Title

> **power repeated —** Power analysis for repeated-measures analysis of variance

Description

power repeated computes sample size, power, or effect size for one-way or two-way repeated-measures analysis of variance (ANOVA). By default, it computes sample size for given power and effect size. Alternatively, it can compute power for given sample size and effect size or compute effect size for given sample size, power, and number of groups. Also see [PSS] **power** for a general introduction to the power command using hypothesis tests.

Quick start

Sample size for a repeated measures design with one 3-level within-subject factor, a correlation of 0.3 between measurements, and an error variance of 42 with default power of 0.8 and significance level $\alpha = 0.05$

    power repeated 25 27 22, varerror(42) corr(.3)

Same as above, specified as 3 measurements on 1 group with within-subject variance of 4.22

    power repeated, varerror(42) corr(.3) nrepeated(3) ngroups(1) ///
        varwithin(4.22)

Same as above, specified as cell means in matrix cm

    matrix cm = (25,27,22)
    power repeated cm, corr(.3) varerror(42)

As above, and show the mean and covariance matrices in the output

    power repeated cm, corr(.3) varerror(42) showmatrices

Sample size for the between effect in a design with a 3-level within-subject factor and a 2-level between-subject factor

    power repeated 18 14 12\14 13 10, covmatrix(24 9 9\9 24 9\9 9 24)

Same as above, specified as cell means in matrix cm2 and covariances in matrix cov

    matrix cm2 = (18,14,12\14,13,10)
    matrix cov = (24,9,9\9,24,9\9,9,24)
    power repeated cm2, covmatrix(cov)

As above, but for the within effect

    power repeated cm2, covmatrix(cov) factor(within)

As above, but for the between−within effect

    power repeated cm2, covmatrix(cov) factor(bwithin)

Power for a design with one within-subject factor, a sample size of 25, and $\alpha = 0.01$

    power repeated 25 27 22, varerror(42) corr(.3) n(25) alpha(.01)

As above, but for sample sizes of 20, 24, 28, and 32
```
power repeated 25 27 22, varerror(42) corr(.3) n(20(4)32)
```

As above, but show results in a graph of sample size versus power
```
power repeated 25 27 22, varerror(42) corr(.3) n(20(4)32) graph
```

Power for the between effect of a design with a 3-level within-subject factor and a 2-level between-subject factor with a sample size of 160
```
power repeated cm2, covmatrix(cov) n(160)
```

As above, but specify sample sizes of 100 and 140 for groups 1 and 2, respectively
```
power repeated cm2, covmatrix(cov) n1(100) n2(140)
```

Effect size for a one-group repeated-measures design
```
power repeated, varerror(42) corr(.3) n(24) ngroups(1)  ///
      nrepeated(3) power(.8)
```

Effect size for the within-subject effect of a design with a 3-level within-subject factor, a 2-level between-subject factor, and a sample size of 160
```
matrix cov = (24,9,9\9,24,9\9,9,24)
power repeated, covmatrix(cov) n(160) power(.8) ngroups(2)  ///
      factor(within)
```

Menu

Statistics > Power and sample size

Syntax

Compute sample size

> power repeated *meanspec*, *corrspec* [power(*numlist*) *options*]

Compute power

> power repeated *meanspec*, n(*numlist*) *corrspec* [*options*]

Compute effect size

> power repeated, n(*numlist*) power(*numlist*) ngroups(#) *corrspec* [*options*]

where *meanspec* is either a matrix *matname* containing cell means or individual cell means in a matrix form:

$$m_{1,1} \ m_{1,2} \ [\ldots \ m_{1,K}] \ [\backslash \ldots [\backslash \ m_{J,1} \ m_{J,2} \ [\ldots \ m_{J,K}]]]$$

m_{jk}, where $j = 1, 2, \ldots, J$ and $k = 1, 2, \ldots, K$, is the alternative cell mean or the cell mean of the jth row (group) and kth column (repeated measure) under the alternative hypothesis.

matname is the name of a Stata matrix with J rows and K columns containing values of alternative cell means.

At least one group, $J = 1$, and two repeated measures, $K = 2$, must be specified.

where *corrspec* for computing power and sample size is {corr(*numlist*) | covmatrix(*matname*)}, and *corrspec* for computing effect size is {nrepeated(#) corr(*numlist*) | covmatrix(*matname*)}.

options	Description

[Main]
Main

* <u>alpha</u>(*numlist*)	significance level; default is alpha(0.05)
* <u>power</u>(*numlist*)	power; default is power(0.8)
* <u>beta</u>(*numlist*)	probability of type II error; default is beta(0.2)
* n(*numlist*)	total sample size; required to compute power or effect size
<u>nfrac</u>tional	allow fractional sample sizes
* <u>nper</u>group(*numlist*)	number of subjects per group; implies balanced design
* n#(*numlist*)	number of subjects in group *#*
<u>grw</u>eights(*wgtspec*)	group weights; default is one for each group, meaning equal group sizes
<u>ngro</u>ups(*#*)	number of groups
<u>nrep</u>eated(*#*)	number of repeated measures
* <u>corr</u>(*numlist*)	correlation between repeated measures; one of corr() or covmatrix() is required
<u>covm</u>atrix(*matname*)	covariance between repeated measures; one of corr() or covmatrix() is required
<u>fac</u>tor(<u>b</u>etween│<u>w</u>ithin│<u>bw</u>ithin)	tested effect: between, within, or between–within; default is factor(between)
* <u>vareff</u>ect(*numlist*)	variance explained by the tested effect specified in factor()
* <u>varb</u>etween(*numlist*)	variance explained by the between-subjects effect; synonym for factor(between) and vareffect(*numlist*)
* <u>varw</u>ithin(*numlist*)	variance explained by the within-subject effect; synonym for factor(within) and vareffect(*numlist*)
* <u>varbw</u>ithin(*numlist*)	variance explained by the between–within effect; synonym for factor(bwithin) and vareffect(*numlist*)
* <u>varer</u>ror(*numlist*)	error variance; default is varerror(1) when corr() is specified; not allowed with covmatrix()
<u>showm</u>atrices	display cell-means matrix and covariance matrix
<u>showm</u>eans	display cell means
<u>para</u>llel	treat number lists in starred options or in command arguments as parallel when multiple values per option or argument are specified (do not enumerate all possible combinations of values)

[Table]
Table

[<u>no</u>]<u>table</u> [(*tablespec*)]	suppress table or display results as a table; see [PSS] **power, table**
<u>sav</u>ing(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*

[Graph]
Graph

<u>graph</u> [(*graphopts*)]	graph results; see [PSS] **power, graph**

Iteration
<u>init</u>(#)	initial value for sample size or effect size;
	default is to use a bisection algorithm to bound the solution
<u>iter</u>ate(#)	maximum number of iterations; default is iterate(500)
<u>tol</u>erance(#)	parameter tolerance; default is tolerance(1e-12)
<u>ftol</u>erance(#)	function tolerance; default is ftolerance(1e-12)
[no]log	suppress or display iteration log
[no]dots	suppress or display iterations as dots
<u>noti</u>tle	suppress the title

