Chapter 850

# Cox Regression

## Introduction

Cox proportional hazards regression models the relationship between the hazard function $\lambda(t|X)$ of survival time and *k* covariates using the following formula

$$\log\left(\frac{\lambda(t|X)}{\lambda_0(t)}\right) = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

where $\lambda_0(t)$ is the baseline hazard. Note that the covariates may be discrete or continuous.

This procedure calculates power and sample size for testing the hypothesis that $\beta_1 = 0$ versus the alternative that $\beta_1 = B$. Note that $\beta_1$ is the change in log hazards for a one-unit change in $X_1$ when the rest of the covariates are held constant. The procedure assumes that this hypothesis will be tested using the Wald (or score) statistic

$$z = \frac{\hat{\beta}_1}{\sqrt{\mathrm{Var}(\hat{\beta}_1)}}$$

## Power Calculations

Suppose you want to test the null hypothesis that $\beta_1 = 0$ versus the alternative that $\beta_1 = B$. Hsieh and Lavori (2000) gave a formula relating sample size, $\alpha$, $\beta$, and *B* when $X_1$ is normally distributed. The sample size formula is

$$D = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{(1 - R^2)\sigma^2 B^2}$$

where *D* is the number of events, $\sigma^2$ is the variance of $X_1$, and $R^2$ is the proportion of variance explained by the multiple regression of $X_1$ on the remaining covariates. It is interesting to note that the number of censored observations does not enter into the power calculations. To obtain a formula for the sample size, *N*, we inflate *D* by dividing by *P*, the proportion of subjects that fail.

Thus, the formula for *N* is

$$N = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{P(1 - R^2)\sigma^2 B^2}$$

This formula is an extension of an earlier formula for the case of a single, binary covariate derived by Schoenfeld (1983). Thus, it may be used with discrete or continuous covariates.

## Assumptions

It is important to note that this formulation assumes that proportional hazards model with $k$ covariates is valid. However, it does not assume exponential survival times.

# Example 1 – Power for Several Sample Sizes

Cox regression will be used to analyze the power of a survival time study. From past experience, the researchers want to evaluate the sample size needs for detecting regression coefficients of 0.2 and 0.3 for the independent variable of interest. The variable has a standard deviation of 1.20. The *R*-squared of this variable with seven other covariates is 0.18.

The event rate is thought to be 70% over the 3-year duration of the study. The researchers will test their hypothesis using a 5% significance level with a two-sided Wald test. They decide to calculate the power at sample sizes between 5 and 250.

## 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) | **5 to 250 by 40** |
| P (Overall Event Rate) | **0.70** |
| B (Log Hazard Ratio) | **0.2 0.3** |
| R-Squared of X1 with Other X's | **0.18** |
| S (Standard Deviation of X1) | **1.2** |

# Output

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

## Numeric Reports

### Numeric Results

Solve For: Power
Alternative Hypothesis: Two-Sided

| Power | Sample Size N | Regression Coefficient B | Standard Deviation of X1 SD | Event Rate P | R-Squared of X1 with Other X's $R^2$ | Alpha |
|---|---|---|---|---|---|---|
| 0.06017 | 5 | 0.2 | 1.2 | 0.7 | 0.18 | 0.05 |
| 0.22959 | 45 | 0.2 | 1.2 | 0.7 | 0.18 | 0.05 |
| 0.38837 | 85 | 0.2 | 1.2 | 0.7 | 0.18 | 0.05 |
| 0.52908 | 125 | 0.2 | 1.2 | 0.7 | 0.18 | 0.05 |
| 0.64643 | 165 | 0.2 | 1.2 | 0.7 | 0.18 | 0.05 |
| 0.74004 | 205 | 0.2 | 1.2 | 0.7 | 0.18 | 0.05 |
| 0.81223 | 245 | 0.2 | 1.2 | 0.7 | 0.18 | 0.05 |
| 0.08849 | 5 | 0.3 | 1.2 | 0.7 | 0.18 | 0.05 |
| 0.44815 | 45 | 0.3 | 1.2 | 0.7 | 0.18 | 0.05 |
| 0.71043 | 85 | 0.3 | 1.2 | 0.7 | 0.18 | 0.05 |
| 0.86202 | 125 | 0.3 | 1.2 | 0.7 | 0.18 | 0.05 |
| 0.93865 | 165 | 0.3 | 1.2 | 0.7 | 0.18 | 0.05 |
| 0.97412 | 205 | 0.3 | 1.2 | 0.7 | 0.18 | 0.05 |
| 0.98953 | 245 | 0.3 | 1.2 | 0.7 | 0.18 | 0.05 |

