power logistic — Power analysis for logistic regression+

+This command is part of StataNow.

Description Menu Syntax Remarks and examples Reference Also see

Description

`power logistic` computes sample size, power, or effect size for a test of one coefficient in logistic regression. There are three main uses of `power logistic`, each of which is described in its own manual entry. See [PSS-2] **power logistic onebin** for when the logistic regression has only one binary covariate. See [PSS-2] **power logistic twobin** for when the logistic regression has two binary covariates, one of which is the covariate of interest and the other is a nuisance covariate. See [PSS-2] **power logistic general** for a description of the general case that can accommodate covariates from any of 11 distributions while allowing for up to 20 nuisance covariates.

By default, `power logistic` computes sample size for a given power and effect size, where the effect size may be specified as a coefficient or an odds ratio. Alternatively, it can compute power given sample size and effect size, or it can compute the effect size given power and sample size.

Menu

Statistics > Power, precision, and sample size

Syntax

Sample-size computation with one binary covariate

`power logistic` *oratio*$_X$ `,` { `px(`*numlist*`)` | `oddsx(`*numlist*`)` } [`power(`*numlist*`)` *onebinopts*]

Sample-size computation with two binary covariates

`power logistic` *oratio*$_X$ `,` { `px(`*numlist*`)` | `oddsx(`*numlist*`)` } { `pz(`*numlist*`)` | `oddsz(`*numlist*`)` }
[`power(`*numlist*`)` *twobinopts*]

Sample-size computation with arbitrary covariates

`power logistic` *oratio*$_X$ `,` `x(`*xzspec*`)` [`z1(`*xzspec*`)` [`z2(`*xzspec*`)` [...]]
`power(`*numlist*`)` *generalopts*]

oratio$_X$ is the odds ratio for covariate of interest X under the alternative hypothesis H_a. Argument *oratio*$_X$ may be specified either as one number or as a list of values in parentheses (see [U] **11.1.8 numlist**). Power and effect-size calculations follow similar syntax; see [PSS-2] **power logistic onebin**, [PSS-2] **power logistic twobin**, and [PSS-2] **power logistic general** for details.

Remarks and examples

Remarks are presented under the following headings:

This entry provides an overview of the `power logistic` command and demonstrates how sample-size calculations can be specified for logistic regression with one or two binary covariates as well as the general syntax for logistic regression with arbitrary covariates. For further details and information about performing power and effect-size calculations, see the individual entries [PSS-2] **power logistic onebin**, [PSS-2] **power logistic twobin**, and [PSS-2] **power logistic general**. See [PSS-2] **Intro (power)** for a general introduction to power and sample-size analysis, and see [PSS-2] **power** for a general introduction to the `power` command using hypothesis tests.

Introduction to power logistic

Logistic regression is a commonly used statistical method for analyzing binary outcome variables. Researchers often need to determine the appropriate sample size to ensure sufficient power for detecting an association between a binary outcome variable and a covariate of interest, potentially in the presence of nuisance covariates.

For example, consider a study that examines factors influencing whether migratory birds return to the same nesting site from one year to the next. Binary outcome Y is an indicator of whether the nesting site was reused, where the observed $y_i = 1$ if bird i returns to the nesting site it used last year and $y_i = 0$ if bird i does not. We will use logistic regression to test whether birds of one sex are more likely to return to the same nesting site than birds of the other sex. We define binary covariate of interest X as an indicator for female birds, where the observed $x_i = 1$ if bird i is female and $x_i = 0$ if bird i is male. Previous observations suggest that birds who mate with the same partner as last year are more likely to return to the same nesting site, as are heavier birds. We are not interested in studying the effect of a bird's mate or weight, but it would be foolhardy to ignore these effects. We include mate in our logistic regression as nuisance covariate Z_1, where the observed $z_{1i} = 1$ if bird i has the same mate as last year and $z_{1i} = 0$ if not. And we include weight as nuisance covariate Z_2, where the observed z_{2i} is the weight of bird i. Taken together, we write $\mathbf{Z} = (Z_1, Z_2)$. The logistic regression can be written as

$$\Pr(y_i = 1|x_i, \mathbf{z}_i) = H(\beta_X x_i + \zeta_0 + \zeta_1 z_{1i} + \zeta_2 z_{2i}) \qquad i = 1, 2, \ldots, n$$

where β_X is the coefficient quantifying the effect of a bird's sex on nesting-site reuse, ζ_0 is the logistic intercept, ζ_1 is the effect of partnering with the same mate, ζ_2 is the effect of the bird's weight, and n is the sample size. Function $H(\eta) = \{1 + \exp(-\eta)\}^{-1}$ is the logistic distribution function.

