power cmh — Power and sample size for the Cochran–Mantel–Haenszel test

Description

power cmh computes sample size, power, or effect size (the minimum detectable odds ratio) for a Cochran–Mantel–Haenszel (CMH) test of association in stratified 2×2 tables. The command accommodates unbalanced stratum sizes and unbalanced group sizes within each stratum.

Quick start

Sample size for a two-sided CMH test with success probabilities in the control group of 0.2, 0.3, and 0.4 and hypothesized common odds ratio of 3.5 given default power of 0.8 and significance level $\alpha = 0.05$

    power cmh .2 .3 .4, oratio(3.5)

Sample size for an unbalanced design with numbers of subjects in the experimental group as 25% of the subjects in each stratum

    power cmh .2 .3 .4, oratio(3.5) grratios(.25 .25 .25)

Sample size for a one-sided test with the continuity correction applied

    power cmh .2 .3 .4, oratio(3.5) onesided continuity

Power for a total sample size of 120 subjects

    power cmh .2 .3 .4, oratio(3.5) n(120)

Plot of power against total sample size for samples of 100 to 150 in increments of 10

    power cmh .2 .3 .4, oratio(3.5) n(100(10)150) graph

Same as above, but display results in a table instead of a graph

    power cmh .2 .3 .4, oratio(3.5) n(100(10)150)

Minimum detectable odds ratio with 80% power

    power cmh .2 .3 .4, power(.8) n(120)

Same as above, but with 50% of subjects from the first stratum, 30% of subjects from the second stratum, and 20% of subjects from the third stratum

    power cmh .2 .3 .4, power(.8) n(120) strweights(50 30 20)

Menu

Statistics > Power, precision, and sample size

Syntax

Compute sample size

> power cmh *probspec* , <u>or</u>atio(*numlist*) [<u>p</u>ower(*numlist*) *options*]

Compute power

> power cmh *probspec* , <u>or</u>atio(*numlist*) n(*numlist*) [*options*]

Compute target odds ratio

> power cmh *probspec* , n(*numlist*) <u>p</u>ower(*numlist*) [*options*]

where *probspec* is either a matrix *matname* containing the probability of a success in a control group for each stratum or a list of individual stratum probabilities:

$p_{11}\ p_{12}\ \cdots\ p_{1K}$

p_{1k}, where $k = 1,\ 2,\ \ldots,\ K$, is the control-group probability of a success in the kth stratum. Each p_{1k} may be specified either as one number or as a list of values in parentheses (see [U] **11.1.8 numlist**).

matname is the name of a Stata matrix with K columns containing control-group success probabilities. Multiple rows are allowed, in which case each row corresponds to a different set of K strata probabilities or, equivalently, column k corresponds to *numlist* for the control-group success probability in the kth stratum.

options	Description
Main	
∗ <u>a</u>lpha(*numlist*)	significance level; default is alpha(0.05)
∗ power(*numlist*)	power; default is power(0.8)
∗ <u>b</u>eta(*numlist*)	probability of type II error; default is beta(0.2)
∗ n(*numlist*)	total sample size; required to compute power or effect size
nfractional	allow fractional sample sizes
∗ nperstratum(*numlist*)	number of subjects per stratum; implies balanced design
∗ n#(*numlist*)	number of subjects in stratum #
strweights(*wgtspec*)	stratum weights; default is one for each stratum, meaning equal stratum sizes
∗ grratios(*grspec*)	stratum-specific group ratios of the experimental-group size to the stratum size, n_{2k}/n_k
∗ <u>or</u>atio(*numlist*)	common odds ratio of the experimental group to the control group; required to compute power or sample size
<u>contin</u>uity	apply continuity correction; default is no continuity correction
direction(<u>upper</u>|<u>lower</u>)	direction of the effect for effect-size determination; default is direction(upper), which means that the postulated value of the parameter is larger than the hypothesized value
onesided	one-sided test; default is two sided
parallel	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	
[<u>no</u>]<u>table</u>[(*tablespec*)]	suppress table or display results as a table; see [PSS-2] **power, table**
<u>sav</u>ing(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*
Graph	
<u>g</u>raph[(*graphopts*)]	graph results; see [PSS-2] **power, graph**
Iteration	
init(#)	initial value for sample size or effect size
<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
Reporting	
[no]showgrstrsizes	suppress or display group-per-stratum sizes
showasmatrix	display all sample sizes in a matrix
<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.

collect is allowed; see [U] 11.1.10 Prefix commands.

notitle does not appear in the dialog box.

wgtspec	Description
$\#_1$ $\#_2$ ... $\#_K$	K stratum weights. Weights must be positive and must be integers unless option nfractional is specified. Multiple values for each stratum weight $\#_k$ can be specified as a *numlist* enclosed in parentheses.
matname	matrix with K columns containing K stratum weights. Multiple rows are allowed, in which case each row corresponds to a different set of K weights or, equivalently, column k corresponds to a *numlist* for the kth 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_stratum	number of subjects per stratum	N/N_s
N_avg	average number of subjects per stratum	N_{avg}
N#	number of subjects in stratum #	$N_\#$
N_per_group	number of subjects per group	$N/2$
G1	number of subjects in control group	G_1
G2	number of subjects in experimental group	G_2
N_per_grstr	number of subjects per group and stratum	$N/(2N_s)$
G1_#	number of subjects in control group and stratum #	$G_{1,\#}$
G2_#	number of subjects in experimental group and stratum #	$G_{2,\#}$
delta	effect size	δ
N_s	number of strata	N_s
oratio	odds ratio	θ
p1_#	#th stratum control-group success probability	$p_{1,\#}$
strwgt#	weight for stratum #	$w_\#$
grratio#	ratio of experimental-group size to stratum size for stratum #	$G_{2,\#}/N_\#$
target	target parameter; synonym for oratio	
_all	display all supported columns	

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

Column N_per_stratum is shown in the default table only for equal-strata designs; otherwise, columns N# are displayed.

