Klein baum Survival Analysis
Self learning Text

3 The Cox Proportional Hazards Model and Its Characteristics

Contents

Introduction

We begin by discussing some computer results using the Cox PH model, without actually specifying the model; the purpose here is to show the similarity between the Cox model and standard linear regression or logistic regression.

We then introduce the Cox model and describe why it is so popular. In addition, we describe its basic properties, including the meaning of the proportional hazards assumption.

Abbreviated Outline

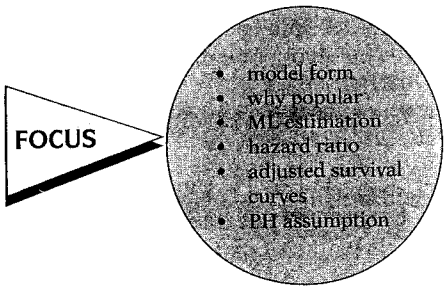
The outline below gives the user a preview of the material to be covered by the presentation. A detailed outline for review purposes follows the presentation.

Objectives

Upon completing the module, the learner should be able to:

1. State or recognize the general form of the Cox PH model.
2. State the specific form of a Cox PH model appropriate for the analysis, given a survival analysis scenario involving one or more explanatory variables.
3. State or recognize the form and properties of the baseline hazard function in the Cox PH model.
4. Give three reasons for the popularity of the Cox PH model.
5. State the formula for a designated hazard ratio of interest given a scenario describing a survival analysis using a Cox PH model, when
   a. there are confounders but no interaction terms in the model;
   b. there are both confounders and interaction terms in the model.
6. State or recognize the meaning of the PH assumption.
7. Determine and explain whether the PH assumption is satisfied when the graphs of the hazard functions for two groups cross each other over time.
8. State or recognize what is an adjusted survival curve.
9. Compare and/or interpret two or more adjusted survival curves.
10. Given a computer printout involving one or more fitted Cox PH models,
    a. compute or identify any hazard ratio(s) of interest;
    b. carry out and interpret a designated test of hypothesis;
    c. carry out, identify or interpret a confidence interval for a designated hazard ratio;
    d. evaluate interaction and confounding involving one or more covariates.

Presentation



This presentation describes the **Cox proportional hazards (PH) model,** a popular mathematical model used for analyzing survival data. Here, we focus on the model form, why the model is popular, maximum likelihood (ML) estimation of the model parameters, the formula for the hazard ratio, how to obtain adjusted survival curves, and the meaning of the PH assumption.

I. A Computer Example Using the Cox PH Model

EXAMPLE

Leukemia Remission Data

Group 1 ($n = 21$)		Group 2 ($n = 21$)	
(weeks)	log WBC	(weeks)	log WBC
6	2.31	1	2.80
6	4.06	1	5.00
6	3.28	2	4.91
7	4.43	2	4.48
10	2.96	3	4.01
10	2.88	4	4.36
13	3.60	4	2.42
16	2.32	5	3.49
17	2.57	5	3.97
19	3.20	8	3.52
20	2.80	8	3.05
22	2.70	8	2.32
23	2.60	8	3.26
25	2.16	11	3.49
32	2.05	11	2.12
32	2.01	12	1.50
34	1.78	12	3.06
35	2.20	15	2.70
37	2.53	17	2.95
41	1.47	22	2.73
43	1.45	23	1.97

