

means, d . In the mixed model formulation presented here, β_1, β_2 , etc. represent individual treatment effects. If there are no random effects, then we can use this routine for TAD to calculate the approximate power for testing pairwise contrasts of fixed effects in mixed models designs.

Brown and Prescott (2006) presents an example on page 228 of an experiment for which the power for testing pairwise contrasts can be calculated using this procedure. To determine the relative efficacy of three treatments in controlling hypertension, patients are assigned to one of the three treatments and blood pressure is measured at four follow-up visits. The study aims to determine the differences in average blood pressure among the three treatments.

Testing Fixed Effects

Significance tests for fixed effects can be done using tests based on the t distribution. We can define tests of fixed and random effects as contrasts

$$\mathbf{C} = \mathbf{L}'\hat{\boldsymbol{\beta}} = \mathbf{0},$$

respectively. For example, in a trial containing three treatments A, B, and C, a pairwise comparison of treatments A and C is given by the contrast

$$\mathbf{C}_{AC} = \mathbf{L}'\hat{\boldsymbol{\beta}} = (0 \ 1 \ 0 \ -1)\hat{\boldsymbol{\beta}} = \hat{\beta}_A - \hat{\beta}_C,$$

where the first term in $\boldsymbol{\beta}$ is the intercept term, and the other three terms are the treatment effects.

For a single comparison, the test statistic is given by

$$\begin{aligned} t_{df} &= \frac{\mathbf{L}'\hat{\boldsymbol{\beta}}}{\text{SE}(\mathbf{L}'\hat{\boldsymbol{\beta}})} \\ &= \frac{\hat{\beta}_j - \hat{\beta}_h}{\text{SE}(\hat{\beta}_j - \hat{\beta}_h)}, \end{aligned}$$

where df is the degrees of freedom, usually determined using the Satterthwaite approximation, and $\hat{\beta}_j$ and $\hat{\beta}_h$ ($j \neq h$) are estimated treatment effects.

Contrasts such as this can be tested in SAS[®] using the ESTIMATE statement or by including the PDIFFF option in an LSMEANS statement. For example, if the variable designating three treatments, A, B, and C, were called "treat", then I could use the following statements in PROC MIXED to test for a difference between A and C

```
ESTIMATE 'A-C' treat 1 0 -1;
```

or

```
LSMEANS treat/ PDIFFF;
```

The latter statement would produce tests of all pairwise comparisons of the levels of the treatment variable. The former would only test the difference between groups A and C. Of course, these comparison statements must be used in conjunction with appropriate model and class statements (see pages 233-237 of Brown and Prescott (2006) for an example analyzed using SAS[®] PROC MIXED).

Estimates of the correlation (ρ) and the standard deviation (σ) for use in power calculations can be found using SAS[®] PROC MIXED. For a model fit using compound symmetry, σ^2 and ρ can be estimated as the sum of the variance parameters, and the compound symmetry variance parameter divided by the sum of the variance parameters, respectively. For AR(1), Banded(1), and Simple covariance models, σ^2 and ρ can be estimated as the residual variance, and the correlation between adjacent measurements, respectively. Alternatively, the R and RCORR options may be used within the REPEATED statement to display the covariance and correlation matrices, from which the parameter estimates can be determined.

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 beta criterion you have specified for each of the terms. The "solve for" parameter is displayed on the vertical axis of the plot.

Test

Alternative Hypothesis

This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always $H_0 : d = 0$.

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

Possible selections are:

- **One-Sided**

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

- **Two-Sided**

This is the most common selection. It yields the *two-tailed* test. Use this option when you are testing whether the means are different, but you do not want to specify beforehand which mean is larger.

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.

Tests for Two Means in a Repeated Measures Design

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 (When Solving for Sample Size)

Group Allocation

Select the option that describes the constraints on $N1$ or $N2$ or both.

The options are

- **Equal ($N1 = N2$)**

This selection is used when you wish to have equal sample sizes in each group. Since you are solving for both sample sizes at once, no additional sample size parameters need to be entered.