Column N_per_group is shown in the default table only when group sizes are the same; otherwise, columns G1 and G2 are displayed.

Column N_per_grstr is shown in the default table only for balanced designs.

Columns G1_# and G2_# are shown in the table only if requested or if option showgrstrsizes is specified.

Columns strwgt# are shown in the default table only if option strweights() is specified.

Columns grratio# are shown in the default table only if option grratios() is specified.

Options

> Main

alpha(), power(), beta(), n(), nfractional; see [PSS-2] **power**.

nperstratum(*numlist*) specifies the stratum size. Only positive integers are allowed. This option implies a balanced-strata design with equal strata sizes. nperstratum() cannot be specified with n(), n#(), or strweights().

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

strweights(*wgtspec*) specifies K stratum weights for an unequal-strata 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. strweights() cannot be specified with nperstratum() or n#().

grratios(*grspec*) specifies K ratios, one for each stratum, of the number of subjects in the experimental group to the number of subjects in the corresponding stratum, n_{2k}/n_k. By default, a balanced group design (or equal numbers of subjects in the control and experimental groups in each stratum) is assumed, which corresponds to setting the K ratios to 0.5.

grspec is similar to *wgtspec* but allows noninteger numbers.

oratio(*numlist*) specifies the alternative value of the common odds ratio of the experimental group to the control group. This option specifies the magnitude of an effect size. It is required to compute power or sample size.

continuity requests that the continuity correction be applied. By default, no continuity correction is applied.

direction(), onesided, parallel; see [PSS-2] **power**.

> Table

table, table(), notable; see [PSS-2] **power, table**.

saving(); see [PSS-2] **power**.

> Graph

graph, graph(); see [PSS-2] **power, graph**. Also see the *column* table for a list of symbols used by the graphs.

init(#) specifies the initial value for the estimated sample size or effect size when an iterative search is required. When computing the sample size for the two-sided test, the sample-size estimate from the one-sided test is used. The initial estimate for computing the effect size is obtained from a bisection search.

iterate(), tolerance(), ftolerance(), log, nolog, dots, nodots; see [PSS-2] **power**.

Reporting

showgrstrsizes and noshowgrstrsizes displays or suppresses the display of sample sizes in each group and stratum. The default for general output is to display group-per-stratum sizes in a matrix. The default for table output is to suppress the display of group-per-stratum sizes. If you specify this option with table output, group-per-stratum sizes will be displayed in a table as columns. This option has no effect on graphical output.

showasmatrix requests that reported sample sizes be displayed as a matrix containing group-per-stratum sizes as cells, and total strata sizes, total group sizes, and a total sample size as marginal totals. This option is not allowed with table or graphical output.

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

notitle; see [PSS-2] **power**.

Remarks and examples

Remarks are presented under the following headings:

This entry describes the power cmh command and the methodology for power and sample-size analysis for a CMH test of association in $2 \times 2 \times K$ tables. See [PSS-2] **Intro (power)** for a general introduction to power and sample-size analysis and [PSS-2] **power** for a general introduction to the power command using hypothesis tests.

Introduction

Many studies are designed to ascertain the relationship between a binary exposure (exposed or unexposed) and a binary outcome (success of failure). Sometimes the relationship between the two binary variables is influenced by another variable (or variables). One way to adjust for such influence is to stratify on that variable and perform stratified analysis.

For example, a health researcher might conduct a case–control study of birth defects (cases) and mother's medication use during pregnancy (exposure) stratified by mother's age group. A law firm might want to know if there is a relationship between a defendant receiving the death penalty and the defendant's race stratified by the murder victim's race. An education researcher might like to know if students' promotion to the next grade level is more likely after participating in a remedial reading program after stratifying by grade level.

In stratified analysis, a separate 2×2 contingency table is constructed for each stratum $k = 1, 2, \ldots, K$.

	Group		
	Control	Experimental	Total
Failure	a_k	b_k	m_{0k}
Success	c_k	d_k	m_{1k}
Total	n_{1k}	n_{2k}	n_k

The corresponding stratum-specific probabilities of success are π_{1k} and π_{2k} in the control and experimental groups, respectively, and their respective estimates are $p_{1k} = c_k/n_{1k}$ and $p_{2k} = d_k/n_{2k}$. The strength of the association in each table is described by the stratum-specific odds ratio of the experimental group to control group θ_k, estimated as $p_{2k}/(1 - p_{2k})/\{p_{1k}/(1 - p_{1k})\} = (a_k d_k)/(b_k c_k)$. An overall measure of association between the exposure groups and the outcome across all strata is formed to provide stratified-adjusted inference about the association.