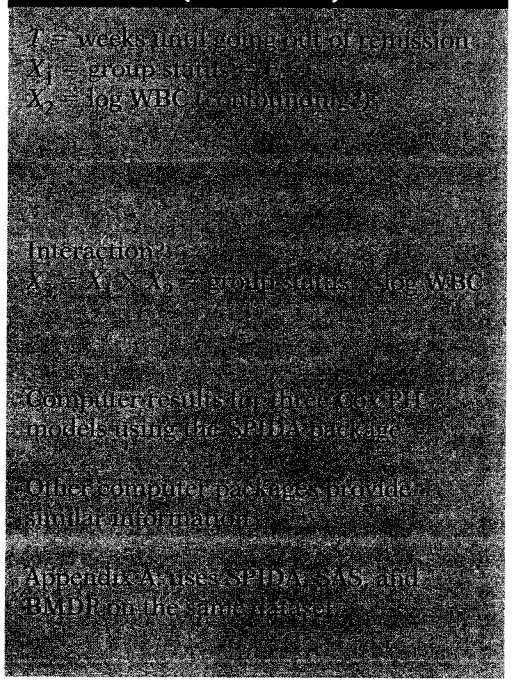
+ denotes censored observation

We introduce the Cox PH model using computer output from the analysis of remission time data (Freireich et al., *Blood,* 1963), which we previously discussed in Chapters 1 and 2. The data set is listed here at the left.

These data involve two groups of leukemia patients, with 21 patients in each group. Group 1 is the treatment group, and group 2 is the placebo group. The data set also contains the variable log WBC, which is a well-known prognostic indicator of survival for leukemia patients.

For this example, the basic question of interest concerns comparing the survival experience of the two groups adjusting for the possible confounding and/or interaction effects of log WBC.

EXAMPLE (continued)



We are thus considering a problem involving two explanatory variables as predictors of survival time T, where T denotes "weeks until going out of remission." We label the explanatory variables X_1 (for group status) and X_2 (for log WBC). The variable X_1 is the primary study or exposure variable of interest. The variable X_2 is an extraneous variable that we are including as a possible confounder or effect modifier.

Note that if we want to evaluate the possible interaction effect of log WBC on group status, we would also need to consider a third variable, that is, the product of X_1 and X_2.

For this dataset, the computer results from fitting three different Cox proportional hazards models are presented below. The computer package used is **SPIDA.** This is one of several packages that have procedures for carrying out a survival analysis using the Cox model. The information printed out by different packages will not have exactly the same format, but they will provide similar information. A comparison of output using SPIDA, SAS, and BMDP procedures on the same dataset is provided in Appendix A at the back of this text.

Output from Spida

Model 1:

Column name	Coeff	StErr	p-value	*HR*	0.95	CI	*P(PH)*
Rx	1.509	0.410	0	4.523	2.027	10.094	0.794
n:42	%Cen: 28.571		–2 log *L*: 172.759				

Model 2:

Column name	Coeff	StErr	p-value	*HR*	0.95	CI	*P(PH)*
Rx	1.294	0.422	0.002	3.648	1.505	8.343	0.944
log WBC	1.604	0.329	0.000	4.975	2.609	9.486	0.917
n:42	%Cen: 28.571		–2 log *L*: 144.559				

Model 3:

Column name	Coeff	StErr	p-value	*HR*	0.95	CI	*P(PH)*
Rx	2.355	1.681	0.161	10.537	0.391	284.200	0.628
log WBC	1.803	0.447	0.000	6.067	2.528	14.561	0.996
Rx × log WBC	–0.342	0.520	0.510	0.710	0.256	1.967	0.410
n:42	%Cen: 28.571		–2 log *L*: 144.131				

OUTPUT FROM SPIDA

Model 1:

Column

Name	Coeff	StErr	p-value	*HR*
Rx	1.509	0.410	0	4.523

%Cen 28.571 −2 log *L*: 172.759

Model 2: **Hazard ratios**

Column

Name	Coeff	StErr	p-value	*HR*
Rx	1.294	0.422	0.002	3.648
log WBC	1.604	0.329	0.000	4.975

%Cen 28.571 −2 log *L*: 144.559

Model 3:

Column

Name	Coeff	StErr	p-value	*HR*
Rx	2.355	1.681	0.161	10.537
log WBC	1.803	0.447	0.000	6.067
Rx × log WBC	−0.142	0.520	0.510	0.710

%Cen 28.571 −2 log *L*: 144.131

We now describe how to use the computer printout to evaluate the possible effect of treatment status on remission time adjusted for the potential confounding and interaction effects of the covariate log WBC. For now, we focus only on the first five columns of information provided in the printout, as presented at the left for all three models.