- **Enter $N2$, solve for $N1$**

Select this option when you wish to fix $N2$ at some value (or values), and then solve only for $N1$. Please note that for some values of $N2$, there may not be a value of $N1$ that is large enough to obtain the desired power.

- **Enter $R = N2/N1$, solve for $N1$ and $N2$**

For this choice, you set a value for the ratio of $N2$ to $N1$, and then PASS determines the needed $N1$ and $N2$, with this ratio, to obtain the desired power. An equivalent representation of the ratio, R , is

$$N2 = R * N1.$$

- **Enter percentage in Group 1, solve for $N1$ and $N2$**

For this choice, you set a value for the percentage of the total sample size that is in Group 1, and then PASS determines the needed $N1$ and $N2$ with this percentage to obtain the desired power.

$N2$ (Sample Size, Group 2)

This option is displayed if Group Allocation = "Enter $N2$, solve for $N1$ "

$N2$ is the number of items or individuals sampled from the Group 2 population.

$N2$ must be ≥ 2 . You can enter a single value or a series of values.

R (Group Sample Size Ratio)

This option is displayed only if Group Allocation = "Enter $R = N2/N1$, solve for $N1$ and $N2$."

R is the ratio of $N2$ to $N1$. That is,

$$R = N2 / N1.$$

Use this value to fix the ratio of $N2$ to $N1$ while solving for $N1$ and $N2$. Only sample size combinations with this ratio are considered.

$N2$ is related to $N1$ by the formula:

$$N2 = [R \times N1],$$

where the value $[Y]$ is the next integer $\geq Y$.

Tests for Two Means in a Repeated Measures Design

For example, setting $R = 2.0$ results in a Group 2 sample size that is double the sample size in Group 1 (e.g., $N1 = 10$ and $N2 = 20$, or $N1 = 50$ and $N2 = 100$).

R must be greater than 0. If $R < 1$, then $N2$ will be less than $N1$; if $R > 1$, then $N2$ will be greater than $N1$. You can enter a single or a series of values.

Percent in Group 1

This option is displayed only if Group Allocation = "Enter percentage in Group 1, solve for N1 and N2."

Use this value to fix the percentage of the total sample size allocated to Group 1 while solving for $N1$ and $N2$. Only sample size combinations with this Group 1 percentage are considered. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single or a series of values.

Sample Size (When Not Solving for Sample Size)

Group Allocation

Select the option that describes how individuals in the study will be allocated to Group 1 and to Group 2.

The options are

- **Equal ($N1 = N2$)**
This selection is used when you wish to have equal sample sizes in each group. A single per group sample size will be entered.
- **Enter $N1$ and $N2$ individually**
This choice permits you to enter different values for $N1$ and $N2$.
- **Enter $N1$ and R , where $N2 = R * N1$**
Choose this option to specify a value (or values) for $N1$, and obtain $N2$ as a ratio (multiple) of $N1$.
- **Enter total sample size and percentage in Group 1**
Choose this option to specify a value (or values) for the total sample size (N), obtain $N1$ as a percentage of N , and then $N2$ as $N - N1$.

Sample Size Per Group

This option is displayed only if Group Allocation = "Equal ($N1 = N2$)."

The Sample Size Per Group is the number of items or individuals sampled from each of the Group 1 and Group 2 populations. Since the sample sizes are the same in each group, this value is the value for $N1$, and also the value for $N2$.

The Sample Size Per Group must be ≥ 2 . You can enter a single value or a series of values.

$N1$ (Sample Size, Group 1)

*This option is displayed if Group Allocation = "Enter $N1$ and $N2$ individually" or "Enter $N1$ and R , where $N2 = R * N1$."*

$N1$ is the number of items or individuals sampled from the Group 1 population.

$N1$ must be ≥ 2 . You can enter a single value or a series of values.

Tests for Two Means in a Repeated Measures Design

N2 (Sample Size, Group 2)

This option is displayed only if Group Allocation = "Enter N1 and N2 individually."

N2 is the number of items or individuals sampled from the Group 2 population.

N2 must be ≥ 2 . You can enter a single value or a series of values.

