

## Chapter 18

# The Cox Proportional Hazards Model

We just encountered the Kaplan-Meier estimate of the survival function in Chapter 11. Now we are going to talk about models that essentially treat the survival function as the *outcome* in a supervised learning problem. These models are called **Cox proportional hazards models**.

### 18.1 Survival and Hazard Functions

Consider a situation where we have some process that generates events, and we're trying to model the time to first event. Assume the probability of the event's occurring at each time,  $t$ , is given by the function  $f(t)$ . The cumulative probability of the event's having occurred by time  $t$  is

$$F(t) = \int_0^t f(t)dt$$

and the probability of an individual not having experienced the event by time  $t$  is

$$S(t) = 1 - F(t),$$

the **survival function**. The probability of experiencing the event in an infinitesimally small interval starting at  $t$ , given that one has not experienced it

by time  $t$ , is:

$$\lambda(t) = \frac{f(t)}{S(t)}$$

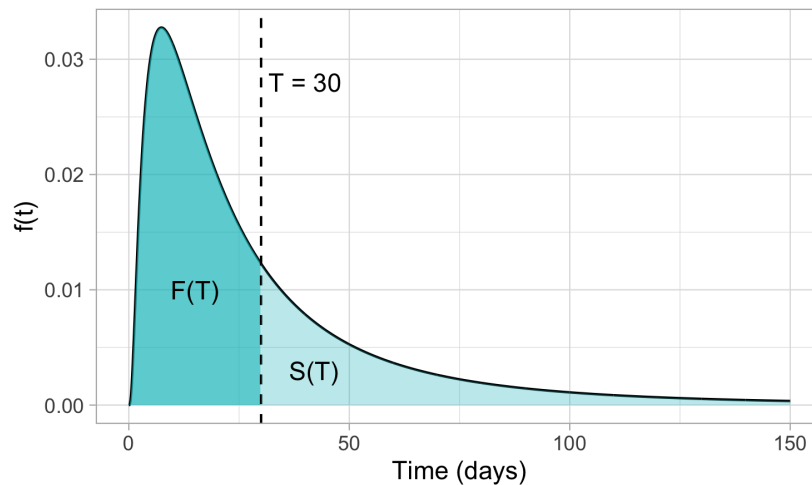
and is called the **hazard**. The **cumulative hazard** function is equal to

$$\Lambda(t) = \int_0^t \lambda(t') dt' = -\log S(t)$$

so  $S(t) = \exp(-\Lambda(t))$ .

None of these expressions should be immediately obvious to you, because deriving them requires calculus. It's a great exercise to go through the derivations, but for now, let's focus on capturing the intuition.

Here is a graphical representation of some of these quantities. Remember that the probability distribution  $f(t)$  must integrate to one.



#### Question 18.1

What's the interpretation of the hazard,  $\lambda(t)$ , on this image? What happens to the hazard if  $S(t)$  is low vs. high for the same  $f(x)$ ?

### Question 18.2

Assume  $f(t)$  is exponential:  $f(t) = b \cdot \exp(-bt)$ , where  $b$  is constant. Then  $F(t) = 1 - \exp(-bt)$  and  $S(t) = \exp(-bt)$ . What is the hazard,  $\lambda(t)$ ? What is the cumulative hazard,  $\Lambda(t)$ ?

### Question 18.3

The concept of a “cumulative hazard” is pretty weird. How should this quantity be interpreted?

## 18.2 Estimating Survival and Cumulative Hazard

The Kaplan-Meier estimate of survival (Chapter 11) is the most common estimate of the survival function. One can estimate the cumulative hazard using a couple of different methods.

