Chapter 861

Tests for the Odds Ratio in Logistic Regression with One Binary X (Wald Test)

Introduction

Logistic regression expresses the relationship between a binary response variable and one or more independent variables called *covariates*. This procedure is for the case when there is only one, binary covariate (X) in the logistic regression model and a Wald test is used to test its significance. Often, Y is called the *response* variable and X is referred to as the *exposure* variable. For example, Y might refer to the presence or absence of cancer and X might indicate whether the subject smoked or not.

Power Calculations

Using the *logistic model*, the probability of a binary event is

$$\Pr(Y = 1|X) = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)} = \frac{1}{1 + \exp(-\beta_0 - \beta_1 X)}$$

This formula can be rearranged so that it is linear in *X* as follows

$$\log\left(\frac{\Pr(Y = 1|X)}{1 - \Pr(Y = 1|X)}\right) = \beta_0 + \beta_1 X$$

Note that the left side is the logarithm of the odds of a response event (Y = 1) versus a response non-event (Y = 0). This is sometimes called the *logit* transformation of the probability. In the logistic regression model, the magnitude of the association of *X* and *Y* is represented by the slope β_1. Since X is binary, only two cases need be considered: X = 0 and X = 1.

The logistic regression model lets us define two quantities

$$P_0 = \Pr(Y = 1|X = 0) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$$

$$P_1 = \Pr(Y = 1|X = 1) = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}$$

These values are combined in the odds ratio (OR) of P_1 to P_0 resulting in

$$OR_{yx} = \exp(\beta_1)$$

or, by taking the logarithm of both sides, simply

$$\log(OR_{yx}) = \log\left(\frac{\frac{P_1}{(1-P_1)}}{\frac{P_0}{(1-P_0)}}\right) = \beta_1$$

Hence the relationship between Y and X can be quantified as a single regression coefficient. It well known that the distribution of the maximum likelihood estimate of β_1 is asymptotically normal. The significance of this slope is commonly tested with the Wald test

$$z = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}$$

It is considered good practice to base the power analysis on the same test statistic that is used for analysis, so we base our power analysis on the Wald test.

Demidenko (2007) gives the following formula for the power of the Wald test in the case of a logistic regression model. His formula for the power of a two-sided Wald test is

$$\text{Power} = \Phi\left(-z_{1-\frac{\alpha}{2}} + \frac{\beta_1\sqrt{N}}{\sqrt{V}}\right) + \Phi\left(-z_{1-\frac{\alpha}{2}} - \frac{\beta_1\sqrt{N}}{\sqrt{V}}\right)$$

where z is the usual quantile of the standard normal distribution and V is calculated as follows.

Let p_x be the probability that X = 1 in the sample. The information matrix for this model is

$$I = \begin{bmatrix} \frac{p_x\exp(\beta_0+\beta_1)}{\left(1+\exp(\beta_0+\beta_1)\right)^2} + \frac{(1-p_x)\exp(\beta_0)}{\left(1+\exp(\beta_0)\right)^2} & \frac{p_x\exp(\beta_0+\beta_1)}{\left(1+\exp(\beta_0+\beta_1)\right)^2} \\ \frac{p_x\exp(\beta_0+\beta_1)}{\left(1+\exp(\beta_0+\beta_1)\right)^2} & \frac{p_x\exp(\beta_0+\beta_1)}{\left(1+\exp(\beta_0+\beta_1)\right)^2} \end{bmatrix}$$

The value of V is the (2,2) element of the inverse of I.

The values of β_0 and β_1 are calculated from P_1 or OR_{yx} and P_0 using

$$\beta_0 = \log\left(\frac{P_0}{1-P_0}\right)$$

$$\beta_1 = \log(OR_{yx}) = \log\left(\frac{\frac{P_1}{(1-P_1)}}{\frac{P_0}{(1-P_0)}}\right)$$

Thus, the effect size is calculated in terms of β_0 and β_1 but specified in terms of P_1 or OR_{yx} and P_0.

