Project Proposal

Alex Brown, David Geyfman, Faraz Farooq, Jason Freeberg

28 October 2016

Questions

Our team would like to investigate how a couple's race and years of education relate to the time until a couple's divorce. The full explanations of all covariates are listed below. We will begin with the simplest model solely based on time and event status. Then we will add covariates to the model, check their significance and arrive at a final model.

Explanation of variables:

1) *id*... A couple's identification number during the study

2) *edu*... The husband's education level.

- 0 = less than 12 years (only high school)
- 1 = 12 to 15 years (only bachelors or equivalent)
- 2 = 16 or more years (some form of graduate studies)

3) *hblack*... Coded 1 if the husband is black, 0 otherwise

4) *mixed*... Coded 1 if the couple is of different ethnicities, 0 otherwise

- ethnicities are defined as either being "black" or "other"

5) *years*... The number of years from wedding to either divorce or censorship

6) *div*... The failure indicator

- 1 = divorced
- 0 = censored

```
# Attach packages
library(ggplot2)
library(survival)
library(cowplot)
```

```
## Warning: package 'cowplot' was built under R version 3.3.2
```

```
##
## Attaching package: 'cowplot'
```

```
## The following object is masked from 'package:ggplot2':
##
##     ggsave
```

```r
#setwd("~/Documents/Fall/PSTAT175/SurvivalAnalysis")

# Load data
colNames <- c("id", "edu", "hblack", "mixed", "years", "div")
divorce <- read.table(file = "divorce.txt", header = F, col.names = colNames)
head(divorce)
```

```
##   id edu hblack mixed  years div
## 1  9   1      0     0 10.546   0
## 2 11   0      0     0 34.943   0
## 3 13   0      0     0  2.834   1
## 4 15   0      0     0 17.532   1
## 5 33   1      0     0  1.418   0
## 6 36   0      0     0 48.033   0
```

Code to Make Graphs Look Nicer

Exploratory Analysis

The plots show the following estimates (clockwise from top left):

1) The baseline KM estimate for all observations

2) The KM estimates for the level of education

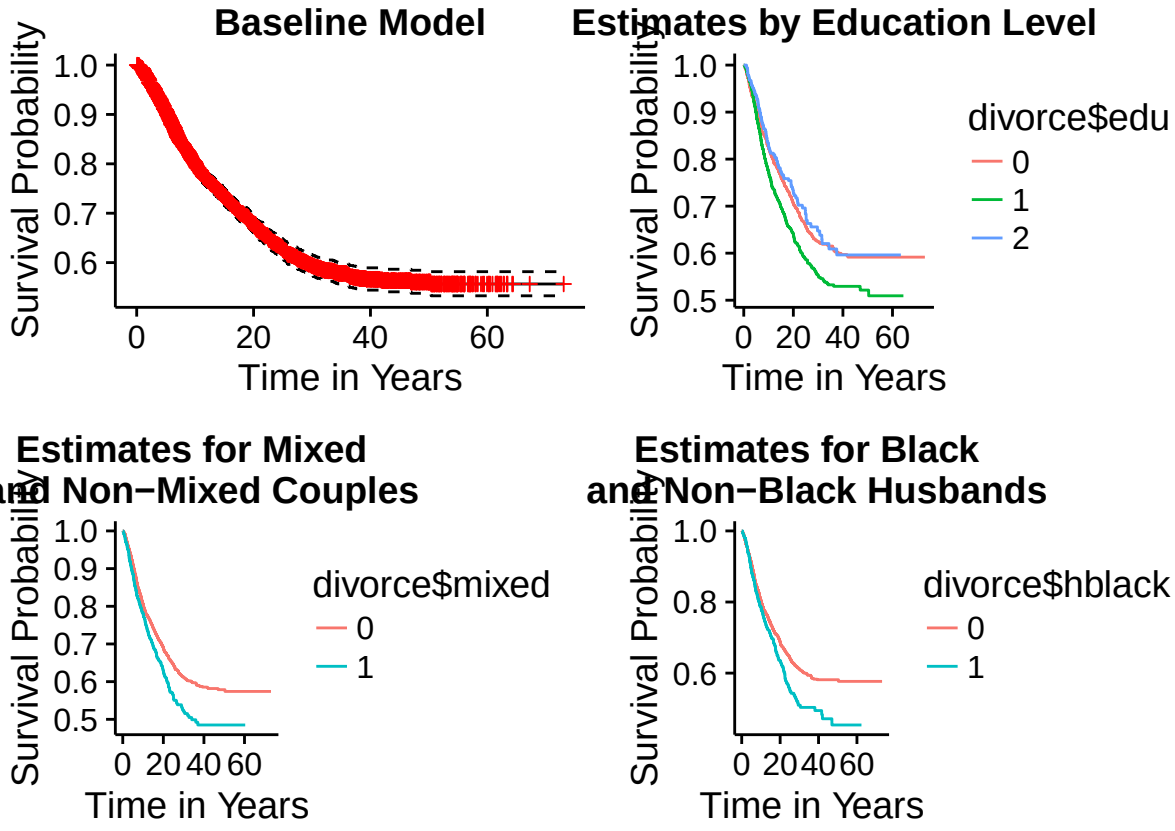
3) The KM estimates for mixed and non-mixed couples