The effect of covariate X can also be expressed in terms of an odds ratio,

$$\text{OR}_X = \exp(\beta_X \text{U}_X) = \Pr(Y = 1|X = 1)\Pr(Y = 0|X = 0)/\{\Pr(Y = 0|X = 1)\Pr(Y = 1|X = 0)\}$$

where U_X is the unit change in X for the odds ratio and $\text{U}_X > 0$. In this example, U_X must equal 1 because X is a Bernoulli random variable. The null hypothesis is $H_0: \beta_X = 0$, which can also be expressed as $H_0: \text{OR}_X = 1$. The alternative hypothesis is $H_a: \beta_X \neq 0$, or equivalently, $H_a: \text{OR}_X \neq 1$.

The `power logistic` command provides power and sample-size analysis for the test of $\beta_X = 0$ in logistic regression. The formula for power, sample-size, and effect-size calculations is based on the likelihood-ratio test of $\beta_X = 0$. However, Bush (2015) demonstrates that sample-size requirements for Wald and score tests are generally equivalent to the sample-size requirement for the likelihood-ratio test.

Thus, these calculations may be used to plan studies that will be analyzed using the logit command, which conducts a Wald test of $\beta_X = 0$, or the logistic command, which conducts an equivalent Wald test of $\text{OR}_X = 1$. If you prefer a likelihood-ratio test, you can use the lrtest command.

Computing sample size

In the following examples, we perform sample-size calculations for the study of migratory birds. To compute sample size, you must specify the effect size and other parameters of the logistic regression, and, optionally, the power of the test using either the power() or the beta() option. A default power of 0.8 is assumed if power() is not specified. For further details and information about performing power and effect-size calculations, see the individual entries [PSS-2] **power logistic onebin**, [PSS-2] **power logistic twobin**, and [PSS-2] **power logistic general**.

▷ Example 1: Sample size with one binary covariate

To demonstrate, suppose we ignored the nuisance covariates in the study of migratory birds, leaving only the covariate of interest X. With a single binary covariate, we can use the one binary syntax of power logistic. We wish to compute the sample size required to detect an odds ratio of 1.5 for the effect of X on Y. We specify the effect size, $\text{OR}_X = 1.5$, as an argument. Female birds compose approximately 57% of our target population, which we specify in the px() option. A previous study of male birds found a 20% rate of nest-site reuse, so we specify $\Pr(Y = 1|X = 0) = 0.2$ using the pycondx0() option. Below, we calculate the sample size required to achieve 80% power with a 5% level test, using the default values of alpha and power.

```
. power logistic 1.5, px(0.57) pycondx0(0.2)
Estimated sample size for logistic regression odds-ratio test
Likelihood-ratio test
H0: OR_X = 1  versus  Ha: OR_X != 1
Study parameters:
        alpha =    0.0500
        power =    0.8000
        delta =    1.5000  (odds ratio)
      oratiox =    1.5000
           px =    0.5700
     pycondx0 =    0.2000
Estimated sample size:
            N =     1,092
```

When power logistic is called, the output lists the study parameters we specified, followed by the estimated sample size. A sample of 1,092 birds is required to detect an odds ratio of 1.5 with 80% power using a 5% level test.

The same sample-size calculation can be specified using the general syntax of power logistic. We still specify the effect size as an argument, but now we explicitly specify the distribution of X in the x() option. Covariate X follows a Bernoulli distribution with parameter $p_X = 0.57$. Instead of specifying $\Pr(Y = 1|X = 0)$ in the pycondx0() option, we specify $\Pr\{Y = 1|X = 0, \mathbf{Z} = E(\mathbf{Z})\}$ in the pycondx0zm() option. This logistic regression does not contain nuisance covariates, so in this case, $\Pr(Y = 1|X = 0) = \Pr\{Y = 1|X = 0, \mathbf{Z} = E(\mathbf{Z})\}$.

```
. power logistic 1.5, x(distribution(bernoulli 0.57)) pycondx0zm(0.2)
Estimated sample size for logistic regression odds-ratio test
Likelihood-ratio test
H0: OR_X = 1  versus  Ha: OR_X != 1
Study parameters:
        alpha =    0.0500
        power =    0.8000
        delta =    1.5000  (odds ratio)
   pycondx0zm =    0.2000
Covariate of interest X: Bernoulli(px), bins = 2
      oratiox =    1.5000
           px =    0.5700
Estimated sample size:
            N =     1,092
```

The result is the same as before: a sample of 1,092 birds is required to detect an odds ratio of 1.5 with 80% power using a 5% level test.