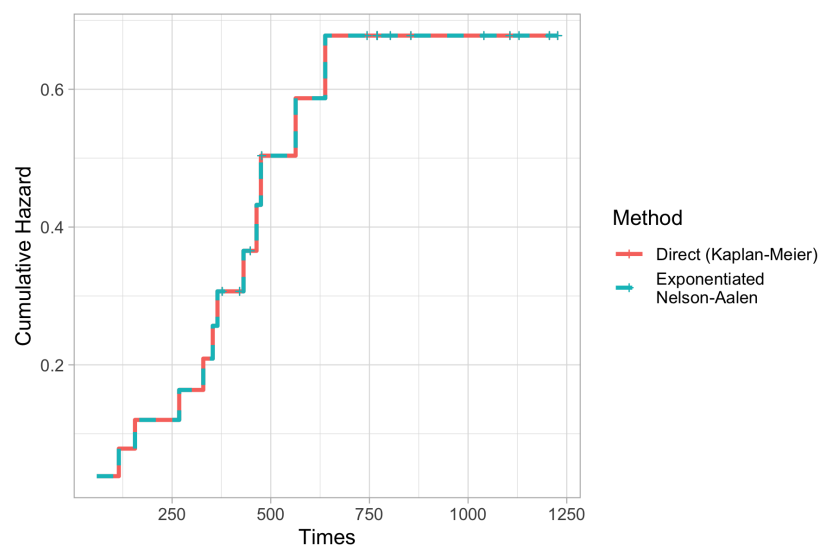
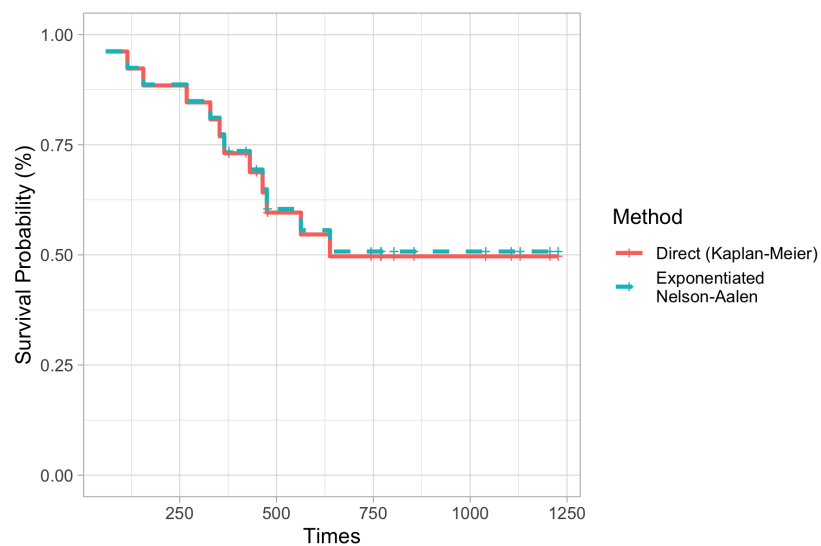
1. Take the negative log of the Kaplan-Meier estimate of survival:

$$\begin{aligned}\hat{\Lambda}_{KM}(t) &= -\log \hat{S}_{KM}(t) \\ &= -\sum_{i:t_i < t} \log \left(1 - \frac{d_i}{n_i}\right)\end{aligned}$$

2. Use the **Nelson-Aalen estimator**

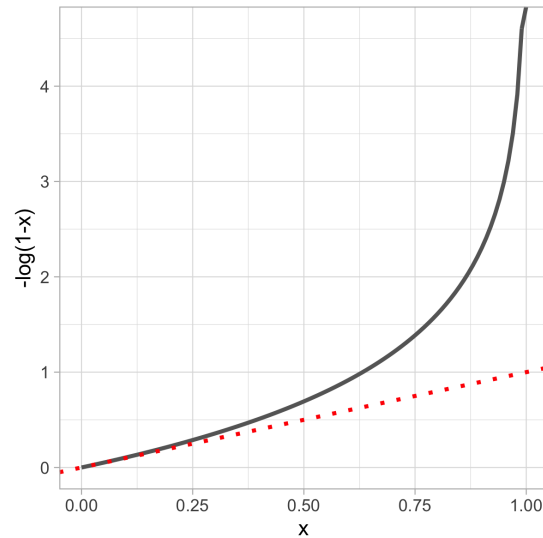
$$\hat{\Lambda}_{NA}(t) = \sum_{i:t_i < t} \frac{d_i}{n_i}$$

One thing that is confusing about the R `survival` package’s output is that it uses the Kaplan-Meier estimate for survival by default, but it uses the Nelson-Aalen estimator for cumulative hazard by default. So if you exponentiate the negative cumulative hazard that comes out of `survfit`, it won’t match the survival estimate. The two are close in most cases, though. Here are some pictures for the ovarian cancer survival dataset we discussed in Chapter 11:



#### Question 18.4

Here is a plot showing the function  $-\log(1-x)$  vs.  $x$ . Look at the expressions for the two estimators for the cumulative hazard, above. Under what conditions will they be similar? Under what conditions will they be different?



#### Question 18.5

The curvature of the Nelson-Aalen estimator gives you an idea of how the hazard varies with time. A concave shape is an indication of *deceleration* of the hazard; for example, if the event in question is death and time is patient age, this would represent higher infant/childhood mortality than adult mortality. A convex shape is an indication of *acceleration* of the hazard; in the death/age example, this would represent a process that accelerates as one ages (so called “wear-out mortality”).

Looking at the graph of the overall cumulative hazard, what do you notice about how the hazard for ovarian cancer changes since the initiation of treatment?

### Question 18.6

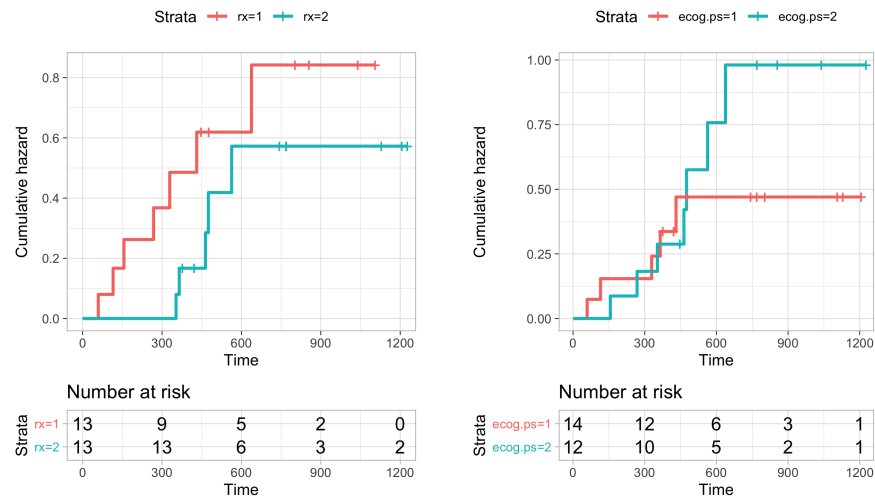
Here are the raw data from treatment group 1 of the `ovarian` dataset. These are the same data we looked at when building the Kaplan-Meier curve in Chapter 11, Question 11.1. Using these data, fill in the remaining cells of the table below. Here we are using the Nelson-Aalen estimator for the cumulative hazard.

	rx	futime	fustat
1	1	59	1
2	1	115	1
3	1	156	1
4	1	268	1
5	1	329	1
6	1	431	1
7	1	448	0
8	1	477	0
9	1	638	1
10	1	803	0
11	1	855	0
12	1	1040	0
13	1	1106	0

$j$	$t_j$	$n_j$	$d_j$	$\hat{\Lambda}(t_j)$	Calculation
0	0	13	0	0.000	$\frac{0}{13}$
1	59	13	1	0.077	$\hat{\Lambda}(t_0) + \frac{1}{13}$
2	115	12	1	0.160	$\hat{\Lambda}(t_1) + \frac{1}{12}$
3	156				
4	268				
5	329	9	1	0.462	$\hat{\Lambda}(t_4) + \frac{1}{9}$
6	431	8	1	0.587	$\hat{\Lambda}(t_5) + \frac{1}{8}$
7	448	7	0	0.587	$\hat{\Lambda}(t_6) + \frac{0}{7}$
8	477	6	0	0.587	$\hat{\Lambda}(t_7) + \frac{0}{6}$
9	638	5	1	0.787	$\hat{\Lambda}(t_8) + \frac{1}{5}$
10	803	4	0		
11	855	3	0		
12	1040	2	0		
13	1106	1	0		

### Question 18.7

Here are plots of the cumulative hazard (Nelson-Aalen estimator) for patients by sex and by ECOG performance score status:



How does treatment group appear to impact the cumulative hazard? What about ECOG score? Do the cumulative hazards appear proportional (i.e., related by a common multiplier) in each case?

## 18.3 Deriving the Cox Model

We have spent considerable time on the hazard and cumulative hazard because these play important roles in what is perhaps the most famous survival analysis tool: the Cox proportional hazards model (“Cox model”, for short), developed by D.R. Cox in 1972. The model has the following form:

$$\lambda(t|x) = \lambda_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$$

and another way of writing it is:

$$\log \left( \frac{\lambda(t|x)}{\lambda_0(t)} \right) = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p.$$

Note that for now we are assuming that the covariates do not depend on time. There is a variant of the Cox model called the **extended Cox model** that allows time-dependent covariates. For now, we will just consider the fixed covariate case.

#### Question 18.8

Compare the Cox model to a logistic regression model. What is the same? What is different?

Here  $\lambda_0(t)$  is called the **baseline hazard**. As usual, we symbolize the linear sum of the  $\beta$ s as  $\beta^T x$ . Importantly, the baseline hazard can take any shape as a function of time. The  $\lambda_0(t)$  part of the equation is, therefore, referred to as the “nonparametric” part, while the  $\beta^T x$  part is called the “parametric” part. The overall model is referred to as **semiparametric**.

The ratio of the hazards for two different sets of covariates,  $x$  and  $z$ , is

$$\frac{\lambda(t|x)}{\lambda(t|z)} = \frac{\lambda_0(t) \exp(\beta^T x)}{\lambda_0(t) \exp(\beta^T z)} = \exp(\beta^T (x - z)).$$

Taking the log of this, we arrive at

$$\log \left( \frac{\lambda(t|x)}{\lambda(t|z)} \right) = \beta^T (x - z).$$

A single coefficient,  $\beta_j$ , is therefore the **hazard ratio** when the corresponding predictor,  $x_j$ , increases by one. This ratio is assumed to be constant over time. The hazard ratio is also called the **relative risk**.

#### Question 18.9

Why doesn't the Cox model have an intercept,  $\beta_0$ ?

#### Question 18.10

Compare the interpretation of the coefficients in a Cox model to their interpretation in a logistic regression model.



## 18.4 Fitting the Cox Model

Fitting a Cox model means taking a sample of possibly right-censored data and deriving estimates for the parameters,  $\beta_1, \dots, \beta_p$ . Cox models are fit using a variant of maximum likelihood estimation called **partial likelihood estimation**.

Let's consider all of the unique times,  $t_i$ , that events are observed. For now, we will assume that exactly one event is observed at each of these times (no ties). We will use  $R(t_i)$  to refer to the set of subjects who are “at risk” (i.e., not censored) just prior to time  $t_i$ . At each failure time,  $t_i$ , the contribution to the partial likelihood is:

$$\begin{aligned}\mathcal{L}_i(\beta) &= \frac{P(\text{person } i \text{ experiences event} \mid \text{still around at } t_i)}{\sum_{l \in R(t_i)} P(\text{person } l \text{ experiences event} \mid \text{still around at } t_i)} \\ &= \frac{\lambda(t_i | x_i)}{\sum_{l \in R(t_i)} \lambda(t_i | x_l)} = \frac{\exp(\beta^T x^{(i)})}{\sum_{l \in R(t_i)} \exp(\beta^T x^{(l)})}\end{aligned}$$

The complete partial likelihood over all  $K$  observed event times,  $t_i = t_1, \dots, t_K$ , is:

$$\mathcal{L}(\beta) = \prod_{i=1}^K \frac{\exp(\beta^T x^{(i)})}{\sum_{l \in R(t_i)} \exp(\beta^T x^{(l)})}.$$

This is the thing that the model fitting process is trying to maximize. It is optimized numerically, using the same types of optimization procedures used by logistic regression and other generalized linear models.

### Question 18.11

Why is this quantity called a “partial likelihood”, instead of just a “likelihood”?

## 18.5 Interpreting Cox Models

Here is a simple Cox model for survival time vs. age, residual disease, ECOG score, and treatment group in the ovarian cancer dataset.

```

```{r}
m <- coxph(Surv(futime, fustat) ~ age + resid.ds + ecog.ps + rx, data = d)
summary(m)
```

```

Call:  
coxph(formula = Surv(futime, fustat) ~ age + resid.ds + ecog.ps +  
rx, data = d)

n= 26, number of events= 12

|           | coef     | exp(coef) | se(coef) | z      | Pr(> z )   |
|-----------|----------|-----------|----------|--------|------------|
| age       | 0.12481  | 1.13294   | 0.04689  | 2.662  | 0.00777 ** |
| resid.ds2 | 0.82619  | 2.28459   | 0.78961  | 1.046  | 0.29541    |
| ecog.ps2  | 0.33621  | 1.39964   | 0.64392  | 0.522  | 0.60158    |
| rx2       | -0.91450 | 0.40072   | 0.65332  | -1.400 | 0.16158    |

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

|           | exp(coef) | exp(-coef) | lower .95 | upper .95 |
|-----------|-----------|------------|-----------|-----------|
| age       | 1.1329    | 0.8827     | 1.0335    | 1.242     |
| resid.ds2 | 2.2846    | 0.4377     | 0.4861    | 10.738    |
| ecog.ps2  | 1.3996    | 0.7145     | 0.3962    | 4.945     |
| rx2       | 0.4007    | 2.4955     | 0.1114    | 1.442     |

Concordance= 0.807 (se = 0.068 )  
Likelihood ratio test= 17.04 on 4 df, p=0.002  
Wald test = 14.25 on 4 df, p=0.007  
Score (logrank) test = 20.81 on 4 df, p=3e-04

### Question 18.12

Interpret the coefficients, exponentiated coefficients, standard errors of the coefficients, Z scores, and *p*-values in this model. Also try your hand at interpreting the second block of model output, which includes the exponentiated coefficients (again), the exponentiated negative coefficients, and the lower and upper bounds of a 95% confidence interval for the exponentiated coefficients.

## 18.6 Making Predictions with Cox Models

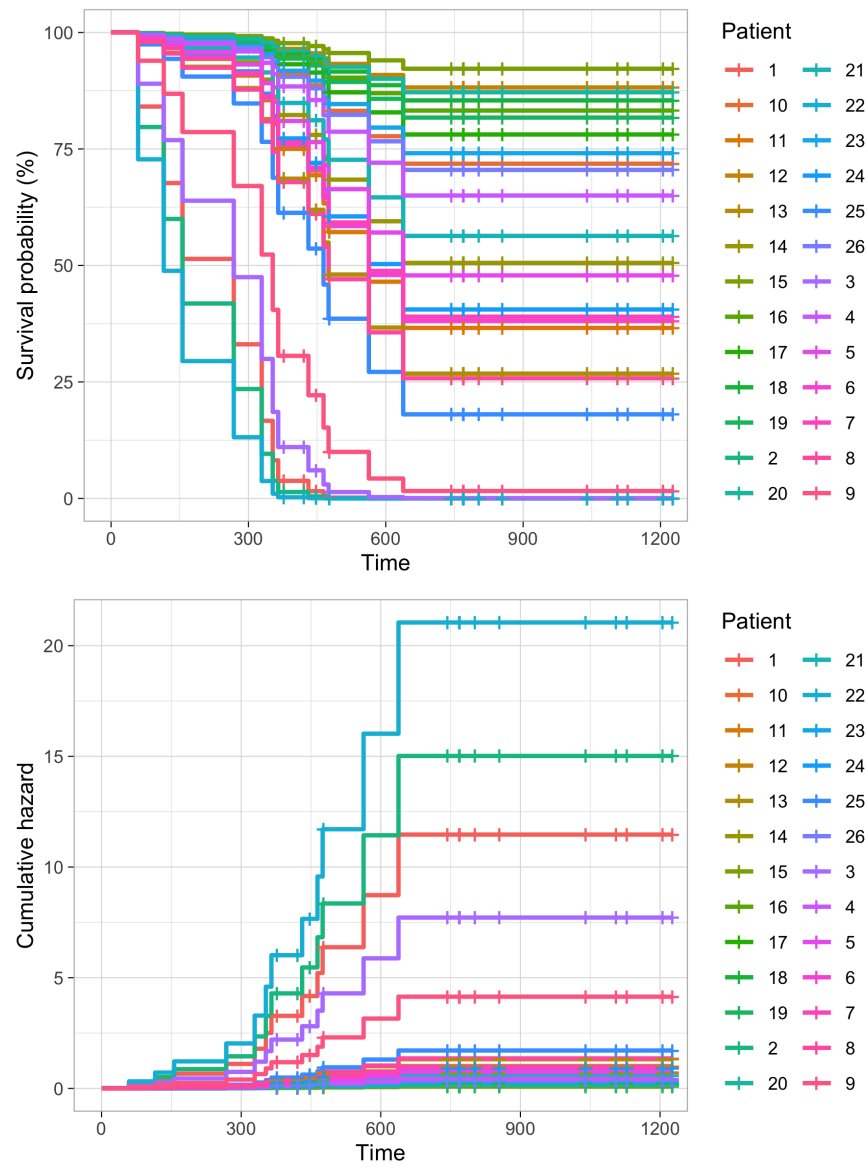
The Cox model fitted using the `coxph` function can be used to make various predictions on the original dataset. This yields a lot of output. All of the available outputs for the `ovarian` dataset are shown below.

|    | age   | lp_age | resid.ds | lp_resid.ds | ecog.ps | lp_ecog.ps | rx | lp_rx | lp    | risk  | expected | surv |
|----|-------|--------|----------|-------------|---------|------------|----|-------|-------|-------|----------|------|
| 1  | 72.33 | 9.03   | 2        | 0.83        | 1       | 0.00       | 1  | 0.00  | 2.67  | 14.43 | 0.16     | 0.85 |
| 2  | 74.49 | 9.30   | 2        | 0.83        | 1       | 0.00       | 1  | 0.00  | 2.94  | 18.90 | 0.46     | 0.63 |
| 3  | 66.47 | 8.30   | 2        | 0.83        | 2       | 0.34       | 1  | 0.00  | 2.27  | 9.71  | 0.40     | 0.67 |
| 4  | 53.36 | 6.66   | 2        | 0.83        | 1       | 0.00       | 2  | -0.91 | -0.61 | 0.54  | 0.11     | 0.89 |
| 5  | 50.34 | 6.28   | 2        | 0.83        | 1       | 0.00       | 1  | 0.00  | -0.08 | 0.93  | 0.25     | 0.78 |
| 6  | 56.43 | 7.04   | 1        | 0.00        | 2       | 0.34       | 1  | 0.00  | 0.19  | 1.21  | 0.33     | 0.72 |
| 7  | 56.94 | 7.11   | 2        | 0.83        | 2       | 0.34       | 2  | -0.91 | 0.17  | 1.18  | 0.40     | 0.67 |
| 8  | 59.85 | 7.47   | 2        | 0.83        | 2       | 0.34       | 2  | -0.91 | 0.53  | 1.71  | 0.70     | 0.49 |
| 9  | 64.18 | 8.01   | 2        | 0.83        | 1       | 0.00       | 1  | 0.00  | 1.65  | 5.21  | 2.15     | 0.12 |
| 10 | 55.18 | 6.89   | 1        | 0.00        | 2       | 0.34       | 2  | -0.91 | -0.88 | 0.42  | 0.24     | 0.79 |
| 11 | 56.76 | 7.08   | 1        | 0.00        | 2       | 0.34       | 1  | 0.00  | 0.24  | 1.27  | 0.94     | 0.39 |
| 12 | 50.11 | 6.25   | 1        | 0.00        | 1       | 0.00       | 2  | -0.91 | -1.84 | 0.16  | 0.12     | 0.89 |
| 13 | 59.63 | 7.44   | 2        | 0.83        | 2       | 0.34       | 2  | -0.91 | 0.51  | 1.66  | 1.23     | 0.29 |
| 14 | 57.05 | 7.12   | 2        | 0.83        | 1       | 0.00       | 2  | -0.91 | -0.15 | 0.86  | 0.63     | 0.53 |
| 15 | 39.27 | 4.90   | 1        | 0.00        | 1       | 0.00       | 1  | 0.00  | -2.28 | 0.10  | 0.08     | 0.93 |
| 16 | 43.12 | 5.38   | 1        | 0.00        | 2       | 0.34       | 1  | 0.00  | -1.47 | 0.23  | 0.17     | 0.84 |
| 17 | 38.89 | 4.85   | 2        | 0.83        | 2       | 0.34       | 1  | 0.00  | -1.17 | 0.31  | 0.23     | 0.79 |
| 18 | 44.60 | 5.57   | 1        | 0.00        | 1       | 0.00       | 1  | 0.00  | -1.62 | 0.20  | 0.15     | 0.86 |
| 19 | 53.91 | 6.73   | 1        | 0.00        | 1       | 0.00       | 2  | -0.91 | -1.37 | 0.25  | 0.19     | 0.83 |
| 20 | 44.21 | 5.52   | 2        | 0.83        | 1       | 0.00       | 2  | -0.91 | -1.76 | 0.17  | 0.13     | 0.88 |
| 21 | 59.59 | 7.44   | 1        | 0.00        | 2       | 0.34       | 2  | -0.91 | -0.33 | 0.72  | 0.53     | 0.59 |
| 22 | 74.50 | 9.30   | 2        | 0.83        | 2       | 0.34       | 1  | 0.00  | 3.28  | 26.49 | 1.65     | 0.19 |
| 23 | 43.14 | 5.38   | 2        | 0.83        | 1       | 0.00       | 1  | 0.00  | -0.97 | 0.38  | 0.04     | 0.96 |
| 24 | 63.22 | 7.89   | 1        | 0.00        | 2       | 0.34       | 2  | -0.91 | 0.13  | 1.14  | 0.18     | 0.84 |
| 25 | 64.42 | 8.04   | 2        | 0.83        | 1       | 0.00       | 2  | -0.91 | 0.77  | 2.16  | 0.45     | 0.64 |
| 26 | 58.31 | 7.28   | 1        | 0.00        | 1       | 0.00       | 2  | -0.91 | -0.82 | 0.44  | 0.09     | 0.91 |

The term `age` is the raw value for age for each patient, and the term `lp_age` is the linear predictor,  $\beta_{\text{age}}x_{\text{age}}^{(i)}$  for patient  $i$ . The same is true for the other predictors. The term `lp` is the entire linear predictor,  $\beta^T x$ , for each  $x^{(i)}$ . Confusingly, it has been centered, so its value is shifted from the sum of columns 2, 4, 6, and 8 by a fixed amount (Exercise: What is this amount?). The `risk` term is just the overall risk score,  $\exp(\text{lp})$ . The `expected` term is the expected number of events given the covariates and follow-up time. The survival probability, `surv`, for each subject is  $\exp(-\text{expected})$ .

Cox models make no assumptions about the shape of the baseline hazard, so to make predictions, they simply use the empirical survival (or cumulative hazard) curve for the entire dataset and then adjust it up or down depending on the values of the covariates. The way R does this is super confusing - it estimates the baseline hazard at the means of the covariates after centering, so the baseline hazard is not very interpretable. You can get it out of the model using the `basehaz` function. In any case, here's what you get when you ask

the model to predict the survival and cumulative hazard curves for all of the patients in the training set:



### Question 18.13

Which patients have the lowest and highest risk scores? Where do they appear on the patient-level survival and cumulative hazard graphs?

## 18.7 Testing the Proportional Hazards Assumption

The Cox model makes three important assumptions:

1. *Common baseline hazard.* At any time,  $t$ , all individuals experience the same baseline hazard,  $\lambda_0(t)$ .
2. *Proportional hazards.* The hazard for one individual is proportional to the hazard of any other individual.
3. *Time-invariance.* The constant of proportionality between the hazards of any two individuals does not depend on time.

All of them are potentially problematic. In particular, it's hard to come up with situations in which the hazards for *any* two individuals can reasonably be assumed to be proportional. Thus, it's important to check this assumption.

There are whole book chapters and papers devoted to model diagnostics for the Cox model. I will present a couple of common methods here and leave the others to the course website.

### 18.7.1 Schoenfeld Residuals

In Section 18.4, we saw how the Cox model was fit using maximization of the partial likelihood. The quantity

$$\frac{\exp(\beta^T x^{(i)})}{\sum_{l \in R(t_i)} \exp(\beta^T x^{(l)})}$$

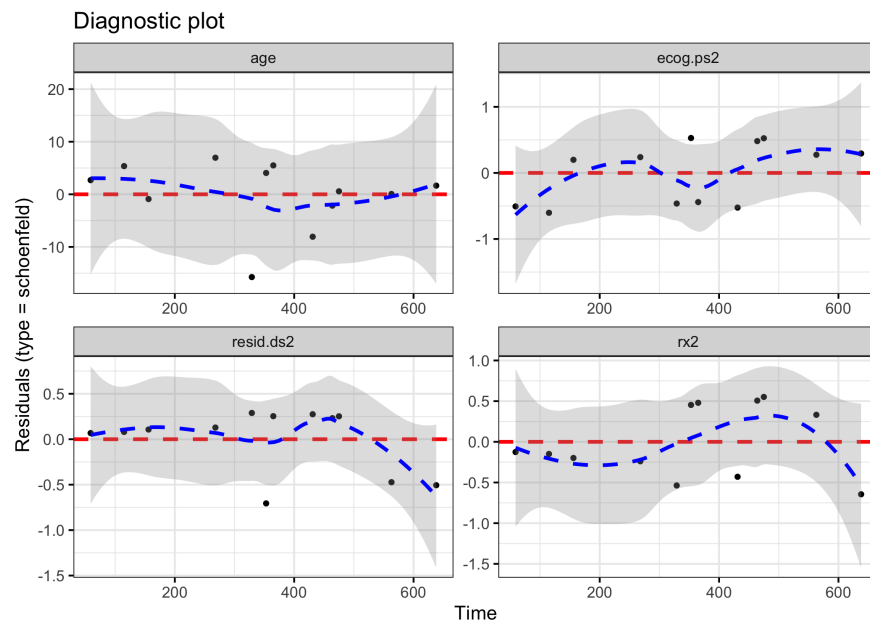
was important because it gave us the probability, according to the model, that the person observed to experience the event at time  $t_i$  would experience

it, given all the people in the risk set just prior to time  $t_i$ . The **Schoenfeld residual** capitalizes on this idea.

**Schoenfeld residual:** The covariate value  $x_j^{(i)}$  for the person ( $i$ ) who actually experienced the event at time  $t_i$ , minus the expected value of the covariate for the risk set at  $t_i$ . Or:

$$\text{residual} = x_j^{(i)} - \sum_{l \in R(t_i)} x_j^{(l)} \frac{\exp(\beta^T x^{(l)})}{\sum_{m \in R(t_i)} \exp(\beta^T x^{(m)})}$$

There is one Schoenfeld residual for each combination of observed event and covariate. Typically, they will be plotted against time to assess if there is a trend. The test for trend comes from a simple linear regression model of the residuals against time, conducted separately for each covariate.



#### Question 18.14

How would you conduct the test for trend for each covariate using a simple linear regression model?

**Question 18.15**

Try calculating the Schoenfeld residual for a single covariate and failure time (your choice). All of the information you need is in the table below. The `risk` column is the same as in the previous table. It is  $\exp(\beta^T x)$ .

|    | futime | fustat | age   | resid.ds | ecog.ps | rx | risk  |
|----|--------|--------|-------|----------|---------|----|-------|
| 1  | 59     | 1      | 72.33 | 2        | 1       | 1  | 14.43 |
| 2  | 115    | 1      | 74.49 | 2        | 1       | 1  | 18.90 |
| 3  | 156    | 1      | 66.47 | 2        | 2       | 1  | 9.71  |
| 22 | 268    | 1      | 74.50 | 2        | 2       | 1  | 26.49 |
| 23 | 329    | 1      | 43.14 | 2        | 1       | 1  | 0.38  |
| 24 | 353    | 1      | 63.22 | 1        | 2       | 2  | 1.14  |
| 25 | 365    | 1      | 64.42 | 2        | 1       | 2  | 2.16  |
| 26 | 377    | 0      | 58.31 | 1        | 1       | 2  | 0.44  |
| 4  | 421    | 0      | 53.36 | 2        | 1       | 2  | 0.54  |
| 5  | 431    | 1      | 50.34 | 2        | 1       | 1  | 0.93  |
| 6  | 448    | 0      | 56.43 | 1        | 2       | 1  | 1.21  |
| 7  | 464    | 1      | 56.94 | 2        | 2       | 2  | 1.18  |
| 8  | 475    | 1      | 59.85 | 2        | 2       | 2  | 1.71  |
| 9  | 477    | 0      | 64.18 | 2        | 1       | 1  | 5.21  |
| 10 | 563    | 1      | 55.18 | 1        | 2       | 2  | 0.42  |
| 11 | 638    | 1      | 56.76 | 1        | 2       | 1  | 1.27  |
| 12 | 744    | 0      | 50.11 | 1        | 1       | 2  | 0.16  |
| 13 | 769    | 0      | 59.63 | 2        | 2       | 2  | 1.66  |
| 14 | 770    | 0      | 57.05 | 2        | 1       | 2  | 0.86  |
| 15 | 803    | 0      | 39.27 | 1        | 1       | 1  | 0.10  |
| 16 | 855    | 0      | 43.12 | 1        | 2       | 1  | 0.23  |
| 17 | 1040   | 0      | 38.89 | 2        | 2       | 1  | 0.31  |
| 18 | 1106   | 0      | 44.60 | 1        | 1       | 1  | 0.20  |
| 19 | 1129   | 0      | 53.91 | 1        | 1       | 2  | 0.25  |
| 20 | 1206   | 0      | 44.21 | 2        | 1       | 2  | 0.17  |
| 21 | 1227   | 0      | 59.59 | 1        | 2       | 2  | 0.72  |

**Question 18.16**

The R function `cox.zph` performs the test of trend for all predictors as well as a global test of trend using ANOVA (don't worry, we'll get to this later). Here is the output for this model:

|          | chisq | df | p     |
|----------|-------|----|-------|
| age      | 0.170 | 1  | 0.680 |
| resid.ds | 1.155 | 1  | 0.282 |
| ecog.ps  | 2.928 | 1  | 0.087 |
| rx       | 0.595 | 1  | 0.440 |
| GLOBAL   | 4.455 | 4  | 0.348 |

For which predictors is there a potentially worrying association between the Schoenfeld residuals and time?

### 18.7.2 What to do if Violations are Found

Interpretation of the Cox model is relatively insensitive to deviations from proportionality, especially for large sample sizes. However, if nonproportionality is a huge issue for one or more predictors, there are a few strategies to deal with it.

1. *Stratify.* One can stratify the model by different levels of the problematic predictor(s), essentially building separate models for the other covariates at each different level of the problematic predictor(s). This only works for predictors that have discrete levels, however; otherwise, one would need to discretize. A potential downside is that stratification eliminates the model's ability to quantify the effect of the stratification variable(s).
2. *Partition the time axis.* Sometimes proportionality holds for the first part of the time axis but falls apart at the end. In that case, one can analyze the data from the first part of the study separately. The disadvantage, of course, is that one must throw out information from later parts of the study.
3. *Add a nonlinear effect term.* Continuous covariates with nonlinear effects on the outcome may lead to nonproportional hazards. Including transformations of these covariates may help to alleviate the nonproportional hazards.