

# Chapter 18: The Cox Proportional Hazards Model

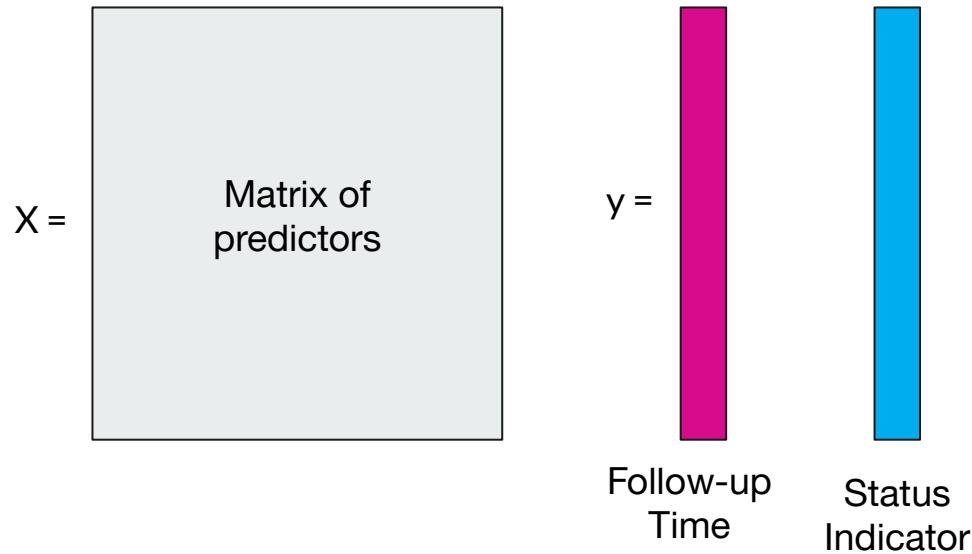
Modern Clinical Data Science  
Chapter Guides  
Bethany Percha, Instructor

---

# How to Use this Guide

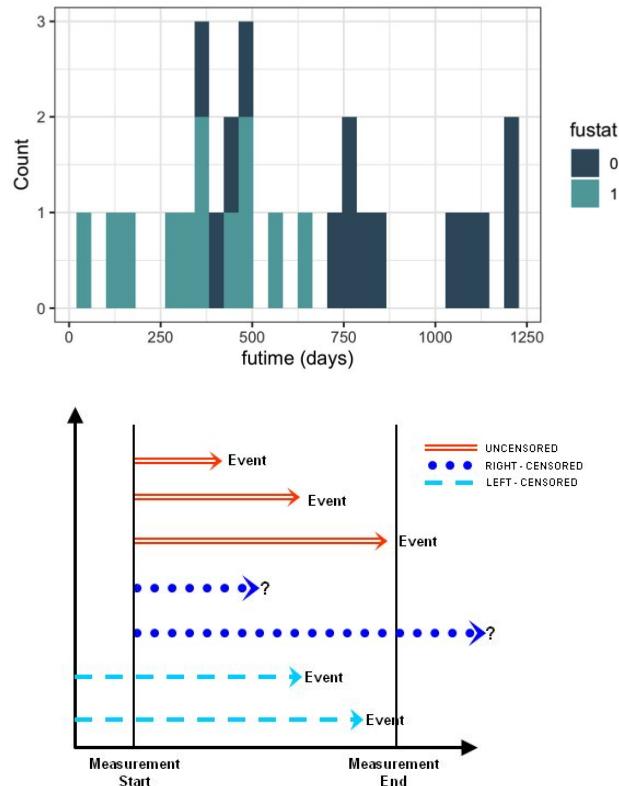
- Read the corresponding notes chapter first
- Try to answer the discussion questions on your own
- Listen to the chapter guide (should be 30 min, max) while following along in the notes

# Survival analysis

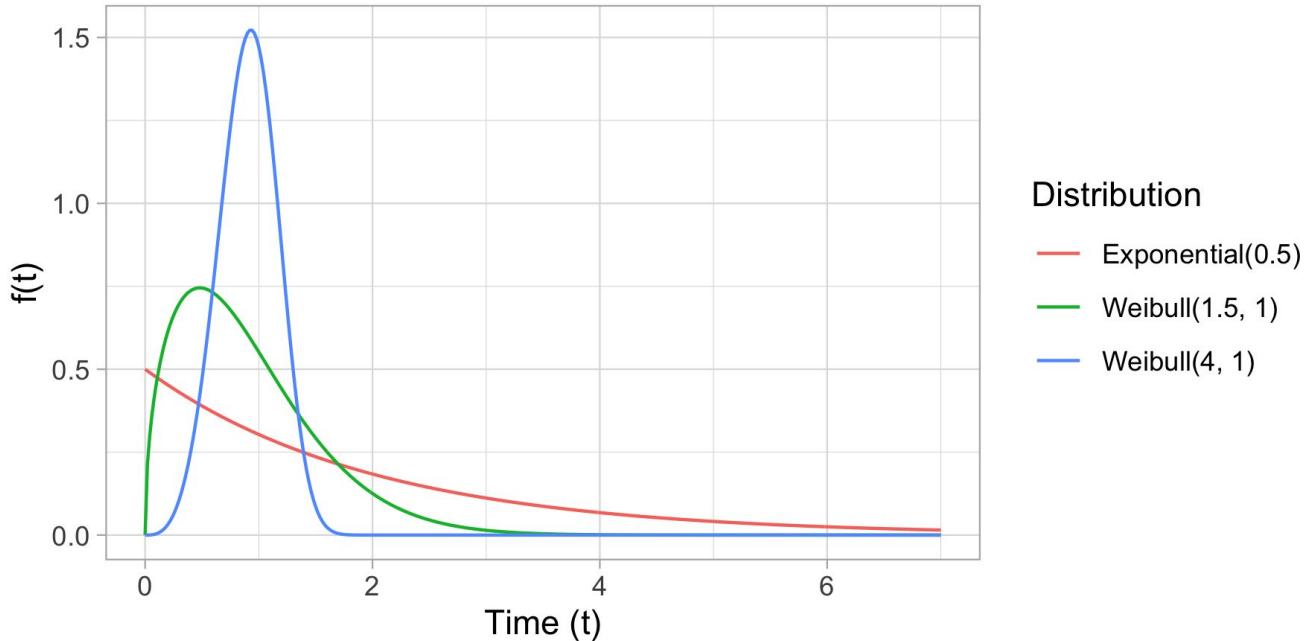


# Ovarian Cancer Dataset

- futime: The number of days from enrollment in the study until death or censoring, whichever came first
- fustat: An indicator of death (1) or censoring (0)
- age: The patient's age in years at the time of treatment administration
- resid.ds: Residual disease present at the time of treatment administration (1 = no, 2 = yes)
- rx: Treatment group (1 = cyclophosphamide, 2 = cyclophosphamide + adriamycin)
- ecog.ps: A measure of performance score or functional status at the time of treatment administration, using the Eastern Cooperative Oncology Group's (ECOG) scale. It ranges from 0 (fully functional) to 4 (completely disabled). Level 4 subjects are usually considered too ill to enter a randomized trial such as this. The patients in this dataset are all at Levels 1 and 2.



# Problem: Tricky to decide how to model outcome.



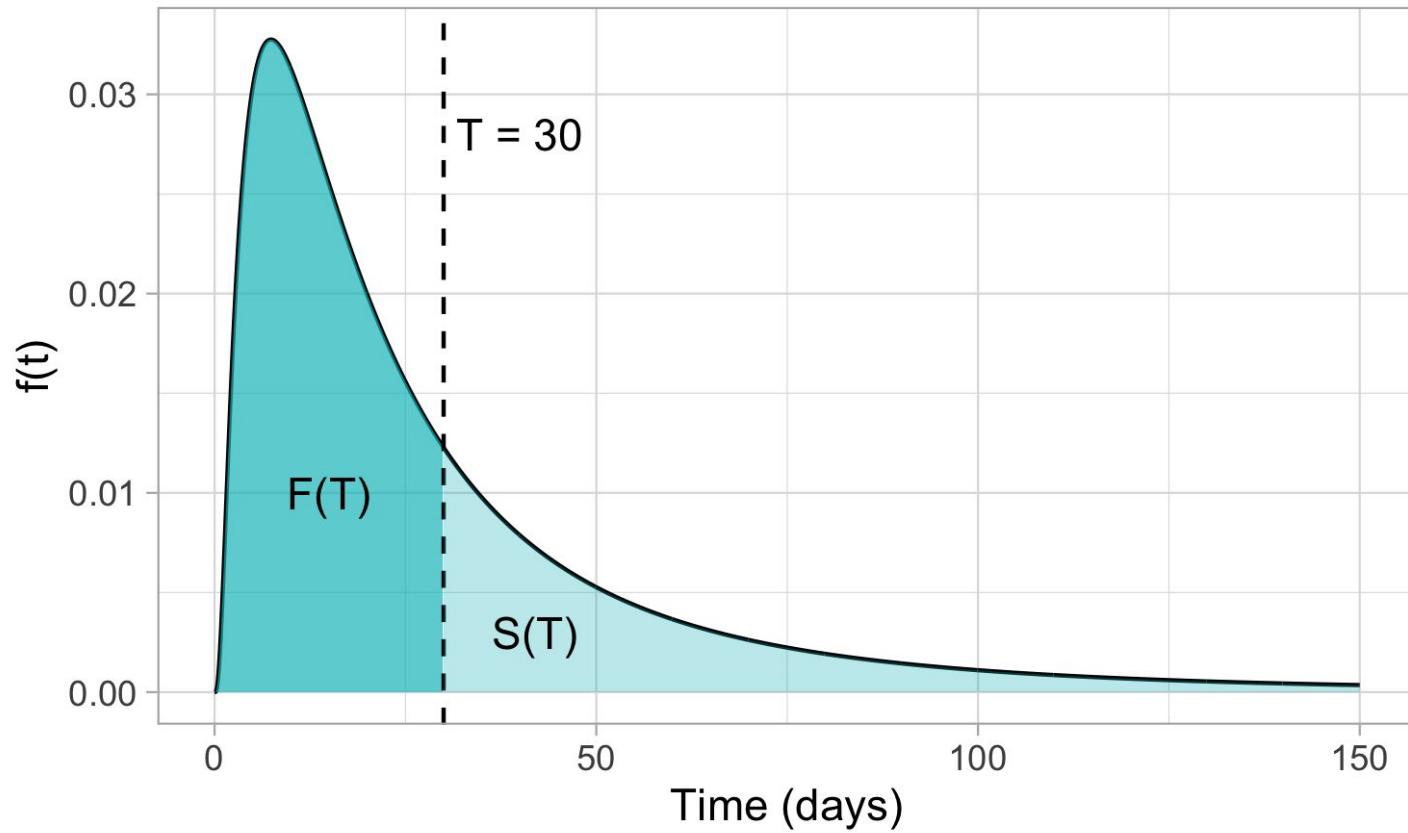
## **Solution: Semiparametric model (Cox model).**

The model gives you a way to quantify the effects of predictors...

- without saying anything about the underlying distribution of event times
- even in the presence of censoring