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N The size of the sample drawn from the population.
B The size of the regression coefficient to be detected.
SD The standard deviation of X1.
P The event rate.
$R^2$ The R-squared achieved when X1 is regressed on the other covariates.
Alpha The probability of rejecting a true null hypothesis.

### Summary Statements

A Cox regression (survival response Y versus X's) design will be used to test whether the (natural) log hazard ratio (B) is different from 0, where B is the coefficient for X1, the variable of interest (H0: B = 0 versus H1: B ≠ 0). The comparison will be made using a two-sided Cox regression slope test of B, with a Type I error rate (α) of 0.05. The standard deviation of X1 is assumed to be 1.2. The R-squared of X1 with the other X's in the model is assumed to be 0.18. The anticipated overall event rate is 0.7. To detect a log hazard ratio (regression coefficient B) of 0.2 with a sample size of 5, the power is 0.06017.

**Dropout-Inflated Sample Size**

| Dropout Rate | Sample Size N | Dropout-Inflated Enrollment Sample Size N' | Expected Number of Dropouts D |
|---|---|---|---|
| 20% | 5 | 7 | 2 |
| 20% | 45 | 57 | 12 |
| 20% | 85 | 107 | 22 |
| 20% | 125 | 157 | 32 |
| 20% | 165 | 207 | 42 |
| 20% | 205 | 257 | 52 |
| 20% | 245 | 307 | 62 |

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, 7 subjects should be enrolled to obtain a final sample size of 5 subjects.

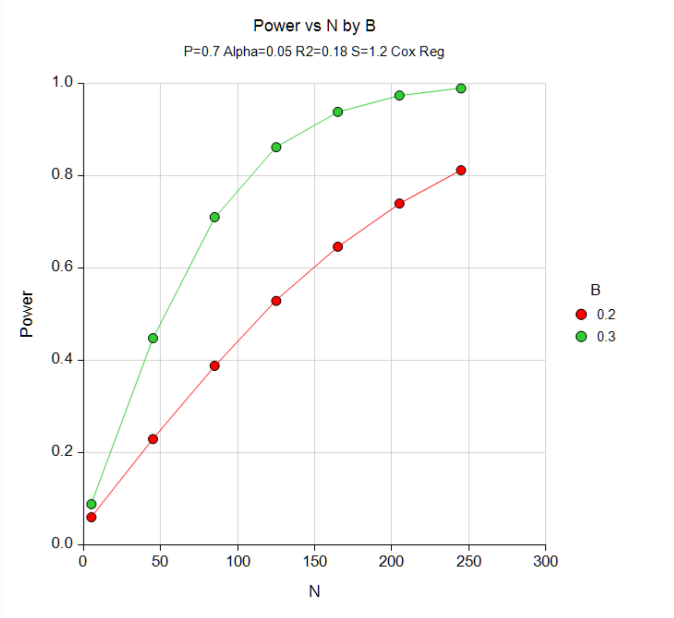
**References**

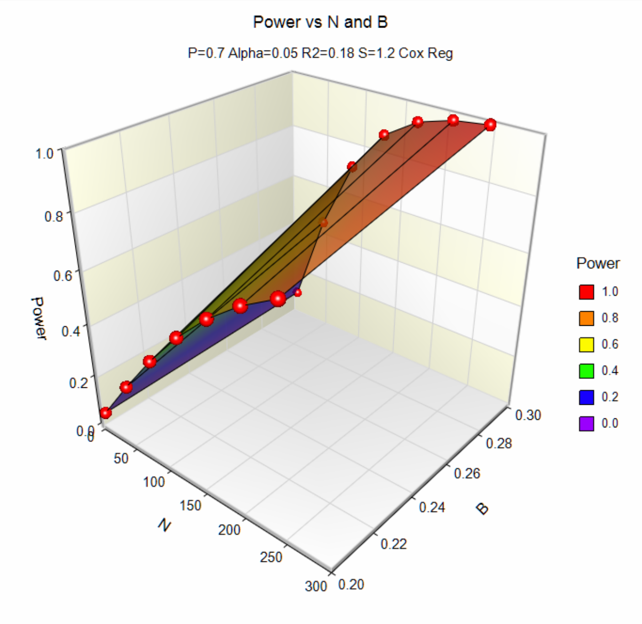
Hsieh, F.Y. and Lavori, P.W. 2000. 'Sample-Size Calculations for the Cox Proportional Hazards Regression Model with Nonbinary Covariates', Controlled Clinical Trials, Volume 21, pages 552-560.
Schoenfeld, David A. 1983. 'Sample-Size Formula for the Proportional-Hazards Regression Model', Biometrics, Volume 39, pages 499-503.

