

Chapter 6 Exercises

Estimating survival, cause-specific mortality and competing risks from continuous-time observations

Dennis L. Murray
Guillaume Bastille-Rousseau

Department of Biology, Trent University, Peterborough, ON, K9J 7B8, Canada

Exercise 1: Preparing survival databases and graphing survival curves

In this first exercise, we will implement basic survival functions in R, run preliminary analyses of survival using Kaplan-Meier and Nelson-Aalen estimates, and graphically display survival curves using a simulated dataset.

We will first open a csv file containing the simulated data for the exercise and display them in R using the following R code:

```
# setwd("../") # Specify working directory
data<-read.csv("Surv_exerc1.csv", sep=";") # Open csv file
data      # Display data
```

##	id	enter	exit	death	x1	x2	x3
## 1	1	0	16	1	5.8	0	2
## 2	2	0	75	0	6.9	1	1
## 3	3	0	4	1	2.3	0	3
## 4	4	2	56	1	7.9	1	2
## 5	5	3	8	1	2.0	0	2
## 6	6	7	49	1	8.2	1	3
## 7	7	14	71	1	4.4	0	1
## 8	8	15	16	1	2.3	0	2
## 9	9	16	100	0	7.6	1	3
## 10	10	16	18	1	2.9	0	2
## 11	11	27	99	0	8.4	1	1
## 12	12	28	33	1	4.6	1	3
## 13	13	29	85	0	6.7	1	2
## 14	14	30	33	1	3.6	0	3
## 15	15	30	34	1	3.1	0	1
## 16	16	30	35	1	2.4	0	2
## 17	17	31	42	0	5.5	1	3
## 18	18	32	44	1	5.0	0	1
## 19	19	32	100	0	7.1	0	2
## 20	20	32	36	1	4.1	0	3
## 21	21	37	41	1	2.2	0	2
## 22	22	37	38	1	1.8	0	3
## 23	23	38	79	0	7.2	1	1
## 24	24	40	43	1	1.7	0	2
## 25	25	41	100	0	9.8	1	3
## 26	26	42	43	1	2.6	0	2
## 27	27	44	100	0	9.1	1	3
## 28	28	45	67	1	7.8	1	2
## 29	29	47	50	1	3.2	0	1
## 30	30	50	100	0	8.2	1	3

The first six rows of the table are as follows:

```
head(data)

##   id enter exit death  x1 x2 x3
## 1  1     0   16     1 5.8 0  2
## 2  2     0   75     0 6.9 1  1
## 3  3     0    4     1 2.3 0  3
## 4  4     2   56     1 7.9 1  2
## 5  5     3    8     1 2.0 0  2
## 6  6     7   49     1 8.2 1  3
```

This table is comprised of a subject ID (**id**), day of entry into the study (**enter**), day of exit from the study (**exit**) the failure binary variable (**death**: 1=death, 0=censor), and 3 predictors or covariates that are intrinsic to the animal and that we suspect might be related to mortality risk (**x1**, **x2**, **x3**). **x1** is a continuous variable and hypothetically-speaking could represent body mass or size of individuals, **x2** is binary and might reflect gender, and **x3** is categorical and could stand for age class. Of course, these variables also could be extrinsic to the subjects and might represent features like habitat use patterns, home range size, exposure to high risk environments, or group size.

We need to create a variable representing absolute time to allow us to analyze survival according to study time (i.e., we need to get rid of calendar time to standardize for the different entry times). We will add this new variable to our dataset:

```
data$abstime<-data$exit-data$enter
```

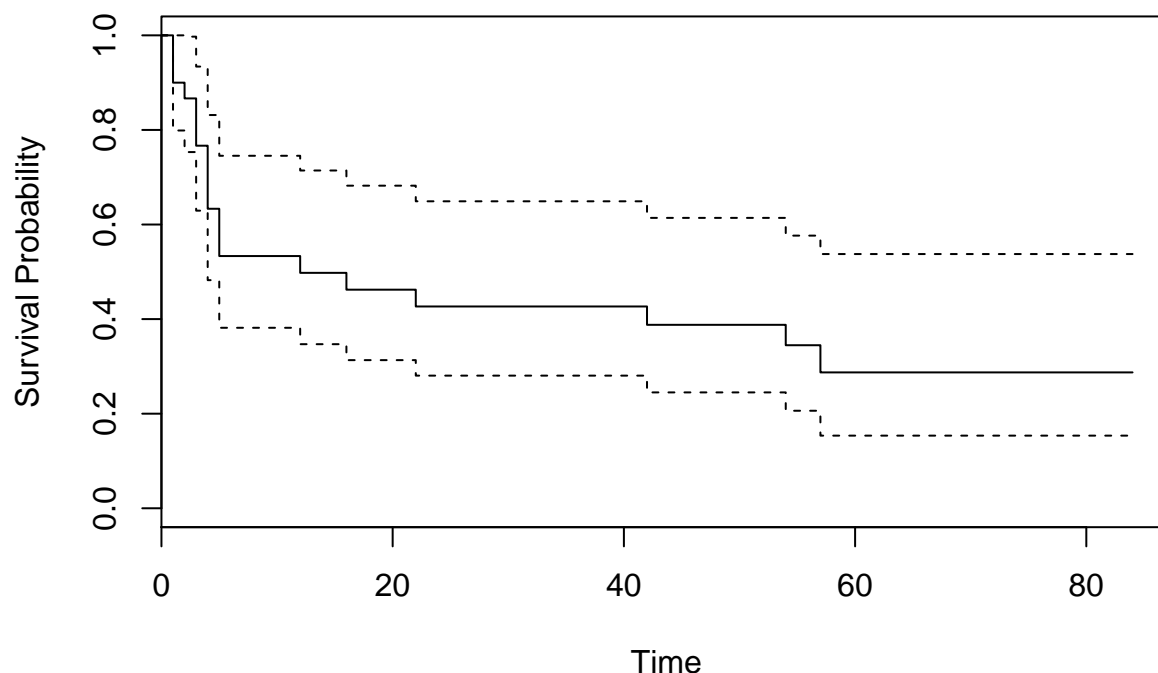
We are now ready to determine survival functions. The library “survival” contains several descriptive statistics and models that are commonly used in survival analysis. We will first generate a Kaplan-Meier (KM) curve using the following lines and display them as a graph:

```
# install.packages("Survival") # Note: only need to run prior to first use
library(survival)
attach(data)
model0 <- survfit(Surv(abstime, death) ~ 1)
summary(model0) # Results of the KM
```

```
## Call: survfit(formula = Surv(abstime, death) ~ 1)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    1     30      3    0.900  0.0548    0.799    1.000
##    2     27      1    0.867  0.0621    0.753    0.997
##    3     26      3    0.767  0.0772    0.629    0.934
##    4     23      4    0.633  0.0880    0.482    0.832
##    5     19      3    0.533  0.0911    0.382    0.745
##   12     15      1    0.498  0.0917    0.347    0.714
##   16     14      1    0.462  0.0918    0.313    0.682
##   22     13      1    0.427  0.0913    0.280    0.649
##   42     11      1    0.388  0.0909    0.245    0.614
##   54      9      1    0.345  0.0904    0.206    0.577
##   57      6      1    0.287  0.0918    0.154    0.538
```

```
plot(model0, main="Kaplan-Meier estimate with 95% confidence bounds",
      xlab="Time", ylab="Survival Probability")
```

Kaplan–Meier estimate with 95% confidence bounds



The KM estimator describes the survival probability of an individual through time. From the figure or model output, we can obtain survival rates at any given point in time. It is possible to relate output from `summary(model1)` to the figure, for example, the survival rate after 54 days is 0.345.

It could also be valuable to represent the survival probability of an individual from the beginning of the study rather than the time since capture (see Chapter 6, section 6.3.3). In this case, the survival time is calculated not from the capture date but from a specific calendar date. In our example, because we only have one year of data, modelling time since the beginning of the study is similar to using a recurrent approach (see Fieberg & Delgiudice 2009). The following code will calculate KM estimates using time since start of the study and display them with the previous KM curve:

```
model1 <- survfit(Surv(exit, death) ~ 1, conf.type="none" )
model1 # Give information about the survival function

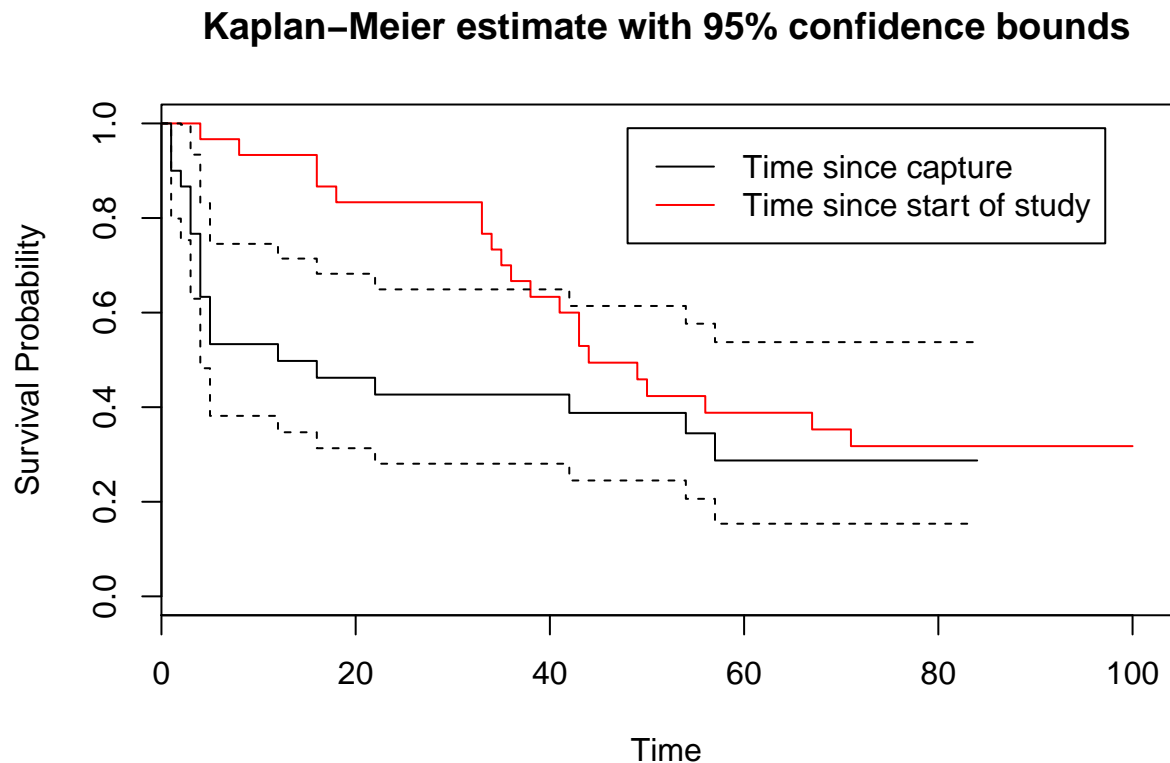
## Call: survfit(formula = Surv(exit, death) ~ 1, conf.type = "none")
##
##      n events median
##    30      20     44

summary(model1) # Results of the Kaplan-Meier

## Call: survfit(formula = Surv(exit, death) ~ 1, conf.type = "none")
##
##  time n.risk n.event survival std.err
##    4     30      1    0.967  0.0328
##    8     29      1    0.933  0.0455
##   16     28      2    0.867  0.0621
##   18     26      1    0.833  0.0680
##   33     25      2    0.767  0.0772
##   34     23      1    0.733  0.0807
##   35     22      1    0.700  0.0837
##   36     21      1    0.667  0.0861
```

```
##      38      20      1      0.633 0.0880
##      41      19      1      0.600 0.0894
##      43      17      2      0.529 0.0918
##      44      15      1      0.494 0.0922
##      49      14      1      0.459 0.0921
##      50      13      1      0.424 0.0916
##      56      12      1      0.388 0.0905
##      67      11      1      0.353 0.0889
##      71      10      1      0.318 0.0867
```

```
plot(model1, main="Kaplan-Meier estimate with 95% confidence bounds",
      xlab="Time", ylab="Survival Probability", col=2) # Graphical display
lines(model0)
legend(48, 0.99, c("Time since capture", "Time since start of study"), lty=1, col=c(1,2) )
```



As you see from the graph, changing the format of the time variable drastically alters the shape of the KM estimates. Using time since capture, the survival probability declines markedly during the first few days whereas using time since start of the study we see a more gradual decline. This illustrates the importance of knowing exactly what questions we want to answer when selecting an appropriate time unit for analysis.

If we are interested in examining the effects of gender (**x2**) on survival, we can type the following lines to obtain a gender specific KM curve. In this case, we use a non-parametric log-rank test to compare survival estimates:

```
modelx2 <- survfit(Surv(abstime, death) ~ x2)
summary(modelx2)
```

```
## Call: survfit(formula = Surv(abstime, death) ~ x2)
##
##              x2=0
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    1      17       3   0.8235  0.0925   0.66087      1.000
```

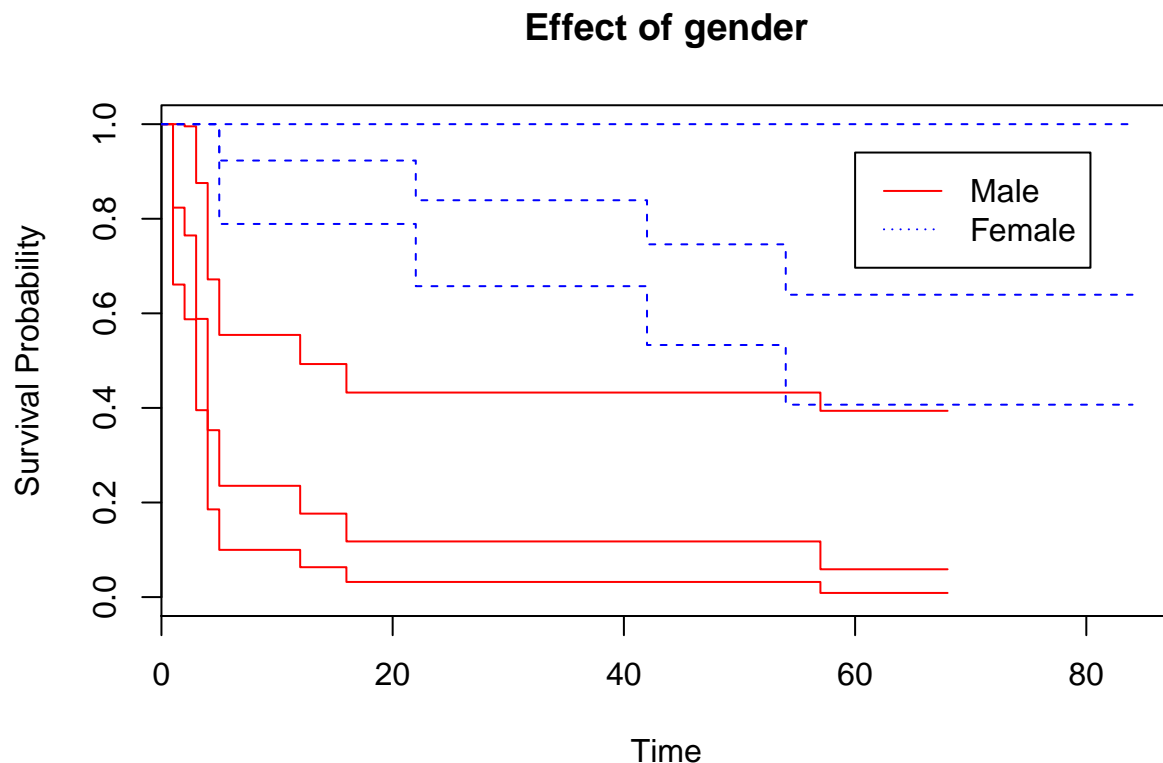
```
##      2      14      1  0.7647  0.1029      0.58746      0.995
##      3      13      3  0.5882  0.1194      0.39521      0.876
##      4      10      4  0.3529  0.1159      0.18543      0.672
##      5       6      2  0.2353  0.1029      0.09987      0.554
##     12       4      1  0.1765  0.0925      0.06320      0.493
##     16       3      1  0.1176  0.0781      0.03200      0.432
##     57       2      1  0.0588  0.0571      0.00879      0.394
```

```
##
##              x2=1
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##      5      13       1   0.923  0.0739      0.789      1
##     22      11       1   0.839  0.1045      0.657      1
##     42       9       1   0.746  0.1279      0.533      1
##     54       7       1   0.639  0.1475      0.407      1
```

```
survdif(Surv(abstime, death) ~ x2) # Log-rank test
```

```
## Call:
## survdiff(formula = Surv(abstime, death) ~ x2)
##
##      N Observed Expected (O-E)^2/E (O-E)^2/V
## x2=0 17      16      7.78      8.69     16.5
## x2=1 13       4     12.22      5.53     16.5
##
## Chisq= 16.5 on 1 degrees of freedom, p= 5e-05
```

```
plot(modelx2, lty=c(1,2), main="Effect of gender", col=c("red", "blue"),
      xlab="Time", ylab="Survival Probability", conf.int=T)
legend(60, 0.94, c("Male", "Female"), lty=c(1,3), col=c("red", "blue"))
```



This new figure illustrates that survival probability is significantly different across most of the range as is

evident from the non-overlapping confidence intervals. The significant P-value produced by the **survdiff** further supports these differences, although for most research in ecology the result of graphical comparisons and nonparametric tests should not be considered as the terminal analysis (see Chapter 11, section 11.5.1).