*Specifying a list of values in at least two starred options, or at least two command arguments, or at least one starred option and one argument results in computations for all possible combinations of the values; see [U] **11.1.8 numlist**. Also see the parallel option.

notitle does not appear in the dialog box.

wgtspec	Description
$\#_1 \#_2 \ldots \#_J$	J group weights. Weights must be positive and must be integers unless option nfractional is specified. Multiple values for each group weight $\#_j$ can be specified as a *numlist* enclosed in parentheses.
matname	matrix with J columns containing J group weights. Multiple rows are allowed, in which case each row corresponds to a different set of J weights or, equivalently, column j corresponds to a *numlist* for the jth weight.

where *tablespec* is

   column[:*label*] [*column*[:*label*] [. . .]] [, *tableopts*]

column is one of the columns defined below, and *label* is a column label (may contain quotes and compound quotes).

column	Description	Symbol
alpha	significance level	α
power	power	$1 - \beta$
beta	type II error probability	β
N	total number of subjects	N
N_per_group	number of subjects per group	N/N_g
N_avg	average number of subjects per group	N_{avg}
N#	number of subjects in group #	$N_{\#}$
delta	effect size	δ
N_g	number of groups	N_g
N_rep	number of repeated measurements	N_{rep}
m#$_1$_#$_2$	cell mean (#$_1$, #$_2$): group #$_1$, occasion #$_2$	$\mu_{\#_1, \#_2}$
Var_b	between-subjects variance	σ_b^2
Var_w	within-subject variance	σ_w^2
Var_bw	between–within (group-by-occasion) variance	σ_{bw}^2
Var_be	between-subjects error variance	σ_{be}^2
Var_we	within-subject error variance	σ_{we}^2
Var_bwe	between–within (group-by-occasion) error variance	σ_{bwe}^2
Var_e	error variance	σ_e^2
corr	correlation between repeated measures	ρ
grwgt#	group weight #	$w_{\#}$
target	target parameter; synonym for target effect variance	
_all	display all supported columns	

Column beta is shown in the default table in place of column power if specified.

Column N_per_group is available and is shown in the default table only for balanced designs.

Columns N_avg and N# are shown in the default table only for unbalanced designs.

Columns m#$_1$_#$_2$ are not shown in the default table.

Columns Var_b and Var_be are shown in the default table for the between-subjects test, Var_w and Var_we for the within-subjects test, and Var_bw and Var_bwe for the between–within test.

Columns grwgt# are not shown in the default table.

Options

> ### Main

alpha(), power(), beta(), n(), nfractional; see [PSS] **power**.

npergroup(*numlist*) specifies the group size. Only positive integers are allowed. This option implies a balanced design. npergroup() cannot be specified with n(), n#(), or grweights().

n#(*numlist*) specifies the size of the #th group. Only positive integers are allowed. All group sizes must be specified. For example, all three options n1(), n2(), and n3() must be specified for a design with three groups. n#() cannot be specified with n(), npergroup(), or grweights().

grweights(*wgtspec*) specifies J group weights for an unbalanced design. The weights may be specified either as a list of values or as a matrix, and multiple sets of weights are allowed; see *wgtspec* for details. The weights must be positive and must also be integers unless the nfractional option is specified. grweights() cannot be specified with npergroup() or n#().

ngroups(*#*) specifies the number of groups. This option is required if *meanspec* is not specified. This option is also required for effect-size determination unless grweights() is specified. For a one-way repeated-measures ANOVA, specify ngroups(1).

nrepeated(*#*) specifies the number of repeated measurements within each subject. At least two repeated measurements must be specified. This option is required if the corr() option is specified and *meanspec* is not specified. This option is also required for effect-size determination unless covmatrix() is specified.

corr(*numlist*) specifies the correlation between repeated measurements. corr() cannot be specified with covmatrix(). This option requires the nrepeated() option unless *meanspec* is specified.

covmatrix(*matname*) specifies the covariance matrix between repeated measurements. covmatrix() cannot be specified with corr() or varerror().

factor(between | within | bwithin) specifies the effect of interest for which power and sample-size analysis is to be performed. For a one-way repeated-measures ANOVA, only factor(within) is allowed and is implied when only one group is specified. In a two-way repeated-measures ANOVA, the tested effects include the between effect or main effect of a between-subjects factor, the within effect or main effect of a within-subject factor, and the between–within effect or interaction effect of the between-subjects factor and the within-subject factor. The default for a two-way repeated design is factor(between).

vareffect(*numlist*) specifies the variance explained by the tested effect specified in factor(). For example, if factor(between) is specified, vareffect() specifies the variance explained by the between-subjects factor. This option is required if the factor() option is specified and *meanspec* is not specified. This option is not allowed with the effect-size determination. Only one of vareffect(), varbetween(), varwithin(), or varbwithin() may be specified.

varbetween(*numlist*) specifies the variance explained by the between-subjects factor. This option is equivalent to specifying factor(between) and vareffect(*numlist*) and thus cannot be combined with factor(). This option is not allowed with the effect-size determination. Only one of vareffect(), varbetween(), varwithin(), or varbwithin() may be specified. This option is not allowed when only one group is specified.

varwithin(*numlist*) specifies the variance explained by the within-subject factor. This option is equivalent to specifying factor(within) and vareffect(*numlist*) and thus cannot be combined with factor(). This option is not allowed with the effect-size determination. Only one of vareffect(), varbetween(), varwithin(), or varbwithin() may be specified.

varbwithin(*numlist*) specifies the variance explained by the interaction between a between-subjects factor and a within-subject factor. This option is equivalent to specifying factor(bwithin) and vareffect(*numlist*) and thus cannot be combined with factor(). This option is not allowed with the effect-size determination. Only one of vareffect(), varbetween(), varwithin(), or varbwithin() may be specified. This option is not allowed when only one group is specified.

varerror(*numlist*) specifies the error variance if covmatrix() is not specified. This option is allowed only if corr() is specified. When corr() is specified, the default is varerror(1).

showmatrices specifies that the cell-means matrix and the covariance matrix be displayed, when applicable.

showmeans specifies that the cell means be reported. For a text or graphical output, this option is equivalent to showmatrices except only the cell-mean matrix will be reported. For a tabular output, the columns containing cell means will be included in the default table.

parallel; see [PSS] **power**.

    Table

table, table(), notable; see [PSS] **power, table**.

saving(); see [PSS] **power**.

    Graph

graph, graph(); see [PSS] **power, graph**. Also see the *column* table for a list of symbols used by the graphs.

    Iteration

init(#) specifies the initial value of the sample size for the sample-size determination or the initial value of the effect size δ for the effect-size determination. The default uses a bisection algorithm to bracket the solution.

iterate(), tolerance(), ftolerance(), log, nolog, dots, nodots; see [PSS] **power**.

The following option is available with power repeated but is not shown in the dialog box:

notitle; see [PSS] **power**.

