Stat538 HW4 - Cox Models

Arthur Lui

February 7, 2014

Introduction:

To study the association between the covariates *Age, Race, Poverty, Smoing, Alcohol, Education, and Prenatal*, as defined in the description text file provided, analysis was done and a Cox Proportional Hazards Model was fitted to the data. Some variable selection was performed to determine which covariates were included in the model.

Results:

The hazard rate for weaning mothers in poverty is 81% that of mothers not in poverty (significant at the .05 level).

The hazard rate for weaning black mothers is 1.2 times that of white mothers. The hazard rate for weaning mothers that are not black is 1.36 times that of white mothers (significantly higher at the 95% confidence level).

The hazard rate for weaning mothers who smoked around the time of giving birth is 1.3 times that of mothers that did not smoke around the time of giving birth (significant at the 95% confidence level).

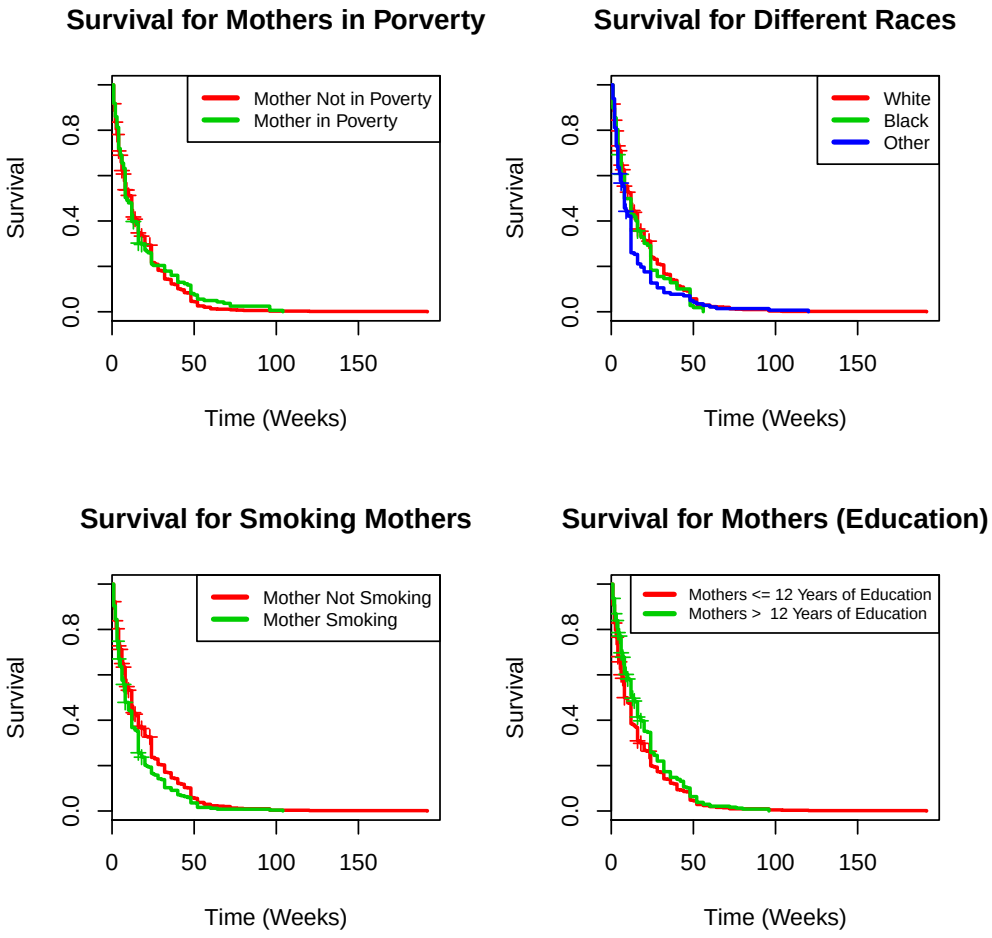
The Kaplan-Meier curve for Education was plotted using an indicator variable for whether the mother had received 12 or fewer years of education (approximately the number of years of elementary & high school education) or more than 12 years of education (e.g. tertiary education, trade school). The hazard rate for mothers with more than 12 years of education was 96% that of mothers with 12 or fewer years of education (marginally not significant at the 95% confidence level).

I checked for the significance of interaction terms and found that only one interaction pair, education and poverty, had a marginally insignificant effect at the 95% confidence level. To keep the model simple, it was excluded from the model.