The CMH test (Cochran 1954; Mantel and Haenszel 1959) is commonly used to test for association in stratified 2×2 tables or, equivalently, in $2 \times 2 \times K$ tables. It forms a common odds ratio θ, which can be viewed as a weighted aggregate of stratum-specific odds ratios θ_k to quantify the strength of the association between the exposure and outcome in such stratified analysis. The null hypothesis for this test is $H_0: \pi_{1k} = \pi_{2k}$ for all $k = 1, 2, \ldots, K$. This is equivalent to $H_0: \theta_k = 1$ for all k. The alternative hypothesis is such that there is common odds ratio that is different from one, that is, $H_a: \theta_1 = \theta_2 = \cdots = \theta_K = \theta \neq 1$. Thus, we can rewrite our hypotheses simply as $H_0: \theta = 1$ versus the two-sided alternative $H_a: \theta \neq 1$, an upper one-sided alternative $H_a: \theta > 1$, or a lower one-sided alternative $H_a: \theta < 1$. Statistical significance is assessed using a weighted CMH χ^2 test statistic. Under the null, the distribution of the test statistic is approximately a χ^2 distribution with one degree of freedom. See, for example, Lachin (2011) for more details about this test.

power cmh provides power and sample-size computations based on the asymptotic distribution of the test statistic according to Nam (1992); see *Methods and formulas* for details.

Using power cmh

power cmh computes sample size, power, or effect size (the minimum detectable odds ratio) for the CMH test of association in $2 \times 2 \times K$ tables. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying option alpha(). You can specify the onesided option to request a one-sided test.

To compute the total sample size, you must specify the probabilities of success p_{1k} in the control group in each of the K strata following the command name; the common odds ratio in the oratio() option; and, optionally, the power of the test in option power(). The default power is set to 0.8.

To compute power, you must specify the total sample size in the n() option, the common odds ratio in the oratio() option, and the control-group success probabilities p_{1k} following the command name.

To compute effect size, the minimum detectable odds ratio, or the target odds ratio, you must specify the total sample size in n(); the power in power(); the control-group success probabilities following the command name; and, optionally, the direction of the effect. The direction is upper by default, direction(upper), which means that the common odds ratio is assumed to be larger than one. You can change the direction to be lower, which means that the common odds ratio is assumed to be smaller than one, by specifying the direction(lower) option.

There are multiple ways to specify the control-group success probabilities; see *Alternative ways of specifying probabilities*.

By default, all computations assume a design with equal numbers of subjects in each group and each stratum. To change group-specific proportions of subjects in each stratum, use the grratios(*grspec*) option. A common proportion may be used for all strata or you may specify different proportions for each stratum. Regardless of whether you choose a common proportion or choose to let the proportions vary across strata, you must specify K ratios of cases to the stratum sample size within grratios().

To accommodate unequal stratum sizes for power and effect-size computations, you can specify either individual stratum sizes in options n1(), n2(), ..., nK() or a combination of the total sample size in n() and integer stratum weights in strweights(*wgtpec*). For equal stratum sizes, you can also specify the nperstratum() option to specify a common stratum size instead of a total sample size in n().

By default, all computations assume no continuity correction. Use the continuity option to change that.

power cmh reports group-per-stratum sizes as a matrix in the output. To suppress this matrix, use the noshowgrstrsizes option. Alternatively, for the table output, you can use the showgrstrsizes option to include columns containing group-per-stratum sizes in the default table.

To make the output of power cmh more compact, you may consider using the showasmatrix option to display all sample sizes in a matrix.

Sample-size determination for the two-sided test and effect-size determination require iteration. The default initial value for the sample size uses the estimate for the one-sided test. The initial estimate for the effect size is obtained using a bisection search. To specify a different starting value, you may use option init(). For more options used to control the iteration process, see [PSS-2] **power**.

Alternative ways of specifying probabilities

There are multiple ways in which you can supply the success probabilities in the control group to power cmh.

You may specify each p_{1k} following the command line as

power cmh p_{11} p_{12} $\cdots$ p_{1K} , $\cdots$

At least two probabilities must be specified.

When you have many strata, you may find it more convenient to first define a Stata matrix as a row or column vector and use it with power cmh. The dimension of the matrix must be at least 2. For example,

matrix define *probmat* = (p_{11}, p_{12}, ..., p_{1K})

power cmh *probmat*, ...

In some cases, you may wish to examine multiple control-group success probabilities in one or more strata. To do this, you can specify multiple values or *numlist* for each of the stratum probabilities in parentheses:

power cmh ($p_{11,1}$ $p_{11,2}$ $\cdots$ $p_{11,L1}$) ($p_{12,1}$ $p_{12,2}$ $\cdots$ $p_{12,L2}$) $\cdots$, $\cdots$

Each of the *numlists* may contain different numbers of values, $L_1 \neq L_2 \neq \cdots \neq L_K$. power cmh will produce results for all possible combinations of values across *numlists*. Results are presented in a table. If instead you would like to treat each specification as a separate scenario, you may specify the parallel option.

You can accommodate multiple sets of stratum probabilities in a matrix form by adding a row for each specification. The columns of a matrix with multiple rows correspond to K strata probabilities, and values within each column k correspond to multiple specifications of the kth stratum probability or a *numlist* for the kth stratum probability.