Remarks and examples   stata.com

Remarks are presented under the following headings:

> *Introduction*
> *Using power repeated*
> *Computing sample size*
> *Computing power*
> *Computing effect size and target variance explained by the tested effect*
> *Testing hypotheses about means from multiple dependent populations*

This entry describes the power repeated command and the methodology for power and sample-size analysis for one-way and two-way repeated-measures ANOVA. See [PSS] **intro** for a general introduction to power and sample-size analysis and [PSS] **power** for a general introduction to the power command using hypothesis tests.

Introduction

Repeated-measures ANOVA models are popular among experimenters because of their increased power compared with regular ANOVA models. Repeated-measures designs allow multiple measurements on the same subject. The repeated measurements often correspond to outcomes measured over time for each subject, but they can also correspond to different measurements of the same treatment or measurements of different treatments. The key point is that multiple measurements are made on the same subject.

One example of a repeated-measures design is a longitudinal study that offers an important alternative to a cross-sectional study because of its ability to establish a temporal relationship between the treatment and the outcome. For example, patients with hypertension might be randomized to receive a new experimental drug or standard care and have their systolic blood pressure measured at baseline and each year for two years.

	Baseline	Year 1	Year 2
Old drug	145	135	130
New drug	145	130	120

What makes repeated-measures designs more powerful? Using each subject as his or her own control reduces subject-to-subject variability that is explained by anything other than the effect of the treatment under study. This may dramatically increase power for detecting the effect of the treatment of interest.

Two classes of methods can be used to analyze repeated-measures data: univariate methods such as regular F tests and multivariate methods such as Wilks's lambda test, Pillai's trace test, and the Lawley–Hotelling trace test. The multivariate methods are more flexible in terms of the assumptions about the repeated-measures covariance structure, but they have lower power than regular F tests. In this entry, we concentrate on the univariate methods.

A repeated-measures design belongs to a class of within-subject designs, designs that contain one or more within-subject factors. A within-subject factor is a factor for which each subject receives several or all levels. A between-subjects factor, on the other hand, is any factor for which each subject receives only one level. In what follows, we consider designs with one fixed within-subject factor—one-way fixed-effects repeated-measures ANOVA models—or designs with one fixed between-subjects factor and one fixed within-subjects factor—two-way fixed-effects repeated-measures ANOVA models.

In a one-way repeated-measures model, the within-subject effect is the effect of interest. In a two-way repeated-measures model, you can choose between the three effects of interest: a main between-subjects effect or the between effect, a main within-subject effect or the within effect, and an interaction effect between the between-subjects factor and within-subject factor or the between–within effect. `power repeated` provides power and sample-size computations for the tests of all of these effects.

Repeated-measures ANOVA assumes that errors are normally distributed with zero mean and constant variance. The measurements between subjects are independent, but the measurements within a subject are assumed to be correlated. The within-subject covariance matrices must be constant between groups defined by the levels of between-subjects factors. The validity of the regular F test also relies on the so-called sphericity assumption (or, more generally, the circularity assumption). You can think of this assumption as all differences between levels of the within-subject factor having the same variances. A special case of this assumption is compound symmetry, a less stringent assumption. A covariance matrix is said to have a compound-symmetry structure if all the variances are equal and all the covariances are equal.

The assumption of sphericity is rather restrictive. When it is violated, the distribution of the test statistic of the regular F test of within and between–within effects is no longer an exact F distribution. To compensate for this, several nonsphericity corrections such as the Greenhouse–Geisser correction or Huynh–Feldt correction for the degrees of freedom of the regular F test are proposed (for example, Geisser and Greenhouse [1958]; Huynh and Feldt [1976]).

The distribution of the test statistic under the alternative hypothesis is a noncentral F distribution for all the considered tests. Thus power is a function of the noncentrality parameter, and the noncentrality parameter is a function of the ratio of the variance of the tested effect to the comparison error variance used in the denominator of the corresponding F test. For example, for a test of the within effect, the comparison error variance is the within-effect error variance. In what follows, by comparison error variance, we will imply one of the between-effect, within-effect, or between–within-effect error variance, whichever is appropriate for the considered test. The effect size for each of the F tests is defined as the square root of the ratio of the variance of the tested effect to the comparison error variance.

This entry describes power and sample-size analysis of repeated-measures ANOVA using the univariate F test with Greenhouse–Geisser correction for the nonsphericity.

Using power repeated

power repeated computes sample size, power, or effect size for one-way and two-way fixed-effects repeated-measures ANOVA models. A one-way repeated-measures ANOVA model includes one fixed within-subject factor. The supported two-way repeated-measures ANOVA includes one fixed between-subjects factor and one fixed within-subject factor. A one-way model is available as a special case of a two-way model with one group. At least one group and two repeated measures must be specified.

All computations are performed assuming a significance level of 0.05. You may change the significance level by specifying the alpha() option.

The computations are performed for an F test of the effect of interest. In a one-way model, the only effect of interest is a within-subject effect. In a two-way model, you can choose between the three effects of interest: between-subjects effect with factor(between) (the default), within-subject effect with factor(within), and between–within effect with factor(bwithin).

All computations require that you specify a residual covariance between repeated measures. You can either specify any unstructured covariance matrix in covmatrix() or specify the correlation between repeated measures in corr() and the error variance in varerror(). If corr() is specified, varerror(1) is assumed. The latter specification implies a residual covariance with compound-symmetry structure.

To compute the total sample size, you must also specify the alternative *meanspec* and, optionally, the power of the test in power(). The default power is set to 0.8.

To compute power, you must specify the total sample size in the n() option and the alternative *meanspec*.

Instead of the alternative cell means, you can specify the number of groups (rows) in the ngroups() option, the number of repeated measures (columns) in the nrepeated() option, and the variance explained by the tested effect in the vareffect() option when computing sample size or power. See *Alternative ways of specifying effect* in [PSS] **power twoway**; substitute ngroups() for nrows(), nrepeated() for ncols(), varbetween() for varrow(), varwithin() for varcolumn(), and varbwithin() for varrowcolumn(). If covmatrix() is specified, the nrepeated() option is not required—the number of repeated measures is determined by the dimensionality of the specified covariance matrix.

To compute effect size, the square root of the ratio of the variance explained by the tested factor to the comparison error variance, and the target variance explained by the tested factor, in addition to the residual covariance, you must specify the total sample size in n(), the power in power(), the number of groups in ngroups(), and the number of repeated measures in nrepeated() if corr() is specified.

By default, all computations assume a balanced- or an equal-allocation design. You can use grweights() to specify an unbalanced design for power, sample-size, or effect-size computations. For power and effect-size computations, you can specify individual group sizes in n1(), n2(), and so on, instead of a combination of n() and grweights() to accommodate an unbalanced design. For a balanced design, you can also specify npergroup() to specify a group size instead of a total sample size in n().

In repeated-measures ANOVA, sample size and effect size depend on the noncentrality parameter of the F distribution, and their estimation requires iteration. The default initial values are obtained from a bisection search that brackets the solution. If you desire, you may change this by specifying your own value in the init() option. See [PSS] **power** for the descriptions of other options that control the iteration procedure.