For each model, the first column identifies the **variables** that have been included in the model. The second column gives **regression coefficients** corresponding to each variable in the model. The third column gives **standard errors** of the regression coefficients. The fourth column gives **p-values** for testing the significance of each coefficient. The fifth column, labeled as ***HR***, gives **hazard ratios** for the effect of each variable adjusted for the other variables in the model.

Except for the *HR* column, these computer results are typical of output found in standard linear regression printouts. As the printout suggests, we can analyze the results from a Cox model in a manner similar to the way we would analyze a linear regression model.

EXAMPLE (continued)

Same dataset for each model:

42 subjects

Time (weeks) until out of remission

Model 1: *Rx* only

Model 2: *Rx* and log WBC

Model 3: *Rx*, log WBC, and *Rx* × log WBC

We now distinguish among the output for the three models shown here. All three models are using the same set of remission time data on 42 subjects. The outcome variable for each model is the same—time in weeks until a subject goes out of remission. However, the independent variables are different for each model. Model 1 contains only the treatment status variable, indicating whether a subject is in the treatment or placebo group. Model 2 contains two variables—treatment status and log WBC. And model 3 contains an interaction term defined as the product of group status and log WBC.

OUTPUT: ML ESTIMATION

Model 3:

Column name	Coeff	StErr	p-value	*HR*
Rx	2.355	1.681	0.161	10.537
log WBC	1.803	0.447	0.000	6.067
Rx× log WBC	−0.342	0.520	(0.510)	0.710

n: 42 %Cen: 28.571 −2 log *L*: 144.131

EXAMPLE (continued)

$P = 0.510$: $\frac{-0.342}{0.520} = -0.66 \approx Z$ Wald statistic

LR statistic: uses $-2 \log L = 144.131$ (i.e., log likelihood value)

We now focus on the output for model 3. The method of estimation used to obtain the coefficients for this model, as well as the other two models, is maximum likelihood (ML) estimation. Note that a p-value of 0.510 is obtained for the coefficient of the product term for the interaction of treatment with log WBC. This p-value indicates that there is no significant interaction effect, so that we can drop the product term from the model and consider the other two models instead.

The p-value of 0.510 that we have just described is obtained by dividing the coefficient −0.342 of the product term by its standard error of 0.520, which gives −0.66, and then assuming that this quantity is approximately a standard normal or *Z* variable. This *Z* statistic is known as a **Wald statistic,** which is one of two test statistics typically used with ML estimates. The other test statistic, called the **likelihood ratio,** or *LR* statistic, makes use of the log likelihood value. This is given by $-2 \log L$ in the output, which has the value 144.131 for model 3.

OUTPUT

Model 2:

Column name	Coeff	StErr	p-value	*HR*
Rx	1.294	0.422	0.002	3.648
log WBC	1.604	0.329	0.000	4.975

n: 42 %Cen: 28.571 −2 log *L*: 144.559

We now look at the printout for model 2, which contains two variables. The treatment status variable (*Rx*) represents the exposure variable of primary interest. The log WBC variable is being considered as a confounder. Our goal is to describe the effect of treatment status adjusted for log WBC.

Notice first that the log likelihood value for model 2 is given by $-2 \log L = 144.559$. We can use this value together with the $-2 \log L$ value from model 3 to obtain the *LR* statistic for testing the significance of the interaction term in model 3.

EXAMPLE (continued)

LR (interaction in model 3)

$= -2 \log L_{\text{model 2}} - (-2 \log L_{\text{model 3}})$

$= 144.559 - 144.131 = 0.428$

(*LR* is χ^2 with 1 d.f. under H_0: no interaction.)

$0.40 < P < 0.50$, **not significant**

Wald test $P = 0.510$

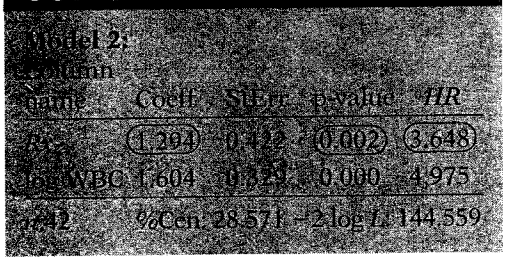
We compute 144.559 minus 144.131 to obtain 0.428. This test statistic has a chi-square distribution under the null hypothesis of no interaction effect. The p-value for this test is between 0.40 and 0.50, which indicates no significant interaction. Although the p-values for the Wald test (0.510) and the *LR* test are not exactly the same, both p-values lead to the same conclusion.

LR≠Wald

When in doubt, use the LR test.

In general, the LR and Wald statistics may not give exactly the same answer. Statisticians have shown that of the two test procedures, the LR statistic has better statistical properties, so when in doubt, you should use the LR test.

OUTPUT



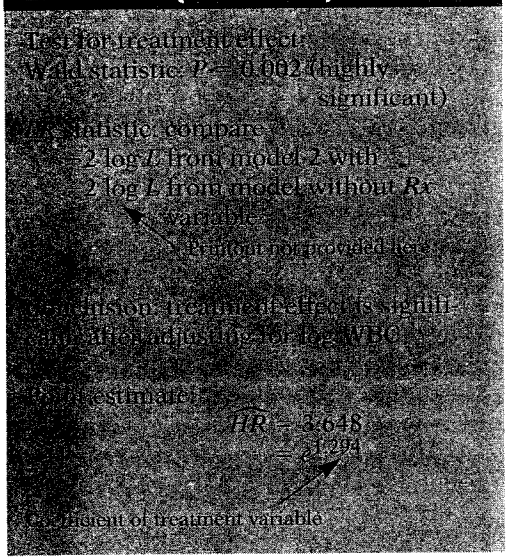
We now focus on how to assess the effect of treatment status adjusting for log WBC using the model 2 output, again shown here.

Three statistical objectives:

1. **test for significance of effect**
2. **point estimate of effect**
3. **confidence interval for effect**

There are three statistical objectives typically considered. One is to **test for the significance** of the treatment status variable, adjusted for log WBC. Another is to obtain a **point estimate of the effect** of treatment status, adjusted for log WBC. And a third is to obtain a **confidence interval for this effect.** We can accomplish these three objectives using the output provided, without having to explicitly describe the formula for the Cox model being used.

EXAMPLE (continued)



To test for the significance of the treatment effect, the p-value provided in the table for the Wald statistic is 0.002, which is highly significant. Alternatively, a likelihood ratio (LR) test could be performed comparing the log likelihood statistic (144.559) for model 2, with the log likelihood for a model which does not contain the treatment variable. This latter model, which should contain only the log WBC variable, is not provided here, so we will not report on it other than to note that the LR test is also very significant. Thus, these test results show that using model 2, the treatment effect is significant, after adjusting for log WBC.

A point estimate of the effect of the treatment is provided in the *HR* column by the value 3.648. This value gives the estimated hazard ratio (HR) for the effect of the treatment; in particular, we see that the hazard for the placebo group is 3.6 times the hazard for the treatment group. Note that the value 3.648 is calculated as e to the coefficient of the treatment variable; that is, e to the 1.294 equals 3.648.

To describe the confidence interval for the effect of treatment status, we consider the output for the extended table for model 2 given earlier.

OUTPUT

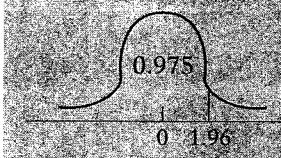
Model 2:

Column name	Coeff	StErr	p-value	*HR*	0.95	CI	*P(PH)*
Rx	1.294	0.422	0.002	3.648	1.505	8.343	0.944
log WBC	1.604	0.329	0.000	4.975	2.609	9.486	0.917
n:42	%Cen: 28.571		-2 log *L*: 144.559				

EXAMPLE (continued)

95% confidence interval for the *HR*: (1.505, 8.343)

95% CI for β_1: $1.294 \pm (1.96)(0.422)$



95% CI for $HR = e^{\beta_1}$:

$$e^{\hat{\beta}_1 \pm 1.96 s_{\hat{\beta}_1}} = e^{1.294 \pm 1.96(0.422)}$$

From the table, we see that a 95% confidence interval for the treatment effect is given by the range of values 1.505–8.343. This is a confidence interval for the hazard ratio (HR), which surrounds the point estimate of 3.648 previously described. Notice that this confidence interval is fairly wide, indicating that the point estimate is somewhat unreliable. As expected from the low p-value of 0.002, the confidence interval for HR does not contain the null value of 1.