Now, say that we want to compare survival functions between age classes (**x3**) and body mass (**x1**). We follow the same approach:

```
modelx3 <- survfit(Surv(abstime, death) ~ x3)
modelx1 <- survfit(Surv(abstime, death) ~ x1)
summary(modelx3)
```

```
## Call: survfit(formula = Surv(abstime, death) ~ x3)
##
##               x3=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     3      7       1   0.857   0.132    0.633      1
##     4      6       1   0.714   0.171    0.447      1
##    12      5       1   0.571   0.187    0.301      1
##    57      3       1   0.381   0.199    0.137      1
##
##               x3=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1     12       2   0.833   0.108    0.6470    1.000
##     2     10       1   0.750   0.125    0.5410    1.000
##     3      9       1   0.667   0.136    0.4468    0.995
##     4      8       1   0.583   0.142    0.3616    0.941
##     5      7       2   0.417   0.142    0.2133    0.814
##    16      5       1   0.333   0.136    0.1498    0.742
##    22      4       1   0.250   0.125    0.0938    0.666
##    54      3       1   0.167   0.108    0.0470    0.591
##
##               x3=3
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1     11       1   0.909  0.0867    0.754    1.000
##     3     10       1   0.818  0.1163    0.619    1.000
##     4      9       2   0.636  0.1450    0.407    0.995
##     5      7       1   0.545  0.1501    0.318    0.936
##    42      5       1   0.436  0.1547    0.218    0.874
summary(modelx1)
```

```
## Call: survfit(formula = Surv(abstime, death) ~ x1)
##
##               x1=1.7
##      time      n.risk      n.event      survival      std.err
##        3          1          1          0          NaN
## lower 95% CI upper 95% CI
##      NA          NA
##
##               x1=1.8
##      time      n.risk      n.event      survival      std.err
##        1          1          1          0          NaN
## lower 95% CI upper 95% CI
##      NA          NA
##
##               x1=2
```

```

##           time          n.risk      n.event      survival      std.err
##           5              1          1           0           NaN
## lower 95% CI upper 95% CI
##           NA              NA
##
##           x1=2.2
##           time          n.risk      n.event      survival      std.err
##           4              1          1           0           NaN
## lower 95% CI upper 95% CI
##           NA              NA
##
##           x1=2.3
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    1     2     1     0.5   0.354     0.125         1
##    4     1     1     0.0   NaN         NA         NA
##
##           x1=2.4
##           time          n.risk      n.event      survival      std.err
##           5              1          1           0           NaN
## lower 95% CI upper 95% CI
##           NA              NA
##
##           x1=2.6
##           time          n.risk      n.event      survival      std.err
##           1              1          1           0           NaN
## lower 95% CI upper 95% CI
##           NA              NA
##
##           x1=2.9
##           time          n.risk      n.event      survival      std.err
##           2              1          1           0           NaN
## lower 95% CI upper 95% CI
##           NA              NA
##
##           x1=3.1
##           time          n.risk      n.event      survival      std.err
##           4              1          1           0           NaN
## lower 95% CI upper 95% CI
##           NA              NA
##
##           x1=3.2
##           time          n.risk      n.event      survival      std.err
##           3              1          1           0           NaN
## lower 95% CI upper 95% CI
##           NA              NA
##
##           x1=3.6
##           time          n.risk      n.event      survival      std.err
##           3              1          1           0           NaN
## lower 95% CI upper 95% CI
##           NA              NA
##
##           x1=4.1
##           time          n.risk      n.event      survival      std.err

```

```

##           4           1           1           0           NaN
## lower 95% CI upper 95% CI
##           NA           NA
##
##           x1=4.4
##           time           n.risk           n.event           survival           std.err
##           57           1           1           0           NaN
## lower 95% CI upper 95% CI
##           NA           NA
##
##           x1=4.6
##           time           n.risk           n.event           survival           std.err
##           5           1           1           0           NaN
## lower 95% CI upper 95% CI
##           NA           NA
##
##           x1=5
##           time           n.risk           n.event           survival           std.err
##           12           1           1           0           NaN
## lower 95% CI upper 95% CI
##           NA           NA
##
##           x1=5.5
##           time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##           x1=5.8
##           time           n.risk           n.event           survival           std.err
##           16           1           1           0           NaN
## lower 95% CI upper 95% CI
##           NA           NA
##
##           x1=6.7
##           time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##           x1=6.9
##           time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##           x1=7.1
##           time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##           x1=7.2
##           time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##           x1=7.6
##           time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##           x1=7.8
##           time           n.risk           n.event           survival           std.err
##           22           1           1           0           NaN
## lower 95% CI upper 95% CI
##           NA           NA
##
##           x1=7.9
##           time           n.risk           n.event           survival           std.err

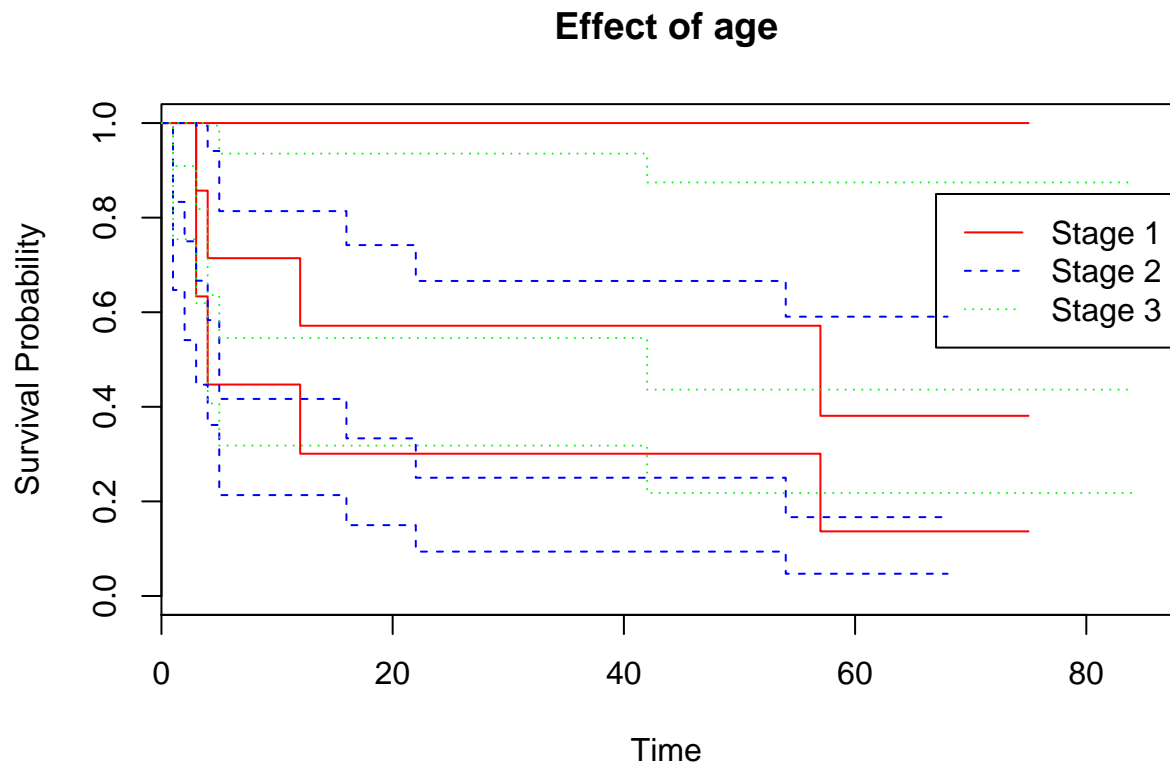
```



```
##          54          1          1          0          NaN
## lower 95% CI upper 95% CI
##          NA          NA
##
##          x1=8.2
##          time      n.risk      n.event      survival      std.err
##          42.000      2.000      1.000      0.500      0.354
## lower 95% CI upper 95% CI
##          0.125      1.000
##
##          x1=8.4
##          time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##          x1=9.1
##          time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##          x1=9.8
##          time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

A brief look at the output reveals that the output of `modelx1` provides little information. This is because KM estimates are suitable for categorical variables, but not continuous variables. We will therefore only display the KM curve for the effect of age stage (`modelx3`):

```
plot(modelx3, lty=c(1,2,3), main="Effect of age", col=c("red", "blue", "green"),
     xlab="Time", ylab="Survival Probability", conf.int=T)
legend(67, 0.85, c("Stage 1", "Stage 2", "Stage 3"), lty=c(1,2,3), col=c("red", "blue", "green"))
```



Question 1:

For the stage variable (x_3), are differences in survival probability between groups significant at Day 10, Day

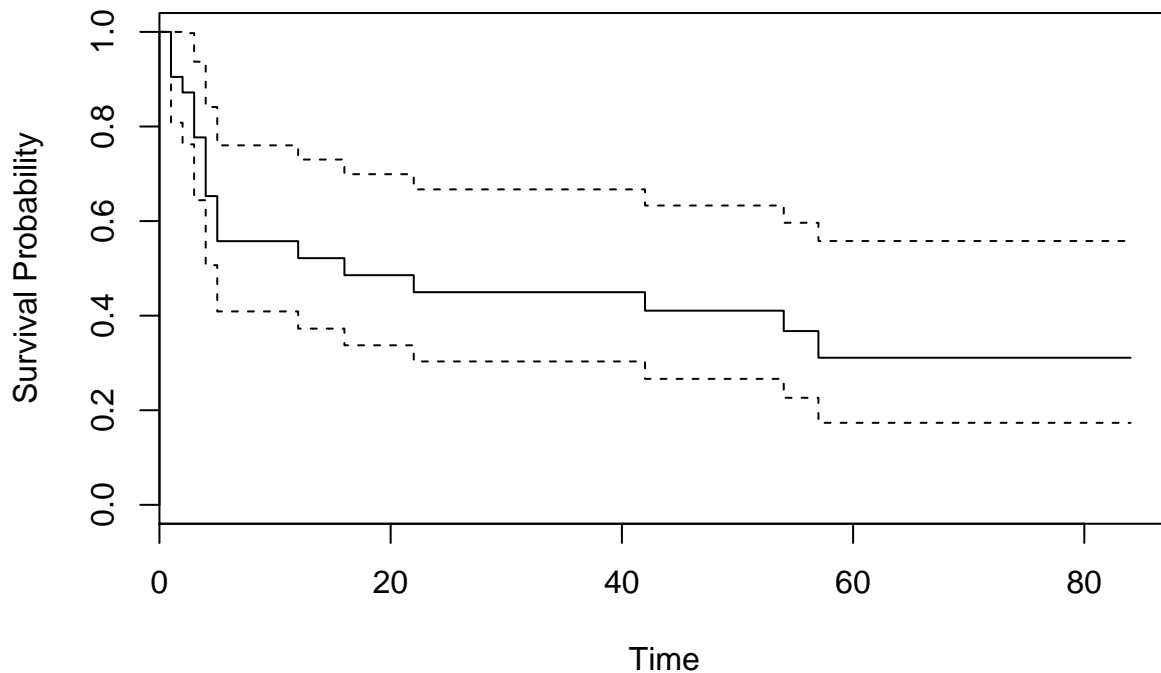
20, and Day 50?

We can also use the Nelson-Aalen (NA) estimator to compare survival and cumulative hazard functions between variables. .

```
model2 <- survfit(coxph(Surv(abstime,death)~1), type="aalen")

plot(model2, main="Nelson-Aalen estimate with 95% confidence bounds",
      xlab="Time", ylab="Survival Probability")
```

Nelson-Aalen estimate with 95% confidence bounds

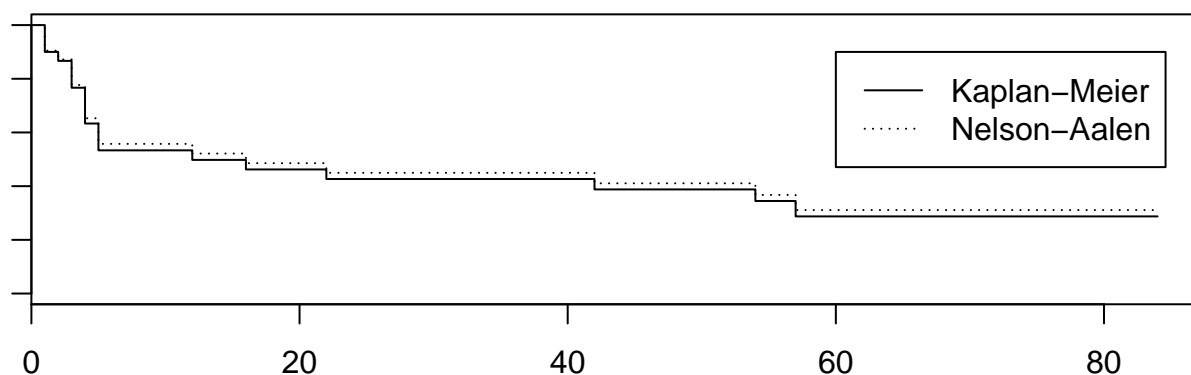


Next, we might want to superimpose both KM and NA on the same set of axes to illustrate their differences in terms of the survivor function. We can also do the same comparison using their respective cumulative hazard functions:

```
model3 <- survfit(Surv(abstime,death)~1)
par(mfcol=c(2,1), mar=c(1,1,2.7,1), oma=c(0,0,0,0))# Graphical parameters; 2 rows, 1 column
plot(model3, conf.int=F, main="Survivor function for KM and NA", xlab="Survival Time (Months)",
      ylab="Survival Probability")
lines(model2, conf.int=F, lty=3)
legend(60, .9, c("Kaplan-Meier", "Nelson-Aalen"), lty=c(1,3))

plot(model3, fun="cumhaz", conf.int=F, main="Cumulative hazard functions for KM and NA",
      xlab="Survival Time (Months)", ylab="Cumulative hazard")
lines(model2, fun="cumhaz", conf.int=F, lty=3)
legend(60, .6, c("Kaplan-Meier", "Nelson-Aalen"), lty=c(1,3))
```

Survivor function for KM and NA



Cumulative hazard functions for KM and NA



Although in both cases the estimators provide comparable results, the estimated difference between the two functions will increase as subjects die and the sample size becomes smaller with time. Although in this case it probably doesn't really matter which estimator is used, note that in data sets with very small sample sizes, we expect the KM and NA estimators to be increasingly different, with the NA survival estimate, and the KM cumulative hazard estimate, being most biased relative to their counterparts (Cleves et al., 2010).

Exercise 2: Nonparametric tests of nest survival in painted and snapping turtles.

Generalist predators may actively search for some prey while others are encountered unexpectedly (Schmidt, Goheen and Naumann, 2001; Bastille-Rousseau et al., 2011). These incidental predation events can have important consequences on prey populations, such as in freshwater turtles. In this exercise, we will look at nest predation by racoon on two species of freshwater turtle. We focus on the effect of landscape features on the rate of detection and consumption of eggs, including the level of predator traffic and distance to habitat edge. The data used in this exercise are modified from Wirsing et al. (2012). We will expand from survival functions seen in the previous exercise by using a simple semi-parametric cox regression, which we later expand to accommodate stratification, mixed-effects, and time-dependent covariates.

First, we open a csv file containing the data for the exercise and display them in R using the following R code:

```
# setwd("C:/...") # Specify working directory

nest<-read.csv("Surv_exerc2_3.csv", sep=",")
head(nest,5)
```

```
##   nestid sp year enter exit edgedist corridor predation
## 1   S12  0   1   160  161    5.71        0         1
## 2   S15  0   1   161  162    5.51        1         1
## 3   S47  0   1   166  167    4.50        1         1
```

```
## 4      S5  0    1   158 159      4.45      1      1
## 5     S179 0    2   160 270      5.71      0      0
```

```
str(nest) # Give the structure of an object
```

```
## 'data.frame':    292 obs. of  8 variables:
## $ nestid   : Factor w/ 292 levels "P1","P10","P11",...: 118 151 235 238 183 185 283 22 33 248 ...
## $ sp       : int  0 0 0 0 0 0 0 1 1 0 ...
## $ year     : int  1 1 1 1 2 2 2 1 1 1 ...
## $ enter    : int  160 161 166 158 160 160 159 157 183 169 ...
## $ exit     : int  161 162 167 159 270 240 160 234 293 170 ...
## $ edgedist : num  5.71 5.51 4.5 4.45 5.71 5.51 4.5 1.21 6.89 3.32 ...
## $ corridor : int  0 1 1 1 0 1 1 1 1 1 ...
## $ predation: int  1 1 1 1 0 0 1 0 0 1 ...
```

The table is comprised of a nest ID (**nestid**), turtle species (**sp**: 0=Snapping, 1=Painted), year (**year**: 0=2004, 1=2005), date of entry into the study (**enter**), date of exit from the study (**exit**), distance to edge habitat (**edgedist**), nest occurrence near a predator travel corridor (**corridor**: 0=off, 1=on) and nest predation event (**predation**: 0=no predation, 1=predated).

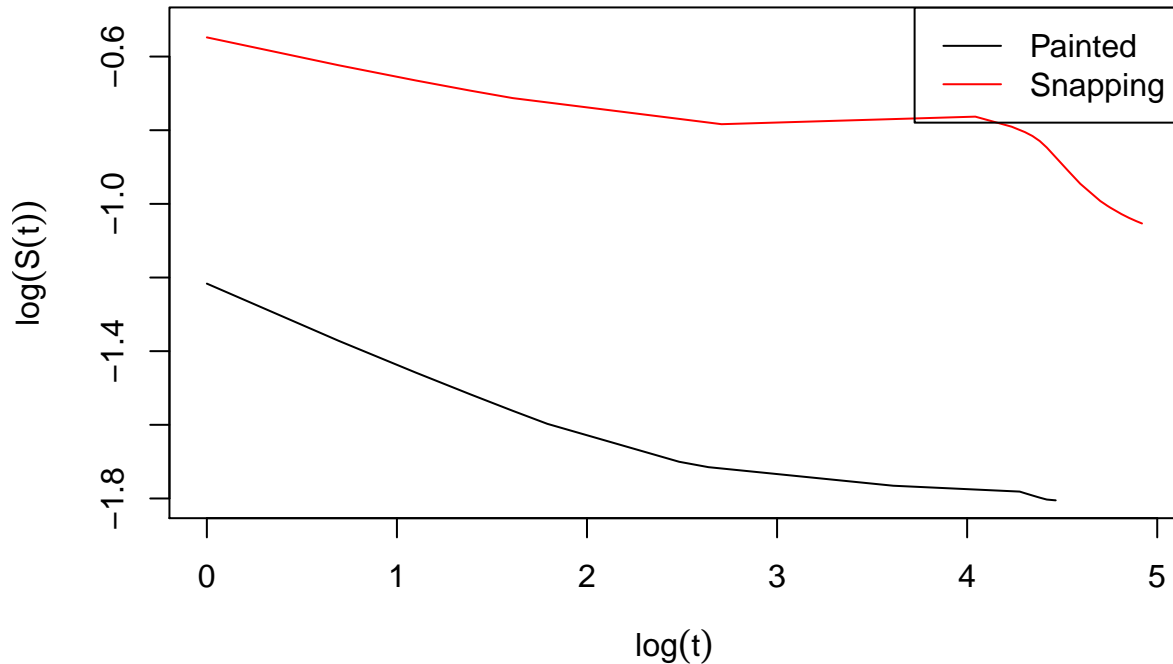
Question 2:

Using the previous exercise as an example, test if there are differences in mortality risk (from predation only) between species (binary variable: **sp**) and year (binary variable: **year**). Hint: Use absolute time.

In this second exercise, we first look at the influence of nest distance to an edge habitat (**edgedist**) and nest occurrence near a predator travel area (**corridor**) on nest probability of being predated (**predation**). Before going further and building on the results of Question 1, we must first determine whether hazards are proportional between species. We begin with a graphical test by plotting $\ln(S(t))$ against $\ln(\text{time})$ separately for both species (see Chapter 10, section 10.5.3). If the curves are parallel, we can surmise that the proportional hazards assumption is upheld. The following lines will produce the figure:

```
# install.packages("AICcmodavg") # Note: only need to run prior to first use
library(survival)
library(AICcmodavg) # Package for model selection
nest$abstime<-nest$exit-nest$enter
km_mod <- survfit(Surv(abstime, predation) ~ sp, data=nest, conf.type="none" )
sp<-rep(0:1, km_mod$strata)
plot(log(km_mod$time),log(km_mod$surv), type='n',
     ylab=expression(log(S(t))),
     xlab=expression(log(t)), main="Proportional hazard assumption")
lines(lowess(log((km_mod$surv[sp==0]))~log(km_mod$time[sp==0])), col=1)
lines(lowess(log((km_mod$surv[sp==1]))~log(km_mod$time[sp==1])), col=2)
legend('topright', c("Painted", "Snapping"), col=1:2, lty=1)
```

Proportional hazard assumption



Even if the curves seem parallel throughout most of the study time, they do diverge at the end, which may indicate that the proportional hazards assumption is not upheld between species. There are several ways of dealing with deviation from the PH assumption, but for now, we can circumvent this issue by first running our analysis with the snapping turtle data alone ($sp = 0$). We will use a nonparametric Cox Proportional Hazards model within an information theoretic model selection framework (see Chapter 3). Our candidate models will be comprised of two univariate models with either edgedist or corridor as an independent variable and a third model including these two variables together. The following script will construct a table calculating AIC and corresponding weights for these three candidate models:

```
snapping<-nest[nest$sp==0,] # select snapping turtle data only
Cand.mod <- list() # set up candidate model list

Cand.mod[[1]] <-coxph(Surv(abstime, predation)~edgedist, data=snapping)
Cand.mod[[2]] <-coxph(Surv(abstime, predation)~corridor, data=snapping)
Cand.mod[[3]] <-coxph(Surv(abstime, predation)~edgedist+corridor, data=snapping) # input models into list
Modnames <- c("EdgeDist", "Corridor", "EdgeDist+Corridor")#set up model names

aictab(cand.set = Cand.mod, modnames = Modnames, sort = TRUE)

##
## Model selection based on AICc:
##
##           K      AICc Delta_AICc AICcWt Cum.Wt      LL
## EdgeDist      1 1529.31      0.00  0.57  0.57 -763.64
## EdgeDist+Corridor 2 1529.93      0.63  0.41  0.98 -762.94
## Corridor      1 1536.19      6.88  0.02  1.00 -767.08
# output summary table of models
```

This model selection exercise shows that the univariate edgedist model is the most strongly supported (the AICcWt column shows the relative weight of each model) followed closely by the edgedist+corridor model.

We can display coefficients of these two models and hazard ratios which are simply the exponent of our coefficient (`exp(coef)`):

```
Cand.mod[[3]]
```

```
## Call:
## coxph(formula = Surv(abstime, predation) ~ edgedist + corridor,
##       data = snapping)
##
##              coef exp(coef) se(coef)      z      p
## edgedist 0.02044   1.02065  0.00694  2.94 0.0032
## corridor 0.25102   1.28533  0.21485  1.17 0.2427
##
## Likelihood ratio test=8.61  on 2 df, p=0.01
## n= 198, number of events= 166
```

```
Cand.mod[[1]]
```

```
## Call:
## coxph(formula = Surv(abstime, predation) ~ edgedist, data = snapping)
##
##              coef exp(coef) se(coef)      z      p
## edgedist 0.0153    1.0154   0.0054  2.84 0.0045
##
## Likelihood ratio test=7.19  on 1 df, p=0.007
## n= 198, number of events= 166
```

Before going further into the interpretation of our results, it is crucial to test for the proportional hazards assumption of this particular model. Because we are testing the effects of a continuous variable (`edgedist`), we cannot test the assumption using the simple graphical test described at the beginning of this exercise. Instead, we use `cox.zph`, a function that determines the correlation coefficient between transformed survival time and the scaled Schoenfeld residuals. We can test the assumption on the two top-models:

```
cox.zph(Cand.mod[[3]])
```

```
##              rho chisq      p
## edgedist -0.0116 0.0207 0.886
## corridor -0.0637 0.7658 0.382
## GLOBAL      NA 1.0495 0.592
```

```
cox.zph(Cand.mod[[1]])
```

```
##              rho chisq      p
## edgedist 0.0475 0.299 0.585
```

Both models satisfy the assumption, as the test result is not significant. As you can see, when a multivariate Cox model is evaluated for hazard proportionality, `cox.zph` performs the test on each variable as well as on the global model.

Now that we are confident that our models are robust, we can use model averaging in order to interpret the results of the two top-models.

```
modavg(cand.set = Cand.mod[c(1,3)], parm = "edgedist", modnames = c("EdgeDist", "EdgeDist+Corridor"))
```

```
##
## Multimodel inference on "edgedist" based on AICc
##
## AICc table used to obtain model-averaged estimate:
##
```

```
##           K      AICc Delta_AICc AICcWt Estimate  SE
## EdgeDist      1 1529.31      0.00   0.58    0.02 0.01
## EdgeDist+Corridor 2 1529.93      0.63   0.42    0.02 0.01
##
## Model-averaged estimate: 0.02
## Unconditional SE: 0.01
## 95% Unconditional confidence interval: 0, 0.03
```

```
#Perform model-averaging
```

```
modavg(cand.set = Cand.mod[c(1,3)], parm = "corridor", modnames = c("EdgeDist", "EdgeDist+Corridor"))
```

```
##
## Multimodel inference on "corridor" based on AICc
##
## AICc table used to obtain model-averaged estimate:
##
##           K      AICc Delta_AICc AICcWt Estimate  SE
## EdgeDist+Corridor 2 1529.93      0      1    0.25 0.21
##
## Model-averaged estimate: 0.25
## Unconditional SE: 0.21
## 95% Unconditional confidence interval: -0.17, 0.67
```

Model averaging indicates a small significant effect of edgeDist, but not of corridor. Based on the model-averaged coefficients, predation risk increased by 20% ($|1-\exp(0.02)| \times 10 \times 100$) for every 10 m increase from distance to an edge habitat.

Let's return to the full dataset and try to satisfy the cox proportional hazards assumption again. Recall that in the formulation of the Cox proportion hazard model, h_0 is the baseline unspecified hazard function. In order to meet the assumption of proportional hazards, we may want to allow the model to estimate different baseline functions for different groups using stratification. To simplify, we will only consider the global model including corridor and edgedist.

```
strat.mod1 <-coxph(Surv(abstime, predation)~edgedist+corridor+strata(sp), data=nest)
cox.zph(strat.mod1)
```

```
##           rho  chisq    p
## edgedist -0.0208 0.0907 0.763
## corridor -0.0973 2.0585 0.151
## GLOBAL      NA 2.5105 0.285
```

```
strat.mod1
```

```
## Call:
## coxph(formula = Surv(abstime, predation) ~ edgedist + corridor +
##       strata(sp), data = nest)
##
##           coef exp(coef) se(coef)      z      p
## edgedist  0.01583   1.01596  0.00564   2.80 0.005
## corridor -0.08530   0.91823  0.17782  -0.48 0.631
##
## Likelihood ratio test=13.36 on 2 df, p=0.001
## n= 292, number of events= 220
```

The stratified model satisfies the proportional hazards assumption. However, had we run the proportional hazards test separately for each species, the outcome would have differed owing to the role of **edgedist** and **corridor** on nest survival being different between species.

Question 3

*How are the results different across species? Using only the painted turtle dataset, conduct a similar cox proportional hazards analysis and model selection for the influence of corridor and **edgedist**. Compare results with the snapping turtle models and the stratified models including both species.*

Even though our preliminary analysis (Question 1) seemed to indicate little effect of year on nest predation risk, it may be more appropriate to isolate observations taken within a year and therefore consider year as a random factor with its own variance term in our Cox proportional hazards analysis. From a biological perspective this makes sense because we expect that different environmental conditions or predator densities between years could fundamentally affect nest survival, but this extra variation is neither a focal point of our study nor a factor that is associated with a specific prediction. Recent developments in Cox proportional hazards modeling now allow us to build random effects models (also termed frailty models) of similar complexity to what is available for generalized linear models, including the presence of both random intercept and slopes. For further details on mixed-effects modelling see Pinheiro & Bates (2000) and Bolker and colleagues (2009).

Note that in a cox model, the random intercept is not of specific interest since it refers to the unspecified baseline hazard ($h_0(t)$). As an example, we first perform a simple mixed-effects model that considers that the effect of distance to edge can vary across year for the snapping turtle data. To do so, we use the function 'coxme' in the coxme package. The syntax of the random effect is similar to the popular lme4 package.