For example, the following two specifications for three strata with two scenarios each are the same:

power cmh $(p_{11,1}\ p_{11,2})\ (p_{12,1}\ p_{12,2})\ (p_{13,1}\ p_{13,2})$, . . .

and

matrix define *probmat* = $(p_{11,1},\ p_{12,1},\ p_{13,1}\ \backslash\ p_{11,2},\ p_{12,2},\ p_{13,2})$

power cmh *probmat*, . . .

In the above specification, if you wish to specify a *numlist* only for the first stratum, you may define your matrix as

matrix define *probmat* = $(p_{11,1},\ p_{12},\ p_{13}\ \backslash\ p_{11,2},\ .,\ .)$

and the results of

power cmh *probmat*, . . .

will be the same as the results of

power cmh $(p_{11,1}\ p_{11,2})\ p_{12}\ p_{13}$, . . .

In the following sections, we describe the use of power cmh accompanied by examples for computing sample size, power, and the minimum detectable odds ratio.

Motivating example

Consider example 4.1 of a clinical trial in duodenal ulcers from Lachin (2011). Blum (1982) describes a hypothetical clinical trial studying the effectiveness of a new drug for treating an ulcer versus placebo. The outcome is whether an ulcer is healed or not, that is, whether the excretion of gastric juices that lead to ulceration of the duodenum is retarded. Three classes of ulcers are considered: acid-dependent (caused by excessive gastric secretion), drug-dependent (caused by excessive use of drugs), and ulcers of intermediate origin. The third class contains ulcers for which the cause is difficult to determine.

The research objective is to evaluate the effectiveness of the treatment on healing ulcers adjusted for ulcer types.

The data are summarized in three 2×2 tables, one for each ulcer type.

```
. use https://www.stata-press.com/data/r19/ulcer
(Duodenal ulcers data)
. table (ulcer healed) (treatment) [fw=weight]
```

	Treatment		
	Placebo	Drug	Total
Ulcer type			
Acid-dependent			
Healing status			
Not healed	27	26	53
Healed	20	16	36
Total	47	42	89
Drug-dependent			
Healing status			
Not healed	5	3	8
Healed	4	9	13
Total	9	12	21
Intermediate			
Healing status			
Not healed	28	18	46
Healed	16	28	44
Total	44	46	90
Total			
Healing status			
Not healed	60	47	107
Healed	40	53	93
Total	100	100	200

The stratum-specific proportions of healed ulcers in each group are as follows:

```
. table (ulcer treatment) [fw=weight], statistic(mean healed) nformat(%8.3g)
> nototals
```

	Mean
Ulcer type	
Acid-dependent	
Treatment	
Placebo	.426
Drug	.381
Drug-dependent	
Treatment	
Placebo	.444
Drug	.75
Intermediate	
Treatment	
Placebo	.364
Drug	.609

The common (marginal) odds ratio can be computed by using the `cc` command. (See [R] **Epitab** and Testing hypotheses about association in $2 \times 2 \times K$ tables for details about this command.)

```
. cc treatment healed [fw=weight]
```

	Exposed	Unexposed	Total	Proportion exposed
Cases	53	47	100	0.5300
Controls	40	60	100	0.4000
Total	93	107	200	0.4650

	Point estimate	[95% conf. interval]
Odds ratio	1.691489	.9299035 3.08095 (exact)
Attr. frac. ex.	.408805	-.0753804 .6754248 (exact)
Attr. frac. pop	.2166667	

```
                          chi2(1) =    3.40  Pr>chi2 = 0.0653
```

The estimate of the marginal odds ratio for these data is 1.691.

Suppose that we would like to conduct a similar study for evaluating a new treatment for healing ulcers. We would like to perform power and sample-size analysis for this new study. In what follows, we demonstrate how to use `power cmh` to perform these analyses, and we use the results from the current study as our pilot estimates.

Computing sample size

To compute sample size, you must specify the control-group probability of success in each stratum following the command name; the common odds ratio in option `oratio()`; and, optionally, the power of the test in option `power()`. A default power of 0.8 is assumed if `power()` is not specified.

▷ Example 1: Sample size for a two-sided CMH test

Consider a study of drug effectiveness for healing ulcers from *Motivating example*. The data are stratified by an ulcer type. We would like to conduct a new study to evaluate another ulcer treatment versus placebo. We need to compute the required sample size for this study. We use the estimates from the previous ulcer study as our pilot estimates.

The proportions of healed ulcers in the placebo (control) group for each of the three strata were estimated to be 0.426, 0.444, and 0.364, respectively. We anticipate an improvement in the new treatment, and we would like to detect an odds ratio of at least 2.5 with 80% power using a two-sided 5%-level test.

To compute the required sample size, we specify the control-group proportions after the command name and an odds ratio of 2.5 in option `oratio()`. We omit options `alpha(0.05)` and `power(0.8)`, because the specified values are their defaults.

```
. power cmh 0.426 0.444 0.364, or(2.5)

Performing iteration ...

Estimated sample size for a test of independence in stratified 2x2 tables
Cochran-Mantel-Haenszel test
H0: OR = 1  versus  Ha: OR != 1;  OR = OR_1 = OR_2 = OR_3

Study parameters:

                   alpha =    0.0500
                   power =    0.8000
                   delta =    2.5000
                     N_s =         3
                    p1_1 =    0.4260
                    p1_2 =    0.4440
                    p1_3 =    0.3640
                   oratio =    2.5000

Estimated sample sizes:

                       N =       156
          N per stratum =        52
            N per group =        78
   N per group/stratum =        26
```

A total sample of 156 subjects (78 per group, 52 per stratum, and 26 per each group and stratum) must be obtained to detect an odds ratio of at least 2.5 with 80% power using a 5%-level two-sided CMH test.