◁

▷ Example 2: Sample size with two binary covariates

Continuing with the study of migratory birds, we now include nuisance covariate Z as an indicator that a bird partners with the same mate. Previous research indicates that 40% of the target population has the same mate as the year before, 16% of male birds with new mates return to the same nesting site they used the previous year, and the odds that a bird with the same mate will return to the same nesting site is 1.8 times the odds that a bird with a new mate will do so. Written mathematically, this is, $\Pr(Z = 1) = 0.4$, $\Pr(Y = 1|X = 0, Z = 0) = 0.16$, and $\text{OR}_Z = 1.8$, respectively.

We can use the two binary syntax of `power logistic` to perform the sample-size calculation. Instead of specifying $\Pr(Y = 1|X = 0)$ in the `pycondx0()` option (as we did when we had only one binary covariate), we now include $\Pr(Y = 1|X = 0, Z = 0)$ using the `pycondx0z0()` option. We specify $\Pr(Z = 1)$ and OR_Z using the `pz()` and `oratioz()` options, respectively.

```
. power logistic 1.5, px(0.57) pz(0.4) oratioz(1.8) pycondx0z0(0.16)
Estimated sample size for logistic regression odds-ratio test
Likelihood-ratio test
H0: OR_X = 1  versus  Ha: OR_X != 1
Study parameters:
        alpha =    0.0500
        power =    0.8000
        delta =    1.5000  (odds ratio)
      oratiox =    1.5000
           px =    0.5700
      oratioz =    1.8000
           pz =    0.4000
   pycondx0z0 =    0.1600
Estimated sample size:
            N =     1,118
```

The output when using the two binary syntax is similar to the output we saw with the one binary syntax, but now it includes information about nuisance covariate Z. Including Z in the logistic regression brings the required sample size to 1,118 birds.

We now demonstrate how this sample-size calculation can be specified using the general syntax of `power logistic`. We still specify the effect size as an argument, but now we specify the distribution of X in the `x()` option and the distribution of Z in the `z1()` option. We specify the odds ratio for Z using the `oratio()` suboption of `z1()`. Instead of specifying $\Pr(Y = 1|X = 0, Z = 0)$, we now specify the logistic intercept ζ_0 directly using the `intercept()` option, where $\zeta_0 = H^{-1}\{\Pr(Y = 1|X = 0, Z = 0)\} = \texttt{logit(0.16)} = -1.658$.

```
. power logistic 1.5, x(distribution(bernoulli 0.57))
> z1(distribution(bernoulli 0.4) oratio(1.8)) intercept(-1.658)
Estimated sample size for logistic regression odds-ratio test
Likelihood-ratio test
H0: OR_X = 1  versus  Ha: OR_X != 1
Study parameters:
        alpha =    0.0500
        power =    0.8000
        delta =    1.5000  (odds ratio)
    intercept =   -1.6580
Covariate of interest X: Bernoulli(px), bins = 2
      oratiox =    1.5000
           px =    0.5700
Nuisance covariate Z1: Bernoulli(pz1), bins = 2
     oratioz1 =    1.8000
          pz1 =    0.4000
Estimated sample size:
            N =     1,118
```

The output of the general syntax is formatted differently than the output from the two binary syntax, but the parameters and estimated sample size are the same.

◁

▷ Example 3: Sample size with arbitrary covariates

Continuing with the study of migratory birds, we now include both nuisance covariates: Z_1 is an indicator that a bird has partnered with the same mate, and Z_2 is the bird's weight in grams. We model covariates X and Z_1 as before, and we model Z_2 as a normal random variable with mean $\mu_{Z_2} = 52$, standard deviation $\sigma_{Z_2} = 4$, and effect $\text{OR}_{Z_2} = 1.1$.