```
# install.packages("coxme") #Note: only need to run prior to first use
library(coxme)

## Loading required package: bdsmatrix
##
## Attaching package: 'bdsmatrix'
## The following object is masked from 'package:base':
##
##      backsolve

mix.mod1 <-coxme(Surv(abstime, predation)~edgedist+(edgedist|year), data=snapping)
mix.mod1

## Cox mixed-effects model fit by maximum likelihood
##   Data: snapping
##   events, n = 166, 198
##   Iterations= 8 42
##
##              NULL Integrated      Fitted
## Log-likelihood -767.2396   -763.6453 -763.6432
##
##              Chisq df          p  AIC   BIC
## Integrated loglik  7.19  2 0.0274800 3.19 -3.04
## Penalized loglik  7.19  1 0.0073527 5.19  2.07
##
## Model:  Surv(abstime, predation) ~ edgedist + (edgedist | year)
## Fixed coefficients
##              coef exp(coef)    se(coef)      z      p
## edgedist 0.0153291  1.015447 0.005400985  2.84 0.0045
##
## Random effects
## Group Variable Std Dev      Variance
## year edgedist 3.081843e-04 9.497756e-08
```

We can compare the coefficient of this model with the previous Cox model including **edgedist** only (**Cand.mod[[1]]**). The two approaches provide similar results. We can also assess between-year variance of the random effects from the previous output, which is low (9.5e-08) and indicates negligible variability

between both years in terms of the effect of **edgedist**. Similarly, it is possible to extract the random coefficients associated with each year with the command `ranef`:

```
ranef(mix.mod1)

## $year
##      edgedist
## 1 -1.198659e-06
## 2  1.198659e-06
```

and the estimate influence of **edgedist** for each year can be reconstructed by adding the fixed effect of **edgedist** with the random coefficients:

```
ranef(mix.mod1)$year+fixef(mix.mod1)

##      edgedist
## 1 0.0153279
## 2 0.0153303
```

This also reveals the low inter-annual variability in the role of **edgedist**.

Mixed-effects modelling can become handy when dealing with multiple groups or treatments where units within treatments are not independent or can be expected to respond similarly. We started question 2 by building separate models for each species testing the effect of **edgedist** and **corridor** on mortality risk. These models only contained a fixed-effect term and therefore provide what is generally referred to as a marginal inference. In some instances, however, when dealing with numerous groups, separating into multiple level-specific analyses may not be an interesting option especially if researchers are interested in both the average between-group response and also group-specific response (referred to as conditional inference). In our case, a slightly more relevant exercise that uses mixed-effects modeling could involve building a single model with species as a random factor that considers each species as a separate group and that allows the model to estimate both an average response (marginal inference) and species-specific responses for a given covariate (conditional inference).

Question 4:

*Based on the results of question 2, it seems like the effect of distance to edge is similar between species but that the effect of distance to corridor is different. Using the full turtle dataset, perform a mixed-effects model with species (*sp*) as a random factor that assesses the effect of **edgedist** and **corridor** as fixed effects and **corridor** as a random coefficient. Extract the random coefficients and compare them with the previous models (Hint: write the structure of the random effect as follows: (*corridor* | *sp*), to compare the random effects model with the species specific model, then sum the fixed effect with the random effects).*

Finally, it is also possible to integrate time-varying, time dependent or latent variables into the cox proportional hazards model (Kleinbaum and Klein, 2011). Time-varying or latent variables allow the Cox model to consider that the effects of predictors vary through time, and these can be intrinsic or extrinsic to the subject. To incorporate time dependency into the turtle nest survival analysis, we use an extended version of the turtle dataset where daily rainfall (*rain*) and daily mean temperature (*temp*) are recorded. Perhaps we suspect that raccoon predators are more (or less) active according to environmental conditions, so it is conceivable that predation risk varies on a day-by-day basis according to these conditions. We can open the new dataset to show that it is identical to the previous dataset except that survival and environmental conditions are now updated on a daily basis. To compare both datasets for one individual nest (P2) we use the following code:

```
expand<-read.csv("Surv_exerc2_expand.csv", sep=",")

#Example of data
expand[expand$nestid=="P2",]

##      nestid sp year enter exit rain temp predation
```

```
## 81      P2  1    1   165  166  3.1 22.5      0
## 82      P2  1    1   166  167  0.0 22.5      0
## 83      P2  1    1   167  168  4.3 21.3      0
## 84      P2  1    1   168  169  4.4 23.0      0
## 85      P2  1    1   169  170  0.0 23.0      1
```

```
nest[nest$nestid=="P2",]
```

```
##      nestid sp year enter exit edgedist corridor predation abstime
## 142      P2  1    1   165  170   20.87      0      1      5
```

As you can see, the extended dataset now has a line for every day that the turtle was alive that shows the amount of daily precipitation and average temperature. Each line also shows the fate of the individual at the end of that day.

We can merge this dataset with the previous one in order to add the variables edgedist and corridor as well as the column indicating the time of entry in the study (enter). Note that edgedist and corridor are not time-dependent variables so they are constant across all entries per nest. We then use a subset of the data (snapping turtle) and perform a Cox regression to test the effect of rain, temperature, corridor and edgedist on survival:

```
expand2<-merge(expand, nest[,c("nestid", "edgedist", "corridor", "enter")], by="nestid")
snapping2<-expand2[expand2$sp==0,]
```

```
cox_mod<-coxph(Surv(enter.x, exit, predation) ~ edgedist+corridor+rain+temp, data=snapping2)
cox_mod
```

```
## Call:
## coxph(formula = Surv(enter.x, exit, predation) ~ edgedist + corridor +
##      rain + temp, data = snapping2)
##
##              coef exp(coef) se(coef)      z      p
## edgedist 0.03928   1.04006  0.00707  5.56 2.8e-08
## corridor 0.47480   1.60769  0.22256  2.13  0.033
## rain     0.00912   1.00916  0.03306  0.28  0.783
## temp     0.01679   1.01693  0.02754  0.61  0.542
##
## Likelihood ratio test=29.82 on 4 df, p=5e-06
## n= 3131, number of events= 166
```

```
cox.zph(cox_mod)
```

```
##              rho chisq      p
## edgedist  0.0622 0.595 0.441
## corridor  0.0608 0.724 0.395
## rain      -0.0258 0.128 0.720
## temp      -0.0614 0.735 0.391
## GLOBAL           NA 2.130 0.712
```

In this example, the “Surv” argument has been modified to include the entry and exit time variables, rather than the absolute time, which we used previously. This modification allows us to tally survival over single-day intervals. This model indicates that all variables increase the probability of being killed by a predator, but only edgedist and corridor do so significantly. The model also satisfies the proportional hazards assumption.

Using an expanded or simplified version of a dataset in a Cox proportional hazards model should not influence the results. For example, we could perform a simpler model for snapping turtles including only edgedist and corridor in the expanded dataset and compare the results to a similar model using the non-expanded snapping turtle dataset:

```
cox_mod1 <-coxph(Surv(enter, exit, predation)~edgedist+corridor, data=snapping) # Not time varying
cox_mod2<- coxph(Surv(enter.x, exit, predation)~edgedist+corridor, data=snapping2)
cox_mod1
```

```
## Call:
## coxph(formula = Surv(enter, exit, predation) ~ edgedist + corridor,
##       data = snapping)
##
##               coef exp(coef) se(coef)      z      p
## edgedist 0.0397    1.0405   0.0071 5.60 2.2e-08
## corridor 0.4841    1.6227   0.2227 2.17  0.03
##
## Likelihood ratio test=29.04 on 2 df, p=5e-07
## n= 198, number of events= 166
```

```
cox_mod2
```

```
## Call:
## coxph(formula = Surv(enter.x, exit, predation) ~ edgedist + corridor,
##       data = snapping2)
##
##               coef exp(coef) se(coef)      z      p
## edgedist 0.0397    1.0405   0.0071 5.60 2.2e-08
## corridor 0.4841    1.6227   0.2227 2.17  0.03
##
## Likelihood ratio test=29.04 on 2 df, p=5e-07
## n= 3131, number of events= 166
```

Reassuringly, both outputs provide the same results. However, you may have noticed that the output of `cox_mod1` differs from that obtained using the original non-expanded dataset. This difference arises because we used absolute time in the beginning of the exercise but our current model uses calendar time. In order to obtain exactly the same results that we generated previously, we need to convert calendar time to absolute time (i.e. each individual starts at 0). This can be done as follows:

```
snapping2$enter_abs<-snapping2$enter.x-snapping2$enter.y
snapping2$exit_abs<-snapping2$exit-snapping2$enter.y
cox_mod3<- coxph(Surv(enter_abs, exit_abs, predation)~edgedist+corridor, data=snapping2)
cox_mod4<- coxph(Surv(abstime, predation)~edgedist+corridor, data=snapping)
cox_mod3
```

```
## Call:
## coxph(formula = Surv(enter_abs, exit_abs, predation) ~ edgedist +
##       corridor, data = snapping2)
##
##               coef exp(coef) se(coef)      z      p
## edgedist 0.02044    1.02065   0.00694 2.94 0.0032
## corridor 0.25102    1.28533   0.21485 1.17 0.2427
##
## Likelihood ratio test=8.61 on 2 df, p=0.01
## n= 3131, number of events= 166
```

```
cox_mod4
```

```
## Call:
## coxph(formula = Surv(abstime, predation) ~ edgedist + corridor,
##       data = snapping)
##
##
```

```
##           coef exp(coef) se(coef)      z      p
## edgedist 0.02044   1.02065  0.00694  2.94 0.0032
## corridor 0.25102   1.28533  0.21485  1.17 0.2427
##
## Likelihood ratio test=8.61  on 2 df, p=0.01
## n= 198, number of events= 166
```

Question 5:

Using the painted turtle dataset, perform a time-dependent cox proportional hazards regression testing the effect of rain, temperature, corridor and edgedist first using calendar time and then using absolute time. Compare the output for each variable between the two approaches.

Exercise 3: Parametric tests of nest survival in freshwater turtles.

Sometimes the underlying survival and hazard and survival functions are known and it is then possible to test hypotheses that are specific to these distributions or to produce models with higher predictive power than those from standard semi-parametric approaches. Exponential and Weibull distributions are commonly used in such instances and probably cover most of the parametric survival needs of ecologists, but other distributions such as logistic and log-normal are also available. It can be informative as a first step to compare survival or hazard curves fitted from a parametric regression against those obtained from a Cox proportional hazards model, which in most cases should be the analysis of choice in ecology. Such a comparison can indicate the suitability of a parametric distribution for a given dataset. We are able to use parametric methods using the turtle dataset specifically because the age (date of deposition) of nests is known, which allows us to establish the functional form of the survival estimate through the lifespan of a nest. In contrast, in most ecological applications the age of the subject is not known, which limits our understanding of the form of the survival timeline and thereby precludes a parametric approach.

For this exercise, we rely on another package (**eha**) in order to perform parametric analysis. While the package **survival** also has a function for parametric survival analysis (**survreg**), the package **eha** offers greater flexibility in model development and structure. We first fit regressions to the turtle nest dataset using Exponential and Weibull distributions (recall that the Exponential distribution assumes that hazards are constant, the Weibull distribution allows hazards to monotonically increase or decrease, see Chapter 11, section 11.2). We generate parametric hazard curves comparing each parametric distribution with the Cox model for the turtle nest survival dataset, where both species are initially pooled for simplicity and using a specific function in **eha** (**check.dist**).

```
# setwd("...")
# install.packages("eha") # Note: only need to run prior to first use

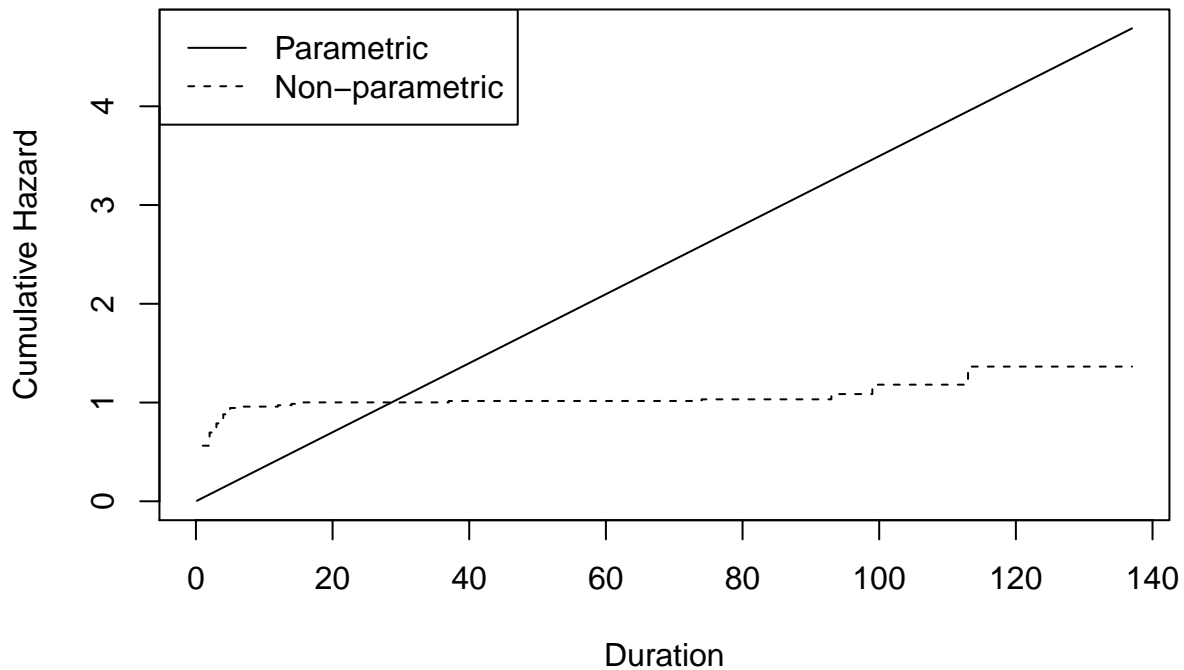
nest<-read.csv("Surv_exerc2_3.csv", sep=",") # Open csv file
library(survival)
library(eha)
nest$abstime<-nest$exit-nest$enter

exp_mod<- phreg(Surv(abstime, predation) ~ sp, dist="weibull", shape=1, data=nest)
wb_mod<- phreg(Surv(abstime, predation) ~ sp, dist="weibull", data=nest)
cox_mod<-coxreg(Surv(abstime, predation) ~ sp, data=nest)

### Calculating prediction of Cox, Exponential and Weibull models

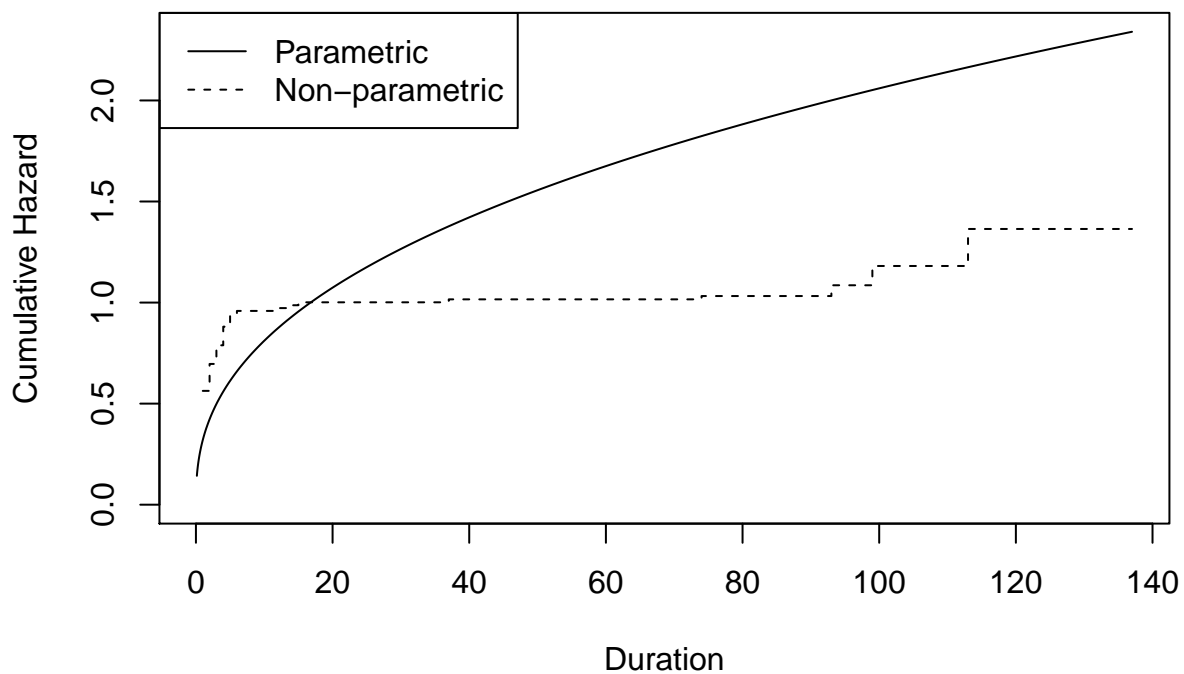
check.dist(cox_mod, exp_mod, main="Exponential")
```

Exponential