R (Group Sample Size Ratio)

*This option is displayed only if Group Allocation = "Enter N1 and R, where $N2 = R * N1$."*

R is the ratio of N2 to N1. That is,

$$R = N2/N1$$

Use this value to obtain N2 as a multiple (or proportion) of N1.

N2 is calculated from N1 using the formula:

$$N2 = [R \times N1],$$

where the value $[Y]$ is the next integer $\geq Y$.

For example, setting $R = 2.0$ results in a Group 2 sample size that is double the sample size in Group 1.

R must be greater than 0. If $R < 1$, then N2 will be less than N1; if $R > 1$, then N2 will be greater than N1. You can enter a single value or a series of values.

Total Sample Size (N)

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This is the total sample size, or the sum of the two group sample sizes. This value, along with the percentage of the total sample size in Group 1, implicitly defines N1 and N2.

The total sample size must be greater than one, but practically, must be greater than 3, since each group sample size needs to be at least 2.

You can enter a single value or a series of values.

Percent in Group 1

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This value fixes the percentage of the total sample size allocated to Group 1. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single value or a series of values.

Effect Size – Difference to Detect

D1 (Difference|H1)

Enter a value for the treatment difference to be detected. This difference represents a contrast of interest between two treatments in the study. You may enter a single value or a range of values such as 5 10 20 or 5 to 25 by 5. The items in the list may be separated with commas or blanks.

Effect Size – Repeated Measurements

M (Number of Time Points)

Enter a value for the number of time points (repeated measurements) at which each subject will be observed. You may enter a single value or a range of values such as 3 5 7 or 2 to 8 by 1. The items in the list may be separated with commas or blanks.

Tests for Two Means in a Repeated Measures Design

Effect Size – Covariance Structure**Covariance Type**

Select the within-subject covariance structure that will be used in the mixed models analysis. The options are:

- **Compound Symmetry**

All variances on the diagonal of the within-subject variance-covariance matrix are equal to σ^2 , and all covariances are equal to $\rho\sigma^2$.

- **AR(1)**

All variances on the diagonal of the within-subject variance-covariance matrix are equal to σ^2 , and the covariance between observations t time periods apart is $\rho^t\sigma^2$.

- **Banded(1)**

All variances on the diagonal of the within-subject variance-covariance matrix are equal to σ^2 , and the covariance between observations one time period apart is $\rho\sigma^2$. Covariances between observations more than one time period apart are equal to zero.

- **Simple**

All variances are equal to σ^2 , and all covariances are equal to zero.

Sigma (Std Dev of a Single Observation)

Enter a value for the standard deviation (the square root of the residual variance). This standard deviation is assumed to be equal for the two groups. This parameter is equal to the square root of the sum of the variance parameters when compound symmetry is fit in a mixed models analysis of repeated measures data. This is equal to the square root of the residual variance parameter when an AR(1), Banded(1), or Simple model is fit in a mixed models analysis. You may enter a single value or a range of values such as *5 10 20* or *5 to 25 by 5*. The items in the list may be separated with commas or blanks.

Rho (Autocorrelation)

Enter a value for the correlation between observations on the same subject. When no previous estimate of the within-patient correlation is available, you should use a conservative prediction of the correlation, i.e. a correlation that is higher than anticipated. You may enter a single value or a range of values such as *0.5 0.6 0.7* or *0.4 to 0.9 by 0.1*. The items in the list may be separated with commas or blanks.

Tests for Two Means in a Repeated Measures Design

Example 1 – Determining Power

Researchers are planning a study of the impact of a new drug on heart rate. They want to evaluate the time-averaged difference in heart rate between subjects who take the new drug, and subjects who take the standard drug. Their experimental protocol calls for a baseline heart rate measurement, followed by administration of a certain level of the drug, followed by three additional measurements 30 minutes apart. They want to be able to detect a 10% difference in heart rate between the two treatments.

Similar studies have found an average heart rate of 93 for individuals taking the standard drug, a standard deviation of 9, and an autocorrelation between adjacent measurements on the same individual of 0.7. The researchers assume that first-order autocorrelation adequately represents the autocorrelation pattern. From a heart rate of 93, a 10% reduction gives 83.7, for a difference of 9.3. The test will be conducted at the 0.05 significance level.

