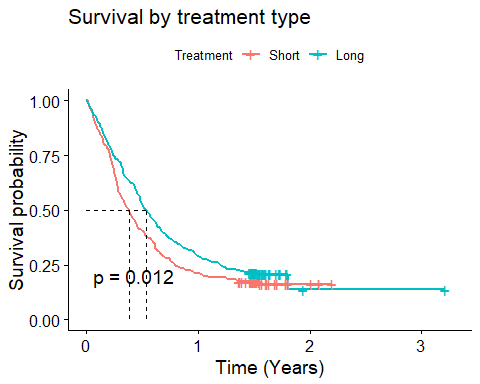
Applied Logistic Regression and Survival Analysis HW11

Colleen Brents

5/7/2019

1. Assuming treatment type (long or short) was randomly assigned, answer the question of interest for the study by:
2. Provide a plot of estimated survival curves for time until drug use comparing long and short treatment (assuming time is in days)



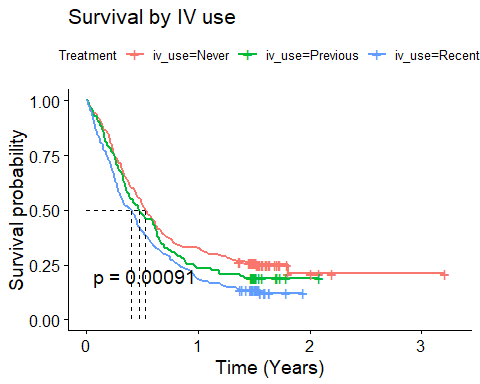
1. Estimate the hazard ratio for time until drug use, comparing long to short treatment

The estimated return to drug use is 0.80 times that of those with short treatment. (aka) We have evidence to reject the null hypothesis that there is no association between time to drug reuse and treatment duration (long or short) [p= 0.012]

1. Is there a treatment effect when adjusting for age and prior treatment status?

The estimated return to drug use of those that have had more than 1 prior drug treatment is 1.25 times that of people who have at most 1 prior treatment, adjusted for age.  
We have evidence to reject the null hypothesis that there is no association between time to drug reuse and prior treatment status, when adjusted for age (p=0.028).

Sidenote - We have evidence to reject the null hypothesis that there is no association between time to drug reuse and age with p=0.05

1. Is there a difference in time until return to drug use by an individual’s history of IV drug use?
2. provide a plot of estimated survival curves 
3. Fit a proportional hazards regression model with IV use history as the predictor of interest and that adjusts for age. Report the estimated hazard ratios for IV usage categories in 1-2 summarizing sentences

The estimated hazard ratio comparing those with previous iv use to those who have never used ivs is 1.32, adjusted for age.

The estimated hazard ratio comparing those with recent iv use to those who have never used ivs is 1.67 , adjusted for age.

The estimated hazard ratio comparing those with recent iv use to previous iv use is 1.27, adjusted for age.

We have evidence to reject the null hypothesis that there is no association between time to drug reuse and previous iv use, when adjusted for age (p<0.001).

1. Is there a difference in time until return to drug use by an individual’s use of heroin or cocaine the 3 months prior to admission? answer by providing and interpreting estimated hazard ratios (can choose what to adjust for)

The estimated hazard ratio comparing those with previous cocaine use to those with no previous drug use is 0.9, adjusted for age at baseline.

The estimated hazard ratio comparing those with previous heroin use to those without previous drug use is 1.35, adjusted for age at baseline.

The estimated hazard ratio comparing those with previous cocaine and heroin use to those without prior drug use is 1.24, adjusted for age at baseline.

The estimated hazard ratio comparing those with previous heroin use to previous cocaine use is 1.02, adjusted for age at baseline.

The estimated hazard ratio comparing those with previous heroine use to those with previous cocaine and heroin use is 1.38, adjusted for age at baseline.

The estimated hazard ratio comparing those with previous cocaine use to those with previous cocaine and heroin use is 0.92, adjusted for age at baseline.

Basically, people with any ‘recent’ heroin experience are more likely to relapse sooner than those with cocaine (or no drug) experience. This seems reasonable given ‘Don’t Do Drugs Campaigns’ and movies like ‘Requiem for a Dream’.