Computing sample size

To compute sample size, you must specify a repeated-measures covariance, an alternative cell means, or the variance of the tested effect and, optionally, the power of the test in the power() option. A power of 0.8 is assumed if power() is not specified.

▷ Example 1: Sample size for a one-way repeated-measures ANOVA

Consider a version of the study described in Winer, Brown, and Michels (1991, 228). Suppose that researchers would like to conduct a similar study to investigate the effects of three drugs on reaction time to a series of standardized tasks. Per design, each subject will receive all three drugs, and a subject's score (mean reaction time to a task) will be recorded for each of the three drugs; that is, there will be three repeated measurements on each subject. This is a simple one-way repeated-measures design in which drug is the within-subject factor. See Winer, Brown, and Michels (1991) for other details of the design.

Before conducting the study, researchers would like to compute the required sample size to detect the effect of interest with 80% power and a 5% significance level. Suppose that the postulated means for the three drug levels are 26.4, 25.6, and 21; the correlation between repeated measurements is 0.6; and the error variance is 77. We use power repeated to compute the sample size:

```
. power repeated 26.4 25.6 21, corr(0.6) varerror(77)
Performing iteration ...

Estimated sample size for repeated-measures ANOVA
F test for within subject with Greenhouse-Geisser correction
Ho: delta = 0   versus   Ha: delta != 0

Study parameters:
        alpha =     0.0500
        power =     0.8000
        delta =     0.7426
          N_g =          1
        N_rep =          3
        means =   <matrix>
        Var_w =     5.6622
       Var_we =    10.2667
        Var_e =    77.0000
          rho =     0.6000

Estimated sample sizes:
            N =         20
  N per group =         20
```

We need to recruit 20 subjects to detect the effect size of $0.7426 = \sqrt{5.6622/10.2667}$ in this study.

Repeated-measures covariance in this study has a compound-symmetry structure by design, so the assumption of sphericity, underlying the F test of means for the within-subject factor, is automatically satisfied. Thus no correction to the degrees of freedom of the test is made.

◁

▷ Example 2: Alternative ways of specifying effect and repeated-measures covariance

Instead of specifying the alternative cell means as in example 1, we can specify the variance between them. Here we also need to specify the number of groups and the number of repeated measures. From example 1, the variance between the means was computed as 5.6622. We specify this value in varwithin(), the number of groups in ngroups(), and the number of repeated measures in nrepeated():

```
. power repeated, ngroups(1) varwithin(5.6622) nrepeated(3) corr(0.6) varerror(77)

Performing iteration ...

Estimated sample size for repeated-measures ANOVA
F test for within subject with Greenhouse-Geisser correction
Ho: delta = 0  versus  Ha: delta != 0

Study parameters:

        alpha =    0.0500
        power =    0.8000
        delta =    0.7426
          N_g =         1
        N_rep =         3
        Var_w =    5.6622
       Var_we =   10.2667
        Var_e =   77.0000
          rho =    0.6000

Estimated sample sizes:

            N =        20
  N per group =        20
```

We obtain the exact same results as in example 1.

Instead of specifying alternative means directly following the command line, we can define a matrix, say, M, containing these means and use it with `power repeated`:

```
. matrix M = (26.4,25.6,21)
. power repeated M, corr(0.6) varerror(77) showmatrices
Performing iteration ...
Estimated sample size for repeated-measures ANOVA
F test for within subject with Greenhouse-Geisser correction
Ho: delta = 0  versus  Ha: delta != 0
Study parameters:
        alpha =    0.0500
        power =    0.8000
        delta =    0.7426
          N_g =         1
        N_rep =         3
        means =   <matrix>
        Var_w =    5.6622
       Var_we =   10.2667
        Var_e =   77.0000
          rho =    0.6000
Study matrices:
  Cell means
                               repeated
                        1          2          3

  groups
              1       26.4       25.6         21

  Covariance
                               repeated
                        1          2          3

  repeated
              1         77
              2       46.2         77
              3       46.2       46.2         77
Estimated sample sizes:
            N =        20
  N per group =        20
```

We used the `showmatrices` option to display the cell-means matrix and the covariance matrix.

We can also use the `covmatrix()` option to specify the repeated-measures covariance matrix. This option allows you to specify unstructured covariance matrices.

We could have typed the values of the covariance matrix displayed above, but instead, we simply retrieve it from the stored result `r(Cov)`. We then display the values of the covariance matrix to verify that we have the correct matrix.

```
. matrix Cov = r(Cov)
. matlist Cov
                 repeated
                        1          2          3

  repeated
              1         77
              2       46.2         77
              3       46.2       46.2         77
```

We specify the covariance matrix in `covmatrix()`:

```
. power repeated M, covmatrix(Cov)
Performing iteration ...

Estimated sample size for repeated-measures ANOVA
F test for within subject with Greenhouse-Geisser correction
Ho: delta = 0  versus  Ha: delta != 0

Study parameters:

        alpha =    0.0500
        power =    0.8000
        delta =    0.7426
         N_g =         1
       N_rep =         3
       means =  <matrix>
       Var_w =    5.6622
      Var_we =   10.2667
         Cov =  <matrix>
    spherical =     true

Estimated sample sizes:

            N =        20
 N per group =        20
```

We obtain the exact same results as before.

◁

▷ Example 3: Sample size for a two-way repeated-measures ANOVA—between effect

A group of researchers would like to design a study to determine whether a new antihypertension medication is more effective than the best medication currently available. They plan their study based on the design and results of the ALLHAT clinical trial (1996, 2002). Average systolic blood pressure (SBP) is assumed to be 145 mm/Hg at baseline in both treatment groups. Using the results of the ALLHAT study, the researchers expect a mean SBP of 135 at year 1 and 130 at year 2 in the old drug group. Using the results of pilot studies, the researchers expect a mean SBP of 130 at year 1 and 120 at year 2 in the new drug group.

	Baseline	Year 1	Year 2
Old drug	145	135	130
New drug	145	130	120

There are two factors in this experiment: treatment group is the between-subjects factor, and measurement time (baseline, year 1, and year 2) is the within-subject factor. Using the ALLHAT study and the pilot data, the researchers assume that the variance of SBP will be 225 for both groups at each of the three measurements. They also assume that the correlation between the repeated measurements is 0.7, so the covariance matrix is

$$\Sigma = \begin{bmatrix} 225 & 157.5 & 157.5 \\ 157.5 & 225 & 157.5 \\ 157.5 & 157.5 & 225 \end{bmatrix}$$

There are potentially three tests of interest here: the test of the main effect of treatment, the test of the main effect of time, and the test of the interaction effect between treatment and time.

Let's compute the required sample size for the test of the between effect, treatment. This is the default test in `power repeated` when there is more than one group.