What power does the study achieve over a range of possible sample sizes?

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Means in a Repeated Measures Design** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Repeated Measures**, and then clicking on **Tests for Two Means in a Repeated Measures Design**. 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	Power
Alternative Hypothesis	Two-Sided
Alpha.....	0.05
Group Allocation	Equal (N1 = N2)
Sample Size Per Group	4 to 20 by 2
D1	9.3
M.....	4
Covariance Type.....	AR(1)
Sigma.....	9
Rho	0.7

Tests for Two Means in a Repeated Measures Design

Annotated Output

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

Numeric Results

Numeric Results

Two-Sided Test. Null Hypothesis: $D = 0$. Alternative Hypothesis: $D \neq 0$.
Covariance Type = AR(1)

Power	N1	N2	N	M	D1	Sigma	Rho	Alpha
0.42660	4	4	8	4	9.300	9.000	0.700	0.050
0.58468	6	6	12	4	9.300	9.000	0.700	0.050
0.70890	8	8	16	4	9.300	9.000	0.700	0.050
0.80135	10	10	20	4	9.300	9.000	0.700	0.050
0.86742	12	12	24	4	9.300	9.000	0.700	0.050
0.91318	14	14	28	4	9.300	9.000	0.700	0.050
0.94407	16	16	32	4	9.300	9.000	0.700	0.050
0.96448	18	18	36	4	9.300	9.000	0.700	0.050
0.97773	20	20	40	4	9.300	9.000	0.700	0.050

References

Brown, H. and Prescott, R., 2006. Applied Mixed Models in Medicine. 2nd ed. John Wiley & Sons Ltd. Chichester, West Sussex, England. Chapter 6.
Liu, H. and Wu, T., 2005. 'Sample Size Calculation and Power Analysis of Time-Averaged Difference.' Journal of Modern Applied Statistical Methods, Vol. 4, No. 2, pages 434-445.
Diggle, P.J., Liang, K.Y., and Zeger, S.L., 1994. Analysis of Longitudinal Data. Oxford University Press. New York, New York. Chapter 2.

Report Definitions

Power is the probability of rejecting a false null hypothesis.
N1 and N2 are the number of items sampled from each population.
N is the total sample size, $N1 + N2$.
M is the number of time points (repeated measurements) at which each subject is observed.
D1 is the difference between the means of groups 1 and 2 under the alternative hypothesis.
Sigma is the standard deviation of a single observation. It is the same for both groups. Rho is the correlation between observations on the same subject.
Alpha is the probability of rejecting a true null hypothesis.

Summary Statements

Group sample sizes of 4 and 4 achieve 43% power to detect a difference of 9.300 in a design with 4 repeated measurements having a AR(1) covariance structure when the standard deviation is 9.000, the correlation between observations on the same subject is 0.700, and the alpha level is 0.050.

This report gives the power for each value of the other parameters.

Power

This is the computed power for detecting the time-averaged difference between the two group means.

Group 1 Sample Size (N1)

The value of $N1$ is the number of subjects in group 1.

Group 2 Sample Size (N2)

The value of $N2$ is the number of subjects in group 2.

Sample Allocation Ratio (R)

This is the ratio of the number of subjects in group 2 to the number in group 1 ($R = N2/N1$).

Time Points (M)

This is the number of repeated measurements taken.

Difference to be Detected (D1)

This is the treatment difference that is to be detected.

Tests for Two Means in a Repeated Measures Design

Standard Deviation (Sigma)

This is the value of σ , the standard deviation or the square root of the residual variance.