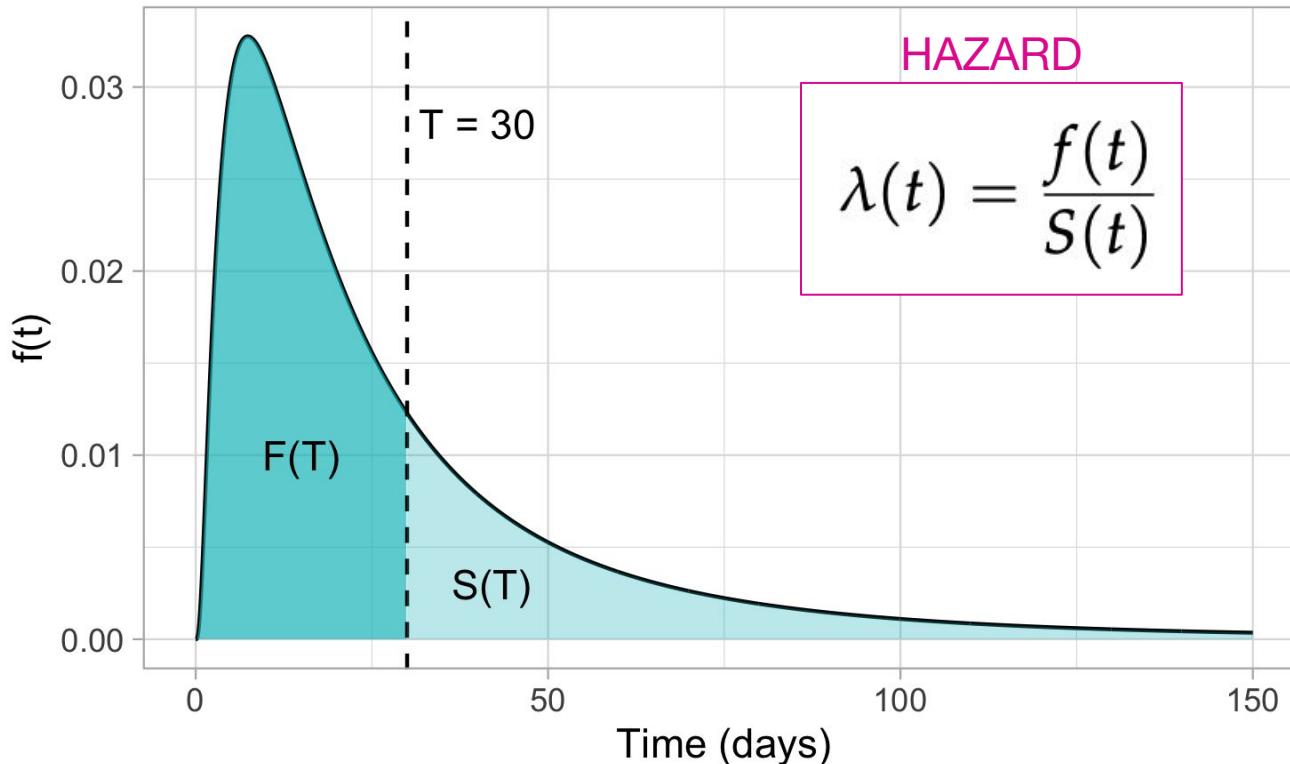
---

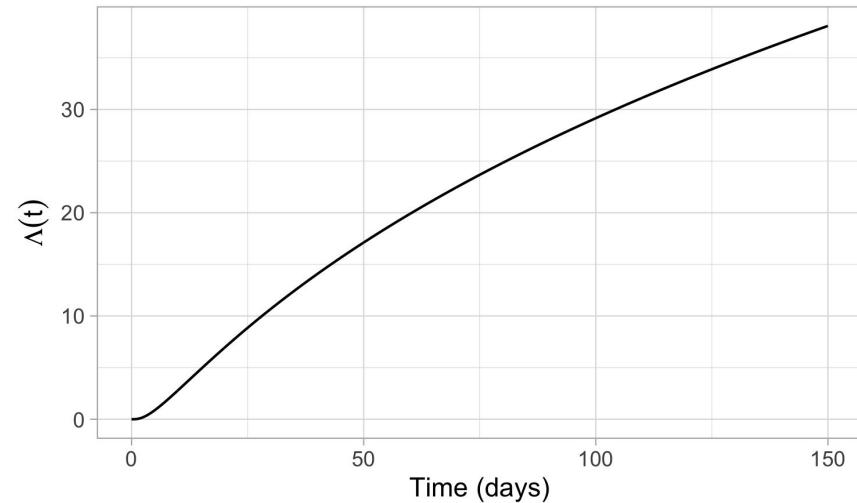
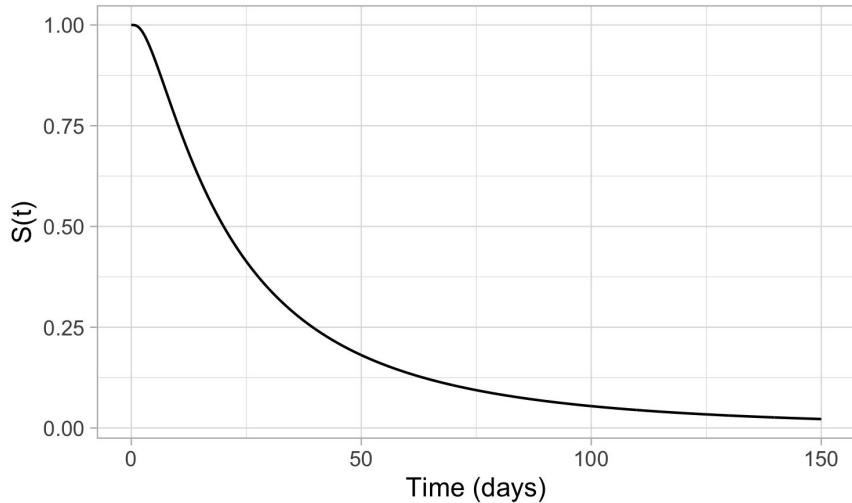
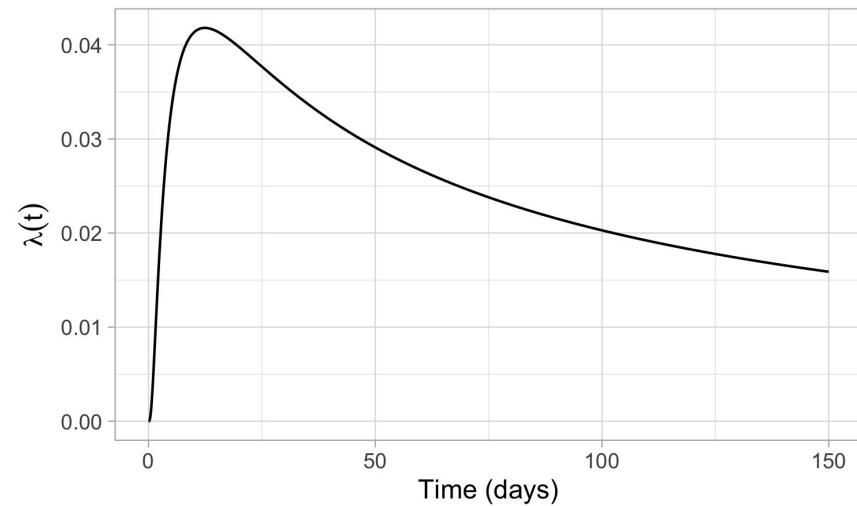
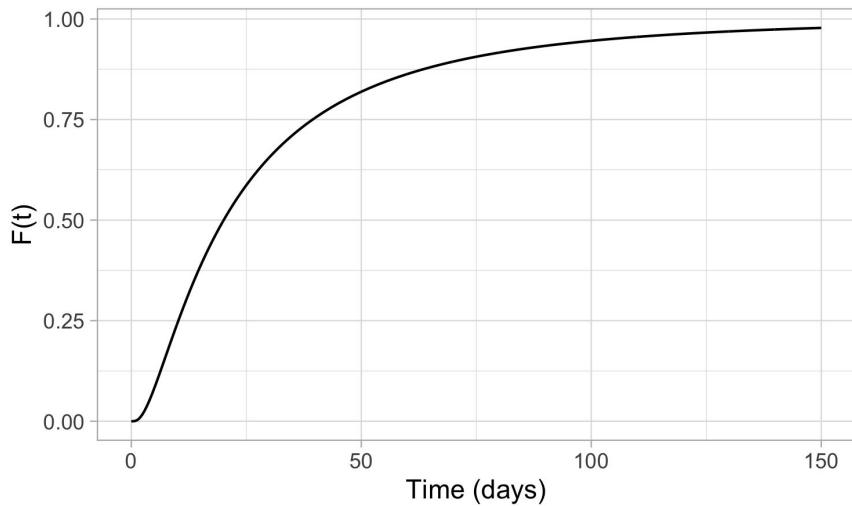
# Part I: Survival and Hazard



### 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  $f(x)$ ?





## 18.1 Survival and Hazard Functions

COMPLETE  
EXPLANATION IN NOTES

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

$$S(t) = 1 - F(t), \quad \text{by time } t, \text{ is:}$$

the **survival function**. The probability of experiencing a finitesimally small interval starting at  $t$ , given that

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

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.

## ONLINE ONLY :)

### 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?

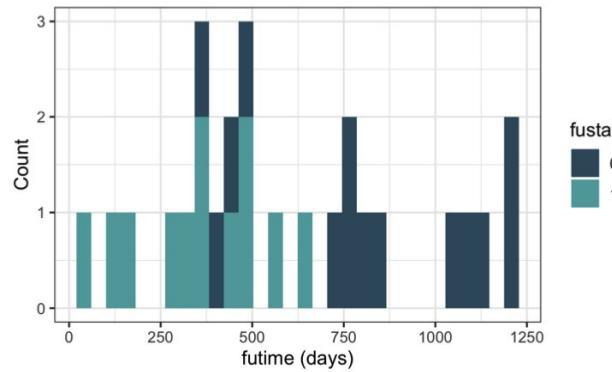
### Question 18.3

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

1. The total amount of risk accumulated up to time  $t$ .
2. The number of times per subject that we would expect to observe failures over a given period if the failure event were repeatable.
3. It just pops out of the modeling.

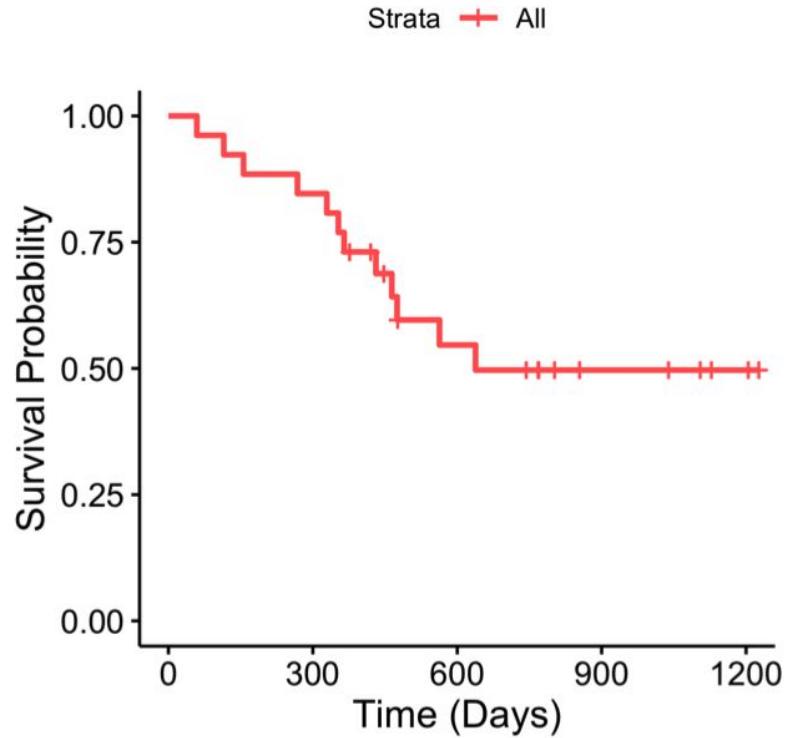
# Return to the ovarian cancer dataset

- futime: The number of days from enrollment in the study until death or censoring, whichever came first
- fustat: An indicator of death (1) or censoring (0)
- age: The patient's age in years at the time of treatment administration
- resid.ds: Residual disease present at the time of treatment administration (1 = no, 2 = yes)
- rx: Treatment group (1 = cyclophosphamide, 2 = cyclophosphamide + adriamycin)
- ecog.ps: A measure of performance score or functional status at the time of treatment administration, using the Eastern Cooperative Oncology Group's (ECOG) scale. It ranges from 0 (fully functional) to 4 (completely disabled). Level 4 subjects are usually considered too ill to enter a randomized trial such as this. The patients in this dataset are all at Levels 1 and 2.