```
check.dist(cox_mod, wb_mod, main="Weibull")
```

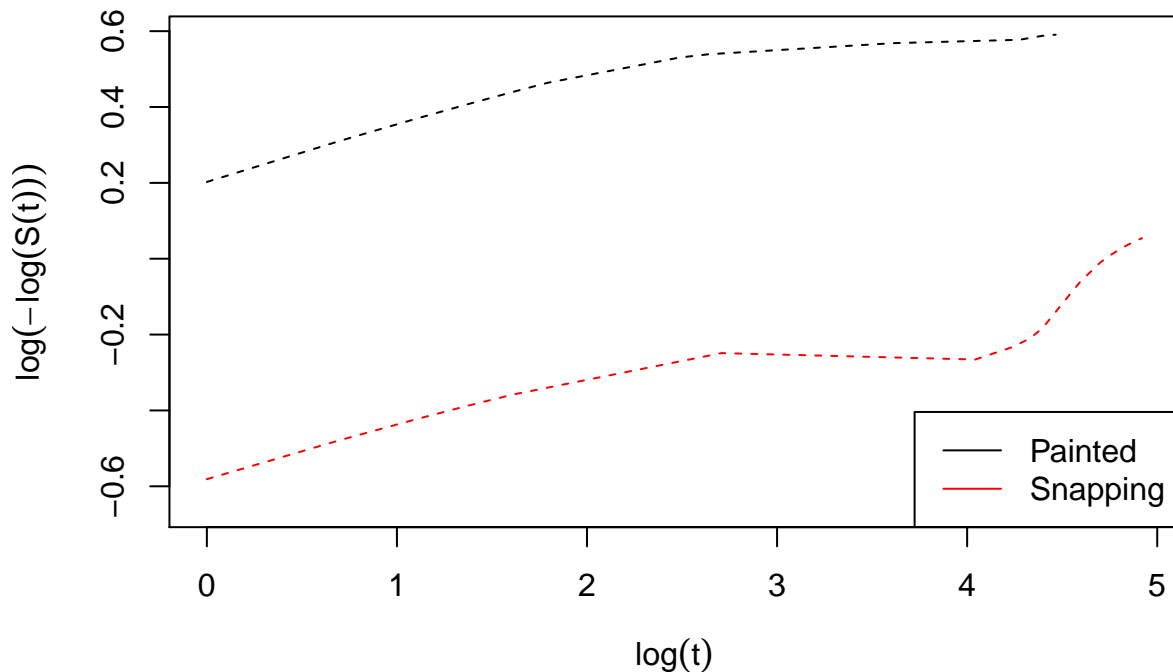
Weibull



As illustrated by the figure, both parametric models perform poorly in reproducing the cumulative hazard with species pooled. The poor fit of the Weibull can be partly explained by the fact that the regression only estimated a single shape parameter for both species of turtle and we could have improved fit by stratifying the model according to species.

Another graphical test of the appropriateness of the Weibull distribution, as well as the proportional hazards assumption (as we saw above), is to plot the $\log(-\log(S(t)))$ versus $\log t$ separately for each factor. The main expectation from this figure is that straight but not necessarily parallel lines support the Weibull distribution whereas parallel curves support the proportional hazards assumption (Kleinbaum and Klein, 2011).

```
km_mod <- survfit(Surv(abstime, predation) ~ sp, data=nest, conf.type="none" )
sp<-rep(0:1, km_mod$strata)
plot(log(km_mod$time), log(-log(km_mod$surv)), type='n', xlab=expression(log(t)),
      ylab=expression(log(-log(S(t)))))
lines(lowess(log(-log(km_mod$surv[sp==0])) ~ log(km_mod$time)[sp==0]), lty=2)
lines(lowess(log(-log(km_mod$surv[sp==1])) ~ log(km_mod$time)[sp==1]), lty=2, col=2)
legend('bottomright', c("Painted", "Snapping"), col=1:2, lty=1)
```



From the figure we see that the lines are neither parallel nor straight, which means that the Weibull model is inappropriate, and that the proportional hazards assumption is likely not satisfied by the data. Recall that we observed a similar pattern in Exercise 2.

For simplicity, we will continue this exercise by focusing on the snapping turtle data only. First, we generate a figure comparing exponential and Weibull regressions to the cumulative hazard curve obtained from Kaplan-Meier, and we will then use AIC to assess which distribution best fits the data:

```
snapping<-nest[nest$sp==0,]
exp_mod2<- phreg(Surv(abstime, predation) ~ 1, dist="weibull", shape=1, data=snapping)
wb_mod2<- phreg(Surv(abstime, predation) ~ 1, dist="weibull", data=snapping)
cox_mod2<-coxreg(Surv(abstime, predation) ~ 1, data=snapping)
```

```
extractAIC(wb_mod2)
```

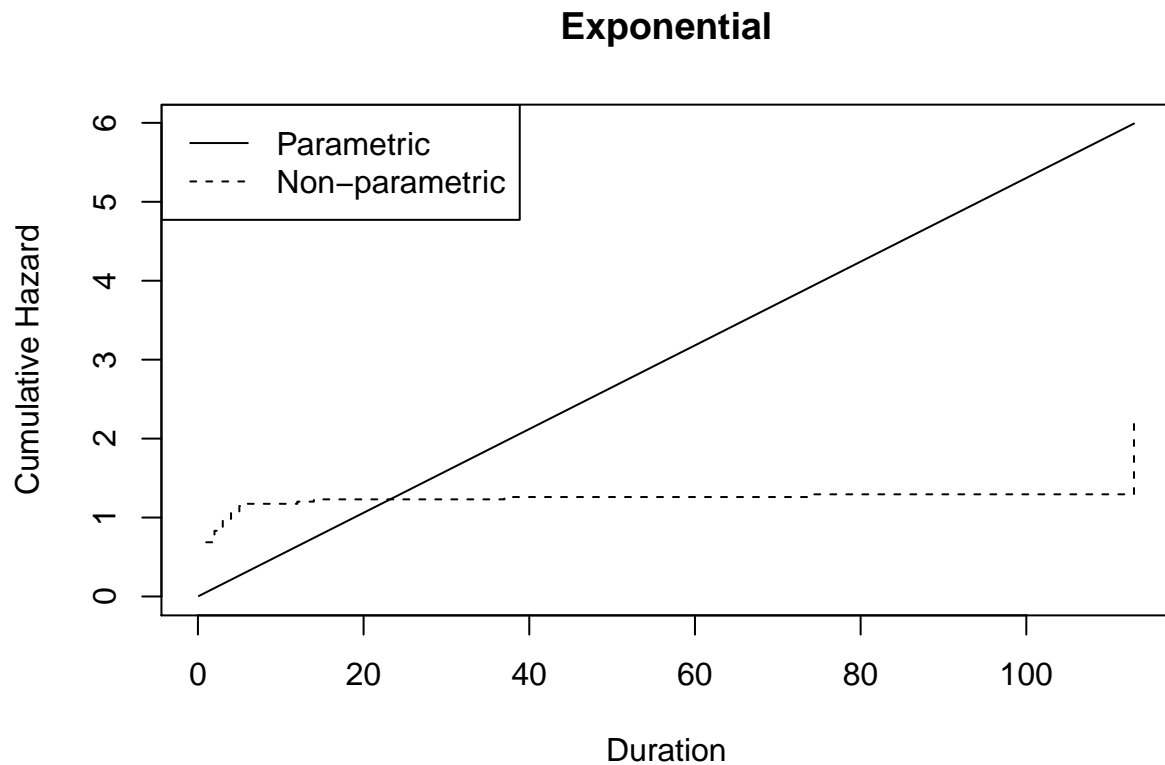
```
## [1] 0.0000 965.6997
```

```
extractAIC(exp_mod2)
```

```
## [1] 0.000 1307.124
```

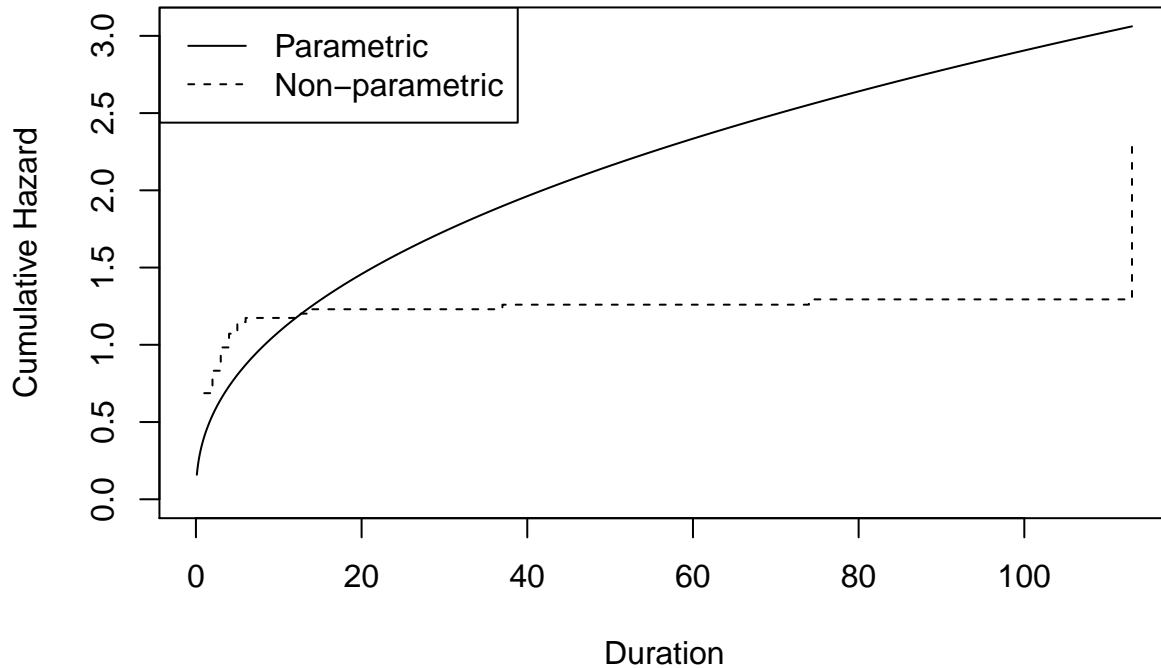
```
summary(wb_mod2)
```

```
## Call:
## phreg(formula = Surv(abstime, predation) ~ 1, data = snapping,
##       dist = "weibull")
##
## Covariate      W.mean      Coef Exp(Coef)  se(Coef)   Wald p
## log(scale)                2.115      0.183    0.000
## log(shape)             -0.848      0.058    0.000
##
## Events                166
## Total time at risk      3131
## Max. log. likelihood   -482.85
check.dist(cox_mod2, exp_mod2, main="Exponential")
```



```
check.dist(cox_mod2, wb_mod2, main="Weibull")
```

Weibull



The AIC values indicate that the Weibull distribution is slightly better than the Exponential fit, which is corroborated by the figure. This makes intuitive sense because we know that nests experience higher hazard early on and thus that the exponential function is unlikely to offer a good fit to the data because it assumes constant hazard. The significance of the $\log(\text{shape})$ parameter in the output from the Weibull regression also reveals that the Weibull model is a better fit to the data. We will proceed with the Weibull model to analyse the influence of edgedist and corridor survival probability for snapping turtle nests. We limit our analysis to the global model including both variables, but model selection could also be performed using parametric methods:

```
wb_mod3<- phreg(Surv(abstime, predation) ~ edgedist+corridor, dist="weibull", data=snapping)
cox_mod3 <-coxreg(Surv(abstime, predation)~edgedist+corridor, data=snapping)
summary(wb_mod3)
```

```
## Call:
## phreg(formula = Surv(abstime, predation) ~ edgedist + corridor,
##       data = snapping, dist = "weibull")
##
## Covariate      W.mean      Coef Exp(Coef)  se(Coef)    Wald p
## edgedist        5.011      0.030    1.031    0.007    0.000
## corridor        0.720      0.291    1.337    0.212    0.171
##
## log(scale)              3.198              0.471    0.000
## log(shape)             -0.792              0.058    0.000
##
## Events              166
## Total time at risk    3131
## Max. log. likelihood -473.46
## LR test statistic     18.78
## Degrees of freedom    2
## Overall p-value       8.35917e-05
```



```
summary(cox_mod3)
```

```
## Call:
## coxreg(formula = Surv(abstime, predation) ~ edgedist + corridor,
##       data = snapping)
##
## Covariate           Mean      Coef    Rel.Risk   S.E.    Wald p
## edgedist           5.011     0.020     1.021    0.007    0.003
## corridor           0.720     0.251     1.285    0.215    0.243
##
## Events                166
## Total time at risk      3131
## Max. log. likelihood   -762.94
## LR test statistic       8.61
## Degrees of freedom      2
## Overall p-value        0.0135101
```

Both approaches provide similar qualitative and quantitative results, indicating that higher values for edgedist increase the risk for snapping turtle.

Question 6:

Using the painted turtle dataset alone, try performing a Weibull regression looking at the effect of corridor on the probability of survival. First, assess if the Weibull is better than the simpler Exponential model. Compare these results with those from the semi-parametric approach developed in Question 2.

Exercise 4: Competing risks analysis for grey wolves.

In this exercise, we use a subset of the data used in Murray et al. (2010) to develop competing risks models analyzing mortality patterns of radio-collared wolves from the Greater Yellowstone Ecosystem area. Our previous survival analyses dealt with a single cause of death, but increasingly ecologists are interested in how individuals incur different risks depending on their exposure to certain mortality agents. We will compare mortality due to anthropogenic sources (legal control, illegal killing, harvest in Canada, vehicle collision) to natural causes (strife, disease, senescence) and look at characteristics of individual wolves (age class, sex, dispersal status) on the probability of succumbing to either anthropogenic or natural cause of death. We start by importing and looking at the data for individual **id** #20:

```
#setwd("...") # Specify working directory
library(survival)
wolf<-read.csv("Surv_exerc4.csv")

#Example of data
wolf[wolf$id==20,]
```

```
##      id sday eday natu anthro sex adult yearl pup dispers reason
## 238 20 6181 6187    0     0   1     0     0   1     0       1
## 239 20 6187 6278    0     0   1     0     1   0     0       1
## 240 20 6278 6370    0     0   1     0     1   0     1       1
## 241 20 6370 6462    0     0   1     0     1   0     1       1
## 242 20 6462 6552    0     0   1     0     1   0     0       1
## 243 20 6633 6715    0     1   1     1     0   0     0       1
```

You can see from the output that the dataset is an extended version with multiple entries for a given individual, allowing for time-varying predictors. **Sday** and **eday** represent the start and exit of the time

interval, respectively, with intervals being defined as 80-90-day biological seasons. Natural or anthropogenic causes of death are indicated by dummy variables **natu** and **anthro**, respectively, and mortality due to either cause is specified within the appropriate interval, as we saw with the extended turtle nest survival dataset (see also Chapter 11, section 11.3.4). Individual characteristics are given by sex (0=Female, 1=Male), dummy codes representing age; **adult**, **yearl**, **pup** (equal to 1 when individual is in that category) and **dispers** (0=resident, 1=disperser). As is evident from the printed dataset, intervals are not always consecutive and the age class and dispersal status of an individual is time-dependent and changes between intervals. Finally, the reason variable indicates why the animal was first recruited to the study (0=targeted and 1=representative), with the targeted group consisting a biased sample of animals that were radio-collared after they had depredated livestock. For this analysis, targeted animals are excluded from the analysis.

