

## 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  $\beta_1$ . Since X is binary, only two cases need be considered: X = 0 and X = 1.

The logistic regression model lets us define two quantities

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

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

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

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  $\beta_1$  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

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

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

The values of  $\beta_0$  and  $\beta_1$  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  $\beta_0$  and  $\beta_1$ , but specified in terms of  $P_1$  or  $OR_{yx}$  and  $P_0$ .

---

## Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

---

### Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

---

#### Solve For

##### Solve For

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Alpha*, *Power*, *Sample Size*, or *ORyx* and *PI*. Select *Sample Size* when you want to calculate the sample size needed to achieve a given power and alpha level. Select *Power* when you want to calculate the power of an experiment.

---

#### Test

##### Alternative Hypothesis

Specify whether the test is one-sided or two-sided. When a two-sided hypothesis is selected, the value of alpha is halved. Everything else remains the same.

Commonly, accepted procedure is to use the Two-Sided option unless you can justify using a one-sided test.

---

#### Power and Alpha

##### Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. A type-II error occurs when you fail to reject the null hypothesis of equal probabilities of the event of interest when in fact they are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

##### Alpha

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis of equal probabilities when in fact they are equal.

Values of alpha must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

---

#### Sample Size

##### N (Sample Size)

This option specifies the total number of observations in the sample. You may enter a single value or a list of values.

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

---

### Baseline Probability

#### **P0 [Pr(Y = 1 | X = 0)]**

This gives the value of the baseline probability of a response, P0, when the exposure is not present.

P0 is a probability, so it must be between zero and one. It cannot be equal to P1.

---

### P1 or Odds Ratio

#### **Use P1 or ORyx**

Indicate whether to specify the odds ratio being tested in terms of P1 or the ORyx. Values of the parameter not selected are ignored.

#### **P1 [Pr(Y = 1 | X = 1)]**

Specify one or more values for P1, the probability of a response when X = 1. This is the response probability at which the power is calculated.

P1 is a probability, so it must be between zero and one. It cannot be equal to P0.

#### **ORyx (Y,X Odds Ratio)**

Specify one or more values of the odds ratio of Y and X, a measure of the effect size (event rate) that is to be detected by the study. This is the ratio of the odds of the outcome Y given that the exposure X = 1 to the odds of Y = 1 given X = 0. That is,  $\text{odds}(Y=1|X=1) / \text{odds}(Y=1|X=0)$ . Note that  $\text{odds}(A) = \text{Pr}(A) / \text{Pr}(\text{Not } A)$

You can enter a single value such as 1.5 or a series of values such as 1.5 2 2.5 or 0.5 to 0.9 by 0.1.

The range of this parameter is  $0 < \text{OR}_{yx} < \infty$  (typically,  $0.1 < \text{OR}_{yx} < 10$ ). Since this is the value under alternative hypothesis,  $\text{OR}_{yx} \neq 1$ .

---

### Prevalence of X

#### **Percent with X = 1**

This is the percentage of the sample in which X = 1. It is often called the prevalence of X.

You can enter a single value or a range of values. The permissible range is 1 to 99.

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

## 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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for the Odds Ratio in Logistic Regression with One Binary X (Wald Test)** procedure. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

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

### Annotated Output

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

### Numeric Results

**Numeric Results for Two-Sided Wald Test**  
Alternative Hypothesis:  $OR_{yx} \neq 1$

| Power  | N    | Percent |  | P0    | P1    | ORyx  | Alpha | Beta   |
|--------|------|---------|--|-------|-------|-------|-------|--------|
|        |      | X=1     |  |       |       |       |       |        |
| 0.2008 | 200  | 50.0    |  | 0.070 | 0.116 | 1.750 | 0.050 | 0.7992 |
| 0.3522 | 400  | 50.0    |  | 0.070 | 0.116 | 1.750 | 0.050 | 0.6478 |
| 0.4902 | 600  | 50.0    |  | 0.070 | 0.116 | 1.750 | 0.050 | 0.5098 |
| 0.6082 | 800  | 50.0    |  | 0.070 | 0.116 | 1.750 | 0.050 | 0.3918 |
| 0.7049 | 1000 | 50.0    |  | 0.070 | 0.116 | 1.750 | 0.050 | 0.2951 |
| 0.2917 | 200  | 50.0    |  | 0.070 | 0.131 | 2.000 | 0.050 | 0.7083 |
| 0.5138 | 400  | 50.0    |  | 0.070 | 0.131 | 2.000 | 0.050 | 0.4862 |
| 0.6854 | 600  | 50.0    |  | 0.070 | 0.131 | 2.000 | 0.050 | 0.3146 |
| 0.8053 | 800  | 50.0    |  | 0.070 | 0.131 | 2.000 | 0.050 | 0.1947 |
| 0.8837 | 1000 | 50.0    |  | 0.070 | 0.131 | 2.000 | 0.050 | 0.1163 |
| 0.3880 | 200  | 50.0    |  | 0.070 | 0.145 | 2.250 | 0.050 | 0.6120 |
| 0.6588 | 400  | 50.0    |  | 0.070 | 0.145 | 2.250 | 0.050 | 0.3412 |
| 0.8268 | 600  | 50.0    |  | 0.070 | 0.145 | 2.250 | 0.050 | 0.1732 |
| 0.9178 | 800  | 50.0    |  | 0.070 | 0.145 | 2.250 | 0.050 | 0.0822 |
| 0.9629 | 1000 | 50.0    |  | 0.070 | 0.145 | 2.250 | 0.050 | 0.0371 |

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

### Report Definitions

Logistic regression equation:  $\text{Log}(P/(1-P)) = \beta_0 + \beta_1 \times X$ , where  $P = \text{Pr}(Y = 1|X)$  and  $X$  is binary.