# What's this?

$$\hat{S}(t) = \prod_{j|t_j \leq t} \frac{n_j - d_j}{n_j}$$



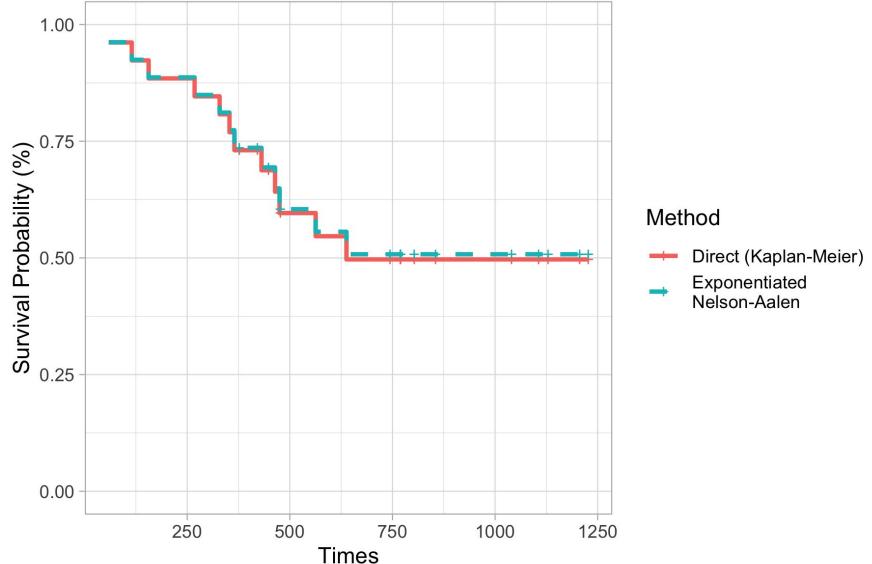
# Estimating the cumulative hazard

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

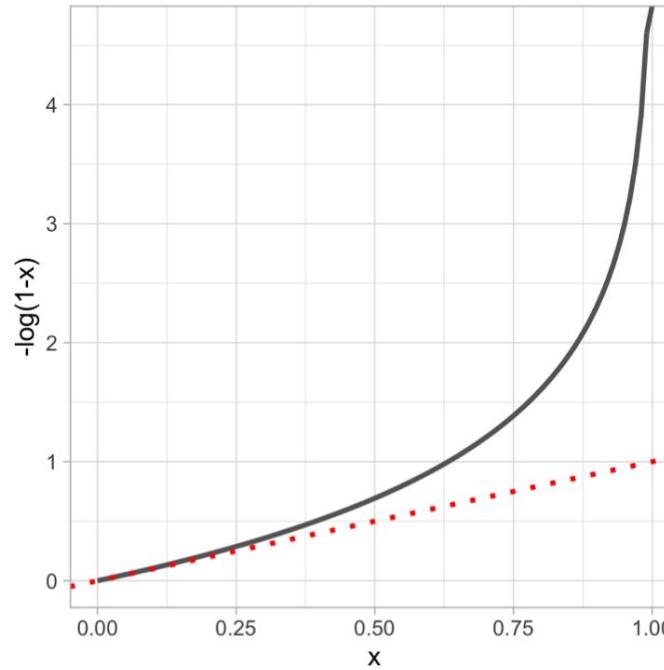
$$\hat{\Lambda}_{NA}(t) = \sum_{i:t_i < t} \frac{d_i}{n_i}$$



## ONLINE ONLY :)

### Question 18.4

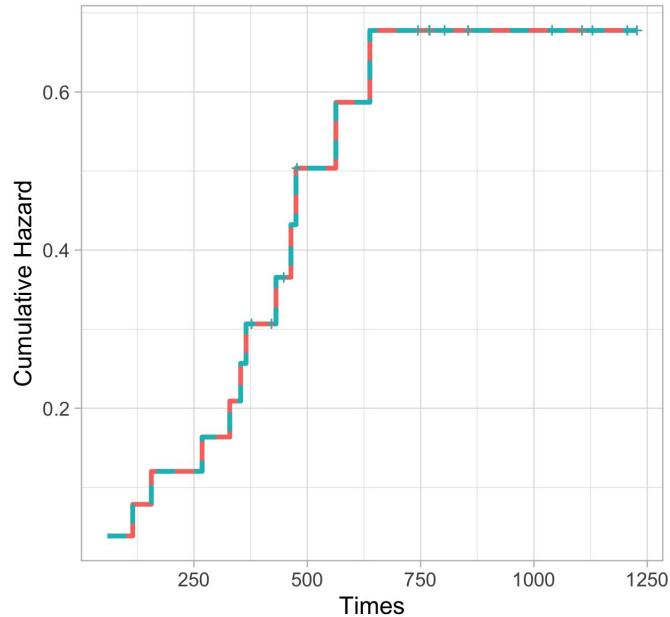
Here is a plot showing the function  $-\log(1 - x)$  vs.  $x$ . Look at the expressions for the two estimators for the cumulative hazard, above. Under what conditions will they be similar? Under what conditions will they be different?



### 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 18.6

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	11	1	0.251	$\hat{\Lambda}(t_2) + \frac{1}{11}$

### 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	11	1	0.251	$\hat{\Lambda}(t_2) + \frac{1}{11}$
4	268	10	1	0.351	$\hat{\Lambda}(t_3) + \frac{1}{10}$

### 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	11	1	0.251	$\hat{\Lambda}(t_2) + \frac{1}{11}$
4	268	10	1	0.351	$\hat{\Lambda}(t_3) + \frac{1}{10}$
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	$\hat{\Lambda}(t_8) + \frac{1}{5}$

### Question 18.6

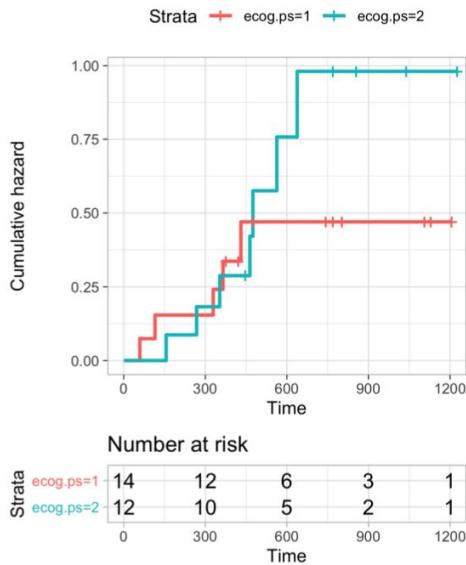
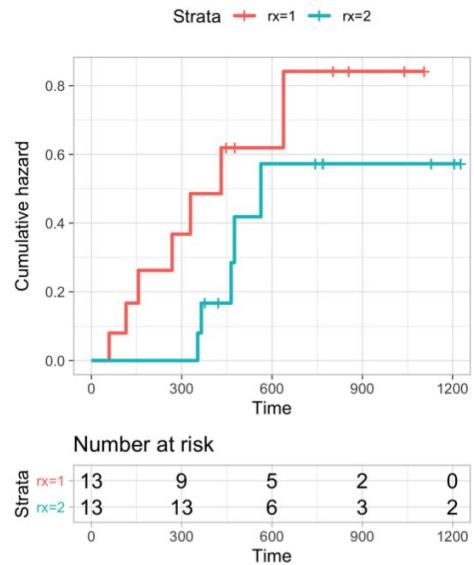
Here are the raw data from treatment group 1 of the ovarian dataset. These are the same data we looked at when building the Kaplan-Meier curve in Chapter 11, Question 11.1. Using these data, fill in the remaining cells of the table below. Here we are using the Nelson-Aalen estimator for the cumulative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