We have evidence to reject the null hypothesis that there is no association between time to drug reuse and previous drug use within the last 3 months, when adjusting for age (p=0.01)

## Appendix

knitr::opts\_chunk$set(  
 echo = FALSE,  
 message = FALSE,  
 warning = FALSE  
)  
options(scipen=1, digits=2)  
library(readr)  
library(ggplot2)  
library(splines)  
library(dplyr)  
library(tidyr)  
library(ggthemes)  
library(knitr)  
library(broom) #tidy   
library(stringr)  
library(lmtest)  
library(sandwich)  
library(car)  
library(survival)  
library(ROCR)  
library(gridExtra) #to make graphs side by side   
library(caret)  
library(survminer)  
  
url <- paste0("https://raw.githubusercontent.com/cbrents/stat796logreg/master/hw11/uis.csv")  
uis<-read\_csv(url)  
  
#heart$mi\_recurr <- factor(heart$mi\_recurr, levels=c(0, 1), labels=c("First heart attack", "Recurring heart attack"))  
df<- uis %>%   
 mutate(id=as.factor(id),  
 white=as.factor(white),  
 tx\_long=as.numeric(tx\_long),  
 iv\_use=as.factor(iv\_use)) #doesn't matter if numeric or factor   
 #status=as.numeric(status)) #event or outcome needs to be numeric?   
  
df<- subset(df, !is.na(hercoc\_use))  
df<-subset(df, !is.na(age))  
  
df<- subset(df, !is.na(tx\_long))  
df<- subset(df, !is.na(iv\_use))  
df<- subset(df, !is.na(nprior\_tx))  
  
  
df\_surv<-Surv(df$time, df$status) #not right censored data?! also continuous data?   
df\_survfit<-survfit(df\_surv ~ tx\_long, data=df)  
print(ggsurvplot(df\_survfit, data=df, break.x.by=365.25,  
 surv.median.line="hv",  
 title="Survival by treatment type",  
 xscale="d\_y", xlab="Time (Years)",  
 conf.int = FALSE, pval=TRUE,   
 legend.labs=c("Short", "Long"),  
 legend.title="Treatment"))  
coxph(Surv(df$time, df$status)~tx\_long, data=df)   
#curious about time to reuse by trt type  
basic<-coxph(Surv(df$time, df$status)~1, data=df)  
trt<- coxph(Surv(df$time, df$status)~tx\_long, data=df)  
anova(basic,trt, test="LRT") #works p<0.05 so tx\_long is significant  
df<- df %>%   
 mutate(nprior\_new=nprior\_tx,  
 nprior\_new=recode(nprior\_new, "'0' = '0'; '1' = '0'; else='1'"))   
nprior<- coxph(Surv(df$time, df$status)~nprior\_new + age, data=df) #higher hr is make nprior binary   
coxph(Surv(df$time, df$status)~nprior\_tx + age, data=df)   
basic\_age<-coxph(Surv(df$time, df$status)~age, data=df)  
anova(basic\_age, nprior,test="LRT")  
anova(basic, basic\_age, test="LRT")  
df\_surv<-Surv(df$time, df$status) #not right censored data?! also continuous data?   
df\_survfit<-survfit(df\_surv ~ iv\_use, data=df)  
print(ggsurvplot(df\_survfit, data=df, break.x.by=365.25,  
 surv.median.line="hv",  
 title="Survival by IV use",  
 xscale="d\_y", xlab="Time (Years)",  
 conf.int = FALSE, pval=TRUE,   
 legend.title="Treatment"))  
blurg<- coxph(Surv(df$time, df$status)~factor(iv\_use) + age, data=df) #iv\_use is a categorical predictor variable  
threea<- exp(coef(blurg)[1]) # compare previous to never   
threeb<- exp(coef(blurg)[2])  
threec<-exp(coef(blurg)[2]-coef(blurg)[1])  
anova(basic\_age, blurg, test="LRT")  
  
df$hercoc\_use<- as.factor(df$hercoc\_use)  
levels(df$hercoc\_use)  
df$hercoc\_use<- factor(df$hercoc\_use, levels = c("Neither", "Cocaine", "Heroin", "Heroin & Cocaine")) #reorder characters   
levels(df$hercoc\_use)  
  
blah<- coxph(Surv(df$time, df$status)~factor(hercoc\_use) + age, data=df)   
foura<- exp(coef(blah)[1]) #cocaine to neither  
fourb<- exp(coef(blah)[2]) # heroin to neither  
fourc<- exp(coef(blah)[3]) # cocain and heroin to neither  
fourd<-exp(coef(blah)[2]-coef(blurg)[1]) #heroin to cocaine  
foure<- exp(coef(blah)[2]-coef(blurg)[3]) #heroin to heroin and cocine   
fourf<-exp(coef(blah)[1]-coef(blurg)[3]) #cocaine to heroin and cocoin   
anova(basic\_age,blah, test="LRT") #compare age and preious use