In example 1, which did not have nuisance covariates, we observed that $\Pr\{Y = 1|X = 0, \mathbf{Z} = E(\mathbf{Z})\}$ was equal to $\Pr(Y = 1|X = 0) = 0.2$. Now that we have nuisance covariates, the two probabilities are slightly different. $\Pr(Y = 1|X = 0)$ marginalizes out the effect of Z_1 and Z_2 by averaging over the $\mathbf{Z}$ values in the population, yielding the average probability of nesting-site reuse for male birds in the population. On the other hand, $\Pr\{Y = 1|X = 0, \mathbf{Z} = E(\mathbf{Z})\}$ is the probability of nesting-site reuse for a male bird with "average" characteristics. For a detailed description of the difference between the two probabilities, see *Average response versus response at average* in [R] **margins**. We do not know the precise value of $\Pr\{Y = 1|X = 0, \mathbf{Z} = E(\mathbf{Z})\}$, so we perform a sensitivity analysis by specifying multiple values of `pycondx0zm()`.

```
. power logistic 1.5, x(distribution(bernoulli 0.57))
> z1(distribution(bernoulli 0.4) oratio(1.8))
> z2(distribution(normal 52 4) oratio(1.1))
> pycondx0zm(0.18 0.2 0.22)
Estimated sample size for logistic regression odds-ratio test
Likelihood-ratio test
H0: OR_X = 1  versus  Ha: OR_X != 1
Covariate of interest X: Bernoulli(px)
Nuisance covariates:
   Z1: Bernoulli(pz1)
   Z2: Normal(muz2, sigmaz2)
```

alpha	power	N	delta	oratiox	px	oratioz1	pz1	oratioz2	muz2
.05	.8	1,179	1.5	1.5	.57	1.8	.4	1.1	52
.05	.8	1,106	1.5	1.5	.57	1.8	.4	1.1	52
.05	.8	1,047	1.5	1.5	.57	1.8	.4	1.1	52

sigmaz2	pycondx0zm
4	.18
4	.2
4	.22

`power logistic` displays a table when we provide *numlist* of values for `pycondx0zm()`. We see that a sample of 1,179 birds would be necessary if $\Pr\{Y = 1|X = 0, \mathbf{Z} = E(\mathbf{Z})\} = 0.18$. To draw a power curve for a sample of 1,179, we add options `n(1179)` and `graph`.

```
. power logistic 1.5, x(distribution(bernoulli 0.57))
> z1(distribution(bernoulli 0.4) oratio(1.8))
> z2(distribution(normal 52 4) oratio(1.1))
> pycondx0zm(0.18 0.2 0.22) n(1179) graph
Covariate of interest X: Bernoulli(px)
Nuisance covariates:
   Z1: Bernoulli(pz1)
   Z2: Normal(muz2, sigmaz2)
```

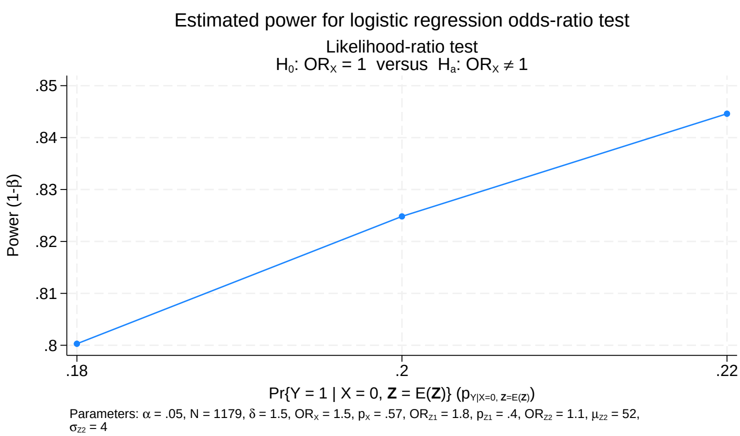

Figure 1. Power curve for a sample of 1,179

As the value of $\Pr\{Y = 1|X = 0, \mathbf{Z} = E(\mathbf{Z})\}$ increases, the power of the test increases as well, reaching a power of nearly 85% when $\Pr\{Y = 1|X = 0, \mathbf{Z} = E(\mathbf{Z})\} = 0.22$.

◁

Reference

Bush, S. 2015. Sample size determination for logistic regression: A simulation study. *Communications in Statistics—Simulation and Computation* 44: 360–373. https://doi.org/10.1080/03610918.2013.777458.

Also see

[PSS-2] **power logistic general** — Power analysis for logistic regression: General case+

[PSS-2] **power logistic onebin** — Power analysis for logistic regression with one binary covariate+

[PSS-2] **power logistic twobin** — Power analysis for logistic regression with two binary covariates+

[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] **logistic** — Logistic regression, reporting odds ratios

[R] **logit** — Logistic regression, reporting coefficients

[R] **lrtest** — Likelihood-ratio test after estimation