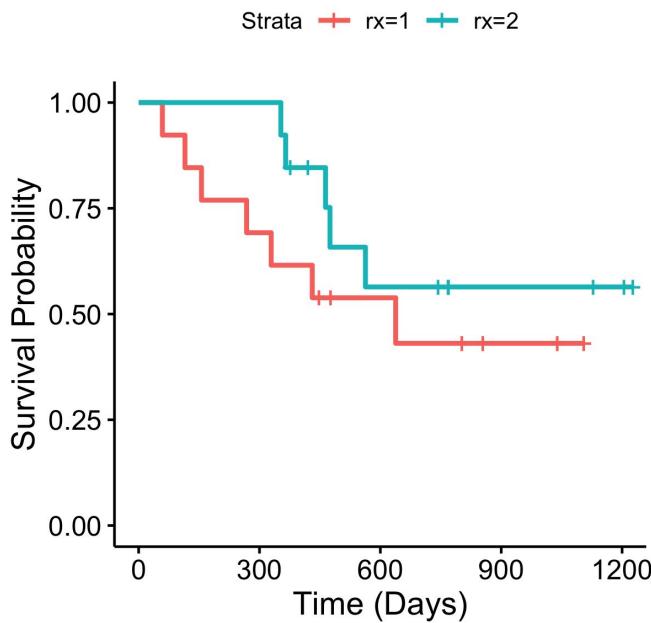
j	$t_j$	$n_j$	$d_j$	$\hat{\Lambda}(t_j)$	Calculation
0	0	13	0	0.000	$\frac{0}{13}$
1	59	13	1	0.077	$\hat{\Lambda}(t_0) + \frac{1}{13}$
2	115	12	1	0.160	$\hat{\Lambda}(t_1) + \frac{1}{12}$
3	156	11	1	0.251	$\hat{\Lambda}(t_2) + \frac{1}{11}$
4	268	10	1	0.351	$\hat{\Lambda}(t_3) + \frac{1}{10}$
5	329	9	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	0.787	$\hat{\Lambda}(t_9) + \frac{0}{4}$
11	855	3	0	0.787	$\hat{\Lambda}(t_{10}) + \frac{0}{3}$
12	1040	2	0	0.787	$\hat{\Lambda}(t_{11}) + \frac{0}{2}$
13	1106	1	0	0.787	$\hat{\Lambda}(t_{12}) + \frac{0}{1}$

### Question 18.7

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



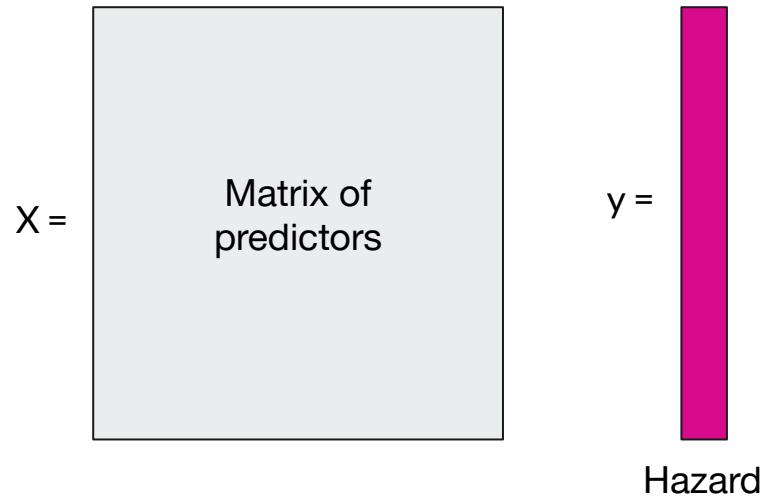
For comparison:



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?

---

# Part II: The Cox Model



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

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

# Cox Model 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.

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

```
```{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

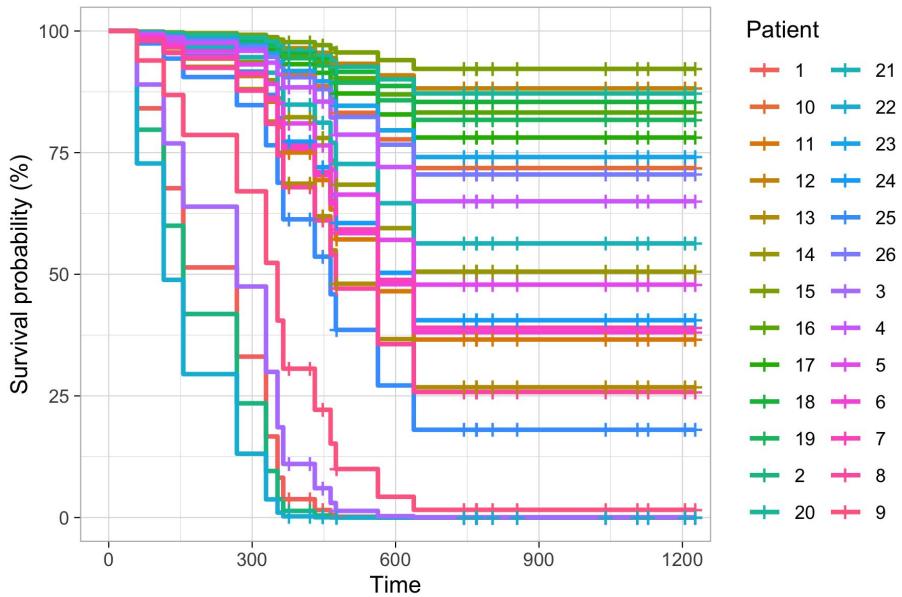
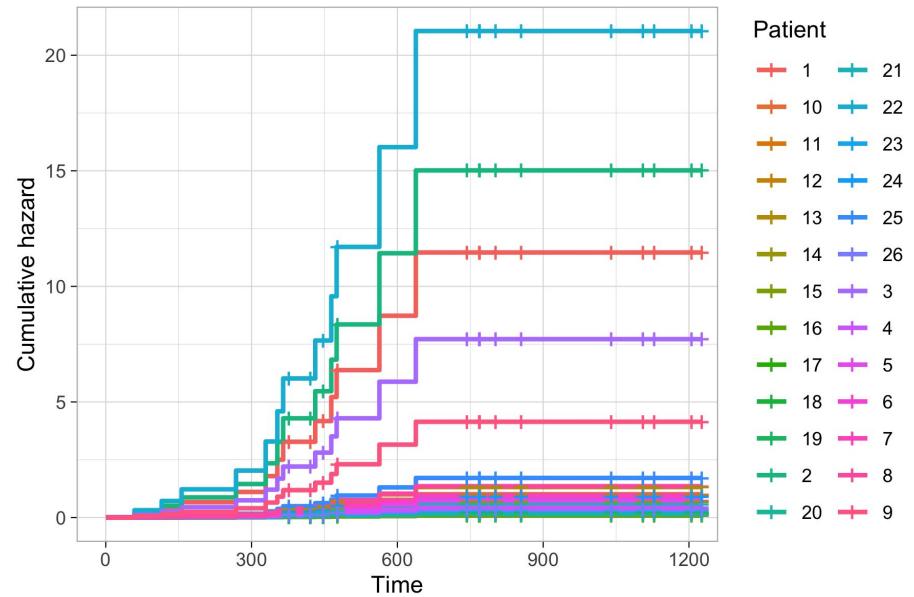
|           | exp(coef) | exp(-coef) | lower .95 | upper .95 |
|-----------|-----------|------------|-----------|-----------|
| age       | 1.1329    | 0.8827     | 1.0335    | 1.242     |
| resid.ds2 | 2.2846    | 0.4377     | 0.4861    | 10.738    |
| ecog.ps2  | 1.3996    | 0.7145     | 0.3962   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



|    | 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 |

### **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?

**Important feature: no dependence of parameters on baseline hazard**

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

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

**Question 18.8**

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

$$\log \left( \frac{\lambda(t|x)}{\lambda_0(t)} \right) = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p. \quad \log \frac{\mu}{1-\mu} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

**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.

# Fitting a Cox Model (please don't kill me)

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

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

## 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.

Call:

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

n= 26, number of events= 12

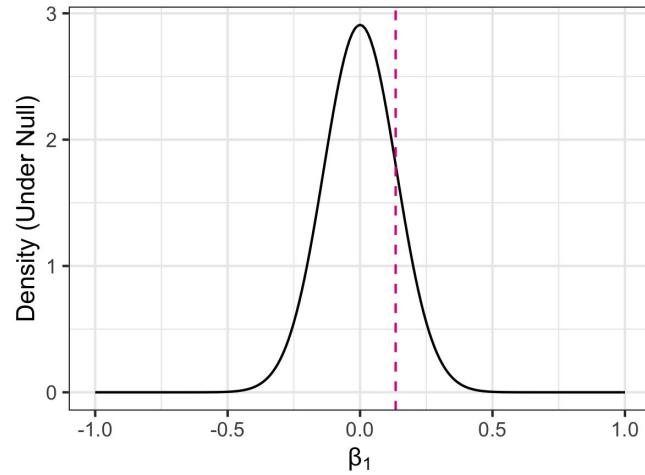
|           | coef     | exp(coef) | se(coef) | z      | Pr(> z )   |
|-----------|----------|-----------|----------|--------|------------|
| age       | 0.12481  | 1.13294   | 0.04689  | 2.662  | 0.00777 ** |
| resid.ds2 | 0.82619  | 2.28459   | 0.78961  | 1.046  | 0.29541    |
| ecog.ps2  | 0.33621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