This report shows the power for each of the scenarios.

## Plots Section

### Plots



# Example 2 – Validation using Hsieh and Lavori (2000)

Hsieh and Lavori (2000) present an example which we will use to validate this program. In this example, $B$ = 1.0, $S$ = 0.3126, $R2$ = 0.1837, $P$ = 0.738, one-sided alpha = 0.05, and power = 0.80. They calculated $N$ = 107.

## 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 .................................. **One-Sided**
Power............................................................ **0.80**
Alpha............................................................ **0.05**
P (Overall Event Rate).................................. **0.738**
B (Log Hazard Ratio) .................................... **1.0**
R-Squared of X1 with Other X's.................... **0.1837**
S (Standard Deviation of X1) ........................ **0.3126**

## Output

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

### Numeric Reports R2 = 0.1837

**Numeric Results**

Solve For: Sample Size
Alternative Hypothesis: One-Sided

| Power | Sample Size N | Regression Coefficient B | Standard Deviation of X1 SD | Event Rate P | R-Squared of X1 with Other X's $R^2$ | Alpha |
|---|---|---|---|---|---|---|
| 0.80321 | 106 | 1 | 0.3126 | 0.738 | 0.1837 | 0.05 |

Note that **PASS** calculated 106 rather than the 107 calculated by Hsieh and Lavori (2000). The discrepancy is due to the intermediate rounding that they did. To show this, we will run a second example from Hsieh and Lavori in which $R2$ = 0 and $P$ = 1.0. In this case, $N$ = 64.

## Numeric Reports with R2 = 0 and P = 1.0

**Numeric Results**

Solve For: Sample Size
Alternative Hypothesis: One-Sided

| Power | Sample Size N | Regression Coefficient B | Standard Deviation of X1 SD | Event Rate P | R-Squared of X1 with Other X's $R^2$ | Alpha |
|---|---|---|---|---|---|---|
| 0.80399 | 64 | 1 | 0.3126 | 1 | 0 | 0.05 |

Note that **PASS** also calculated 64. Hsieh and Lavori obtained the 107 by adjusting this 64 for *P* first and then for *R*2. **PASS** does both adjustments at once, obtaining the 106. Thus, the difference is due to intermediate rounding.

# Example 3 – Validation for Binary X1 using Schoenfeld (1983)

Schoenfeld (1983), page 502, presents an example for the case when X1 is binary. In this example, *B* = ln(1.5) = 0.4055, *S* = 0.5, *R*2 = 0.0, *P* = 0.71, one-sided alpha = 0.05, and power = 0.80. Schoenfeld calculated *N* = 212.

## 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 ................................. **One-Sided**
Power........................................................... **0.80**
Alpha............................................................ **0.05**
P (Overall Event Rate) .................................. **0.71**
B (Log Hazard Ratio) .................................... **0.4055**
R-Squared of X1 with Other X's..................... **0.0**
S (Standard Deviation of X1) ........................ **0.5**

## Output

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

**Numeric Results**

Solve For: Sample Size
Alternative Hypothesis: One-Sided

| Power | Sample Size N | Regression Coefficient B | Standard Deviation of X1 SD | Event Rate P | R-Squared of X1 with Other X's $R^2$ | Alpha |
|---|---|---|---|---|---|---|
| 0.80028 | 212 | 0.4055 | 0.5 | 0.71 | 0 | 0.05 |

Note that **PASS** also obtains *N* = 212.