Autocorr. (Rho)

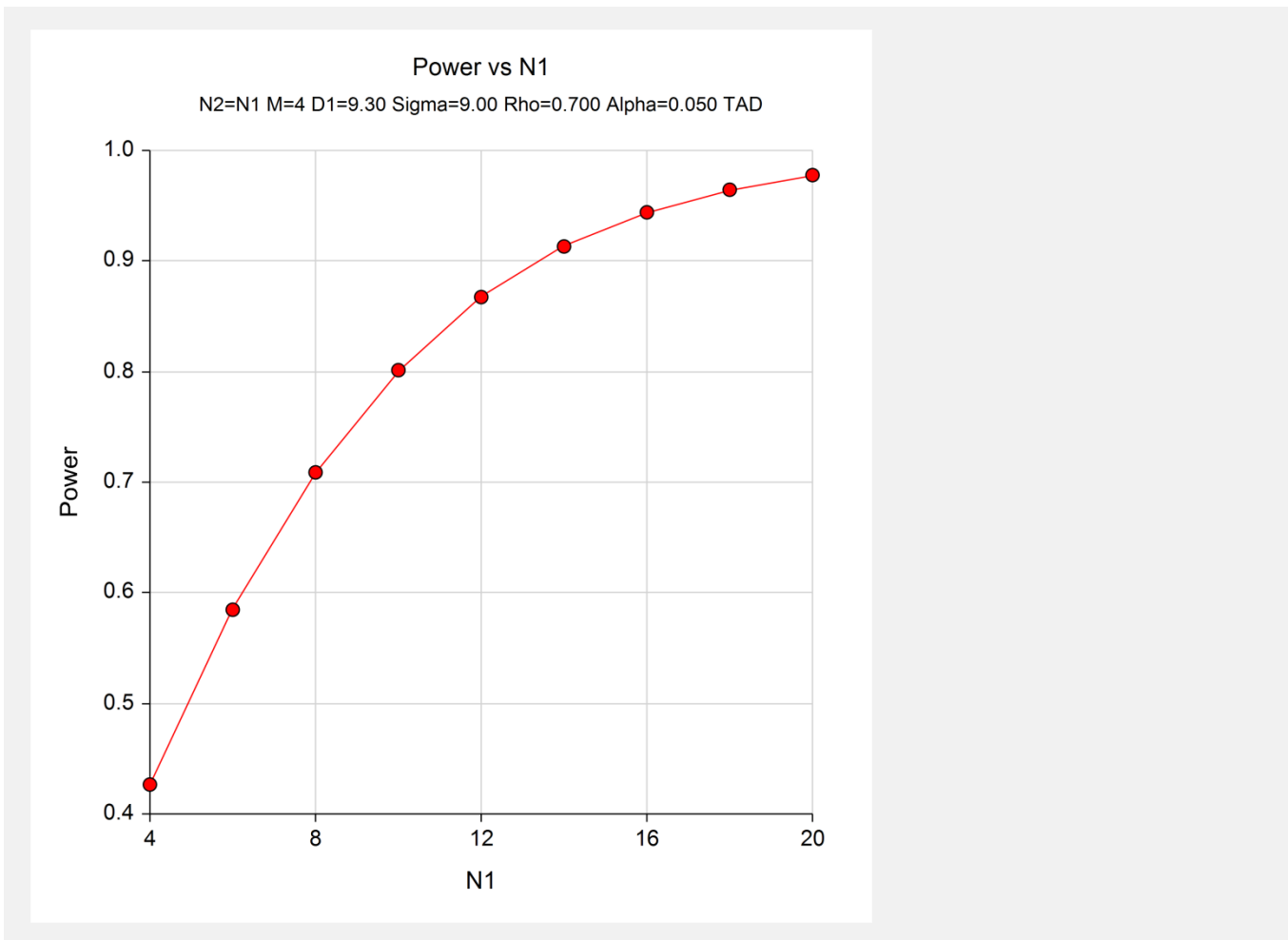
This is the correlation between observations from the same subject.

Alpha

Alpha is the significance level of the test.

Beta

Beta is the probability of failing to reject the null hypothesis when the alternative hypothesis is true.

Plots Section

The chart shows the relationship between power and N1 when the other parameters in the design are held constant.

Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to determine the exact sample size necessary to achieve at least 80% power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Means in a Repeated Measures Design** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Repeated Measures**, and then clicking on **Tests for Two Means in a Repeated Measures Design**. 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	Sample Size
Alternative Hypothesis	Two-Sided
Power	0.80
Alpha	0.05
Group Allocation	Equal (N1 = N2)
D1	9.3
M	4
Covariance Type	AR(1)
Sigma	9
Rho	0.7

Output

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

Numeric Results

Target Power	Actual Power	N1	N2	N	M	D1	Sigma	Rho	Alpha
0.80	0.80135	10	10	20	4	9.300	9.000	0.700	0.050

A group sample size of 10 is required to achieve at least 80% power.

Tests for Two Means in a Repeated Measures Design

Example 3 – Varying the Difference between the Means

Continuing with Examples 1 and 2, the researchers want to evaluate the impact on power of varying the size of the difference between the means for a range of sample sizes from 2 to 8 per group. In the output to follow, we only display the plots. You may want to display the numeric reports as well, but we do not here in order to save space.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Means in a Repeated Measures Design** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Repeated Measures**, and then clicking on **Tests for Two Means in a Repeated Measures Design**. 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	Power
Alternative Hypothesis	Two-Sided
Alpha	0.05
Group Allocation	Equal (N1 = N2)
Sample Size Per Group	4 to 8 by 1
D1	4 to 11 by 1
M	4
Covariance Type.....	AR(1)
Sigma.....	9
Rho	0.7

Output

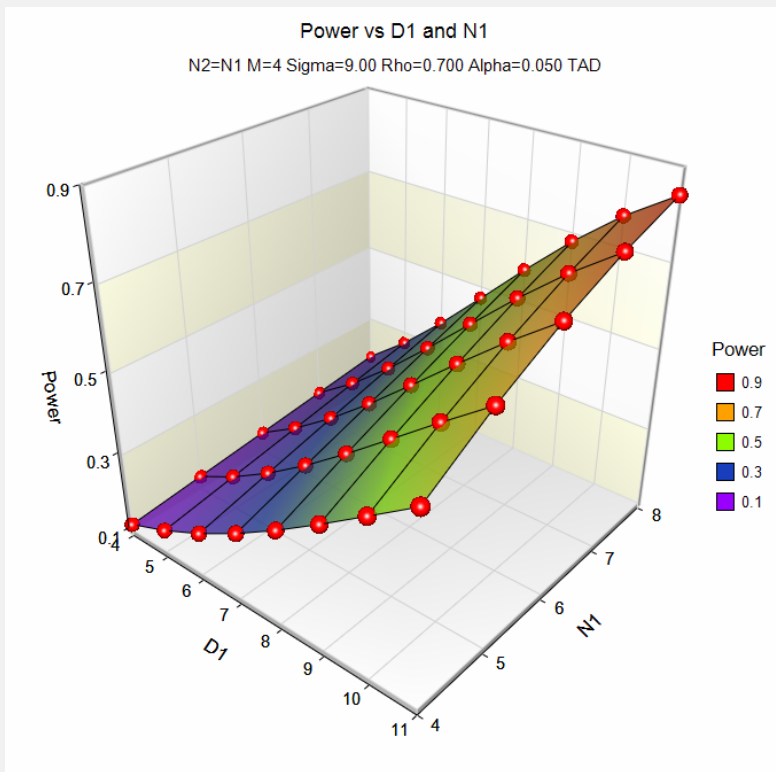
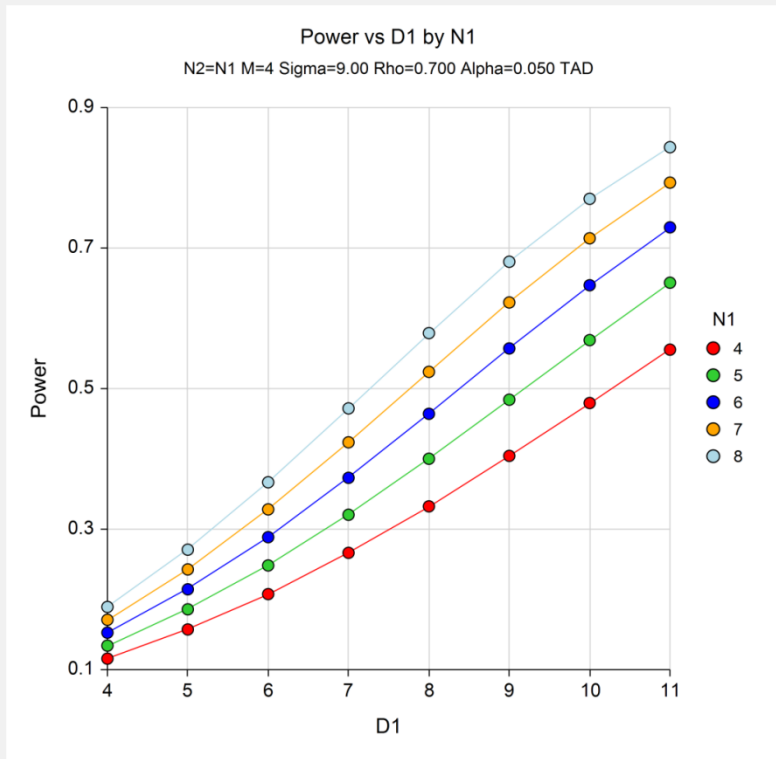
Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Power	N1	N2	N	M	D1	Sigma	Rho	Alpha
0.11574	4	4	8	4	4.000	9.000	0.700	0.050
0.15728	4	4	8	4	5.000	9.000	0.700	0.050
0.20758	4	4	8	4	6.000	9.000	0.700	0.050
0.26631	4	4	8	4	7.000	9.000	0.700	0.050
0.33245	4	4	8	4	8.000	9.000	0.700	0.050
0.40428	4	4	8	4	9.000	9.000	0.700	0.050
0.47949	4	4	8	4	10.000	9.000	0.700	0.050
0.55544	4	4	8	4	11.000	9.000	0.700	0.050
0.13427	5	5	10	4	4.000	9.000	0.700	0.050
0.18591	5	5	10	4	5.000	9.000	0.700	0.050
0.24835	5	5	10	4	6.000	9.000	0.700	0.050
0.32050	5	5	10	4	7.000	9.000	0.700	0.050
0.40016	5	5	10	4	8.000	9.000	0.700	0.050
0.48423	5	5	10	4	9.000	9.000	0.700	0.050
(report continues)								

