Hw10

Xinrui Hu

2023-12-11

library(haven)  
library(survival)  
library(ggsurvfit)  
df <- read\_dta("event1.dta")

#1  
summary(df$marriage)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.00 1.00 34.00 95.51 84.50 1071.00

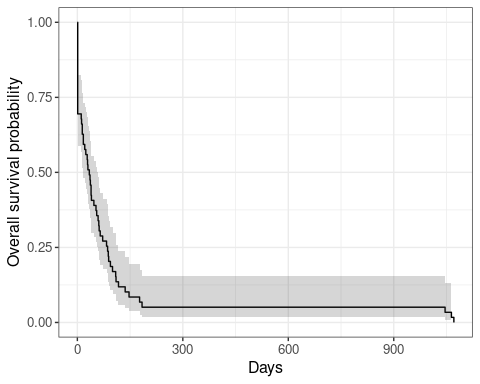
#The min is 1, mean is 95.51, and max is 1071.

#2  
summary((df$evermarr))

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1 1 1 1 1 1

# No subjects were cencered

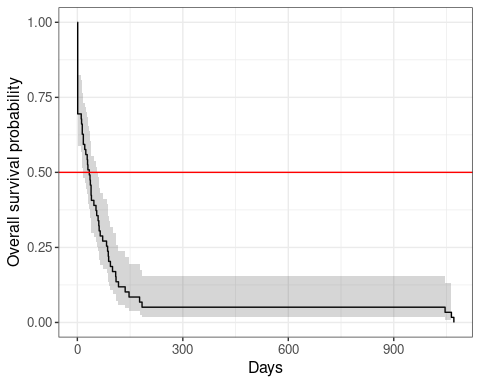
#3  
survfit(Surv(marriage, evermarr) ~ 1, data = df) %>%   
ggsurvfit() + labs( x = "Days", y = "Overall survival probability" ) +  
 add\_confidence\_interval()



#4  
survfit(Surv(marriage, evermarr) ~ 1, data = df)

## Call: survfit(formula = Surv(marriage, evermarr) ~ 1, data = df)  
##   
## n events median 0.95LCL 0.95UCL  
## [1,] 59 59 34 17 59

survfit(Surv(marriage, evermarr) ~ 1, data = df) %>%   
 ggsurvfit() + labs( x = "Days", y = "Overall survival probability" ) +  
 add\_confidence\_interval()+geom\_hline(yintercept = 0.5, color = "red")



# The median survival time is 34

# 5  
summary(survfit(Surv(marriage, evermarr) ~ 1, data = df), times = 1070)

## Call: survfit(formula = Surv(marriage, evermarr) ~ 1, data = df)  
##   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 1070 1 58 0.0169 0.0168 0.00243 0.118

# The rate at 1070 days are 0.0169%

# 6  
survdiff(Surv(marriage, evermarr) ~ cohab, data = df)

## Call:  
## survdiff(formula = Surv(marriage, evermarr) ~ cohab, data = df)  
##   
## N Observed Expected (O-E)^2/E (O-E)^2/V  
## cohab=0 34 34 31.5 0.192 0.487  
## cohab=1 25 25 27.5 0.220 0.487  
##   
## Chisq= 0.5 on 1 degrees of freedom, p= 0.5

#There was a insignificant difference in overall survival among the cohab categories, as chisq = 0.487 p-value = 0.5

# 7  
fit\_cox <- coxph(Surv(marriage, evermarr) ~ cohab, data = df)  
fit\_cox

## Call:  
## coxph(formula = Surv(marriage, evermarr) ~ cohab, data = df)  
##   
## coef exp(coef) se(coef) z p  
## cohab -0.2386 0.7877 0.2705 -0.882 0.378  
##   
## Likelihood ratio test=0.78 on 1 df, p=0.3764  
## n= 59, number of events= 59

# The hazard ratio is 0.78 indicates that cohabit participants have a lower hazard of participants who not cohabit and the p-value = 0.3764, we concluded that cohabitation doesn’t predict the time to marry.

# 8  
fit\_cox1 <- coxph(Surv(marriage, evermarr) ~ .-caseid, data = df)  
fit\_cox1

## Call:  
## coxph(formula = Surv(marriage, evermarr) ~ . - caseid, data = df)  
##   
## coef exp(coef) se(coef) z p  
## attend14 -0.57139 0.56474 0.37851 -1.510 0.131  
## cohab -0.08747 0.91624 0.28726 -0.305 0.761  
## educate -0.10003 0.90481 0.42000 -0.238 0.812  
## race -0.70730 0.49297 0.47425 -1.491 0.136  
##   
## Likelihood ratio test=3.64 on 4 df, p=0.4571  
## n= 59, number of events= 59

# The p-value of educate is greater than 0.05, we can concluded that educate cannot predict time to marriage.