4) The KM estimates for couples with and without a black husband

GGSURV : This is like plot.surv but makes plot look much nicer

```r
divorcefit <- Surv(time = divorce$years, event = divorce$div)
divorceKM <- survfit(divorcefit ~ 1)
educateKM <- survfit(divorcefit ~ divorce$edu)
mixedKM <- survfit(divorcefit ~ divorce$mixed)
hblackKM <- survfit(divorcefit ~ divorce$hblack)

a <- ggsurv(divorceKM) +
  ggtitle("Baseline Model") +
  ylab("Survival Probability") +
  xlab("Time in Years")

b <- ggsurv(educateKM, plot.cens = F) +
  ggtitle("Estimates by Education Level") +
  ylab("Survival Probability") +
  xlab("Time in Years")

c <- ggsurv(mixedKM, plot.cens = F) +
```

```
  ggtitle("Estimates for Mixed \n and Non-Mixed Couples") +
  ylab("Survival Probability") +
  xlab("Time in Years")

d <- ggsurv(hblackKM, plot.cens = F) +
  ggtitle("Estimates for Black \n and Non-Black Husbands") +
  ylab("Survival Probability") +
  xlab("Time in Years")

plot_grid(a,b,c,d, nrow = 2, ncol = 2)
```

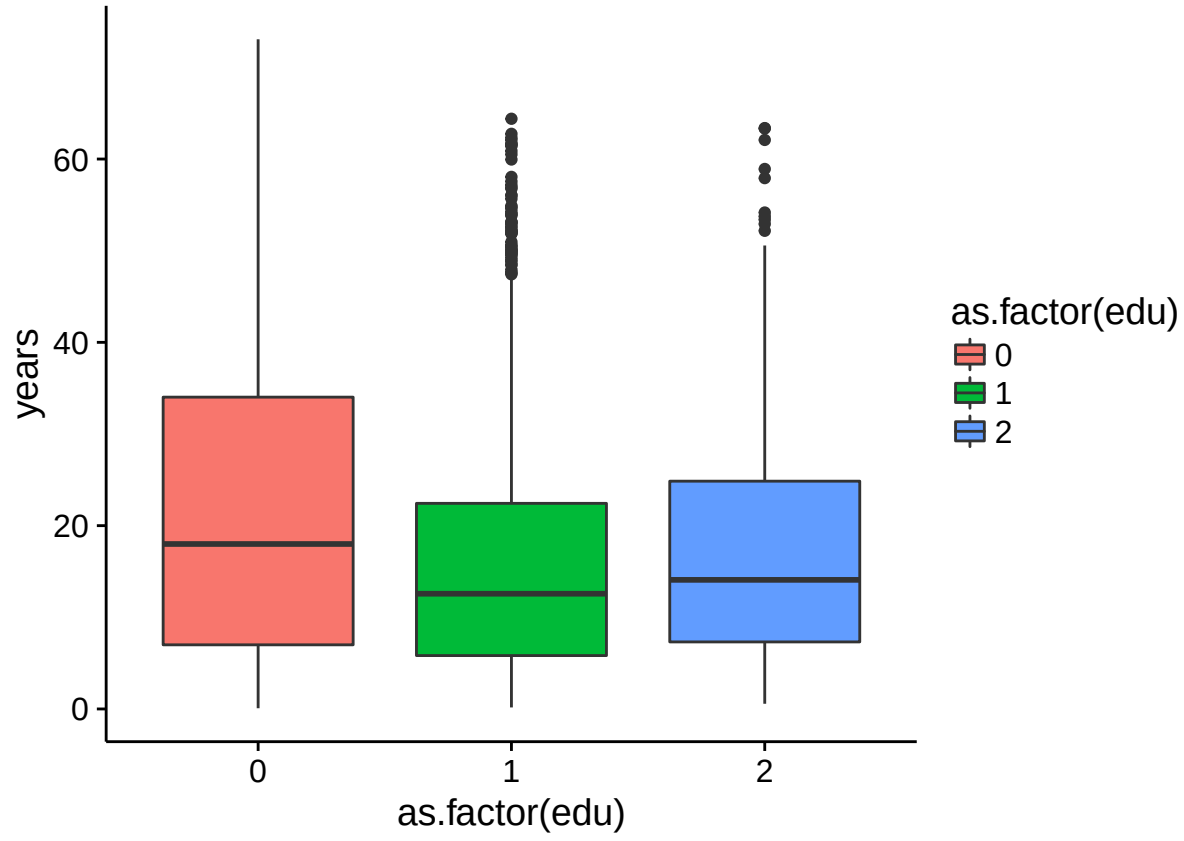


Cox Proportional Hazards Model

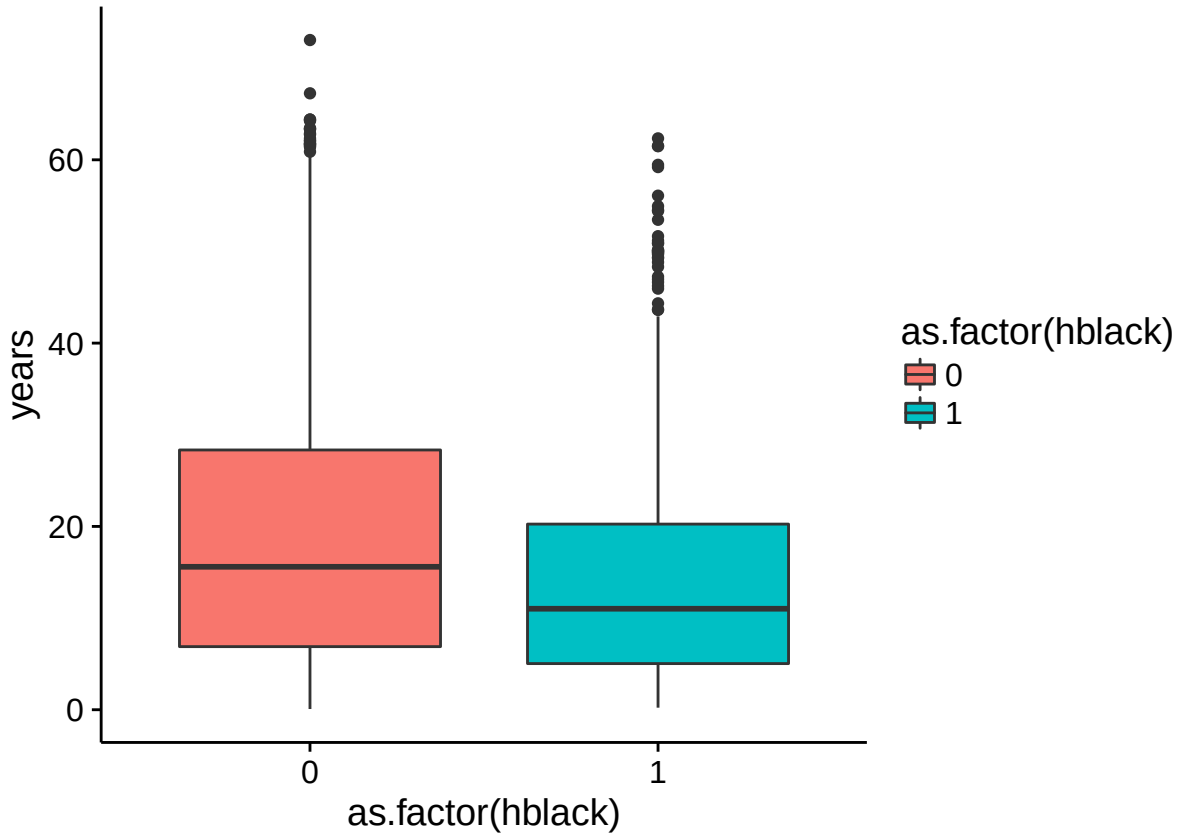
```
## Call:  coxph(formula = divorcefit ~ 1)
##
## Null model
##   log likelihood= -7844.309
##   n= 3371
```

Furthur Exploratory Analysis

```
ggplot(divorce) +
  geom_boxplot(mapping = aes(x = as.factor(edu), y = years,
                             group = edu, fill = as.factor(edu)))
```




```
ggplot(divorce) +
  geom_boxplot(mapping = aes(x = as.factor(hblack), y = years,
                             group = hblack, fill = as.factor(hblack)))
```




```
## Number of censored observations = 2339
## Number of observed divorces = 1032
```