◁

▷ Example 2: Unbalanced design, unequal stratum sizes

In example 1, we assumed equal numbers of subjects in each stratum. Based on our pilot data, there are 89 subjects in stratum 1, 21 subjects in stratum 2, and 90 subjects in stratum 3. In other words, the first and the third strata have about 4 times as many subjects as the second stratum.

Let's see how unequal stratum sizes affect our required sample size. We specify strata weights in the strweights() option.

```
. power cmh 0.426 0.444 0.364, or(2.5) strweights(4 1 4)

Performing iteration ...

Estimated sample size for a test of independence in stratified 2x2 tables
Cochran-Mantel-Haenszel test
HO: OR = 1  versus  Ha: OR != 1;  OR = OR_1 = OR_2 = OR_3

Study parameters:

        alpha =    0.0500
        power =    0.8000
        delta =    2.5000
          N_s =         3
         p1_1 =    0.4260
         p1_2 =    0.4440
         p1_3 =    0.3640
        oratio =   2.5000

Estimated sample sizes:

            N =       162
           N1 =        72
           N2 =        18
           N3 =        72
  N per group =        81

Group-per-stratum sample sizes:
```

		stratum		
		1	2	3
group				
	1	36	9	36
	2	36	9	36

The required total sample size increases slightly from 156 to 162 with 81 subjects now in each group, which is a slightly larger number than before. Because of unbalanced stratum sizes, the number of subjects per group differs across strata with 36 subjects per group in the first and the third strata and 9 subjects per group in the second stratum.

◁

▷ Example 3: Unbalanced design, unequal group ratios

Continuing with example 2 and our *Motivating example*, we notice that our pilot study also had different numbers of subjects in each group across strata. The actual ratios of the experimental group sizes to the stratum sizes were $42/89 = 0.47$ in stratum 1, $12/21 = 0.57$ in stratum 2, and $46/90 = 0.51$ in stratum 3.

To accommodate unequal group sizes across strata, we can specify stratum-specific group ratios in the grratios() option.

```
. power cmh 0.426 0.444 0.364, or(2.5) strweights(4 1 4) grratios(0.47 0.57 0.51)
Performing iteration ...
Estimated sample size for a test of independence in stratified 2x2 tables
Cochran-Mantel-Haenszel test
H0: OR = 1  versus  Ha: OR != 1;  OR = OR_1 = OR_2 = OR_3
Study parameters:
         alpha =    0.0500
         power =    0.8000
         delta =    2.5000
           N_s =         3
          p1_1 =    0.4260
          p1_2 =    0.4440
          p1_3 =    0.3640
        oratio =    2.5000
Estimated sample sizes:
             N =       162
            N1 =        72
            N2 =        18
            N3 =        72
     N group 1 =        80
     N group 2 =        82
Group-per-stratum sample sizes:
```

		stratum		
		1	2	3
group				
	1	38	7	35
	2	34	11	37

Because all the group ratios are close enough to 0.5 (equal-sized groups), the estimate of the total sample size does not change, and the group sizes change only slightly.

We can investigate the impact of group ratios that are appreciably different from 0.5. For example, for the following three group ratios, the total sample size increases to 207 with 88 subjects in the control group and 119 subjects in the experimental group.

```
. power cmh 0.426 0.444 0.364, or(2.5) strweights(4 1 4) grratios(0.8 0.7 0.3)

Performing iteration ...

Estimated sample size for a test of independence in stratified 2x2 tables
Cochran-Mantel-Haenszel test
H0: OR = 1  versus  Ha: OR != 1;  OR = OR_1 = OR_2 = OR_3

Study parameters:
        alpha =    0.0500
        power =    0.8000
        delta =    2.5000
          N_s =         3
         p1_1 =    0.4260
         p1_2 =    0.4440
         p1_3 =    0.3640
        oratio =   2.5000

Estimated sample sizes:
            N =       207
           N1 =        92
           N2 =        23
           N3 =        92
    N group 1 =        88
    N group 2 =       119

Group-per-stratum sample sizes:
```

		stratum		
		1	2	3
group				
	1	18	6	64
	2	74	17	28

◁

Computing power

To compute power, you must specify the total sample size in the n() option, the common odds ratio in option oratio(), and the probabilities of success in the control group following the command name.

▷ Example 4: Power of a two-sided CMH test

Returning to example 1, suppose that we have resources to recruit twice as many subjects, say, 300. To compute power, we add the total number of subjects, 300, in the n() option to the syntax from example 1:

```
. power cmh 0.426 0.444 0.364, or(2.5) n(300)
Estimated power for a test of independence in stratified 2x2 tables
Cochran-Mantel-Haenszel test
H0: OR = 1  versus  Ha: OR != 1;  OR = OR_1 = OR_2 = OR_3
Study parameters:
                        alpha =    0.0500
                            N =       300
              N per stratum =       100
                N per group =       150
       N per group/stratum =        50
                        delta =    2.5000
                          N_s =         3
                        p1_1 =    0.4260
                        p1_2 =    0.4440
                        p1_3 =    0.3640
                      oratio =    2.5000
Estimated power:
                        power =    0.9759
```

For such an increase in sample size, the power increases dramatically to 0.976 to detect an odds ratio of 2.5.