```
wolf<-wolf[wolf$reason==1,]
```

Competing-risks survival analysis can be easily implemented using semi-parametric (e.g., Murray et al., 2010) or parametric (Griffin et al., 2011) approaches. The easiest way to implement competing risks analysis in R is through a data augmentation method (Lunn and McNeil, 1995). This approach can be applied using the functions mentioned in the previous exercises, making competing-risks analysis a logical extension of our previous regression models that now include multiple fates.

To prepare a survival dataset for competing risks analysis, we need to duplicate the dataset for each possible cause of mortality and create a new variable to distinguish between causes of death (**ft**: failure type). Therefore, we have one dataset for each cause of mortality. Subsequently, a new variable (**fail**) needs to be created which assumes the value '1' when a wolf dies of the cause in question and '0' otherwise. Therefore, **fail** assumes the value 0 in all entries that have not failed from the cause (i.e., entries that are censored). The following lines will do this with our wolf dataset:

```
wolf1<-data.frame(wolf,ft=1) #ft1=Natural
wolf2<-data.frame(wolf,ft=2) # ft2=Anthro
expanded<-rbind(wolf1, wolf2)
expanded$event<-1*ifelse(expanded$natu==1,1,0)+2*ifelse(expanded$anthro==1,1,0)
expanded$fail<-ifelse(expanded$event==expanded$ft,1,0)
```

It may be easier to understand the data augmentation process by looking at the output of a few individuals:

```
wolf[wolf$id==1 | wolf$id==7 ,
c("id", "sday", "eday", "natu", "anthro")]
```

```
##      id sday eday natu anthro
## 1      1 5081 5091    0      0
## 2      1 5116 5141    0      1
## 99     7 5977 6006    0      0
## 100    7 6006 6098    0      0
## 101    7 6098 6188    0      0
## 102    7 6262 6338    1      0
```

```
expanded[expanded$id==1 | expanded$id==7 ,
c("id", "sday", "eday", "natu", "anthro", "ft", "event", "fail")]
```

```
##      id sday eday natu anthro ft event fail
## 1      1 5081 5091    0      0  1      0    0
## 2      1 5116 5141    0      1  1      2    0
## 99     7 5977 6006    0      0  1      0    0
## 100    7 6006 6098    0      0  1      0    0
## 101    7 6098 6188    0      0  1      0    0
## 102    7 6262 6338    1      0  1      1    1
## 1114   1 5081 5091    0      0  2      0    0
## 2143   1 5116 5141    0      1  2      2    1
```

```
## 9910    7 5977 6006    0    0 2    0    0
## 10010   7 6006 6098    0    0 2    0    0
## 10110   7 6098 6188    0    0 2    0    0
## 10210   7 6262 6338    1    0 2    1    0
```

Recall that `ft` indicates the causes of death and `fail` indicates whether animals died of the cause of not. Following the data augmentation approach, the survival analysis can be conducted in two ways: one is to stratify by causes of death (e.g., Murray et al., 2010; Melis et al., 2012), the other is to include causes of death as a covariate. The latter method offers some advantages since causes of mortality can be compared directly, but it assumes proportionality in baseline hazards for the different causes. In both approaches, interaction terms between the risk types and predictors can be used to provide insight as to whether a cause of death is additive or compensatory to others (Heisey and Patterson, 2006).

The following code will perform a cox proportional regression looking at the effect of wolf sex, adult status and dispersal status on each case of death. We conduct the analysis using both the stratification and covariate methods, and determine if they satisfy the cox proportional hazards assumptions.

```
comprisk1<-coxph(Surv(sday,eday, fail)~(sex+adult+dispers)*I(ft==2)+ strata(ft), data= expanded)
comprisk2<-coxph(Surv(sday,eday, fail)~(sex+adult+dispers)*I(ft==2), data= expanded)
```

```
cox.zph(comprisk1)
```

```
## Warning in cor(xx, r2): the standard deviation is zero
```

```
##              rho  chisq    p
## sex          -0.0777 0.6674 0.414
## adult         0.0176 0.0360 0.850
## dispers       0.0373 0.1558 0.693
## I(ft == 2)TRUE      NA    NaN   NaN
## sex:I(ft == 2)TRUE  -0.0231 0.0586 0.809
## adult:I(ft == 2)TRUE -0.0245 0.0663 0.797
## dispers:I(ft == 2)TRUE -0.0932 0.9636 0.326
## GLOBAL              NA 4.0098 0.779
```

```
cox.zph(comprisk2)
```

```
##              rho  chisq    p
## sex          -0.06543 0.469276 0.493
## adult         0.00194 0.000423 0.984
## dispers       0.02389 0.063030 0.802
## I(ft == 2)TRUE  0.01770 0.032359 0.857
## sex:I(ft == 2)TRUE -0.03828 0.161972 0.687
## adult:I(ft == 2)TRUE -0.01222 0.016379 0.898
## dispers:I(ft == 2)TRUE -0.07728 0.649008 0.420
## GLOBAL              NA 4.300970 0.745
```

Both models satisfy the assumptions for all of the variables included in our regression. Note that `comprisk1`, the stratified version, displays 'NA' for the variable `I(ft==2) TRUE` which represents the anthropogenic causes of mortality. Because the non-stratified model satisfies our assumptions, we will follow the law of parsimony (i.e., including fewer parameters in the model) and first look at these results:

```
comprisk2
```

```
## Call:
## coxph(formula = Surv(sday, eday, fail) ~ (sex + adult + dispers) *
##       I(ft == 2), data = expanded)
##
##              coef exp(coef) se(coef)      z      p
```

```
## sex                0.37791    1.45923    0.29093    1.30 0.194
## adult              0.00711    1.00713    0.30356    0.02 0.981
## dispers            0.70584    2.02556    0.39113    1.80 0.071
## I(ft == 2)TRUE     -0.32134    0.72518    0.43596   -0.74 0.461
## sex:I(ft == 2)TRUE -0.17918    0.83596    0.39294   -0.46 0.648
## adult:I(ft == 2)TRUE 0.56116    1.75270    0.43437    1.29 0.196
## dispers:I(ft == 2)TRUE 0.82043    2.27149    0.48118    1.71 0.088
##
## Likelihood ratio test=33.55 on 7 df, p=2e-05
## n= 4324, number of events= 109
```

In this model, the first variables represent their effect on the probability of succumbing of natural causes. The term `I(ft == 2)TRUE` indicates if the risk of dying from anthropogenic causes is higher or lower than of natural causes, whereas the last three coefficients indicate the effect of the variables on the probability of succumbing to anthropogenic causes.

If we examine the hazard ratios, this model indicates an increase in the probability of succumbing to natural causes of death for male individuals ($HR = 1.459$, 1=male), for adults ($HR = 1.007$) and individuals that disperse ($HR=0.071$). Resident non-adult females (reference category) are on average less likely to succumb to anthropogenic than to natural causes ($HR = 0.725$). It is possible to reconstitute the hazards ratios of adult male dispersers that succumb to natural causes by adding the coefficients, taking their exponent ($\exp(0.377+0.007+0.705)=2.97$), and comparing this to the hazard ratios of the same types of individuals succumbing to anthropogenic causes ($\exp(0.377+0.007+0.705-0.321-0.179+0.561+0.820)=7.17$). This indicates that dispersing adult males are actually >2 times more at risk of dying to anthropogenic causes than from natural causes. However, these results are to be taken with a grain of salt since most coefficients are not significant in the global model. We can also look at the output of the stratified model:

```
comprisk1
```

```
## Call:
## coxph(formula = Surv(sday, eday, fail) ~ (sex + adult + dispers) *
##       I(ft == 2) + strata(ft), data = expanded)
##
##               coef exp(coef) se(coef)      z      p
## sex                0.405      1.499   0.292   1.38 0.166
## adult             -0.161      0.852   0.310  -0.52 0.604
## dispers            0.782      2.186   0.396   1.97 0.048
## I(ft == 2)TRUE         NA         NA   0.000    NA   NA
## sex:I(ft == 2)TRUE    -0.224      0.799   0.397  -0.56 0.572
## adult:I(ft == 2)TRUE  0.846      2.331   0.447   1.89 0.058
## dispers:I(ft == 2)TRUE 0.665      1.944   0.493   1.35 0.177
##
## Likelihood ratio test=33.8 on 6 df, p=7e-06
## n= 4324, number of events= 109
```

Here, our results differ slightly from the previous model, likely due to the influence of stratification. This model indicates that dispersing individuals are significantly more at risk of dying of natural causes. Also, since this is a stratified model and therefore each cause has its own baseline survival function, the model does not provide an estimate of the coefficient `I(ft ==2)` representing the difference in baseline risk between causes of death.

Note that the output of competing risks models can be used to investigate additive and compensatory effects by examining the interaction terms between causes of death and the predictors. For example, from the previous model we conclude that the effect of sex or adult predictors are partially compensatory because their coefficients trend differently between causes of death. However, since both coefficients are not of the same magnitude, the effect is only partially compensatory. In contrast, Disperser status may play an additive role on risk, given that model coefficients are positive for both causes of death. However, keep in mind that in

this example the output table provides model coefficients and not hazard ratios, so that we need to convert the units using $HR = e^{\text{coefficient}}$ if we want to compare actual hazard ratios between causes of death.

Finally, it is possible to display survival curves with competing risk s analysis, but recall that the standard approach is to use the cumulative incidence functions (CIF) rather than the Kaplan-Meier survival estimates because of the non-independence between risk types (see Chapter 11, section 11.6.2). CIF curves for competing risks can be drawn easily in R using the package **cmprsk** and the function **cuminc**, but other packages such as **wild1** also offer similar functions. However, when calculating CIF it is computationally more straightforward to use a simplified version of our dataset containing a single line per individual. This can be performed quickly in R if our data are already ordered in time:

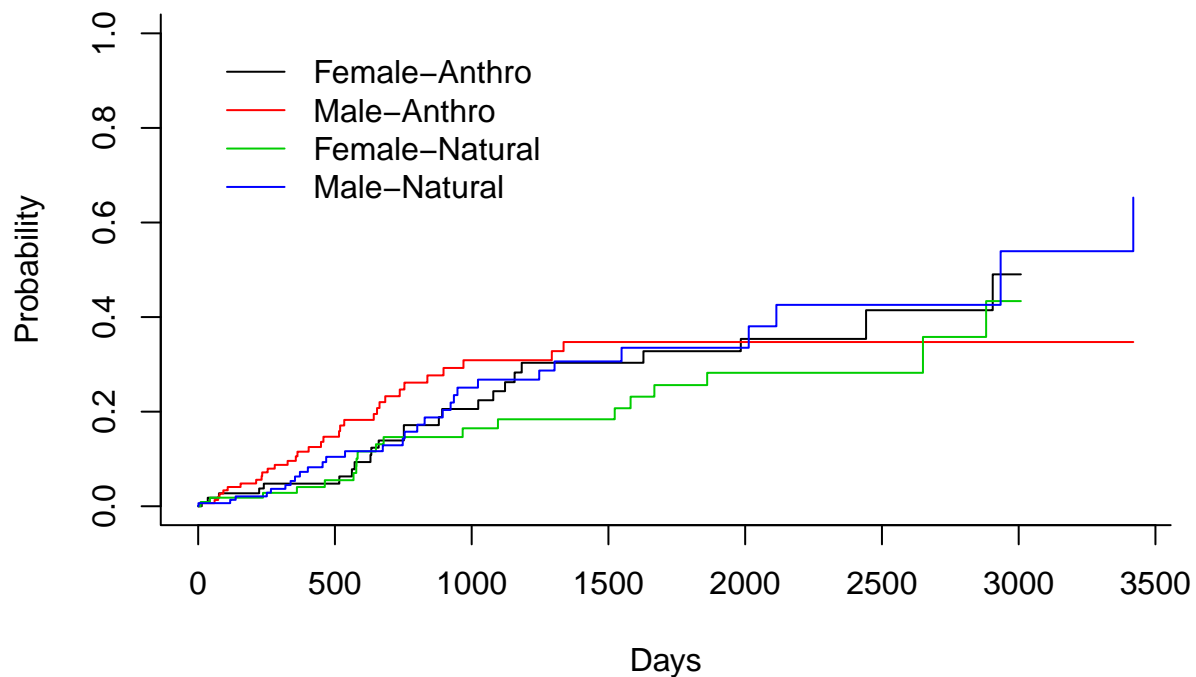
```
id<-unique(as.character(wolf$id))
for (i in id) {
  sub<-wolf[wolf$id==i,]
  out<-sub[nrow(sub),]
  out$sday<-sub[1, "sday"]
  if (i == id[1]) {wolf2<-out}
  if (i != id[1]) {wolf2<-rbind(wolf2,out)}
}
```

We will then create a variable indicating in one column whether the individual survived and was right-censored (**event**=0), died of natural causes (**event** =1) or of anthropogenic causes (**event** =2). We can also create a variable calculating the absolute survival time and then display our cumulative incidence graph with curves identifying each cause and sex.

```
# install.packages("cmprsk") #Note: only need to run prior to first use
library(cmprsk)

wolf2$event<-ifelse(wolf2$natu==1,"Natural", ifelse(wolf2$anthro==1,"Anthro",0))
wolf2$time<-wolf2$eday-wolf2$sday
xx <- cuminc(wolf2$time, wolf2$event, wolf2$sex)
plot(xx,lty=1,color=1:4, curvlab=c("Female-Anthro", "Male-Anthro", "Female-Natural", "Male-Natural"),
     xlab="Days", main="Cumulative incidence function")
```

Cumulative incidence function



As can be seen from this figure, the probability of dying overall is higher for anthropogenic causes, with males being slightly more at risk than females.

Question 7

Using the targeted sample of individuals (`reason=0`), perform an unstratified competing risks cox regression analysis and compare results with the sample for representative individuals.

Answers to questions

Question 1:

For the stage variable (`x3`), are differences in survival probability between groups significant at Day 10, Day 20, and Day 50?

No, confidence intervals are always overlapping throughout the time interval.

Question 2:

Using the previous exercise as an example, test if there are differences in mortality risk **from predation only** between species (binary variable: `sp`) and year (binary variable: `year`). Hint: Use absolute time.

Painted turtles have higher survival than snapping turtles. There is no significant difference in survivorship between 2004 and 2005.