---

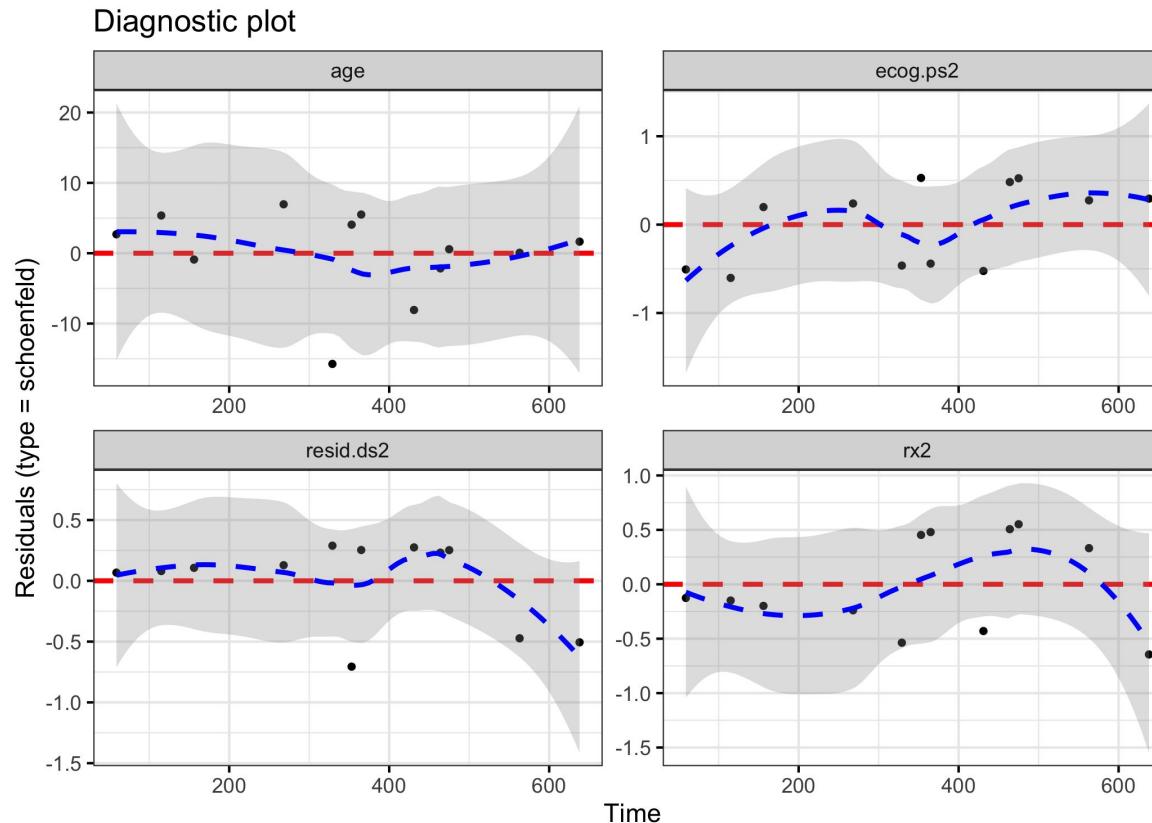
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

|           | 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     |

Remember these graphs from previous chapters?

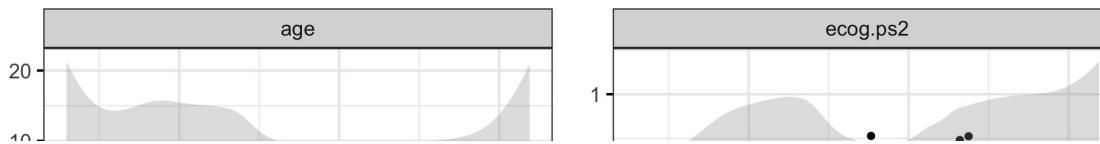


# Model diagnostics: Schoenfeld residuals



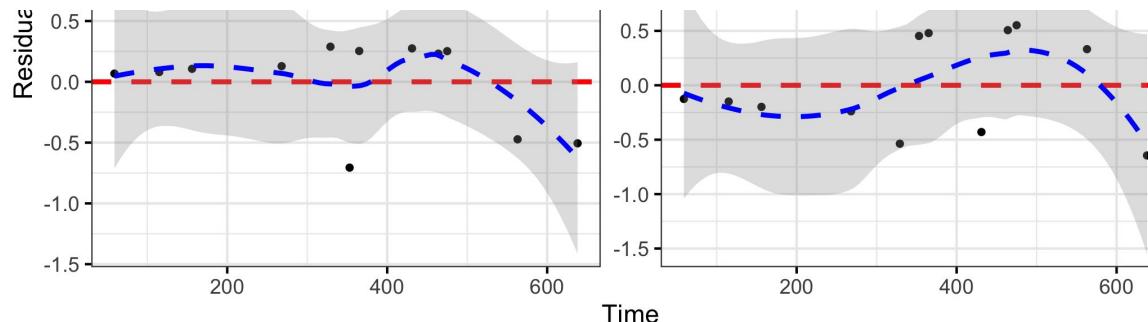
# Model diagnostics: Schoenfeld residuals

Diagnostic plot



## Question 18.14

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



# Model diagnostics: Schoenfeld residuals

**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:

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

**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?

# What to do if proportionality violated:

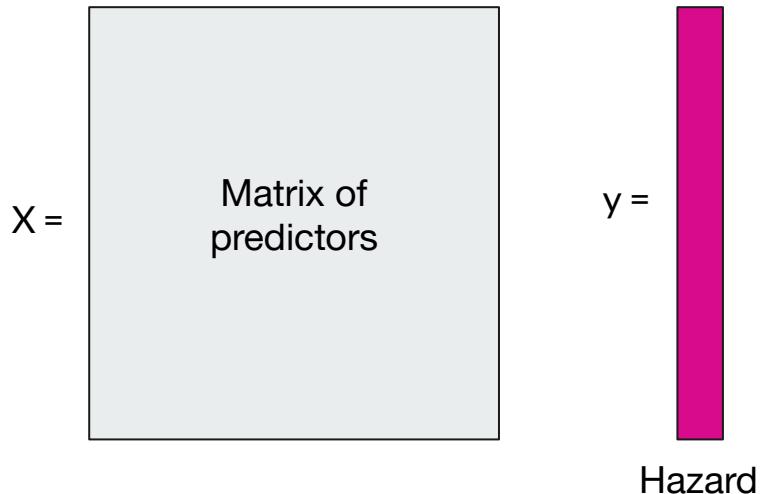
1. *Stratify.* One can stratify the model by different levels of the problematic predictor(s), essentially building separate models for the other covariates at each different level of the problematic predictor(s). This only works for predictors that have discrete levels, however; otherwise, one would 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.