The calculation of the confidence interval for HR is carried out as follows:

1. Compute a 95% confidence interval for the regression coefficient of the *Rx* variable ($\hat{\beta}_1$). The large sample formula is 1.294 plus or minus 1.96 times the standard error 0.422, where 1.96 is the 97.5 percentile of the standard normal or *Z* distribution.
2. Exponentiate the two limits obtained for the confidence interval for the regression coefficient of *Rx*.

SPIDA: provides CI directly

Other packages: provide $\hat{\beta}$'s and $s_{\hat{\beta}}$'s

The SPIDA program output provides the required confidence interval directly, so that the user does not have to carry out the computations required by the large sample formula. Other computer packages may not provide the confidence interval directly, but, rather, may provide only the regression coefficients and their standard errors.

OUTPUT

Model 1:

Column name	Coeff	StErr	p-value	*HR*
Rx	1.509	0.410	0	4.523

(crude model)

n 42 %Cen 28.571 −2 log *L*: 172.759

Model 2:

Column name	Coeff	StErr	p-value	*HR*
Rx	1.294	0.422	0.002	3.648
log WBC	1.604	0.329	0.000	4.975

n 42 %Cen 28.571 −2 log *L*: 144.559

EXAMPLE (continued)

HR for model 1 (4.523) is higher than *HR* for model 2 (3.648).

Confounding: crude versus adjusted *HR*'s are meaningfully different.

Confounding due to log WBC.

Must control for log WBC, i.e., prefer model 2 to model 1.

If no confounding, then consider precision; e.g., if 95% CI is narrower for model 2 than model 1, we prefer model 2.

OUTPUT: Confidence Intervals

Column name	0.95	CI
Rx model 1	2.027	10.094
	width = **8.067**	
	width = **6.838**	
Rx model 2	1.505	8.343
log WBC	2.609	9.486

To this point, we have made use of information from outputs for models 2 and 3, but have not yet considered the model 1 output, which is shown again here. Note that model 1 contains only the treatment status variable, whereas model 2, shown below, contains log WBC in addition to treatment status. Model 1 is sometimes called the "crude" model because it ignores the effect of potential covariates of interest, like log WBC.

Model 1 can be used in comparison with model 2 to evaluate the potential confounding effect of the variable log WBC. In particular, notice that the value in the *HR* column for the treatment status variable is 4.523 for model 1, but only 3.648 for model 2. Thus, the crude model yields an estimated hazard ratio that is somewhat higher than the corresponding estimate obtained when we adjust for log WBC. If we decide that the crude and adjusted estimates are meaningfully different, we then say that there is confounding due to log WBC.

Once we decide that confounding is present, we then **must** control for the confounder—in this case, log WBC—in order to obtain a valid estimate of the effect. Thus, we prefer model 2, which controls for log WBC, to model 1, which does not.

Note that if we had decided that there is no "meaningful" confounding, then we would not need to control for log WBC to get a valid answer. Nevertheless, we might wish to control for log WBC anyhow, to obtain a more precise estimate of the hazard ratio. That is, if the confidence interval for the HR is narrower when using model 2 than when using model 1, we would prefer model 2 to model 1 for **precision** gain.

The confidence intervals for *Rx* in each model are shown here at the left. The interval for *Rx* in model 1 has width equal to 10.094 minus 2.027, or **8.067**; for model 2, the width is 8.343 minus 1.505, or **6.838**. Therefore, model 2 gives a more precise estimate of the hazard ratio than does model 1.