```
nest<-read.csv("Surv_exerc2_3.csv", sep=",")
nest$abstime<-nest$exit-nest$enter
attach(nest)
```

```
## The following object is masked _by_ .GlobalEnv:
##
```

```
##      sp
## The following objects are masked from data:
##
##      abstime, enter, exit
```

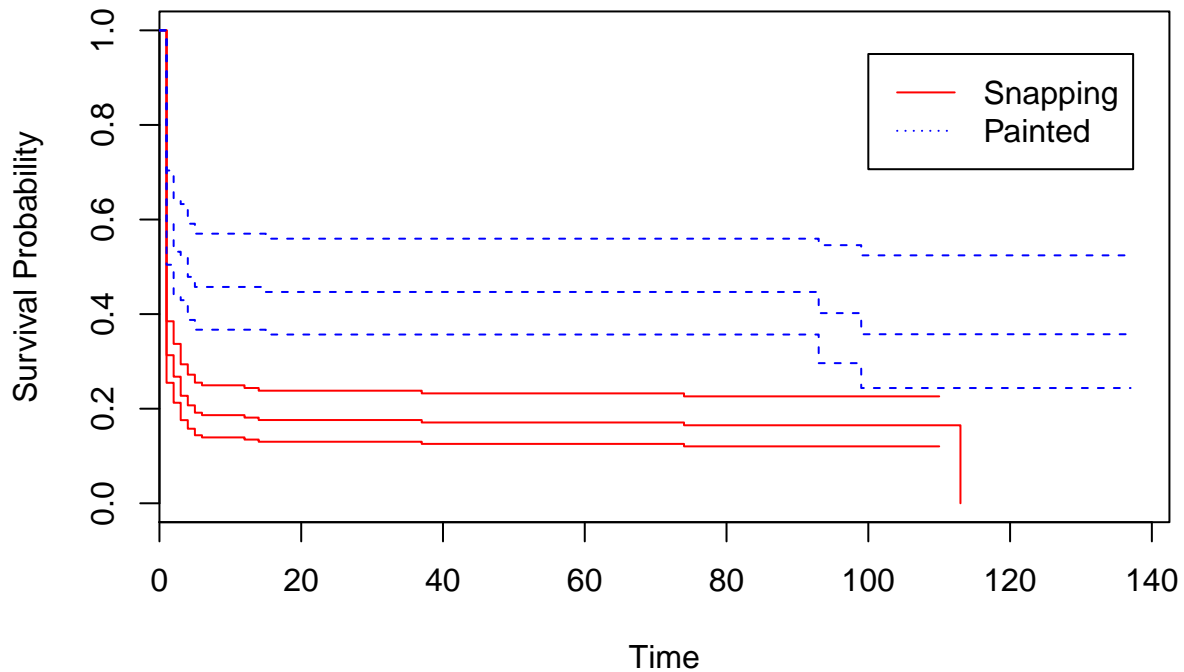
```
library(survival)
```

```
### Difference in species
modelSP <- survfit(Surv(abstime, predation) ~ sp, data=nest)
summary(modelSP)
```

```
## Call: survfit(formula = Surv(abstime, predation) ~ sp, data = nest)
##
##
##      sp=0
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    1    198    136   0.313  0.0330    0.255    0.385
##    2     62     9   0.268  0.0315    0.213    0.337
##    3     53     8   0.227  0.0298    0.176    0.294
##    4     45     4   0.207  0.0288    0.158    0.272
##    5     40     3   0.192  0.0280    0.144    0.255
##    6     37     1   0.186  0.0277    0.139    0.249
##   12     36     1   0.181  0.0274    0.135    0.244
##   14     35     1   0.176  0.0271    0.130    0.238
##   37     34     1   0.171  0.0268    0.126    0.232
##   74     29     1   0.165  0.0265    0.120    0.226
##  113      1      1   0.000    NaN      NA      NA
##
##      sp=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    1     94     38   0.596  0.0506    0.504    0.704
##    2     56      6   0.532  0.0515    0.440    0.643
##    3     50      1   0.521  0.0515    0.429    0.633
##    4     49      4   0.479  0.0515    0.388    0.591
##    5     45      2   0.457  0.0514    0.367    0.570
##   15     43      1   0.447  0.0513    0.357    0.560
##   93     10      1   0.402  0.0627    0.296    0.546
##   99      9      1   0.357  0.0698    0.244    0.524
```

```
plot(modelSP, lty=c(1,2), main="Effect of species", col=c("red", "blue"), xlab="Time",
      ylab="Survival Probability", conf.int=T)
legend(100, 0.95, c("Snapping", "Painted") , lty=c(1,3), col=c("red", "blue") )
```

Effect of species



```
### Difference in year
modelYEAR <- survfit(Surv(abstime, predation) ~ year, data=nest)
summary(modelYEAR)
```

```
## Call: survfit(formula = Surv(abstime, predation) ~ year, data = nest)
```

```
##
```

```
##          year=1
```

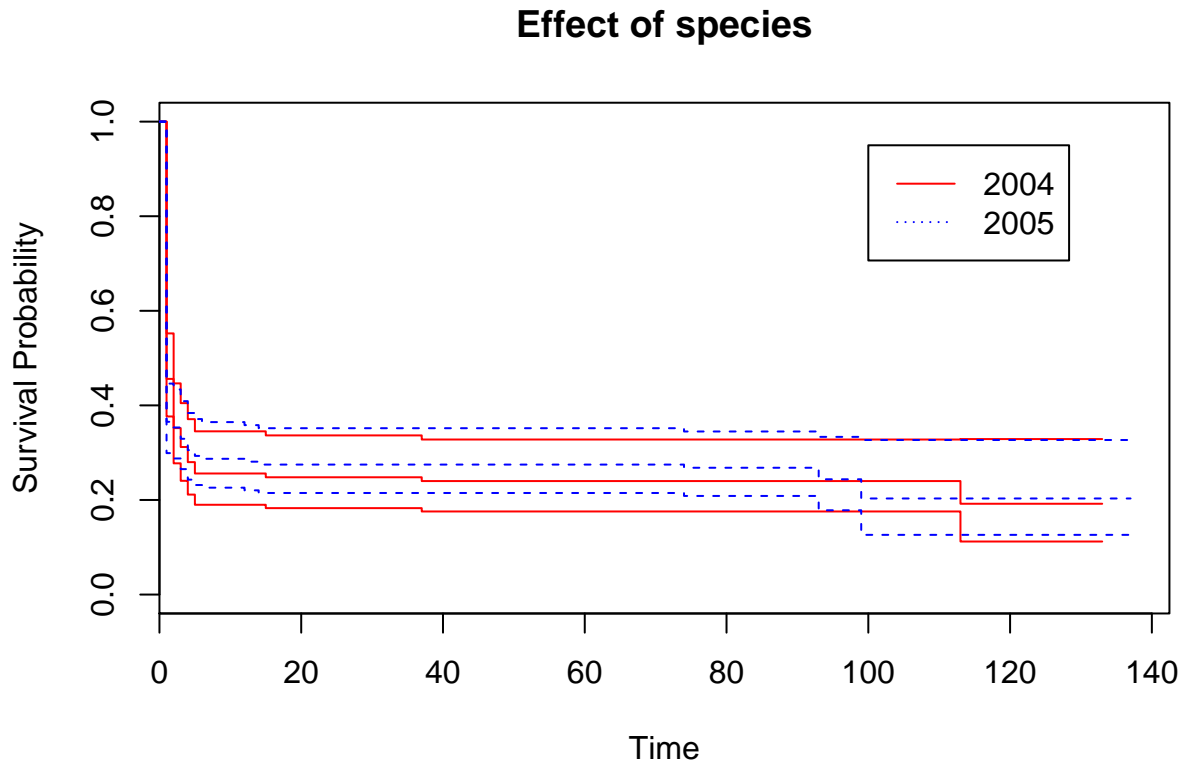
##	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
##	1	125	68	0.456	0.0445	0.377	0.552
##	2	57	13	0.352	0.0427	0.277	0.447
##	3	44	5	0.312	0.0414	0.240	0.405
##	4	39	4	0.280	0.0402	0.211	0.371
##	5	35	3	0.256	0.0390	0.190	0.345
##	15	32	1	0.248	0.0386	0.183	0.337
##	37	31	1	0.240	0.0382	0.176	0.328
##	113	5	1	0.192	0.0527	0.112	0.329

```
##
```

```
##          year=2
```

##	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
##	1	167	106	0.365	0.0373	0.299	0.446
##	2	61	2	0.353	0.0370	0.288	0.434
##	3	59	4	0.329	0.0364	0.265	0.409
##	4	55	4	0.305	0.0356	0.243	0.384
##	5	50	2	0.293	0.0352	0.232	0.371
##	6	48	1	0.287	0.0350	0.226	0.365
##	12	47	1	0.281	0.0348	0.220	0.358
##	14	46	1	0.275	0.0346	0.215	0.352
##	74	41	1	0.268	0.0344	0.209	0.345
##	93	11	1	0.244	0.0390	0.178	0.333
##	99	6	1	0.203	0.0493	0.126	0.327


```
plot(modelYEAR, lty=c(1,2), main="Effect of species", col=c("red", "blue"), xlab="Time",
      ylab="Survival Probability", conf.int=T)
legend(100, 0.95, c("2004", "2005"), lty=c(1,3), col=c("red", "blue"))
```



Question 3

How are the results different across species? Using only the painted turtle dataset, conduct a similar cox proportional hazards analysis and model selection for the influence of corridor and **edgedist**. Compare results with the snapping turtle models and the stratified models including both species.

The stratified model indicates a smaller increase in risk with **edgedist**, but the effect of corridor is qualitatively inverse but still non-significant. Performing the analysis using only the painted turtle data indicates that **edgedist** does not have a significant effect on painted turtle survival whereas nest occurrence on a corridor actually increases survival.

```
library(survival)
library(AICcmodavg)
nest<-read.csv("Surv_exerc2_3.csv", sep=",")
nest$abstime<-nest$exit-nest$enter
painted<-nest[nest$sp==1,]
Cand.mod <- list() # set up candidate model list
# input models into list
Cand.mod[[1]] <-coxph(Surv(abstime, predation)~edgedist, data=painted)
Cand.mod[[2]] <-coxph(Surv(abstime, predation)~corridor, data=painted)
Cand.mod[[3]] <-coxph(Surv(abstime, predation)~edgedist+corridor, data=painted)
# set up model names
Modnames <- c("EdgeDist", "Corridor", "EdgeDist+Corridor")
# output summary table of models
aictab(cand.set = Cand.mod, modnames = Modnames, sort = TRUE)
```

```
##
## Model selection based on AICc:
##
##           K   AICc Delta_AICc AICcWt Cum.Wt      LL
## Corridor      1 427.35      0.00   0.71   0.71 -212.65
## EdgeDist+Corridor 2 429.18      1.82   0.29   1.00 -212.52
## EdgeDist      1 441.50     14.14   0.00   1.00 -219.73
```

Cand.mod[[2]]

```
## Call:
## coxph(formula = Surv(abstime, predation) ~ corridor, data = painted)
##
##           coef exp(coef) se(coef)      z      p
## corridor -2.004    0.135    0.595 -3.37 0.00076
##
## Likelihood ratio test=20.67 on 1 df, p=5e-06
## n= 94, number of events= 54
```

Cand.mod[[3]]

```
## Call:
## coxph(formula = Surv(abstime, predation) ~ edgedist + corridor,
##       data = painted)
##
##           coef exp(coef) se(coef)      z      p
## edgedist 0.00533    1.00535 0.01028 0.52 0.6041
## corridor -1.91354    0.14756 0.62171 -3.08 0.0021
##
## Likelihood ratio test=20.93 on 2 df, p=3e-05
## n= 94, number of events= 54
```

cox.zph(Cand.mod[[2]])

```
##           rho  chisq      p
## corridor -0.0303 0.0497 0.824
```

cox.zph(Cand.mod[[3]])

```
##           rho  chisq      p
## edgedist -0.0273 0.0463 0.830
## corridor -0.0386 0.0802 0.777
## GLOBAL      NA 0.0997 0.951
```

modavg(cand.set = Cand.mod[c(2,3)], parm = "edgedist", modnames = c("Corridor", "EdgeDist+Corridor"))

```
##
## Multimodel inference on "edgedist" based on AICc
##
## AICc table used to obtain model-averaged estimate:
##
##           K   AICc Delta_AICc AICcWt Estimate  SE
## EdgeDist+Corridor 2 429.18      0      1    0.01 0.01
##
## Model-averaged estimate: 0.01
## Unconditional SE: 0.01
## 95% Unconditional confidence interval: -0.01, 0.03
```

```
modavg(cand.set = Cand.mod[c(2,3)], parm = "corridor", modnames = c("Corridor", "EdgeDist+Corridor"))

##
## Multimodel inference on "corridor" based on AICc
##
## AICc table used to obtain model-averaged estimate:
##
##           K   AICc Delta_AICc AICcWt Estimate   SE
## Corridor      1 427.35      0.00   0.71    -2.00 0.60
## EdgeDist+Corridor 2 429.18      1.82   0.29    -1.91 0.62
##
## Model-averaged estimate: -1.98
## Unconditional SE: 0.6
## 95% Unconditional confidence interval: -3.16, -0.79
```

Question 4:

Based on the results of question 2, it seems like the effect of distance to edge is similar between species but that the effect of distance to corridor is different. Using the full turtle dataset, perform a mixed-effects model with species (*sp*) as a random factor that assesses the effect of **edgedist** and **corridor** as fixed effects and **corridor** as a random coefficient. Extract the random coefficients and compare them with the previous models (Hint: write the structure of the random effect as follow: (corridor | *sp*), to compare random effects model with species specific model, sum the fixed effect with the random effects).

The effect of **edgedist** is similar between the mixed-effects model and previous models (e.g. **Cand.mod[[3]]** and Question 3). Similar to the stratified model, the effect of **corridor** is negative and non-significant. The random effect is, however, able to capture the range in response to **corridor** between the two species with an estimate coefficient of 0.432 (-0.661+1.093) for snapping turtle and -1.754 (-0.661-1.093) for painted turtle, which are closer to estimates from individual models.

```
library(coxme)
nest<-read.csv("Surv_exerc2_3.csv", sep=",")
nest$abstime<-nest$exit-nest$ente
mix.mod1 <-coxme(Surv(abstime, predation)~edgedist+corridor+(corridor|sp), data=nest)
mix.mod1
```

```
## Cox mixed-effects model fit by maximum likelihood
##   Data: nest
##   events, n = 220, 292
##   Iterations= 5 30
##
##           NULL Integrated      Fitted
## Log-likelihood -1124.899  -1105.666 -1101.735
##
##           Chisq  df          p   AIC   BIC
## Integrated loglik 38.47 3.00 2.2515e-08 32.47 22.29
## Penalized loglik 46.33 2.91 4.1998e-10 40.50 30.62
##
## Model:  Surv(abstime, predation) ~ edgedist + corridor + (corridor |      sp)
## Fixed coefficients
##           coef exp(coef)    se(coef)      z      p
## edgedist  0.01526344 1.0153805 0.005849237  2.61 0.0091
## corridor -0.66149653 0.5160784 0.893718156 -0.74 0.4600
##
## Random effects
##   Group Variable Std Dev  Variance
```

```
## sp      corridor 1.194913 1.427817
```

```
ranef(mix.mod1)
```

```
## $sp
```

```
##      corridor
```

```
## 0  1.093419
```

```
## 1 -1.093419
```

Question 5:

Using the painted turtle dataset, perform a time-dependent cox proportional hazards regression testing the effect of rain, temperature, corridor and edgedist first using calendar time and then using absolute time. Compare the output for each variable between the two approaches.

The effect of each variable is qualitatively similar between the two times, except for **temp**. Only the effect of **corridor** reduces the probability of being predated in both models while **temp** only reduces risk significantly using calendar time. Note, however, that this variable does not respect the proportional hazards assumption, so a stratified model should be used if we want to use calendar time to study the painted turtle data.