We begin by defining a matrix of means and a covariance matrix.

```
. matrix M    = (145,135,130\145,130,120)
. matrix Cov = (225,157.5,157.5\157.5,225,157.5\157.5,157.5,225)
```

We can use the `matlist` command to display these matrices to verify that we typed them correctly:

```
. matlist M
                     c1          c2          c3
         r1          145         135         130
         r2          145         130         120
. matlist Cov
                     c1          c2          c3
         r1          225
         r2        157.5         225
         r3        157.5       157.5         225
```

For brevity, we use one of the alternative specifications from example 2 to compute sample size. We specify the cell-means matrix `M` following the command name and the covariance matrix `Cov` in `covmatrix()`:

```
. power repeated M, covmatrix(Cov)
Performing iteration ...
Estimated sample size for repeated-measures ANOVA
F test for between subjects
Ho: delta = 0   versus   Ha: delta != 0
Study parameters:
         alpha =     0.0500
         power =     0.8000
         delta =     0.1863
          N_g =          2
        N_rep =          3
        means =    <matrix>
        Var_b =     6.2500
       Var_be =   180.0000
          Cov =    <matrix>
Estimated sample sizes:
            N =        228
  N per group =        114
```

To detect the treatment effect of the specified magnitude, $\delta = 0.1863 = \sqrt{6.25/180}$, we need to enroll 228 subjects with 114 subjects per treatment. Note that the sphericity requirement is not needed for the F test of between effects, so no correction is done to the degrees of freedom of the test.

By default, `power repeated` does not display the specified matrices. If desired, we can use the `showmatrices` option to display them:

```
. power repeated M, covmatrix(Cov) showmatrices

Performing iteration ...

Estimated sample size for repeated-measures ANOVA
F test for between subjects
Ho: delta = 0   versus   Ha: delta != 0

Study parameters:
        alpha =    0.0500
        power =    0.8000
        delta =    0.1863
          N_g =         2
        N_rep =         3
        means =   <matrix>
        Var_b =    6.2500
       Var_be =  180.0000
          Cov =   <matrix>

Study matrices:

  Cell means

                    |          repeated
                    |     1        2        3
  ------------------+-------------------------------
  groups            |
                1   |    145      135      130
                2   |    145      130      120

  Covariance

                    |          repeated
                    |     1        2        3
  ------------------+-------------------------------
  repeated          |
                1   |    225
                2   |  157.5      225
                3   |  157.5    157.5      225

Estimated sample sizes:
            N =       228
  N per group =       114
```

Similar to the alternative specifications discussed in example 2, all the specifications below will produce identical results:

```
. power repeated 145 135 130 \ 145 130 120, covmatrix(Cov)
(output omitted)
. power repeated M, corr(0.7) varerror(225)
(output omitted)
. power repeated, nrepeated(3) corr(0.7) varerror(225) ngroups(2) varbetween(6.25)
(output omitted)
```

◁

▷ Example 4: Sample size for a two-way repeated-measures ANOVA—within effect

Continuing with example 3, we now compute the required sample size for the test of the main effects of time, the within effects.

```
. power repeated M, covmatrix(Cov) factor(within)
Performing iteration ...
Estimated sample size for repeated-measures ANOVA
F test for within subject with Greenhouse-Geisser correction
Ho: delta = 0  versus  Ha: delta != 0
Study parameters:
        alpha =    0.0500
        power =    0.8000
        delta =    1.7392
          N_g =         2
        N_rep =         3
        means =   <matrix>
        Var_w =   68.0556
       Var_we =   22.5000
          Cov =   <matrix>
    spherical =      true
Estimated sample sizes:
            N =         6
  N per group =         3
```

We only need a total of 6 subjects, 3 per group, to detect the within effect in this study.

We can also obtain identical results by using the following alternative specification:

```
. power repeated, covmatrix(Cov) ngroups(2) varwithin(68.0556)
  (output omitted )
```

◁

▷ Example 5: Sample size for a two-way repeated-measures ANOVA—between–within effect

Continuing with example 3, we can also compute the required sample size for the test of the between–within interaction effects, interaction between treatment and time.

```
. power repeated M, covmatrix(Cov) factor(bwithin)
Performing iteration ...
Estimated sample size for repeated-measures ANOVA
F test for between-within subjects with Greenhouse-Geisser correction
Ho: delta = 0  versus  Ha: delta != 0
Study parameters:
        alpha =    0.0500
        power =    0.8000
        delta =    0.4303
          N_g =         2
        N_rep =         3
        means =   <matrix>
       Var_bw =    4.1667
      Var_bwe =   22.5000
          Cov =   <matrix>
    spherical =      true
Estimated sample sizes:
            N =        54
  N per group =        27
```

For this test, we need a total of 54 subjects with 27 subjects per group.

If we are interested in performing all three tests (between, within, and between–within) during our analysis, we should pick the largest of the three sample sizes as our final sample size. In our examples, the largest sample size is 228 for the test of between effects.

We can also obtain results identical to the above by using the following alternative specification:

```
. power repeated, covmatrix(Cov) ngroups(2) varbwithin(4.1667)
(output omitted )
```

◁

▷ Example 6: Unbalanced design

Continuing with example 2, suppose we anticipate that the first group will have twice as many subjects as the second group. We can accommodate this unbalanced design by specifying the corresponding group weights in grweights():

```
. power repeated M, covmatrix(Cov) grweights(2 1)
Performing iteration ...
Estimated sample size for repeated-measures ANOVA
F test for between subjects
Ho: delta = 0  versus  Ha: delta != 0
Study parameters:
        alpha =    0.0500
        power =    0.8000
        delta =    0.1757
          N_g =         2
        N_rep =         3
        means =   <matrix>
        Var_b =    5.5556
       Var_be =  180.0000
          Cov =   <matrix>
Estimated sample sizes:
            N =       258
    Average N =  129.0000
           N1 =       172
           N2 =        86
```

The required total sample size for this unbalanced design is 258 with 172 subjects in the first group and 86 subjects in the second group. The average number of subjects per group is 129.

We can compute results for multiple sets of group weights. The specification of group weights within grweights() is exactly the same as the specification of group means described in *Alternative ways of specifying effect*. Suppose that we would like to compute sample sizes for two unbalanced designs. The first design has twice as many subjects in the first group, and the second design has twice as many subjects in the second group. We specify multiple group weights for the first and second groups in parentheses. We also specify the parallel option to treat multiple weight values in parallel instead of computing results for all possible combinations of these values that would have been done by default.

```
. local columns alpha power N N1 N2 grwgt1 grwgt2 delta N_rep Var_b Var_be

. power repeated M, covmatrix(Cov) grweights((2 1) (1 2)) parallel
> table(`columns', formats("%6.0g"))

Performing iteration ...

Estimated sample size for repeated-measures ANOVA
F test for between subjects
Ho: delta = 0  versus  Ha: delta != 0

   means = <matrix>
     Cov = <matrix>
```

alpha	power	N	N1	N2	grwgt1	grwgt2	delta	N_rep	Var_b	Var_be
.05	.8	258	172	86	2	1	.1757	3	5.556	180
.05	.8	258	86	172	1	2	.1757	3	5.556	180

The default table does not include group weights, so we request a table with custom columns containing group weights via table(). We also request a smaller format to make the table more compact.

◁

Computing power

To compute power, you must specify a repeated-measures covariance, the total sample size in n(), and the alternative cell means or the variance of the tested effect.