---

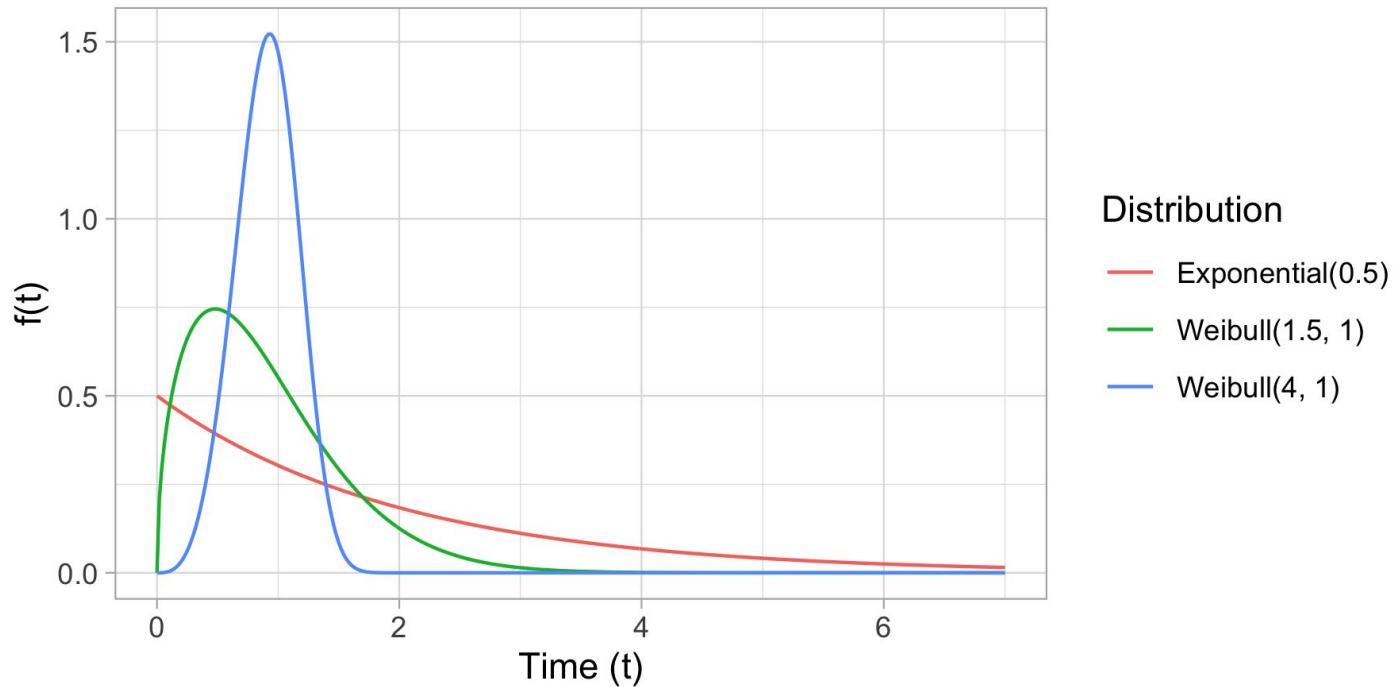
# Part III: Alternatives

# The Cox model is just one tool for this problem.

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



# Parametric regression models



# Penalized Cox models

## Regularized Cox Regression

Kenneth Tay      Noah Simon      Jerome Friedman      Trevor Hastie  
Rob Tibshirani      Balasubramanian Narasimhan

February 17, 2021

### Contents

|   |    |
|---|----|
| Introduction . . . . .                        | 1  |
| Basic usage for right-censored data . . . . . | 2  |
| Cross-validation . . . . .                    | 3  |
| Handling of ties . . . . .                    | 4  |
| Cox models for start-stop data . . . . .      | 6  |
| Stratified Cox models . . . . .               | 8  |
| Plotting survival curves . . . . .            | 9  |
| References . . . . .                          | 12 |

### Introduction

This vignette describes how one can use the `glmnet` package to fit regularized Cox models.

The Cox proportional hazards model is commonly used for the study of the relationship between predictor variables and survival time. In the usual survival analysis framework, we have data of the form  $(y_1, x_1, \delta_1), \dots, (y_n, x_n, \delta_n)$  where  $y_i$ , the observed time, is a time of failure if  $\delta_i$  is 1 or a right-censored time if  $\delta_i$  is 0. We also let  $t_1 < t_2 < \dots < t_m$  be the increasing list of unique failure times, and let  $j(i)$  denote the index of the observation failing at time  $t_i$ .

The Cox model assumes a semi-parametric form for the hazard

$$h_i(t) = h_0(t)e^{x_i^T \beta},$$

where  $h_i(t)$  is the hazard for patient  $i$  at time  $t$ ,  $h_0(t)$  is a shared baseline hazard, and  $\beta$  is a fixed, length  $p$  vector. In the classic setting  $n \geq p$ , inference is made via the partial likelihood

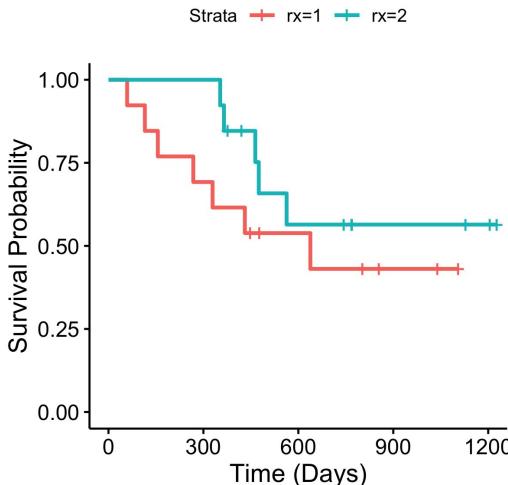
$$L(\beta) = \prod_{i=1}^m \frac{e^{x_{j(i)}^T \beta}}{\sum_{j \in R_i} e^{x_j^T \beta}},$$

# Survival trees

Statistical Methods in Medical Research 1995; **4**: 237–261

## Trees and splines in survival analysis

**Orna Intrator** Department of Statistics, Hebrew University, Jerusalem, Israel and  
**Charles Kooperberg** Department of Statistics, University of Washington, Seattle, USA

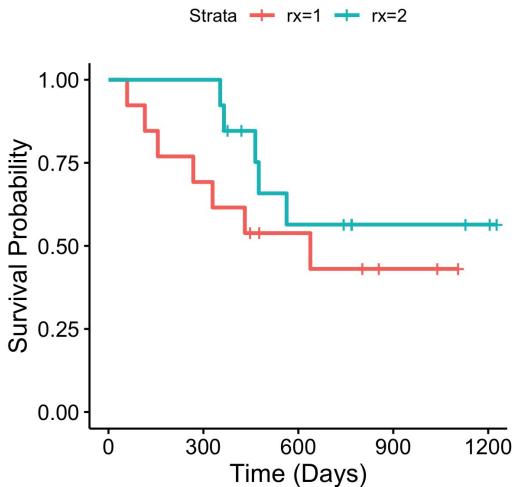


During the past few years several nonparametric alternatives to the Cox proportional hazards model have appeared in the literature. These methods extend techniques that are well known from regression analysis to the analysis of censored survival data. In this paper we discuss methods based on (partition) trees and (polynomial) splines, analyse two datasets using both Survival Trees and HARE, and compare the strengths and weaknesses of the two methods. One of the strengths of HARE is that its model fitting procedure has an implicit check for proportionality of the underlying hazards model. It also provides an explicit model for the conditional hazards function, which makes it very convenient to obtain graphical summaries. On the other hand, the tree-based methods automatically partition a dataset into groups of cases that are similar in survival history. Results obtained by survival trees and HARE are often complementary. Trees and splines in survival analysis should provide the data analyst with two useful tools when analysing survival data.

### 1 Introduction

In this paper we discuss and compare two groups of nonparametric methodologies for the analysis of censored survival data, methods based on recursive partitioning and those based on polynomial splines. Traditional methods for analysing survival data include exploratory methods such as the Kaplan–Meier estimate, the Nelson–Aalen estimate, and various types of tests that summarize differences between two or more survival distributions, and modelling methods such as the Cox proportional hazards model and the accelerated lifetime model. The nonparametric methods discussed in this paper can give insight into data that the traditional methods fail to provide.

# Random survival forests



## Package ‘randomForestSRC’

February 10, 2021

**Version** 2.10.1

**Date** 2021-02-09

**Title** Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC)

**Author** Hemant Ishwaran <hemant.ishwaran@gmail.com>, Udaya B. Kogalur <ubk@kogalur.com>

**Maintainer** Udaya B. Kogalur <ubk@kogalur.com>

**BugReports** <https://github.com>

**Depends** R (>= 3.6.0),

**Imports** parallel, data.tree, Diagram

**Suggests** survival, pec, prodlim, mlcluster

**Description** Fast OpenMP parallelized, unsupervised, survival, classification. Extreme random forests using data. Fast random forests routers for variable importance. Normalize trees on your Safari or Go

**License** GPL (>= 3)

**URL** <http://web.ccs.miami.edu>  
<https://github.com/kogalur>

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2021-02-10 15:01

## Package ‘ranger’

January 10, 2020

**Type** Package

**Title** A Fast Implementation of Random Forests

**Version** 0.12.1

**Date** 2020-01-10

**Author** Marvin N. Wright [aut, cre], Stefan Wager [ctb], Philipp Probst [ctb]

**Maintainer** Marvin N. Wright <cran@rwrig.de>

**Description** A fast implementation of Random Forests, particularly suited for high dimensional data. Ensembles of classification, regression, survival and probability prediction trees are supported. Data from genome-wide association studies can be analyzed efficiently. In addition to data frames, datasets of class 'gwaa.data' (R package 'GenABEL') and 'dgCMatrix' (R package 'Matrix') can be directly analyzed.