Example 1 – Power for a Fixed Sample Size

A study is to be undertaken to study the association between the occurrence of a certain type of cancer (response variable) and the presence of a certain food in the diet. The baseline cancer event rate is 7%. The researchers want a sample size large enough to detect an odds ratio of 2.0 with 80% power at the 0.05 significance level with a two-sided Wald test. They also want to look at the sensitivity of the analysis to the specification of the odds ratio, so they also want to obtain the results for odds ratios of 1.75 and 2.25. They want to begin by considering sample sizes between 200 and 1000. The researchers determine that about 50% of the sample eat the food being studied.

Setup

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

Design Tab

Solve For .... **Power**
Alternative Hypothesis .... **Two-Sided**
Alpha.. **0.05**
N (Sample Size).. **200 to 1000 by 200**
P0 [Pr(Y=1|X=0)]... **0.07**
Use P1 or Odds Ratio .... **ORyx**
Odds Ratio (Odds1/Odds0) **1.75 2.0 2.25**
Percent with X = 1... **50**

Output

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

Numeric Reports

Numeric Results

Solve For: Power
Test Type: Wald Test
Alternative Hypothesis: H1: ORyx ≠ 1

	Sample	Prevalence	Response Probability			
	Size	Percent	Pr(Y = 1 | X = 0)	Pr(Y = 1 | X = 1)	Odds Ratio	
Power	N	with X = 1	P0	P1	ORyx	Alpha
0.2008	200	50	0.07	0.116	1.75	0.05
0.3522	400	50	0.07	0.116	1.75	0.05
0.4902	600	50	0.07	0.116	1.75	0.05
0.6082	800	50	0.07	0.116	1.75	0.05
0.7049	1000	50	0.07	0.116	1.75	0.05
0.2917	200	50	0.07	0.131	2.00	0.05
0.5138	400	50	0.07	0.131	2.00	0.05
0.6854	600	50	0.07	0.131	2.00	0.05
0.8053	800	50	0.07	0.131	2.00	0.05
0.8837	1000	50	0.07	0.131	2.00	0.05
0.3880	200	50	0.07	0.145	2.25	0.05
0.6588	400	50	0.07	0.145	2.25	0.05
0.8268	600	50	0.07	0.145	2.25	0.05
0.9178	800	50	0.07	0.145	2.25	0.05
0.9629	1000	50	0.07	0.145	2.25	0.05

Logistic Regression Equation: Log(P/(1 - P)) = β0 + β1 × X, where P = Pr(Y = 1|X) and X is binary.

Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N	The sample size.
Percent with X = 1	The percent of the sample in which the exposure is 1 (present).
P0	The response probability at X = 0. That is, P0 = Pr(Y = 1|X = 0).
P1	The response probability at X = 1. That is, P1 = Pr(Y = 1|X = 1).
ORyx	The odds ratio under the alternative hypothesis. ORyx = [P1/(1 - P1)] / [P0/(1 - P0)].
Alpha	The probability of rejecting a true null hypothesis.

Summary Statements

An odds ratio (binary response Y versus one binary X) design will be used to test whether the probability that Y equals 1 when X equals 1 (P1 = Pr(Y = 1|X = 1)) is different from the (baseline) probability that Y equals 1 when X equals 0 (P0 = Pr(Y = 1|X = 0)) of 0.07 (H0: ORyx = 1 versus H1: ORyx ≠ 1, given P0 = 0.07). The comparison will be made using a two-sided logistic regression Wald test of β1 (using the model Log(P / (1 - P)) = β0 + β1 × X, where P = Pr(Y = 1|X)), with a Type I error rate (α) of 0.05. Among subjects, 50% are assumed to have the value X = 1 (or be in the X = 1 group), and the remaining 50% are assumed to have the value X = 0. To detect an odds ratio (ORyx) of 1.75 (or P1 of 0.116) with a sample size of 200, the power is 0.2008.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	200	250	50
20%	400	500	100
20%	600	750	150
20%	800	1000	200
20%	1000	1250	250

Dropout Rate The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N The evaluable sample size at which power is computed (as entered by the user). If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.
N' The total number of subjects that should be enrolled in the study in order to obtain N evaluable subjects, based on the assumed dropout rate. N' is calculated by inflating N using the formula N' = N / (1 - DR), with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D The expected number of dropouts. D = N' - N.