Power is the probability of rejecting a false null hypothesis.

$N$  is the sample size.

Percent  $X=1$  is the percent of the sample in which the exposure is 1 (present).

$P_0$  is the response probability at  $X = 0$ . That is,  $P_0 = \text{Pr}(Y = 1|X = 0)$ .

$P_1$  is the response probability at  $X = 1$ . That is,  $P_1 = \text{Pr}(Y = 1|X = 1)$ .

$OR_{yx}$  is the odds ratio under the alternative hypothesis. That is, it is  $[P_1/(1-P_1)]/[P_0/(1-P_0)]$ .

Alpha is the probability of rejecting a true null hypothesis.

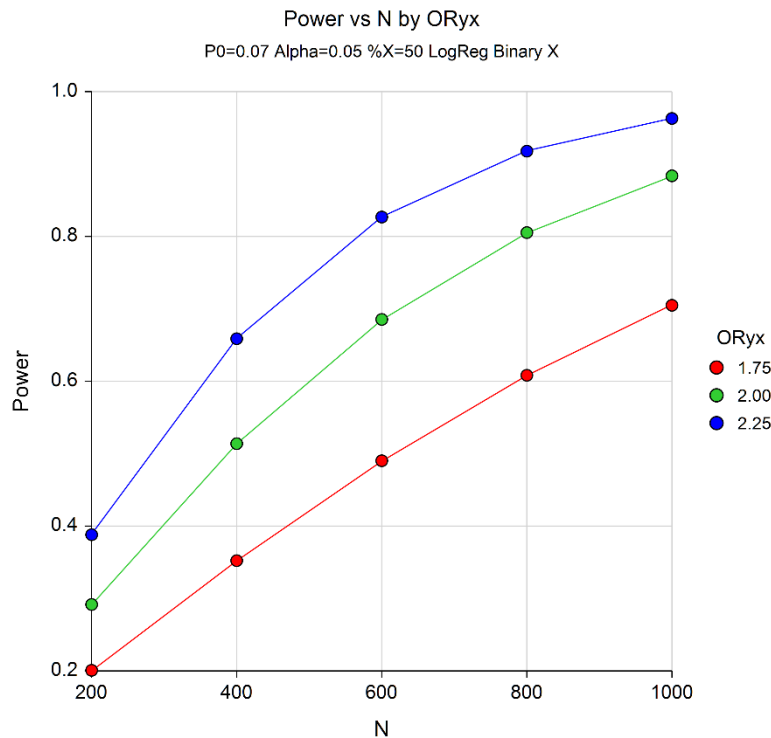
Beta is the probability of accepting a false null hypothesis.