◁

▷ Example 5: Multiple values of study parameters

We may want to check powers for several sample sizes. Continuing with example 4, we simply list desired sample-size values in option n(*numlist*); see [U] **11.1.8 numlist**.

```
. power cmh 0.426 0.444 0.364, or(2.5) n(150(25)300) table(power N)
Estimated power for a test of independence in stratified 2x2 tables
Cochran-Mantel-Haenszel test
H0: OR = 1  versus  Ha: OR != 1;  OR = OR_1 = OR_2 = OR_3
```

power	N
.7904	150
.8473	175
.8902	200
.9253	225
.9475	250
.9634	275
.9759	300

To shorten our default table, we selected only two varying columns, power and total sample size N, by specifying them within the table() option.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [PSS-2] **power, table**. If you wish to produce a power plot, see [PSS-2] **power, graph**.

◁

Computing effect size

Sometimes, we may be interested in determining the smallest odds ratio that yields a statistically significant result for prespecified sample size and power. In this case, we must specify power in option power(), sample size in option n(), and control-group success probabilities following the command

name. We may also choose the direction of the effect by specifying the direction() option. The default is direction(upper), meaning a common odds ratio greater than one, which corresponds to the improvement of the experimental group over the control group. You can use direction(lower) to request an odds ratio less than one.

▷ Example 6: Minimum detectable odds ratio

In example 4, we learned that with a sample of 300 subjects, the power to detect a common odds ratio of 2.5 is very high. We now want to identify the minimum detectable odds ratio for that study with 80% power.

```
. power cmh 0.426 0.444 0.364, power(0.8) n(300)
Performing iteration ...
Estimated odds ratio for a test of independence in stratified 2x2 tables
Cochran-Mantel-Haenszel test
H0: OR = 1  versus  Ha: OR != 1, OR > 1;  OR = OR_1 = OR_2 = OR_3
Study parameters:
                    alpha =     0.0500
                    power =     0.8000
                        N =        300
           N per stratum =        100
             N per group =        150
    N per group/stratum =         50
                     N_s =          3
                    p1_1 =     0.4260
                    p1_2 =     0.4440
                    p1_3 =     0.3640
Estimated effect size and odds ratio:
                    delta =     1.9192
               odds ratio =     1.9192
```

With a sample of 300 subjects, we can detect an odds ratio of 1.92 with 80% power using a two-sided 5%-level CMH test.

◁

Testing hypotheses about association in $2 \times 2 \times K$ tables

You can test hypotheses about association in $2 \times 2 \times K$ tables using the cc command; see [R] **Epitab**. cc conducts a Mantel–Haenszel test, which is asymptotically equivalent to a Cochran test; see, for example, *Methods and formulas*.

▷ Example 7: Performing Mantel–Haenszel test

Consider the ulcer data from *Motivating example*. We would like to test for association between the treatment and healing of an ulcer stratified by a type of ulcer.

```
. use https://www.stata-press.com/data/r19/ulcer
(Duodenal ulcers data)

. describe

Contains data from https://www.stata-press.com/data/r19/ulcer.dta
 Observations:             12                  Duodenal ulcers data
    Variables:              4                  3 Mar 2024 21:40
```

Variable name	Storage type	Display format	Value label	Variable label
treatment	byte	%9.0g	treatlab	Treatment
healed	byte	%10.0g	heallab	Healing status
ulcer	byte	%14.0g	ulcerlab	Ulcer type
weight	byte	%9.0g		Frequency weights

```
Sorted by:
```

The binary variable `healed` indicates whether an ulcer healed. The binary variable `treatment` indicates whether a subject received a treatment (drug) or placebo. The categorical variable `ulcer` is a stratification variable, which records an ulcer type. The `weight` variable contains the numbers of subjects for each combination of values for `healed` and `treatment` for each of the three strata. This variable will be used as a frequency weight on our analysis.

Here are our data:

```
. list, sepby(ulcer) nolabel
```

	treatm~t	healed	ulcer	weight
1.	1	1	1	16
2.	1	0	1	26
3.	0	1	1	20
4.	0	0	1	27
5.	1	1	2	9
6.	1	0	2	3
7.	0	1	2	4
8.	0	0	2	5
9.	1	1	3	28
10.	1	0	3	18
11.	0	1	3	16
12.	0	0	3	28

We use the `cc` command to test for association between `healed` and `treatment` adjusted for `ulcer` types. We specify `healed` as the dependent variable and `treatment` as the exposure variable after the command name and use the `by(ulcer)` option to stratify on `ulcer`. We also specify the `weight` variable as the frequency weight.

```
. cc healed treatment [fweight=weight], by(ulcer)
    Ulcer type │ Odds ratio    [95% conf. interval]   M-H weight

 Acid-dependent │  .8307692       .32516   2.112608   5.842697 (exact)
 Drug-dependent │      3.75     .4283095   35.61362   .5714286 (exact)
   Intermediate │  2.722222     1.069202   6.992856        3.2 (exact)

          Crude │  1.691489     .9299035    3.08095             (exact)
   M-H combined │  1.633836      .934329   2.857044

Test of homogeneity (M-H)      chi2(2) =     4.58  Pr>chi2 = 0.1012
                   Test that combined odds ratio = 1:
                          Mantel-Haenszel chi2(1) =      3.00
                                        Pr>chi2 =     0.0830
```

The combined Mantel–Haenszel odds ratio equals 1.63, which is not significantly different from 1 at the 5% significance level; the *p*-value is 0.0830. Thus, we do not have sufficient evidence to reject the null hypothesis of no association between the treatment and healing of ulcer after adjusting for ulcer type, at least at the 5% significance level.