Dropout Summary Statements

Anticipating a 20% dropout rate, 250 subjects should be enrolled to obtain a final sample size of 200 subjects.

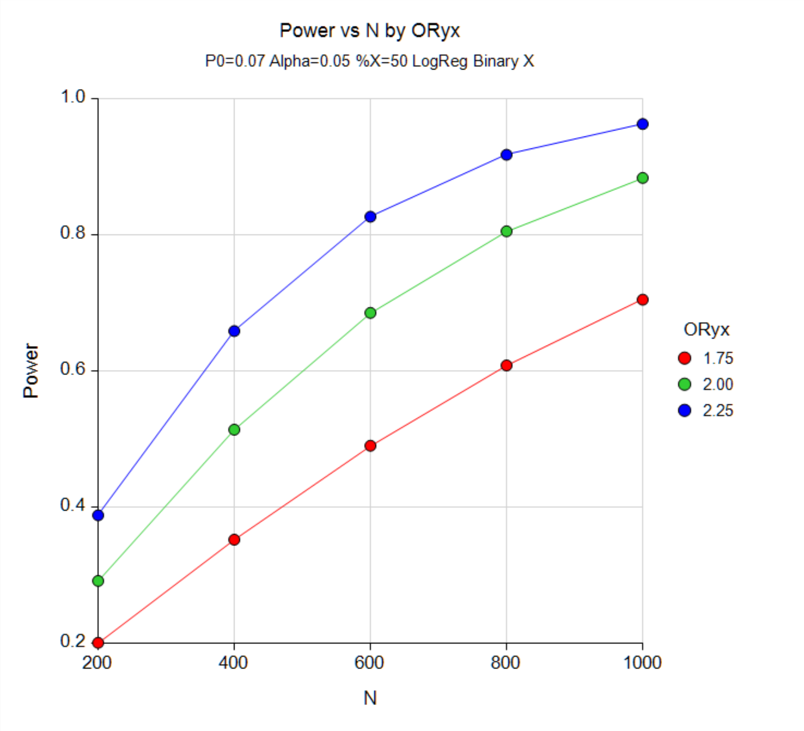
References

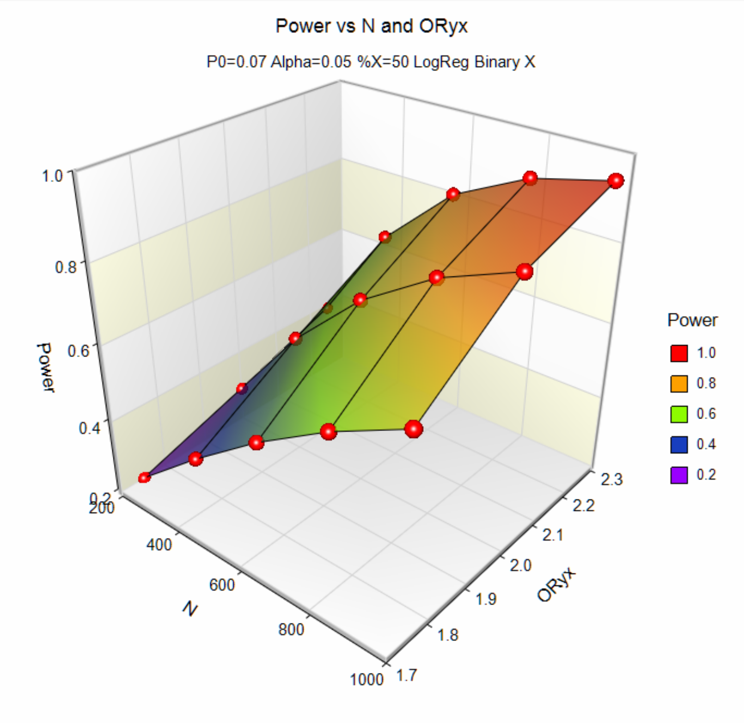
Demidenko, Eugene. 2007. 'Sample size determination for logistic regression revisited', Statistics in Medicine, Volume 26, pages 3385-3397.
Demidenko, Eugene. 2008. 'Sample size and optimal design for logistic regression with binary interaction', Statistics in Medicine, Volume 27, pages 36-46.
Rochon, James. 1989. 'The Application of the GSK Method to the Determination of Minimum Sample Sizes', Biometrics, Volume 45, pages 193-205.

This report shows the power for each of the scenarios.

Plots Section

Plots





These plots show the power versus the sample size for the three values of the odds ratio.

Example 2 – Sample Size for a Fixed Power

Continuing with the settings given in Example 1, the researchers want to obtain the exact sample size necessary for odds ratio value.

Setup

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

Design Tab

Solve For .... **Sample Size**
Alternative Hypothesis .... **Two-Sided**
Power.. **0.8**
Alpha... **0.05**
P0 [Pr(Y=1|X=0)].. **0.07**
Use P1 or Odds Ratio .... **ORyx**
Odds Ratio (Odds1/Odds0) **1.75 2.0 2.25**
Percent with X = 1.. **50**

Output

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

Numeric Results

Solve For: Sample Size
Test Type: Wald Test
Alternative Hypothesis: H1: ORyx ≠ 1

		Prevalence	Response Probability			
Power	Sample Size N	Percent with X = 1	Pr(Y = 1 | X = 0) P0	Pr(Y = 1 | X = 1) P1	Odds Ratio ORyx	Alpha
0.8002	1258	50	0.07	0.116	1.75	0.05
0.8004	790	50	0.07	0.131	2.00	0.05
0.8004	560	50	0.07	0.145	2.25	0.05

Logistic Regression Equation: Log(P/(1 - P)) = β0 + β1 × X, where P = Pr(Y = 1|X) and X is binary.

This report shows the sample size requirement for each odds ratio (ORyx). For example, it shows that a power of 80% is achieved at a sample size of 790 for an odds ratio of 2.0 and 1258 for an odds ratio of 1.75.

Example 3 – Validation for a Binary Covariate using Demidenko (2007)

Demidenko (2007), page 3392, gives an example in which alpha = 0.05, power = 0.8, ORyx = 2, P0 = 0.37, and percent X = 1 is 20. These parameters give an N of 416. We will validate this routine by running the same problem.

Setup

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

Design Tab

Solve For .... **Sample Size**
Alternative Hypothesis .... **Two-Sided**
Power.. **0.8**
Alpha.. **0.05**
P0 [Pr(Y=1|X=0)].. **0.37**
Use P1 or Odds Ratio .... **ORyx**
Odds Ratio (Odds1/Odds0) **2**
Percent with X = 1... **20**

Output

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

Numeric Results

Solve For: Sample Size
Test Type: Wald Test
Alternative Hypothesis: H1: ORyx ≠ 1

		Prevalence	Response Probability			
Power	Sample Size N	Percent with X = 1	Pr(Y = 1 | X = 0) P0	Pr(Y = 1 | X = 1) P1	Odds Ratio ORyx	Alpha
0.8005	417	20	0.37	0.54	2	0.05

Logistic Regression Equation: Log(P/(1 - P)) = β0 + β1 × X, where P = Pr(Y = 1|X) and X is binary.

PASS computes a sample size of 417 which is one more than Demidenko's. Note that an N of 416 achieves a power of 0.7996, slightly less than the 0.8000 requested.