Tests for Two Means in a Repeated Measures Design

Plots Section



These charts show how the power depends on the difference to be detected, D1, as well as the group sample size, N1.

Tests for Two Means in a Repeated Measures Design

Example 4 – Impact of the Number of Repeated Measurements

Continuing with Examples 1 - 3, the researchers want to study the impact on the sample size if they changing the number of measurements made on each individual. Their experimental protocol calls for four measurements that are 30 minutes apart. They want to see the impact of taking twice that many measurements.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Means in a Repeated Measures Design** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Repeated Measures**, and then clicking on **Tests for Two Means in a Repeated Measures Design**. 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
Alternative Hypothesis	Two-Sided
Power	0.80
Alpha	0.05
Group Allocation	Equal (N1 = N2)
D1	9.3
M	4 8
Covariance Type.....	AR(1)
Sigma.....	9
Rho	0.7

Output

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

Numeric Results

Target Power	Actual Power	N1	N2	N	M	D1	Sigma	Rho	Alpha
0.80	0.80135	10	10	20	4	9.300	9.000	0.700	0.050
0.80	0.84737	8	8	16	8	9.300	9.000	0.700	0.050

Doubling the number of repeated measurements per individual decreases the group sample size by 2. This reduction in sample size may not justify the additional four measurements on each subject.

Tests for Two Means in a Repeated Measures Design

Example 5 – Validation using Diggle et al.