▷ Example 7: Power for a two-way repeated-measures ANOVA

The team discovers that they are only able to recruit a maximum of $n = 200$ participants. They would like to calculate the statistical power for the between-subjects effect given this constraint and assuming a balanced design.

```
. power repeated M, covmatrix(Cov) n(200)

Estimated power for repeated-measures ANOVA
F test for between subjects
Ho: delta = 0  versus  Ha: delta != 0

Study parameters:

        alpha =    0.0500
            N =       200
  N per group =       100
        delta =    0.1863
          N_g =         2
        N_rep =         3
        means =  <matrix>
        Var_b =    6.2500
       Var_be =  180.0000
          Cov =  <matrix>

Estimated power:

        power =    0.7462
```

The power corresponding to this design is 75%.

◁

▷ Example 8: Multiple values of study parameters

Continuing with example 7, suppose that the researchers would like to know whether randomizing 60% of the participants to the new drug group and 40% to the old drug group will have an effect on statistical power. For comparison, we will also include the results from a balanced design.

To accommodate this unbalanced design, we could use grweights(), as we demonstrated in example 6. For variety, we instead use n1() and n2() to specify unequal group sizes directly. We also display only a subset of table columns, including power and sample sizes.

```
. power repeated M, covmat(Cov) n1(100 80) n2(100 120) parallel
> table(power N1 N2 N)
Estimated power for repeated-measures ANOVA
F test for between subjects
Ho: delta = 0  versus  Ha: delta != 0

  means = <matrix>
    Cov = <matrix>
```

power	N1	N2	N
.7462	100	100	200
.7289	80	120	200

For the specified unbalanced design, the power decreases slightly to 73% from 75%.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [PSS] **power, table**. If you wish to produce a power plot, see [PSS] **power, graph**.

◁

Computing effect size and target variance explained by the tested effect

Sometimes, we may be interested in determining the smallest effect that yields a statistically significant result for prespecified sample size and power. In this case, repeated-measures covariance, power, sample size, the numbers of groups, and possibly the number of repeated measurements must be specified.

The effect size in power repeated is defined as a square root of the ratio of the variance explained by the tested effect to the comparison error variance. The effect size and the target variance explained by the tested effect are computed.

▷ Example 9: Effect size for a two-way repeated-measures ANOVA

Continuing with example 7, suppose that researchers would like to know how large the between-subjects variance must be to achieve a power of 80% with a total sample size of 200 using a balanced design.

```
. power repeated, covmat(Cov) n(200) power(0.8) ngroups(2)

Performing iteration ...

Estimated between-subjects variance for repeated-measures ANOVA
F test for between subjects
Ho: delta = 0  versus  Ha: delta != 0

Study parameters:

        alpha =    0.0500
        power =    0.8000
            N =       200
  N per group =       100
          N_g =         2
        N_rep =         3
       Var_be =  180.0000
          Cov =   <matrix>

Estimated effect size and between-subjects variance:

        delta =    0.1991
        Var_b =    7.1331
```

We see that to achieve a power of at least 80%, the between-subjects variance must increase to 7.1331 from 6.250, which achieved a power of 0.7462 in example 7. The effect size increases from 0.1863 to 0.1991.

◁

Testing hypotheses about means from multiple dependent populations

After the data are collected, we can use Stata's anova command, for example, to perform inference for repeated-measures ANOVA. We show a quick example of how to do this here; see [R] **anova** for more examples and details.

▷ Example 10: One-way repeated-measures ANOVA

Suppose that researchers conduct their study and collect the data. Consider the data from Winer, Brown, and Michels (1991, 228), a version of which was discussed in example 1.

t43.dta contains 20 observations of scores of 4 repeated measurements identified by the drug variable from 5 people identified by the person variable. We use the anova command to fit a one-way repeated-measures model to these data.

```
. use http://www.stata-press.com/data/r15/t43
(T4.3 -- Winer, Brown, Michels)

. anova score person drug, repeated(drug)
                         Number of obs =         20    R-squared     =  0.9244
                         Root MSE      =    3.06594    Adj R-squared =  0.8803

          Source │  Partial SS         df         MS          F     Prob>F

           Model │       1379           7         197      20.96    0.0000

          person │      680.8           4       170.2      18.11    0.0001
            drug │      698.2           3   232.73333      24.76    0.0000

        Residual │      112.8          12         9.4

           Total │     1491.8          19   78.515789

Between-subjects error term:  person
                    Levels:  5          (4 df)
     Lowest b.s.e. variable:  person
Repeated variable: drug
                                   Huynh-Feldt epsilon         =  1.0789
                                   *Huynh-Feldt epsilon reset to 1.0000
                                   Greenhouse-Geisser epsilon  =  0.6049
                                   Box's conservative epsilon  =  0.3333

                                   ───────── Prob > F ─────────
          Source │    df      F    Regular     H-F      G-G       Box

            drug │     3    24.76   0.0000    0.0000   0.0006    0.0076
        Residual │    12
```

We are interested in the test of the effect of drug. The regular F test reports a significant result. The anova output for the repeated variable drug, however, indicates that the sphericity assumption is not met in these data; for example, the Greenhouse–Geisser epsilon of 0.6049 is different from 1.

When the sphericity assumption is not met, the degrees of freedom of a regular F test must be adjusted. Even after the adjustment, the effect of a drug is still significant according to all tests, at least at the 1% level.

To design a new study based on the results of this experiment, we can use power repeated to compute the required sample size. To perform this computation, we will need the estimates of the repeated-measures covariance and within-drug score means.

anova saves the estimated repeated-measures covariance in e(Srep). We save it to a new matrix Cov and display it:

```
. mat Cov = e(Srep)

. matlist Cov

              │       c1         c2         c3         c4

           r1 │     76.8
           r2 │     53.2       42.8
           r3 │     29.2       15.8       14.8
           r4 │       69         47         27         64
```

We now use the mean command to estimate means for each of the four drug levels. We store the resulting matrix of means in M:

```
. mean score, over(drug)

Mean estimation                    Number of obs    =         20

              1: drug = 1
              2: drug = 2
              3: drug = 3
              4: drug = 4
```

Over	Mean	Std. Err.	[95% Conf. Interval]	
score				
1	26.4	3.919184	18.19705	34.60295
2	25.6	2.925748	19.47634	31.72366
3	15.6	1.720465	11.99903	19.20097
4	32	3.577709	24.51177	39.48823

```
. mat M = e(b)
```

We now specify the obtained matrices with `power repeated` to compute the sample size:

```
. power repeated M, covmatrix(Cov)

Performing iteration ...

Estimated sample size for repeated-measures ANOVA
F test for within subject with Greenhouse-Geisser correction
Ho: delta = 0   versus   Ha: delta != 0

Study parameters:

        alpha =     0.0500
        power =     0.8000
        delta =     3.8543
          N_g =          1
        N_rep =          4
        means =    <matrix>
        Var_w =    34.9100
       Var_we =     2.3500
          Cov =    <matrix>
    spherical =      false

Estimated sample sizes:
            N =          4
  N per group =          4
```

We only need 4 subjects to detect the effect of a drug in a study with 80% power and a 5% significance level.

◁

Stored results

power repeated stores the following in r():

Scalars

r(alpha)	significance level
r(power)	power
r(beta)	probability of a type II error
r(delta)	effect size
r(N)	total sample size
r(N_a)	actual sample size
r(N_avg)	average sample size
r(N#)	number of subjects in group #
r(N_per_group)	number of subjects per group
r(N_g)	number of groups
r(nfractional)	1 if nfractional is specified, 0 otherwise
r(balanced)	1 for a balanced design, 0 otherwise
r(grwgt#)	group weight #
r(N_rep)	number of rows
r(m#$_1$_#$_2$)	cell mean (#$_1$, #$_2$)
r(Var_b)	between-subjects variance
r(Var_w)	within-subject variance
r(Var_bw)	between–within subjects, interaction variance
r(Var_be)	between-subjects error variance
r(Var_we)	within-subject error variance
r(Var_bwe)	between–within subjects, interaction error variance
r(Var_e)	error variance
r(spherical)	1 covariance is spherical, 0 otherwise
r(epsilon)	nonsphericity correction
r(epsilon_m)	mean nonsphericity correction
r(separator)	number of lines between separator lines in the table
r(divider)	1 if divider is requested in the table; 0 otherwise
r(init)	initial value for sample size or effect size
r(maxiter)	maximum number of iterations
r(iter)	number of iterations performed
r(tolerance)	requested parameter tolerance
r(deltax)	final parameter tolerance achieved
r(ftolerance)	requested distance of the objective function from zero
r(function)	final distance of the objective function from zero
r(converged)	1 if iteration algorithm converged, 0 otherwise

Macros

r(type)	test
r(method)	repeated
r(factor)	between, within, or bwithin
r(columns)	displayed table columns
r(labels)	table column labels
r(widths)	table column widths
r(formats)	table column formats

Matrices

r(pss_table)	table of results
r(means)	cell-means matrix
r(Cov)	repeated-measures covariance

Methods and formulas

Consider a sample of n units where each observation comprises q responses based on p predictors. A general linear multivariate model can then be expressed as

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \tag{1}$$

where $\mathbf{Y}$ is an $n \times q$ matrix of dependent variables, $\mathbf{X}$ is an $n \times p$ matrix of fixed predictor variables, $\mathbf{B}$ is a $p \times q$ matrix of coefficients, and the error $\mathbf{E}$ is an $n \times q$ matrix where each row is an independent and identically distributed random variable drawn from a p-dimensional multivariate normal with mean $\mathbf{0}$ and a variance–covariance matrix $\boldsymbol{\Sigma}$. In our repeated measures design, $q = K$ is the number of repeated measures within $p = J$ treatments or groups.

For expositional purposes, consider a two-way repeated-measures design with one between-subjects factor, treatment, and one within-subject factor, time. Suppose we wish to test the effect of a treatment with three levels. The response of each individual is measured at the beginning of the experiment and at three time periods after one of the three treatments is administered. To put this into perspective, we see that $K = 4$ is the number of repeated measures and that $J = 3$ is the number of columns of treatment levels. We can express this model as

$$
\begin{bmatrix}
y_{1,1} & y_{1,2} & y_{1,3} & y_{1,4} \\
y_{2,1} & y_{2,2} & y_{2,3} & y_{2,4} \\
. & . & . & . \\
. & . & . & . \\
. & . & . & . \\
y_{n,1} & y_{n,2} & y_{n,3} & y_{n,4}
\end{bmatrix}
=
\begin{bmatrix}
1 & x_{1,1} & x_{1,2} \\
1 & x_{2,1} & x_{2,2} \\
. & . & . \\
. & . & . \\
. & . & . \\
1 & x_{n,1} & x_{n,2}
\end{bmatrix}
\begin{bmatrix}
\mu_1 & \mu_2 & \mu_3 & \mu_4 \\
\alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} & \alpha_{1,4} \\
\alpha_{2,1} & \alpha_{2,2} & \alpha_{2,3} & \alpha_{2,4}
\end{bmatrix}
+
\begin{pmatrix}
\boldsymbol{\epsilon}_1' \\
\boldsymbol{\epsilon}_2' \\
. \\
. \\
\boldsymbol{\epsilon}_n'
\end{pmatrix}
$$

where $y_{i,k}$ is the response of the ith individual at time period $k = 1, 2, 3, 4$ and

$$
x_{i1} =
\begin{cases}
1 & \text{if subject } i \text{ received treatment 1} \\
0 & \text{if subject } i \text{ received treatment 2} \\
-1 & \text{if subject } i \text{ received treatment 3}
\end{cases}
$$

$$
x_{i2} =
\begin{cases}
0 & \text{if subject } i \text{ received treatment 1} \\
1 & \text{if subject } i \text{ received treatment 2} \\
-1 & \text{if subject } i \text{ received treatment 3}
\end{cases}
$$

represent the effects of a treatment for individual i. The elements in the coefficient matrix $\mathbf{B}$ have the following interpretation: μ_k is the mean-treatment response at time period k, $\alpha_{j,k}$ is the jth treatment effect, $j = 1, 2$, at time period k, and $\alpha_{3,k} = -\alpha_{1,k} - \alpha_{2,k}$. The treatment-by-time means are $\boldsymbol{\mu} = \mathbf{XB}$. The $\boldsymbol{\epsilon}_i$ are independent normal vectors of length K with mean $\mathbf{0}$ and variance–covariance $\boldsymbol{\Sigma}$.

Methods and formulas are presented under the following headings:

> *Hypothesis testing*
> *Computing power*

Hypothesis testing

A hypothesis test for a general linear multivariate model can be formed as

$$
H_0: \boldsymbol{\Theta} = \mathbf{0} \qquad H_a: \boldsymbol{\Theta} \neq \mathbf{0}
$$

where $\boldsymbol{\Theta} = \mathbf{CBU}$ is a $d_c \times d_u$ matrix with arbitrary dimensions d_c and d_u that depend on the specified contrast matrices $\mathbf{C}$ and $\mathbf{U}$. $\mathbf{C}$ is a $d_c \times p$ matrix of full rank, where $\text{rank}(\mathbf{C}) = d_c \leq p$, and $\mathbf{U}$ is a $q \times d_u$ matrix of full rank, where $\text{rank}(\mathbf{U}) = d_u \leq q$.

Each row of $\mathbf{C}$ corresponds to the row of $\boldsymbol{\Theta}$ and forms a contrast to test the between-subjects effects. Similarly, each column of $\mathbf{U}$ corresponds to the column of $\boldsymbol{\Theta}$ and forms a contrast to test the within-subject effect. Together, $\mathbf{C}$ and $\mathbf{U}$ can also be used to test for interaction effects.

The estimates are given by

$$\widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$\widehat{\mathbf{\Theta}} = \mathbf{C}\widehat{\mathbf{B}}\mathbf{U} \tag{2}$$

$$\widehat{\mathbf{H}} = \widehat{\mathbf{\Theta}}'\{\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\}^{-1}\widehat{\mathbf{\Theta}} \tag{3}$$

Define $\widehat{\mathbf{E}} = \mathbf{U}'\widehat{\mathbf{\Sigma}}\mathbf{U}(n-p)$. Then, under the assumption of sphericity, the test statistic is given by

$$F_{C,U} = \frac{\text{tr}(\widehat{\mathbf{H}})/d_c d_u}{\text{tr}(\widehat{\mathbf{E}})/\{d_u(n-p)\}} \tag{4}$$

where the statistic follows an F distribution with $d_c d_u$ numerator and $d_u(n-p)$ denominator degrees of freedom. However, if the assumption is not met, then the test statistic follows an F distribution with $d_c d_u \varepsilon$ numerator and $d_u(n-p)\varepsilon$ denominator degrees of freedom, where

$$\varepsilon = \frac{\text{tr}^2(\widehat{\mathbf{\Sigma}})}{d_u \, \text{tr}(\widehat{\mathbf{\Sigma}}^2)} = \frac{\left(\sum_{k=1}^{d_u} \lambda_k\right)^2}{d_u \sum_{k=1}^{d_u} \lambda_k^2}$$

Under the alternative hypothesis, the power is obtained using a noncentral F distribution with noncentrality parameter equal to

$$\lambda = d_c d_u \epsilon F_{C,U}$$

Computing power

To compute power, we make conjectures about the parameters of interest, $\mathbf{B}$ and $\mathbf{\Sigma}$, and rewrite as

$$\mathbf{\Theta} = \mathbf{C}\mathbf{B}\mathbf{U}$$

$$\mathbf{H} = n\mathbf{\Theta}'\{\mathbf{C}(\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{X}})^{-1}\mathbf{C}'\}^{-1}\mathbf{\Theta}$$

$$= n\mathbf{H}_*$$

$$\mathbf{E} = \mathbf{U}'\mathbf{\Sigma}\mathbf{U}(n-p)$$

$$= \mathbf{\Sigma}_*(n-p)$$

where $\ddot{\mathbf{X}}$ is the $p \times p$ model matrix containing all the unique rows of $\mathbf{X}$ in a special order, and $\mathbf{W}$ is a diagonal matrix containing n_j/n, the sample size for the jth treatment divided by the total sample size. In our three-treatment example, the matrix $\ddot{\mathbf{X}}$ is

$$\ddot{\mathbf{X}} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{pmatrix}$$

The $F_{C,U}$ statistic using the parameter matrices $\mathbf{H}$ and $\mathbf{E}$ is

$$
\begin{aligned}
F_{C,U} &= \frac{\text{tr}(\mathbf{H})/d_c d_u}{\text{tr}(\mathbf{E})/\{d_u(n-p)\}} \\
&= \frac{n}{d_c} \frac{\text{tr}(\mathbf{H}_*)}{\text{tr}(\boldsymbol{\Sigma}_*)}
\end{aligned}
$$

from which we obtain the noncentrality parameter as

$$
\begin{aligned}
\lambda &= d_u d_c \epsilon F_{C,U} \\
&= n\epsilon\delta^2
\end{aligned} \tag{5}
$$

where the effect size δ is defined as $\delta = \sqrt{d_u \text{tr}(\mathbf{H}_*)/\text{tr}(\boldsymbol{\Sigma}_*)}$.

The effect variance (`Var_b`, `Var_w`, or `Var_bw`) reported by `power repeated` is computed as $\text{tr}(\mathbf{H}_*)/d_c$. The effect error variance (`Var_be`, `Var_we`, or `Var_bwe`) is computed as $\text{tr}(\boldsymbol{\Sigma}_*)/(d_c d_u)$.

Under the alternative hypothesis, the test statistic in (4) is distributed as a noncentral F distribution with $d_c d_u \epsilon$ numerator and $d_u(n-p)\epsilon$ denominator degrees of freedom and noncentrality parameter λ from (5).

The power of the overall F test is

$$
1 - \beta = F_{d_c d_u \epsilon, d_u(n-p)\epsilon, \lambda}\left(F_{d_c d_u \epsilon_m, d_u(n-p)\epsilon_m, 1-\alpha}\right) \tag{6}
$$

where $F_{\cdot,\cdot,\lambda}(\cdot)$ is the cdf of a noncentral F distribution, and $\epsilon_m = E(\epsilon)$ is computed as described in Muller and Barton (1989, 551).

Total sample size and effect size are obtained by iteratively solving the nonlinear equation (6). When the `grweights()` option is specified, a constant multiplier n_c is computed and rounded to an integer unless the `nfractional` option is specified. The group sizes are then computed as $\widetilde{w}_j n_c$, where $\widetilde{w}$ is a standardized weight; see *Methods and formulas* of [PSS] **power oneway** for details. The actual sample size, `N_a`, is the sum of the group sizes.

See Muller et al. (1992) for details.

References

ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. 2002. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: The antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). *Journal of the American Medical Association* 288: 2981–2997.

Davis, B. R., J. A. Cutler, D. J. Gordon, C. D. Furberg, J. T. Wright, Jr, W. C. Cushman, R. H. Grimm, J. LaRosa, P. K. Whelton, H. M. Perry, M. H. Alderman, C. E. Ford, S. Oparil, C. Francis, M. Proschan, S. Pressel, H. R. Black, and C. M. Hawkins, for the ALLHAT Research Group. 1996. Rationale and design for the antihypertensive and lipid lowering treatment to prevent heart attack trial (ALLHAT). *American Journal of Hypertension* 9: 342–360.

Geisser, S., and S. W. Greenhouse. 1958. An extension of Box's results on the use of the F distribution in multivariate analysis. *Annals of Mathematical Statistics* 29: 885–891.

Grayling, M. J., J. M. S. Wason, and A. P. Mander. 2018. Group sequential clinical trial designs for normally distributed outcome variables. *Stata Journal* 18: 416–431.

Huynh, H., and L. S. Feldt. 1976. Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics* 1: 69–82.

Muller, K. E., and C. N. Barton. 1989. Approximate power for repeated-measures ANOVA lacking sphericity. *Journal of the American Statistical Association* 84: 549–555.

Muller, K. E., L. M. LaVange, S. Landesman Ramey, and C. T. Ramey. 1992. Power calculations for general linear multivariate models including repeated measures applications. *Journal of the American Statistical Association* 87: 1209–1226.

Winer, B. J., D. R. Brown, and K. M. Michels. 1991. *Statistical Principles in Experimental Design*. 3rd ed. New York: McGraw–Hill.

Also see

[PSS] **power** — Power and sample-size analysis for hypothesis tests

[PSS] **power oneway** — Power analysis for one-way analysis of variance

[PSS] **power pairedmeans** — Power analysis for a two-sample paired-means test

[PSS] **power twoway** — Power analysis for two-way analysis of variance

[PSS] **power, graph** — Graph results from the power command

[PSS] **power, table** — Produce table of results from the power command

[PSS] **Glossary**

[R] **anova** — Analysis of variance and covariance