◁

Stored results

power cmh 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 stratum #
r(N_per_stratum)	number of subjects per stratum
r(N_s)	number of strata
r(nfractional)	1 if nfractional is specified, 0 otherwise
r(balanced)	1 for a balanced design, 0 otherwise
r(strwgt#)	stratum weight #
r(N_per_group)	number of subjects per group
r(G1)	number of subjects in the control group
r(G2)	number of subjects in the experimental group
r(N_per_grstr)	number of subjects per group and stratum
r(G1_#)	number of subjects in the control group in stratum #
r(G2_#)	number of subjects in the experimental group in stratum #
r(grratio#)	ratio of the experimental-group size to stratum size for stratum #
r(p1_#)	control-group probability of success in stratum #
r(oratio)	odds ratio of the experimental group to control group
r(continuity)	1 if continuity correction is used, 0 otherwise
r(c)	continuity-correction value
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)	cmh
r(direction)	upper or lower
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

Methods and formulas

Consider definitions and a 2×2 contingency table for stratum $k = 1, 2, \ldots, K$ from *Introduction*.

Assume a common odds ratio among all the K tables; that is, $\theta_1 = \theta_2 = \cdots = \theta_K = \theta$ (Woolson, Bean, and Rojas 1986; Nam 1992). To test the hypothesis $H_0: \theta = 1$ versus $H_a: \theta > 1$, Cochran's test statistic may be formed as

$$C = \frac{\sum_{k=1}^{K} \omega_k (p_{2k} - p_{1k})}{\sqrt{\sum_{k=1}^{K} \omega_k \overline{p}_k \overline{q}_k}} = \frac{W}{\sqrt{\widehat{\mathrm{Var}_0(W)}}}$$

where

$$\omega_k = \frac{n_{1k} n_{2k}}{n_k}$$

$$\overline{p}_k = \frac{n_{1k} p_{1k} + n_{2k} p_{2k}}{n_k} \qquad \text{and} \qquad \overline{q}_k = 1 - \overline{p}_k$$

Under the null hypothesis, the statistic C converges to a standard normal distribution. We therefore reject the null hypothesis if $C > z_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1 - \alpha)$th quantile of a standard normal distribution.

A similar statistic is the Mantel–Haenszel statistic, which considers a hypergeometric distribution for all fixed marginals for the variance of W under the null hypothesis. Asymptotically, the Cochran test is equivalent to the Mantel–Haenszel test (Armitage, Berry, and Matthews 2002), and thus we refer to this test jointly as the Cochran–Mantel–Haenszel (CMH) test.

Following Nam (1992), we note that the expected value and the variance of W is given, respectively, by

$$E(W) = \sum_{k=1}^{K} \omega_k (\pi_{2k} - \pi_{1k})$$

$$\mathrm{Var}(W) = \sum_{k=1}^{K} \omega_k^2 \left\{ \frac{\pi_{1k}(1 - \pi_{1k})}{n_{1k}} + \frac{\pi_{2k}(1 - \pi_{2k})}{n_{2k}} \right\}$$

The variance under the null hypothesis is given by

$$\text{Var}(W|H_0) = \text{Var}_0(W) = \sum_{k=1}^{K} \omega_k \bar{\pi}_k (1 - \bar{\pi}_k)$$

For the upper one-sided alternative hypothesis, $H_a: \theta > 1$, the asymptotic power is given by

$$1 - \beta \approx P(C > z_{1-\alpha}) = 1 - \Phi(U_u) \tag{1}$$

where

$$U_u = \frac{z_{1-\alpha} \sqrt{\text{Var}_0(W)} - E(W)}{\sqrt{\text{Var}(W)}}$$

and $\Phi(\cdot)$ is the cumulative of the standard normal distribution.

For the lower one-sided alternative hypothesis, $H_a: \theta < 1$, the asymptotic power is given by

$$1 - \beta \approx P(C \le z_\alpha) = \Phi(U_l) \tag{2}$$

where

$$U_l = \frac{z_\alpha \sqrt{\text{Var}_0(W)} - E(W)}{\sqrt{\text{Var}(W)}}$$

When the `onesided` option is used, the direction of the test is determined by the sign of $E(W)$—when $E(W) > 0$, then $H_a: \theta > 1$ and when $E(W) < 0$, then $H_a: \theta < 1$.

Then, the power of the test is given by

$$\pi = \begin{cases} 1 - \Phi(U_u) & \text{for an upper one-sided test} \\ \Phi(U_l) & \text{for a lower one-sided test} \\ 1 - \Phi(U_u) + \Phi(U_l) & \text{for a two-sided test} \end{cases}$$

The two-sided test uses $\alpha/2$ in the definitions of U_u and U_l.

The sample size for a two-sided test and the minimum detectable odds ratio θ must be computed iteratively from the corresponding power equations. The starting value for the sample-size computation is the sample-size estimate for the one-sided test. For the minimum detectable odds ratio, the starting value is obtained from a bisection algorithm.

When strata weights w_k are specified in the `strweights()` option, a constant multiplier n_c is computed and rounded to an integer unless the `nfractional` option is specified. The stratum sizes are then computed as $\tilde{w}_j n_c$, where $\tilde{w}_k$ are normalized strata weights. The actual sample size, N_a, is the sum of the stratum sizes. The numbers of subjects in each group and stratum are computed as $n_k/2$ for equal-group designs and as $n_{2k} = n_k \times \#_k$ and $n_{1k} = n_k - n_{2k}$ when group ratios $\#_k$s are supplied in `grratios()`. Unless the `nfractional` option is specified, the experimental-group sizes corresponding to unbalanced strata are rounded up. For strata that have equal numbers of subjects in each group, the group-per-stratum sizes are not rounded.

Woolson, Bean, and Rojas (1986) provide a formula for the sample size that is required to attain a specific power by substituting $U_u = z_\beta = \Phi^{-1}(\beta)$ and $U_l = z_{1-\beta} = \Phi^{-1}(1-\beta)$ in (1) and (2) and solving for n:

$$n = \begin{cases} \left(\frac{z_{1-\alpha}\sqrt{X}+z_{1-\beta}\sqrt{Y}}{Z}\right)^2 & \text{if } E(W) > 0 \\ \left(\frac{z_\alpha\sqrt{X}+z_\beta\sqrt{Y}}{Z}\right)^2 & \text{if } E(W) < 0 \end{cases}$$

where

$$X = \sum_{k=1}^{K} w_k s_k (1-s_k)\bar{\pi}_k(1-\bar{\pi}_k)$$

$$Y = \sum_{k=1}^{K} w_k s_k (1-s_k)\left\{(1-s_k)\pi_{2k}(1-\pi_{2k}) + s_k\pi_{1k}(1-\pi_{1k})\right\}$$

$$Z = \sum_{k=1}^{K} w_k s_k (1-s_k)(\pi_{2k} - \pi_{1k})$$

and $w_k = n_k/n$ is the fraction of the individual-stratum size to the total sample size, and $s_k = n_{2k}/n_k$ is the fraction of the number of subjects in the experimental group to the sample size of each stratum. Sample size for the two-sided alternative hypothesis is solved iteratively.

The continuity-corrected version of U_u and U_l is

$$U_u = \frac{z_{1-\alpha}\sqrt{\text{Var}_0(W)} - E(W) + \frac{1}{2}}{\sqrt{\text{Var}(W)}} \qquad \text{if } E(W) > 0$$

$$U_l = \frac{z_\alpha\sqrt{\text{Var}_0(W)} - E(W) - \frac{1}{2}}{\sqrt{\text{Var}(W)}} \qquad \text{if } E(W) < 0$$

Nam (1992) derives the sample size for the continuity-corrected version of the CMH test, n_c, as follows:

$$n_c = \begin{cases} \frac{n}{4}\left(1 + \sqrt{1 + \frac{2}{nZ}}\right)^2 & \text{if } E(W) > 0 \\ \frac{n}{4}\left(1 + \sqrt{1 - \frac{2}{nZ}}\right)^2 & \text{if } E(W) < 0 \end{cases}$$

where n is the uncorrected sample size.

For the computation, π_{1k} and π_{2k} are replaced with their estimates p_{1k} and p_{2k} in the above formulas. p_{1k}s are supplied directly to power cmh, and p_{2k}s are computed using the specified p_{1k}s and odds ratio in oratio().

References

Armitage, P., G. Berry, and J. N. S. Matthews. 2002. *Statistical Methods in Medical Research*. 4th ed. Oxford: Blackwell.

Blum, A. L. 1982. "Principles for selection and exclusion". In *The Randomized Clinical Trial and Therapeutic Decisions*, edited by N. Tygstrup, J. M. Lachin, and E. Juhl, 43–58. New York: Dekker.

Cochran, W. G. 1954. Some methods for strengthening the common chi-squared tests. *Biometrics* 10: 417–451. https://doi.org/10.2307/3001616.

Lachin, J. M. 2011. *Biostatistical Methods: The Assessment of Relative Risks*. 2nd ed. Hoboken, NJ: Wiley.

Mantel, N., and W. Haenszel. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 22: 719–748. Reprinted in *Evolution of Epidemiologic Ideas: Annotated Readings on Concepts and Methods*, ed. S. Greenland, pp. 112–141. Newton Lower Falls, MA: Epidemiology Resources.

Nam, J. 1992. Sample size determination for case–control studies and the comparison of stratified and unstratified analyses. *Biometrics* 48: 389–395. https://doi.org/10.2307/2532298.

Woolson, R. F., J. A. Bean, and P. B. Rojas. 1986. Sample size for case–control studies using Cochran's statistic. *Biometrics* 42: 927–932. https://doi.org/10.2307/2530706.

Also see

[PSS-2] **power** — Power and sample-size analysis for hypothesis tests

[PSS-2] **power, graph** — Graph results from the power command

[PSS-2] **power, table** — Produce table of results from the power command

[PSS-5] **Glossary**

[R] **Epitab** — Tables for epidemiologists