**License** GPL-3

**Imports** Rcpp (>= 0.11.2), Matrix

**LinkingTo** Rcpp, RcppEigen

**Depends** R (>= 3.1)

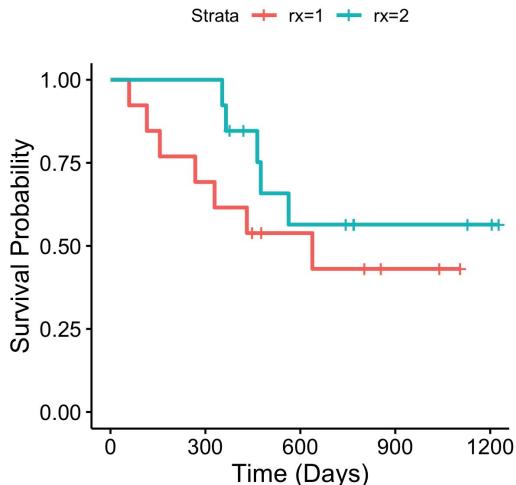
**Suggests** survival, testthat

**Encoding** UTF-8

**RoxygenNote** 7.0.2

**URL** <https://github.com/imbh/b1/ranger>

# Boosted regression models



## Package ‘gbm’

July 15, 2020

**Version** 2.1.8

**Title** Generalized Boosted Regression Models

**Depends** R (>= 2.9.0)

**Imports** lattice, parallel, survival

**Suggests** covr, gridExtra, knitr, pdp, RUnit, splines, tinytest, vip, viridis

**Description** An implementation of extensions to Freund and Schapire's AdaBoost algorithm and Friedman's gradient boosting machine. Includes regression methods for least squares, absolute loss, t-distribution loss, quantile regression, logistic, multinomial logistic, Poisson, Cox proportional hazards partial likelihood, AdaBoost exponential loss, Huberized hinge loss, and Learning to Rank measures (LambdaMart). Originally developed by Greg Ridgeway.

**License** GPL (>= 2) | file LICENSE

**URL** <https://github.com/gbm-developers/gbm>

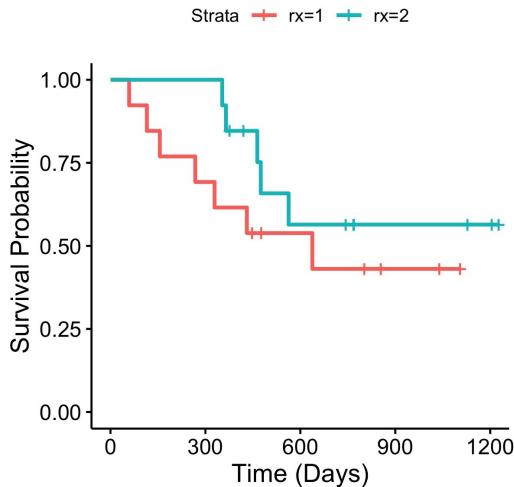
**BugReports** <https://github.com/gbm-developers/gbm/issues>

**Encoding** UTF-8

**RoxygenNote** 7.1.1

**VignetteBuilder** knitr

# And yes... boosted survival trees



## Boosted Trees for Risk Prognosis

**Alexis Bellot**

*Department of Engineering Science  
University of Oxford  
Oxford, United Kingdom*

ALEXIS.BELLOT@ENG.OX.AC.UK

**Mihaela van der Schaar**

*Department of Engineering Science  
University of Oxford  
Oxford, United Kingdom*

MIHAELA.VANDERSCHAAR@ENG.OX.AC.UK

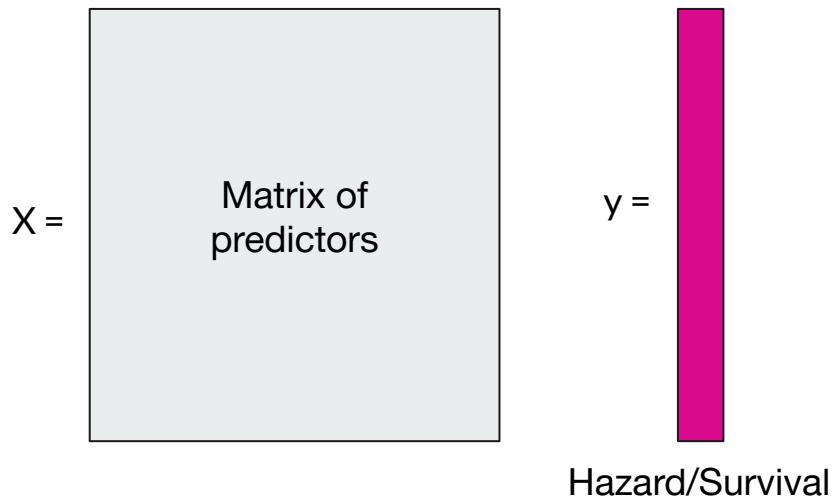
Editor: Editor's name

### Abstract

We present a new approach to ensemble learning for risk prognosis in heterogeneous medical populations. Our aim is to improve overall prognosis by focusing on under-represented patient subgroups with an atypical disease presentation; with current prognostic tools, these subgroups are being consistently mis-estimated. Our method proceeds sequentially by learning nonparametric survival estimators which iteratively learn to improve predictions of previously misdiagnosed patients - a process called *boosting*. This results in fully nonparametric survival estimates, that is, constrained neither by assumptions regarding the baseline hazard nor assumptions regarding the underlying covariate interactions - and thus differentiating our approach from existing boosting methods for survival analysis. In addition, our approach yields a measure of the relative covariate importance that accurately identifies relevant covariates within complex survival dynamics, thereby informing further medical understanding of disease interactions. We study the properties of our approach on a variety of heterogeneous medical datasets, demonstrating significant performance improvements over existing survival and ensemble methods.

### 1. Introduction

**Just remember that these are variations on a theme.**



# Remember this paper from our first class?

Statistical Science  
2001, Vol. 16, No. 3, 199–231

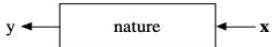
## Statistical Modeling: The Two Cultures

Leo Breiman

**Abstract.** There are two cultures in the use of statistical modeling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown. The statistical community has been committed to the almost exclusive use of data models. This commitment has led to irrelevant theory, questionable conclusions, and has kept statisticians from working on a large range of interesting current problems. Algorithmic modeling, both in theory and practice, has developed rapidly in fields outside statistics. It can be used both on large complex data sets and as a more accurate and informative alternative to data modeling on smaller data sets. If our goal as a field is to use data to solve problems, then we need to move away from exclusive dependence on data models and adopt a more diverse set of tools.

### 1. INTRODUCTION

Statistics starts with data. Think of the data as being generated by a black box in which a vector of input variables  $\mathbf{x}$  (independent variables) go in one side, and on the other side the response variables  $\mathbf{y}$  come out. Inside the black box, nature functions to associate the predictor variables with the response variables, so the picture is like this:



There are two goals in analyzing the data:

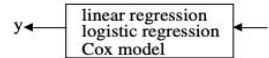
*Prediction.* To be able to predict what the responses are going to be to future input variables;  
*Information.* To extract some information about how nature is associating the response variables to the input variables.

There are two different approaches toward these goals:

#### The Data Modeling Culture

The analysis in this culture starts with assuming a stochastic data model for the inside of the black box. For example, a common data model is that data

The values of the parameters are estimated from the data and the model then used for information and/or prediction. Thus the black box is filled in like this:



*Model validation.* Yes–no using goodness-of-fit tests and residual examination.

*Estimated culture population.* 98% of all statisticians.

#### The Algorithmic Modeling Culture

The analysis in this culture considers the inside of the box complex and unknown. Their approach is to find a function  $f(\mathbf{x})$ —an algorithm that operates on  $\mathbf{x}$  to predict the responses  $\mathbf{y}$ . Their black box looks like this:

