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 β_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

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

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, α, β, and *B* when X_1 is normally distributed. The sample size formula is

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

where *D* is the number of events, σ^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 in to 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

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

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.

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. Under most situations, you will select either *Power* for a power analysis or *Sample Size* for sample size determination.

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 by *PASS*. Everything else remains the same.

Note that the 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. In this procedure, 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.

Note that when the Overall Event Rate is set to 1.0, the sample size becomes the number of events.

P (Overall Event Rate)

Enter one or more values for the event rate. The event rate is the proportion of subjects in which the event of interest occurs during the duration of the study. This is the proportion of non-censored subjects. Since the values entered here are proportions, they must be in the range $0 < P \leq 1$.

Note that when this value is set to 1.0, the sample size is the number of events (deaths).

Effect Size – Hazard Ratio

B (Log Hazard Ratio)

This procedure calculates power or sample size for testing the hypothesis that $\beta_1 = 0$ versus the alternative that $\beta_1 = B$ in a Cox regression. Enter one or more values of B here.

B is the predicted change in log (base e) hazards corresponding to a one unit change in X1 when the other covariates are held constant. Thus, if you want to detect a hazard ratio of 1.5, enter ln(1.5) = 0.4055. Although any non-zero value may be entered, common values are between -3 and 3.

Effect Size – Covariates (X1 is the Variable of Interest)

R-Squared of X1 with Other X's

This is the R-Squared that is obtained when X1 is regressed on the other X's (covariates) in the model. Use this to account for the influence on power and sample size of adding other covariates. Note that the number of additional variables does not matter in this formulation. Only their overall relationship with X1 through this R-Squared value is used.

Of course, this value is restricted to being greater than or equal to zero and less than one. Use zero when there are no other covariates.

S (Standard Deviation of X1)

Enter an estimate of the standard deviation of X1, the predictor variable of interest. The formulation used here assumes that X1 follows the normal distribution. However, you can obtain approximate results for non-normal variables by putting in the correct value here. For example, if X1 is binary, the standard deviation is given by

$$\sqrt{p(1-p)}$$

where p is the proportion of either of the binary values in the population of X1.

If you don't have an estimate, you can press the SD button to obtain a window that will help you determine a rough estimate of the standard deviation.

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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Cox Regression** procedure window by clicking on **Regression**, and then clicking on **Cox Regression**. 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**.

Option	Value
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**

Annotated Output

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

Numeric Results

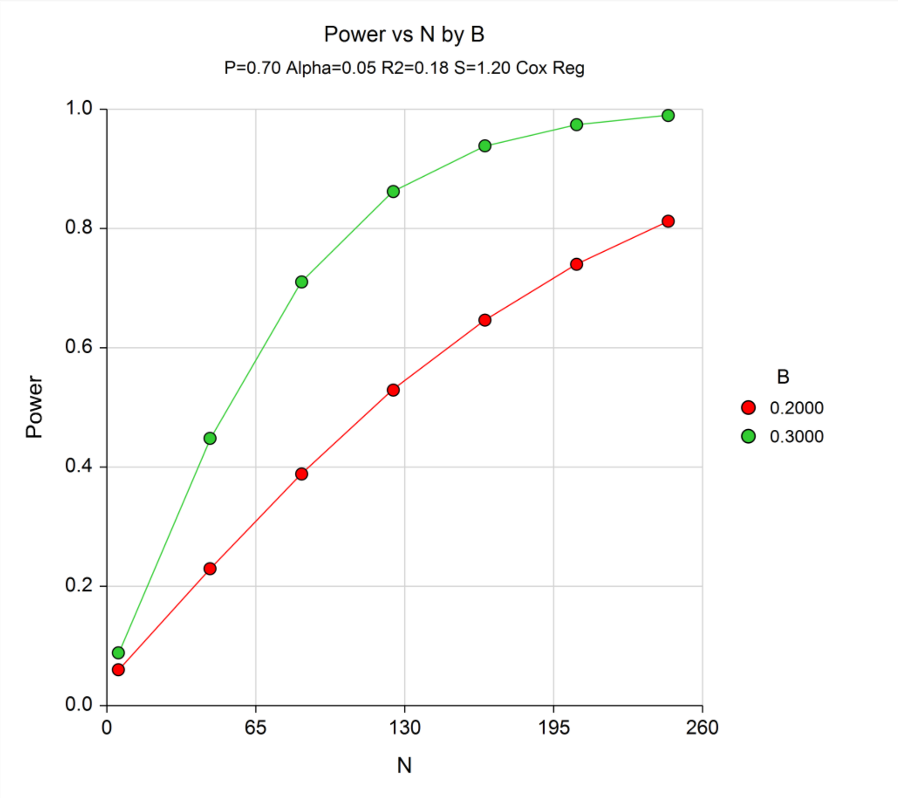
Power	Sample Size (N)	Reg. Coef. (B)	S.D. of X1 (SD)	Event Rate (P)	R-Squared X1 vs Other X's (R2)	Two-Sided Alpha	Beta
0.06017	5	0.2000	1.2000	0.7000	0.1800	0.05000	0.93983
0.22959	45	0.2000	1.2000	0.7000	0.1800	0.05000	0.77041
0.38837	85	0.2000	1.2000	0.7000	0.1800	0.05000	0.61163
0.52908	125	0.2000	1.2000	0.7000	0.1800	0.05000	0.47092
0.64643	165	0.2000	1.2000	0.7000	0.1800	0.05000	0.35357
0.74004	205	0.2000	1.2000	0.7000	0.1800	0.05000	0.25996
0.81223	245	0.2000	1.2000	0.7000	0.1800	0.05000	0.18777
0.08849	5	0.3000	1.2000	0.7000	0.1800	0.05000	0.91151
0.44815	45	0.3000	1.2000	0.7000	0.1800	0.05000	0.55185
0.71043	85	0.3000	1.2000	0.7000	0.1800	0.05000	0.28957
0.86202	125	0.3000	1.2000	0.7000	0.1800	0.05000	0.13798
0.93865	165	0.3000	1.2000	0.7000	0.1800	0.05000	0.06135
0.97412	205	0.3000	1.2000	0.7000	0.1800	0.05000	0.02588
0.98953	245	0.3000	1.2000	0.7000	0.1800	0.05000	0.01047

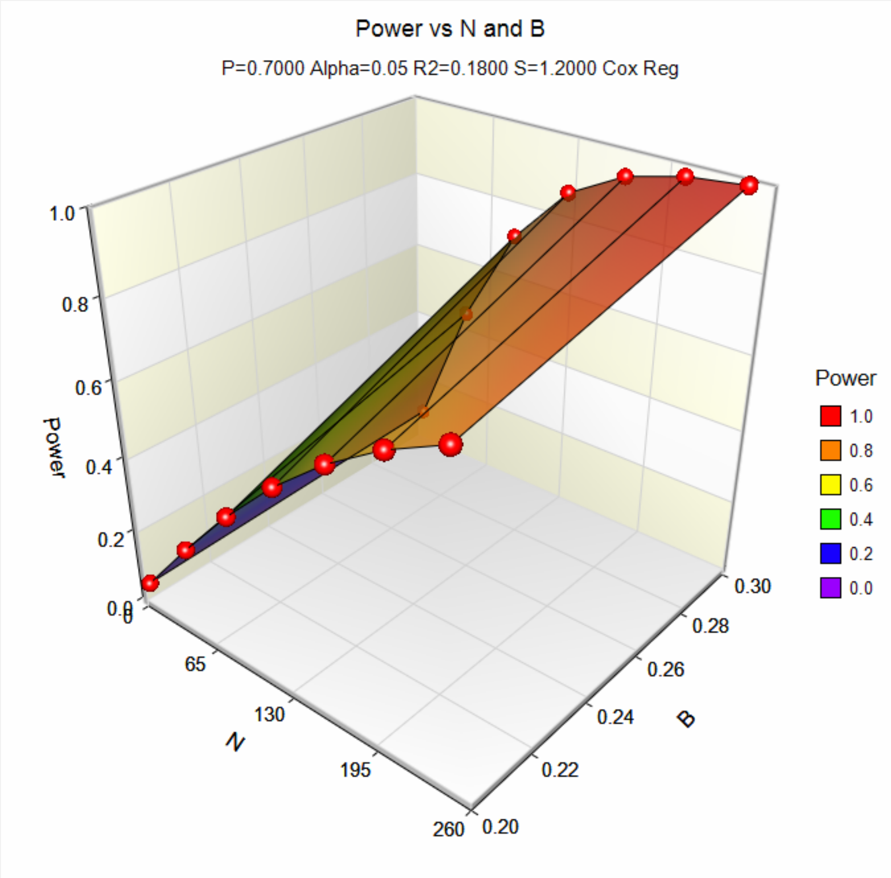
Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the size of the sample drawn from the population.
B is the size of the regression coefficent to be detected.
SD is the standard deviation of X1.
P is the event rate.
R2 is the R-squared achieved when X1 is regressed on the other covariates.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.

Summary Statements
A Cox regression of the log hazard ratio on a covariate with a standard deviation of 1.2000 based on a sample of 5 observations achieves 6% power at a 0.93983 significance level to detect a regression coefficient equal to 0.2000. The sample size was adjusted since a multiple regression of the variable of interest on the other covariates in the Cox regression is expected to have an R-Squared of 0.1800. The sample size was adjusted for an anticipated event rate of 0.7000.

This report shows the power for each of the scenarios.

Plots Section




Power vs N and B
P=0.7000 Alpha=0.05 R2=0.1800 S=1.2000 Cox Reg
Power
1.0
0.8
0.6
0.4
0.2
0.0
N
0
65
130
195
260
B
0.20
0.22
0.24
0.26
0.28
0.30
Power
1.0
0.8
0.6
0.4
0.2
0.0

Example 2 – Validation using Hsieh

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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Cox Regression** procedure window by clicking on **Regression**, and then clicking on **Cox Regression**. 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**.

Option	Value
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 Results

Power	Sample Size (N)	Reg. Coef. (B)	S.D. of X1 (SD)	Event Rate (P)	R-Squared X1 vs Other X's (R2)	One-Sided Alpha	Beta
0.80321	106	1.000	0.313	0.738	0.184	0.05000	0.19679

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 Results with R2 = 0 and P = 1.0

Power	Sample Size (N)	Reg. Coef. (B)	S.D. of X1 (SD)	Event Rate (P)	R-Squared X1 vs Other X's (R2)	One-Sided Alpha	Beta
0.80399	64	1.000	0.313	1.000	0.000	0.05000	0.19601

Note that PASS also calculated 64. Hsieh and Lavori obtained the 107 by adjusting this 64 for P first and then for $R2$. 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

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, $R2$ = 0.0, P = 0.71, one-sided alpha = 0.05, and power = 0.80. Schoenfeld calculated N = 212.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Cox Regression** procedure window by clicking on **Regression**, and then clicking on **Cox Regression**. 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**.

Option	Value
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

Power	Sample Size (N)	Reg. Coef. (B)	S.D. of X1 (SD)	Event Rate (P)	R-Squared X1 vs Other X's (R2)	One-Sided Alpha	Beta
0.80028	212	0.406	0.500	0.710	0.000	0.05000	0.19972

Note that PASS also obtains N = 212.