```
nest<-read.csv("Surv_exerc2_3.csv", sep=",")
nest$abstime<-nest$exit-nest$enter
library(survival)
expand<-read.csv("Surv_exerc2_expand.csv", sep=",")
## Merge corridor and edgedist
expand2<-merge(expand, nest[,c("nestid", "edgedist", "corridor", "enter")], by="nestid")
painted2<-expand2[expand2$sp==1,]
painted2$enter_abs<-painted2$enter.x-painted2$enter.y
painted2$exit_abs<-painted2$exit-painted2$enter.y
cox_mod5<-coxph(Surv(enter.x, exit, predation) ~ edgedist+corridor+rain+temp, data=painted2)
cox_mod6<-coxph(Surv(enter_abs, exit_abs, predation) ~ edgedist+corridor+rain+temp, data=painted2)
cox.zph(cox_mod5)
```

```
##          rho chisq      p
## edgedist  0.327  6.42 0.01130
## corridor  0.218  2.66 0.10315
## rain      0.216  1.84 0.17484
## temp     -0.232  4.46 0.03466
## GLOBAL      NA 13.53 0.00894
```

```
cox.zph(cox_mod6)
```

```
##          rho chisq      p
## edgedist -0.0217 0.0285 0.866
## corridor -0.0440 0.1044 0.747
## rain      0.0484 0.0930 0.760
## temp     -0.0719 0.2425 0.622
## GLOBAL      NA 0.4132 0.981
```

```
cox_mod5
```

```
## Call:
```

```
## coxph(formula = Surv(enter.x, exit, predation) ~ edgedist + corridor +
##      rain + temp, data = painted2)
```

```
##
```

```
##          coef exp(coef) se(coef)      z      p
## edgedist  0.0139    1.0140  0.0106  1.31 0.1890
```

```
## corridor -2.2585    0.1045    0.6249 -3.61 0.0003
## rain      0.0754    1.0783    0.0559  1.35 0.1773
## temp     -0.1400    0.8693    0.0631 -2.22 0.0265
##
## Likelihood ratio test=45.6 on 4 df, p=3e-09
## n= 3714, number of events= 54
```

```
cox_mod6
```

```
## Call:
## coxph(formula = Surv(enter_abs, exit_abs, predation) ~ edgedist +
##       corridor + rain + temp, data = painted2)
##
##               coef exp(coef) se(coef)      z      p
## edgedist  0.00773   1.00776  0.01041   0.74 0.4580
## corridor -1.90965   0.14813  0.62278  -3.07 0.0022
## rain      0.05818   1.05991  0.03308   1.76 0.0786
## temp      0.04287   1.04381  0.03876   1.11 0.2686
##
## Likelihood ratio test=25.06 on 4 df, p=5e-05
## n= 3714, number of events= 54
```

Question 6:

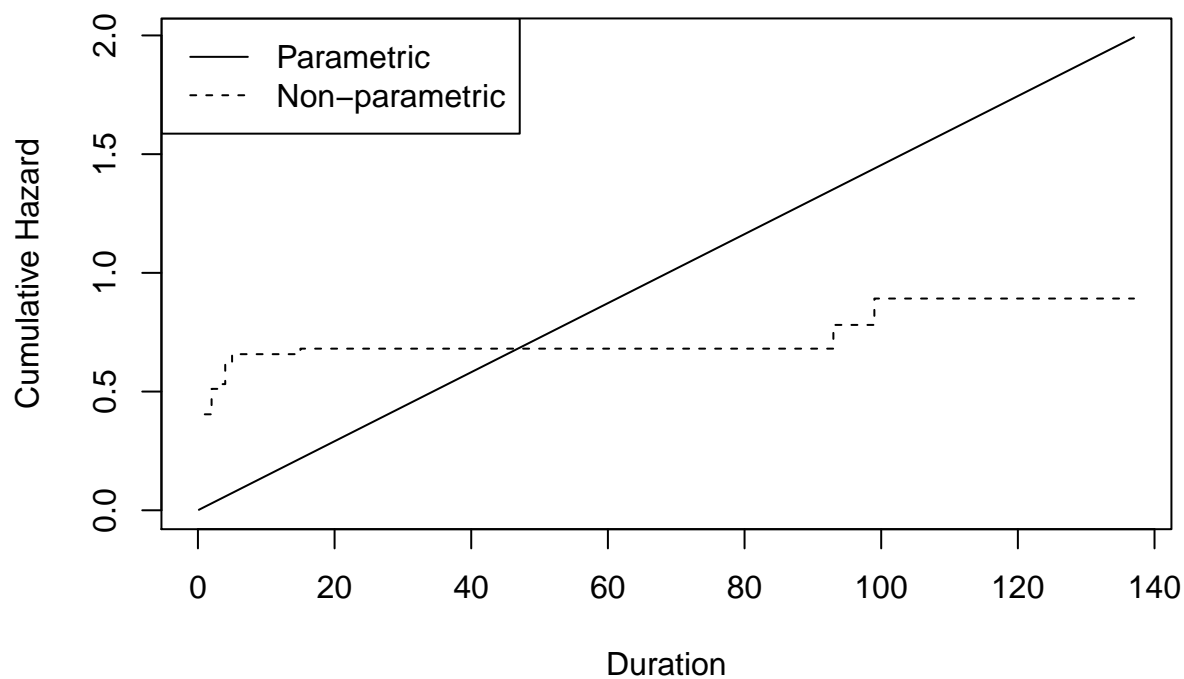
Using the painted turtle dataset alone, try performing a Weibull regression looking at the effect of **corridor** on the probability of survival. First, assess if the Weibull is better than the simpler Exponential model. Compare these results with those from the semi-parametric approach developed in Question 2.

Although, the weibull distribution has poor fit as can be seen using the **check.dist** function, it nevertheless offers a better fit than the exponential model. Both weibull and cox regressions indicate that **corridor** reduces risk of predation for painted turtles.

```
nest<-read.csv("Surv_exerc2_3.csv", sep=",") # Open csv file
library(eha)
nest$abstime<-nest$exit-nest$enter
painted<-nest[nest$sp==1,]
exp_mod4<- phreg(Surv(abstime, predation) ~ 1, dist="weibull", shape=1, data=painted) #
wb_mod4<- phreg(Surv(abstime, predation) ~ 1, dist="weibull", data=painted) #
cox_mod4<-coxreg(Surv(abstime, predation) ~ 1, data=painted) #

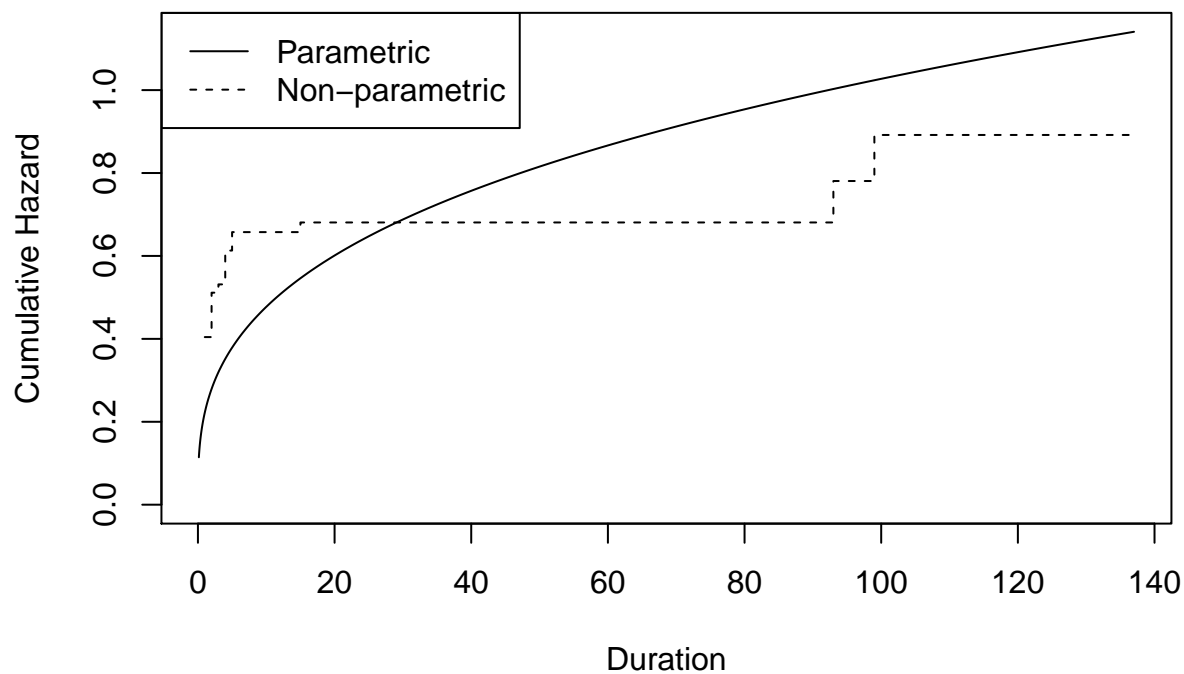
check.dist(cox_mod4, exp_mod4, main="Exponential")
```

Exponential



```
check.dist(cox_mod4, wb_mod4, main="Weibull")
```

Weibull



```
extractAIC(wb_mod4)
```

```
## [1] 0.0000 423.9731
```

```
extractAIC(exp_mod4)
```

```
## [1] 0.0000 564.9351
```

```
summary(wb_mod4)
```

```
## Call:
```

```
## phreg(formula = Surv(abstime, predation) ~ 1, data = painted,  
##       dist = "weibull")
```

```
##
```

## Covariate	W.mean	Coef	Exp(Coef)	se(Coef)	Wald p
## log(scale)		4.525		0.427	0.000
## log(shape)		-1.100		0.118	0.000

```
##
```

## Events	54
## Total time at risk	3714
## Max. log. likelihood	-211.99

```
wb_mod5<- phreg(Surv(abstime, predation) ~ corridor, dist="weibull", data=painted) #  
cox_mod5 <-coxreg(Surv(abstime, predation)~corridor, data=painted)  
summary(wb_mod5)
```

```
## Call:
```

```
## phreg(formula = Surv(abstime, predation) ~ corridor, data = painted,  
##       dist = "weibull")
```

```
##
```

## Covariate	W.mean	Coef	Exp(Coef)	se(Coef)	Wald p
## corridor	0.409	-2.187	0.112	0.596	0.000
## log(scale)		3.438		0.389	0.000
## log(shape)		-1.020		0.113	0.000

```
##
```

## Events	54
## Total time at risk	3714
## Max. log. likelihood	-199.12
## LR test statistic	25.72
## Degrees of freedom	1
## Overall p-value	3.93937e-07

```
summary(cox_mod5)
```

```
## Call:
```

```
## coxreg(formula = Surv(abstime, predation) ~ corridor, data = painted)
```

```
##
```

## Covariate	Mean	Coef	Rel.Risk	S.E.	Wald p
## corridor	0.409	-2.004	0.135	0.595	0.001

```
##
```

## Events	54
## Total time at risk	3714
## Max. log. likelihood	-212.65
## LR test statistic	20.67
## Degrees of freedom	1
## Overall p-value	5.45724e-06

Question 7

Using the targeted sample of individuals (**reason=0**), perform an unstratified competing risks cox regression analysis and compare results with the sample for representative individuals.

Similar to the representative sample, no variables have a significant effect on survival for the targeted individuals although most were qualitatively different. Most notably, targeted individuals that dispersed seem to be at lower risk. Lastly, targeted individuals are more at risk of dying from anthropogenic causes than natural causes (**I (ft ==2) TRUE > 0**) which is the opposite pattern from what was observed for representative individuals.

```
library(survival)
wolf<-read.csv("Surv_exerc4.csv")
wolf<-wolf[wolf$reason==0,]
wolf1<-data.frame(wolf,ft=1) #ft1=Natural
wolf2<-data.frame(wolf,ft=2) # ft2=Anthro
expanded<-rbind(wolf1, wolf2)
expanded$event<-1*ifelse(expanded$natu==1,1,0)+ 2*ifelse(expanded$anthro==1,1,0)
expanded$fail<-ifelse(expanded$event==expanded$ft,1,0)
comprisk3<-coxph(Surv(sday,eday, fail)~(sex+adult+dispers)*I(ft==2), data= expanded)
```

```
## Warning in fitter(X, Y, strats, offset, init, control, weights = weights, :
## Loglik converged before variable 3,7 ; beta may be infinite.
```

```
cox.zph(comprisk3)
```

```
##              rho      chisq      p
## sex              0.1002 2.56e-01 0.613
## adult            -0.0099 2.71e-03 0.958
## dispers           0.2921 1.75e-08 1.000
## I(ft == 2)TRUE     0.0145 3.90e-03 0.950
## sex:I(ft == 2)TRUE -0.0733 1.35e-01 0.714
## adult:I(ft == 2)TRUE 0.0138 4.75e-03 0.945
## dispers:I(ft == 2)TRUE -0.3275 2.79e-08 1.000
## GLOBAL              NA 6.76e-01 0.999
```

```
comprisk3
```

```
## Call:
## coxph(formula = Surv(sday, eday, fail) ~ (sex + adult + dispers) *
##       I(ft == 2), data = expanded)
##
##              coef exp(coef) se(coef)      z      p
## sex             -7.86e-01  4.56e-01  1.23e+00 -0.64 0.52
## adult            -6.70e-01  5.12e-01  1.24e+00 -0.54 0.59
## dispers          -1.76e+01  2.36e-08  7.53e+03  0.00 1.00
## I(ft == 2)TRUE     1.15e+00  3.17e+00  9.06e-01  1.27 0.20
## sex:I(ft == 2)TRUE  4.97e-01  1.64e+00  1.32e+00  0.38 0.71
## adult:I(ft == 2)TRUE 6.44e-01  1.90e+00  1.31e+00  0.49 0.62
## dispers:I(ft == 2)TRUE 1.83e+01  8.72e+07  7.53e+03  0.00 1.00
##
## Likelihood ratio test=19.45 on 7 df, p=0.007
## n= 446, number of events= 24
```

References

Bastille-Rousseau, G, Fortin, D, Dussault, C, *et al.* (2011) Foraging strategies by omnivores: Are black bears actively searching for ungulate neonates or are they simply opportunistic predators? *Ecography*, 34, 588–596.

- Bolker, BM, Brooks, ME, Clark, CJ, *et al.* (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24, 127–135.
- Cleves, M, Gutierrez, RG, Gould, W, *et al.* (2010) *An Introduction to Survival Analysis Using Stata*. Sata Press, College Station, TX.
- Fieberg, J, and Delgiudice, GD (2009) What time is it? Choice of time origin and scale in extended proportional hazards models. *Ecology*, 90, 1687–1697.
- Griffin, KA, Hebblewhite, M, Robinson, HS, *et al.* (2011) Neonatal mortality of elk driven by climate, predator phenology and predator community composition. *Journal of Animal Ecology*, 80, 1246–1257.
- Heisey, DM, and Patterson, BR (2006) A review of methods to estimate cause-specific mortality in presence of competing risks. *Journal of Wildlife Management* 70, 1544–1555.
- Kleinbaum, D. and Klein, M. (2011) *Survival Analysis - A Self Learning Text*. Springer, New York, NY.
- Lunn, M. and McNeil, D. (1995) Applying cox regression to competing risks. *Biometrics*, 51, 524–532.
- Melis, C, Nilsen, EB, Panzacchi, M, *et al.* (2013) Roe deer face competing risks between predators along a gradient in abundance. *Ecosphere* 4, article 111.
- Murray, DL, Smith, DW, Bangs, EE, *et al.* (2010) Death from anthropogenic causes is partially compensatory in recovering wolf populations. *Biological Conservation* 143, 2514–2524.
- Pinheiro, JC, and Bates, DM (2000) *Mixed Effects Models in S and S-Plus*. Springer, New York.
- Schmidt, KA, Goheen, JR and Naumann, R (2001) Incidental nest predation in songbirds: behavioral indicators detect ecological scales and processes. *Ecology* 82, 2937–2947.