Survey, Statistics and Psychometrics Core Research Facility
University of Nebraska-Lincoln

Cox Proportional Hazard Model with a Continuous Covariate

Prepared by Olena Kaminska
Revised by Julia Soulakova

1. What research questions can be answered using the procedure?

When our variable of interest is survival time or hazard rate, one question which can be asked is if it depends on other variables. In this case hazard function of time to event is a response (dependent) variable and other variables are explanatory or controlled variables, also called risk factors. Cox Proportional Hazards Model estimates if each of the covariates has an effect on survival, the direction of the effect and its degree. With Cox Proportional Hazards Model one can model survival time using particular covariates or predict it for new observations based on a set of explanatory and controlled variables.

2. Data considerations

To run the procedure the following variables are used: time to event or survival time, event indicator (also called status, which indicates whether an event of interest happened or not) and one or more covariates.

The covariates can be categorical (ex. gender, race, type of treatment) and continuous (ex. age, height or blood pressure). The covariates also can be fixed, i.e. constant during all the study time (ex. gender, race, type of treatment or height) or time-dependent, i.e. changing during the study (ex. blood pressure, disease status).

This section will discuss the model with continuous fixed-time covariates.

The outcome variable can contain tied and censored times. Ties occur when time points are not distinct and therefore, multiple subjects have an event at the same time. The time is right censored if the event occurs after the study is finished, in such a case the actual time to event is unknown. A general assumption states that the event and censored time are independent.

3. Main ideas and statistics behind the procedure

The main formula was developed by Cox (1972) and is presented as

$h_i(t) = h_0(t)\exp(\beta_1 x_{i1} + \beta_1 x_{i2} + \ldots + \beta_k x_{ik})$, where $h_0(t)$ is a baseline hazard rate, x_{ik} is the value of the k^{th} covariate for the i^{th} participant and β_k is a coefficient of the k^{th} covariate and it indicates the effect of of the covariates on a hazard rate. If β_k is equal to 0, the k^{th} covariate does not have any effect on hazard rate. A positive value of β_k indicates that higher scores of the covariate are associated with higher mortality rates (or rates of events), and negative β_k indicates that higher scores of the corresponding covariate are associated with lower mortality rates (or rates of events).

The value of $\exp(\beta_k)$ provides the change in hazard rate with one unit increase in explanatory variable given that other variables in the model are fixed.

4. Example

4.1. Data Set

Let us consider an example of survival time of patients with cancer. The time variable indicates the time, measured in years, from the time the patient was diagnosed as having the 4^{th} stage of cancer till the death or till the end of the study. We are interested if survival time of patients with cancer depends on age. Age is provided in years. The status variable indicates if the patient died (1) or survived till the end of the study (0). Therefore, for patients with status 1 we have the exact information about their survival time.

Below, SAS syntax is provided for this example. With *data* statement we input the dataset with the information on 14 patients. The dataset has only 14 cases for presentation purposes. The sample size should be considerably higher to use the method.

```
data cancer;
input time age status @@;
datalines;
0.6   86    1     3.5   60    1
1.3   81    1     3.5   58    1
2.4   77    1     4     57    1
2.5   76    0     4     53    1
3.2   71    0     4.3   51    1
3.3   68    1     4.5   48    0
3.3   63    0     5.3   43    1
run;
```

4.2. SAS codes

To run Cox proportional hazard model we use *phreg* procedure in SAS.

```
proc phreg data=cancer ;
  model time*status(0) = age / ties=breslow;
run;
```

By *data= cancer* we indicate that the data Cancer are used in the analysis. In the model we indicate the response variable containing survival time information (*time*), censored information (*status*) with the value indicating censoring (0) and covariates. In our example we use a single covariate (*age*), although multiple covariates can be used.

Our dataset contains ties, i.e. some of the patients have the same survival time points. For example, we have two patients with survival time of 3.3 years, another two with survival time of 3.5 years and two more with survival time of 4 years. If data do not contain ties, the ties statement should be omitted. If data contain ties then *breslow* approach of treating ties is used by default. Other approaches, such as *discrete, efron* or *exact* can also be used. For more information about these approaches please refer to Chapter 49 in SAS User's Guide[1].

4.3. Explanation of SAS output

Below the SAS output is provided and main numbers are discussed.

```
                              The SAS System        11:23 Friday, March 10, 2007  49

                           The PHREG Procedure

                            Model Information

                    Data Set                 WORK.CANCER
                    Dependent Variable       time
                    Censoring Variable       status
                    Censoring Value(s)       0
                    Ties Handling            BRESLOW


                  Number of Observations Read          14
                  Number of Observations Used          14
```

The *model information* part of the output describes the model stated in syntax. Here we can see that *breslow* method was used for dealing with ties.

```
                  Summary of the Number of Event and Censored Values

                                                     Percent
                       Total       Event    Censored    Censored

                          14          10           4       28.57
```

The *Summary of the Number of Event and Censored Values* part indicates that out of total 14 observations, 4 (28.57%) are censored.

```
                            Convergence Status
```

[1] SAS Institute Inc., *SAS/STAT® User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999, Chapter 49, pp.2569-2657

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	36.191	10.915
AIC	36.191	12.915
SBC	36.191	13.217

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	25.2763	1	<.0001
Score	15.9061	1	<.0001
Wald	6.4707	1	0.0110

All three tests, Likelihood Ratio, Score and Wald tests, in *Testing Global Null Hypothesis: BETA=0* have p-value less than 0.05. This suggests that survival curves are different for different age levels.

The SAS System 11:23 Friday, March 10, 2007 50

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
age	1	0.54102	0.21269	6.4707	0.0110	1.718

Analysis of Maximum Likelihood Estimates part provides the information useful for model interpretation. Since the parameter estimate for age is positive (hazard ratio is above 1) we can conclude that mortality rate increases with age.

Hazard ratio is estimated as 1.718, it is the ratio of hazard of patients to hazard of one year younger patients. Thus, with one year increase in age at the time of diagnosis the hazard rate of death increases by 1.718.

4.4. Application

Below we present the calculations for comparison of hazard rates for patients of age 65 and 70. The estimated model for a 65 year old patient is given by

$h_{65}(t) = h_0(t)\exp(0.54102*65)$,

while for a 70-year old patient the model is

$h_{70}(t) = h_0(t)\exp(0.54102*70)$.

To obtain the relative risk, we find the ratio of the hazard rate of 70 year old patient versus hazard rate of 65 year old patient:

$$RR = h_{70}(t) / h_{65}(t) = \frac{h_0(t)\exp(0.54102*70)}{h_0(t)\exp(0.54102*65)} = 14.96$$

This indicates that patients, who are 70 years old have death rate 14.96 times as high as patients who are 65 years old at the time of a diagnosis.