Diggle et al. (1994) page 31 presents an example of calculating the sample size for a TAD study. They calculate the group sample sizes for the cases where d/σ ranges from 0.2 to 0.5, ρ ranges from 0.2 to 0.8, $\alpha = 0.05$, $M = 3$, and power = 0.8. Note that Diggle et al. (1994) uses a one-sided test.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Means in a Repeated Measures Design** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Repeated Measures**, and then clicking on **Tests for Two Means in a Repeated Measures Design**. 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	Sample Size
Alternative Hypothesis	One-Sided
Power	0.80
Alpha	0.05
Group Allocation	Equal (N1 = N2)
D1	0.2 to 0.5 by 0.1
M	3
Covariance Type.....	Compound Symmetry
Sigma.....	1
Rho	0.2 0.5 0.8

Output

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

Numeric Results

Target Power	Actual Power	N1	N2	N	M	D1	Sigma	Rho	Alpha
0.80	0.80178	145	145	290	3	0.200	1.000	0.200	0.050
0.80	0.80154	207	207	414	3	0.200	1.000	0.500	0.050
0.80	0.80012	268	268	536	3	0.200	1.000	0.800	0.050
0.80	0.80475	65	65	130	3	0.300	1.000	0.200	0.050
0.80	0.80154	92	92	184	3	0.300	1.000	0.500	0.050
0.80	0.80270	120	120	240	3	0.300	1.000	0.800	0.050
0.80	0.80885	37	37	74	3	0.400	1.000	0.200	0.050
0.80	0.80321	52	52	104	3	0.400	1.000	0.500	0.050
0.80	0.80012	67	67	134	3	0.400	1.000	0.800	0.050
0.80	0.81343	24	24	48	3	0.500	1.000	0.200	0.050
0.80	0.80028	33	33	66	3	0.500	1.000	0.500	0.050
0.80	0.80109	43	43	86	3	0.500	1.000	0.800	0.050

The sample sizes calculated by **PASS** match the results of Diggle et al. (1994) very closely, with slight differences due to rounding. If you calculate the sample sizes by hand, using the formula given in Diggle et al. (1994), page 31, your answers will match those of **PASS**.

Example 6 – Validation of Sample Size Calculation for Mixed Models Analysis using Brown and Prescott (2006)

Brown and Prescott (2006) pages 268 and 269 presents an example of calculating the sample size for pairwise contrasts in a hypertension trial to be analyzed using mixed models. The analysis of repeated DBP measurements from four post-treatment visits using a compound symmetry covariance pattern resulted in the following covariance matrix for each subject:

$$\mathbf{V}_i = 76 \begin{pmatrix} 1 & 0.53 & 0.53 & 0.53 \\ 0.53 & 1 & 0.53 & 0.53 \\ 0.53 & 0.53 & 1 & 0.53 \\ 0.53 & 0.53 & 0.53 & 1 \end{pmatrix}$$

From this matrix they determine that $\rho = 0.53$ and $\sigma^2 = 76$ ($\sigma = 8.718$).

The trial followed several hundred patients given one of three treatments. Brown and Prescott calculate the group sample size to be 31 for a future study involving four post-treatment visits to detect a difference in DBP of 5 mmHg at the 5% significance level with 80% power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Means in a Repeated Measures Design** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Repeated Measures**, and then clicking on **Tests for Two Means in a Repeated Measures Design**. 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	Sample Size
Alternative Hypothesis	Two-Sided
Power	0.80
Alpha	0.05
Group Allocation	Equal (N1 = N2)
D1	5
M	4
Covariance Type	Compound Symmetry
Sigma	8.718
Rho	0.53

Tests for Two Means in a Repeated Measures Design

Output

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

Numeric Results for M = 4

Target Power	Actual Power	N1	N2	N	M	D1	Sigma	Rho	Alpha
0.80	0.80125	31	31	62	4	5.000	8.718	0.530	0.050

The sample size of 31 calculated by **PASS** matches the results of Brown and Prescott (2006) exactly.

Brown and Prescott further calculate the sample size for the case where no account is taken of repeated measurements and the case of 10 repeated measurements. If we change the number of repeated measurements to 1 and 10, we get the following output (**Example 6b** template):

Numeric Results for M = 1, 10

Target Power	Actual Power	N1	N2	N	M	D1	Sigma	Rho	Alpha
0.80	0.80226	48	48	96	1	5.000	8.718	0.530	0.050
0.80	0.80651	28	28	56	10	5.000	8.718	0.530	0.050

In both cases, the results of **PASS** match those of Brown and Prescott (2006) exactly.