### Summary Statements

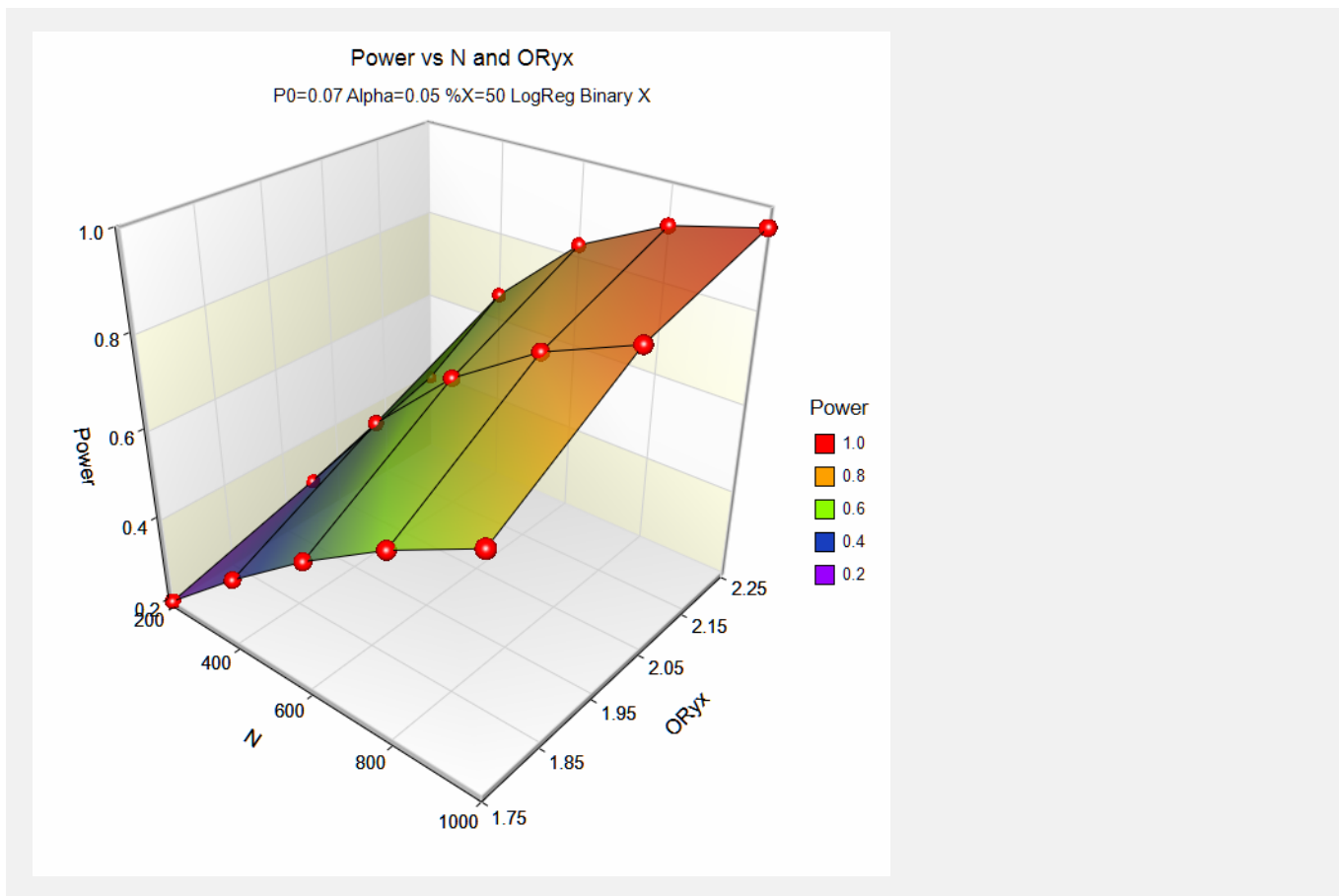
A logistic regression of a binary response variable ( $Y$ ) on a binary independent variable ( $X$ ) with a sample size of 200 observations (of which 50% are in the group  $X=0$  and 50% are in the group  $X=1$ ) achieves 20% power at a 0.050 significance level to detect a change in  $\text{Prob}(Y=1)$  from the baseline value of 0.070 to 0.116. This change corresponds to an odds ratio of 1.750. A two-sided Wald test is used.

This report shows the power for each of the scenarios.

### Plot Section



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



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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for the Odds Ratio in Logistic Regression with One Binary X (Wald Test)** procedure. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

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

### Output

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

#### Numeric Results

| Numeric Results for Two-Sided Wald Test |      |         |       |       |       |       |        |      |
|---|------|---------|-------|-------|-------|-------|--------|------|
| Alternative Hypothesis: ORyx ≠ 1        |      |         |       |       |       |       |        |      |
| Power                                   | N    | Percent |       | P0    | P1    | ORyx  | Alpha  | Beta |
|   |      | X=1     |       |       |       |       |        |      |
| 0.8002                                  | 1258 | 50.0    | 0.070 | 0.116 | 1.750 | 0.050 | 0.1998 |      |
| 0.8004                                  | 790  | 50.0    | 0.070 | 0.131 | 2.000 | 0.050 | 0.1996 |      |
| 0.8004                                  | 560  | 50.0    | 0.070 | 0.145 | 2.250 | 0.050 | 0.1996 |      |

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,  $OR_{yx} = 2$ ,  $P_0 = 0.37$ , and percent  $X = 1$  is 20. These parameters give an  $N$  of 416. We will validate this routine by running the same problem.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Logistic Regression with One Binary Covariate using Wald Test** procedure. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

| <u>Option</u>                  | <u>Value</u>                |
|--------------------------------|-----------------------------|
| <b>Design Tab</b>              |                             |
| Solve For .....                | <b>Sample Size</b>          |
| Alternative Hypothesis .....   | <b>Two-Sided</b>            |
| Power .....                    | <b>0.8</b>                  |
| Alpha .....                    | <b>0.05</b>                 |
| $P_0$ [Pr( $Y=1 X=0$ )] .....  | <b>0.37</b>                 |
| Use $P_1$ or Odds Ratio .....  | <b><math>OR_{yx}</math></b> |
| Odds Ratio (Odds1/Odds0) ..... | <b>2</b>                    |
| Percent with $X = 1$ .....     | <b>20</b>                   |

### Output

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

### Numeric Results

| Numeric Results for Two-Sided Wald Test  |         |      |       |       |           |       |        |
|--|---------|------|-------|-------|-----------|-------|--------|
| Alternative Hypothesis: $OR_{yx} \neq 1$ |         |      |       |       |           |       |        |
| Power                                    | Percent |      | $P_0$ | $P_1$ | $OR_{yx}$ | Alpha | Beta   |
|  | N       | X=1  |       |       |           |       |        |
| 0.8005                                   | 417     | 20.0 | 0.370 | 0.540 | 2.000     | 0.050 | 0.1995 |

PASS calculates 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.