Table 1: Summary Table For Cox Model

	coef	exp(coef)	se(coef)	z	p
race2	0.18	1.20	0.10	1.71	0.09
race3	0.31	1.36	0.10	3.15	0.00
poverty1	-0.21	0.81	0.09	-2.27	0.02
smoked1	0.26	1.30	0.08	3.38	0.00
education	-0.04	0.96	0.02	-1.95	0.05

Plots:



Conclusions:

Further investigation can be conducted to discover the relationship between the significant covariates in the cox model and the survival functions.

A Possible reasons for the significant difference in hazard ratios between mothers in poverty and mothers not in poverty is that mothers in poverty may find it much more difficult to provide solid food rather than breastmilk for their babies. Poverty is associated with a lower hazard rate, but the underlying reason that it takes children longer to be weaned may be that parents have no choice but to feed them solid foods.

SAS Output:

Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
race	2	1	0.15855	0.10409	2.3204	0.1277	1.172	race 2
race	3	1	0.29066	0.09707	8.9658	0.0028	1.337	race 3
poverty		1	-0.19011	0.09218	4.2537	0.0392	0.827	
smoked		1	0.25443	0.07828	10.5655	0.0012	1.290	
education		1	-0.03363	0.01909	3.1041	0.0781	0.967	

Appendix:

R Code:

```
nurse <- read.csv("nursing.csv")

library(survival)
#pairs(nurse)

nurse <- nurse[,-8]
nurse$race <- as.factor(nurse$race)
nurse$poverty <- as.factor(nurse$poverty)
nurse$smoked <- as.factor(nurse$smoked)
nurse$alcohol <- as.factor(nurse$alcohol)
nurse$prenatal <- as.factor(nurse$prenatal)

nurse.full <- coxph(Surv(duration,completion) ~ .,data=nurse,x=T)
# poverty:education kinda significant

nurse.mod <- coxph(Surv(duration,completion) ~ race+poverty+smoked+
                                               education+poverty*education,
                                               data=nurse,x=T)
# poverty:education not significant
nurse.red <- coxph(Surv(duration,completion) ~ race+poverty+smoked+education,data=nurse,x=T)

test.stat <- -2*(nurse.red$loglik[2] - nurse.full$loglik[2])
p.val <- pchisq(test.stat,df=length(nurse.full$coefficients)-
                length(nurse.red$coefficients),lower.tail=F)


# PLOTS:
# Poverty:
plot.pov <- function(cex){
  pov.comp <- survfit(Surv(duration,completion) ~ poverty,
                      type="kaplan-meier",data=nurse)
  plot(pov.comp,col=2:3,lwd=2,main="Survival for Mothers in Porverty",
       xlab="Time (Weeks)",ylab="Survival")
  legend("topright",legend=c("Mother Not in Poverty","Mother in Poverty"),
         col=2:3,lwd=3,cex=cex)
}

# Race:
plot.race <- function(cex){
  race.comp <- survfit(Surv(duration,completion) ~ race,
                       type="kaplan-meier",data=nurse)
  plot(race.comp,col=2:4,lwd=2,main="Survival for Different Races",
       xlab="Time (Weeks)",ylab="Survival")
  legend("topright",legend=c("White","Black","Other"),cex=cex,col=2:4,lwd=3)
}

# Smoked:
plot.smoked <- function(cex){
```

```
  smoked.comp <- survfit(Surv(duration,completion) ~ smoked,
                         type="kaplan-meier",data=nurse)
  plot(smoked.comp,col=2:3,lwd=2,main="Survival for Smoking Mothers",
       xlab="Time (Weeks)",ylab="Survival")
  legend("topright",legend=c("Mother Not Smoking",
                             "Mother Smoking"),
                             cex=cex,col=2:4,lwd=3)
}

# Education:
plot.edu <- function(cex){
  edu <- as.factor((nurse$education > 12) * 1)
  #edu <- nurse$education
  education.comp <- survfit(Surv(duration,completion) ~ edu,
                            type="kaplan-meier",data=nurse)
  plot(education.comp,col=2:3,lwd=2,main="Survival for Mothers (Education)",
       xlab="Time (Weeks)",ylab="Survival")
  legend("topright",legend=c("Mothers <= 12 Years of Education",
                             "Mothers >  12 Years of Education"),
                              col=2:4,lwd=3,cex=cex)
}
```

SAS Code: