PB HLTH 250C: Homework 3

Turn in by 9:30am Tuesday 12 March 2019

Read all questions carefully before answering. You may work in small groups of no more than 3 individuals and turn in a single assignment (and everyone in the group will receive the same grade). Work through the entire assignment individually first, then come together to discuss and collaborate. Please type your responses, **show your work, and please keep answers brief.** Also, **do not submit computer code, or unformatted output.**

Directions:

Parametric survival models:

Use the dataset `frmgham_recoded.Rdata` and code provided herein to explore the relationship between body mass index (BMI) status at baseline and time to death in the Framingham cohort using parametric time-to-event models.

The relevant variables for this analysis are:

- `timedth_yrs` (time of death)
- `death` (indicator of death [=1] or censored [=0])
- `bmi_cat` (categories for BMI <18.5 (underweight), 18.5-<25 (ideal weight), 25-<30 (overweight), 30+ (obese))
- `cursmoke` (indicator of current smoking status: yes (1) vs. no(0))
- `age` (age in years)
- `male` (variable denoting male (1) or female (0))

```
library(survival)
library(car)

load("frmgham_recoded.Rdata")

# Keeps first observation of each subject:
frmgham_recoded <- frmgham_recoded[which(frmgham_recoded$period==1),]

# Makes 18.5-24.9 reference category for BMI
frmgham_recoded$bmi_cat <- relevel(as.factor(frmgham_recoded$bmi_cat),2)
```

- Use the code below to fit a Cox proportional hazards model, an exponential model, and a Weibull model.

```
# Cox model (for reference)
survmod.cox.adj <- coxph(Surv(timedth_yrs, death)~ as.factor(bmi_cat) +
                           age + male + factor(cursmoke),
                         data=frmgham_recoded,
                         method="efron")
```

```
summary(survmod.cox.adj)

# Exponential model
survreg.exp.adj <- survreg(Surv(timedth_yrs, death)~ as.factor(bmi_cat) +
                             age + male + factor(cursmoke),
                           data=frmgham_recoded,
                           dist="exponential")
summary(survreg.exp.adj)

# Weibull model
survreg.weib.adj <- survreg(Surv(timedth_yrs, death)~ as.factor(bmi_cat) +
                              age + male + factor(cursmoke),
                            data=frmgham_recoded, dist="weibull")
summary(survreg.weib.adj)
```

- Calculate hazard ratios and 95% confidence intervals for the parametric survival models:

```
### Calculating log-HRs and SE(log-HR) for Exponential AFT model:
# The functions of the regression parameters from AFT -> PH
g <- c("-a2","-a3","-a4")

# The alpha estimates from the Exponential AFT model
a <- coef(survreg.exp.adj)
v.a <- vcov(survreg.exp.adj) # Covariance matrix of alphas

names(a) <- paste("a", 1:length(a), sep="")
rownames(v.a) <- colnames(v.a) <- names(a)

# Calculate ln(HR) and SEs:
lnhr.se.exp <- rbind(deltaMethod(a, g[1], vcov=v.a),  # UW
                     deltaMethod(a, g[2], vcov=v.a),  # OW
                     deltaMethod(a, g[3], vcov=v.a))  # OB
lnhr.se.exp
# HR and 95% CIs:
round(exp(lnhr.se.exp),2)[,c(1,3,4)]

### Calculating log-HRs and SE(log-HR) for Weibull AFT model:
# The functions of the regression parameters from AFT -> PH
g <- c("-a2/exp(lscale)","-a3/exp(lscale)","-a4/exp(lscale)")

# The alpha estimates from the Exponential AFT model
a <- c(coef(survreg.weib.adj), survreg.weib.adj$icoef[2])
v.a <- vcov(survreg.weib.adj) # Covariance matrix of alphas

names(a) <- c(paste("a", 1:(length(a)-1), sep=""), "lscale")
rownames(v.a) <- colnames(v.a) <- names(a)

# Calculate ln(HR) and SEs:
```

```
lnhr.se.weib <- rbind(deltaMethod(a, g[1], vcov=v.a),  # UW
                      deltaMethod(a, g[2], vcov=v.a),  # OW
                      deltaMethod(a, g[3], vcov=v.a))  # OB
lnhr.se.weib
# HR and 95% CIs:
round(exp(lnhr.se.weib),2)[,c(1,3,4)]
```

On your own - Perform a likelihood ratio test to determine if the Weibull model is more appropriate than the exponential model.

Competing Risks Models[1]

For this analysis we will explore the relationship between death from cardiovascular disease (CVD) and BMI, accounting for other causes of death.

First, load the dataset.

```
load("cvdrisk.Rdata") # data frame called cvdrisk.dat
library(survival)
library(cmprsk) # You may need to install.packages("cmprsk") before this
```

The relevant variables for this analysis are:

- `id` (Study ID)
- `time` (Follow-up time [days])
- `age` (age in years)
- `bmi` (Body mass index, kg/m^2)
- `sex` (variable denoting male (0) or female (1))
- `ev_typ` (Event Type; 1=CVD, 2=Other Cause, 0=Censored)

Adapting the code presented in the lecture, complete the following tasks:

Recode variables

Create a variable for overall mortality (death from any cause), and a variable that represents standard BMI categories:

```
cvdrisk.dat$death.any <- as.integer(cvdrisk.dat$ev_typ==1 |
                                    cvdrisk.dat$ev_typ==2)
# cvdrisk.dat$bmi.cat <- as.integer(cvdrisk.dat$bmi>=25)
cvdrisk.dat <- cvdrisk.dat[cvdrisk.dat$bmi >=18.5,]
cvdrisk.dat$bmi.cat <- cut(cvdrisk.dat$bmi,
                           breaks=c(18.5,25,30,1000), right=FALSE,
                           include.lowest = TRUE)
cvdrisk.dat$bmi.cat <- relevel(cvdrisk.dat$bmi.cat,ref="[18.5,25)")
```

[1] Sample data from Hosmer, Lemeshow and May (2008) *Applied Survival Analysis*. Chapter 9.

Describing survival data

- Referring to the code from the lecture notes and below, **plot the pooled (single group) cumulative incidence function (CIF) and Kaplan-Meier failure function on the same graph** for these data:

```
# Estimate CIF
cif <- cuminc(cvdrisk.dat$time, cvdrisk.dat$ev_typ, cencode=0)
cif.cvd <- cbind(cif["1 1"]$`1 1`$time, cif["1 1"]$`1 1`$est)

# Estimate KM Failure Function (manually calculate to overlay graphs)
km <- survfit(Surv(time, as.integer(ev_typ==1))~1, data=cvdrisk.dat)
km.F <- 1-km$surv
km.time <- km$time

# Plot CIF:
plot(cif.cvd[,1], cif.cvd[,2], bty="l",
     main="CVD Mortality", xlab="Days of Follow-Up",
     ylab="Probability of CVD Death", type="l",
     lty="solid",lwd=2)
# Add KM Failure function:
lines(km.time, km.F, type="l", lty=2, lwd=2, bty="l")
legend("bottomright",
       legend=c("Cumulative Incidence Function for CVD",
                "Kaplan-Meier Failure Function for CVD"),
       lty=c(1,2),
       lwd=2, cex=1, bty="o")
```

Proportional hazards modeling

- Referring to the code from the lecture notes and below, estimate the **cause-specific** CVD and other cause of death hazard ratios using a Cox proportional hazards model with the binary BMI variable as exposure, and adjusting for age and sex.

```
# Create cause-specific death indicators
# censoring at other causes:
cvdrisk.dat$cvd <- as.integer(cvdrisk.dat$ev_typ==1)
cvdrisk.dat$other <- as.integer(cvdrisk.dat$ev_typ==2)

cs.cvd <- coxph(Surv(time, cvd)~bmi.cat+age+sex, ties="efron",
                data=cvdrisk.dat)
summary(cs.cvd)

cs.other <- coxph(Surv(time, other)~bmi.cat+age+sex, ties="efron",
                  data=cvdrisk.dat)
summary(cs.other)
```

- Referring to the code from lecture, estimate the corresponding **subdistribution hazard ratios** from the Fine-Gray model.

```r
# Create design matrix and remove intercept:
X.crr <- model.matrix(~ bmi.cat + age + sex, data=cvdrisk.dat)[,-1]

# Adjusted models:
sd.cvd <- crr(cvdrisk.dat$time, cvdrisk.dat$ev_typ, X.crr,
              failcode=1, cencode=0)
summary(sd.cvd)

sd.other <- crr(cvdrisk.dat$time, cvdrisk.dat$ev_typ, X.crr,
                failcode=2, cencode=0)
summary(sd.other)
```

Questions:

Parametric survival analysis

1. Assuming *Weibull* distributed event times, **(15 points)**
   a. Write out the *general* expression, not substituting any estimated values, and clearly defining any paramters and distributions for any random terms, the **log-hazard function**, including the complete baseline hazard.
   b. Repeat the above for the **log-time function** (in the accelerated failure time metric).
   c. For an arbitrary covariate, show the expression of the hazard ratio from the proportional hazards Weibull model as a function of parameters from the accelerated failure time expression of the same model.
2. Complete the following table using the results from your analyses **(30 points)**:

BMI	Cox HR (95% CI)	Cox SE_β	Exponential HR (95% CI)	Exp SE_β	Weibull HR (95% CI)	Weib SE_β
<18.5						
18.5-<25	1	—	1	—	1	—
25-<30						
≥ 30						

3. Answer the following questions **(15 points)**:
   a. Which model estimates the relationships of interest most precisely? Justify your answer.
   b. Based on the likelihood ratio test, what parameter from the model you outlined in Q1 is being evaluated? Based on the results of this test, would you select the exponential or Weibull model? Justify your answer.
   c. Using the output from the Weibull model, calculate the **time ratio** comparing individuals with BMI ≥ 30 to those with BMI 18.5-<25 (no need for confidence interval). Interpret this parameter. Does this agree or not with the corresponding hazard ratio?

Competing risks

4. Conceptually, what is the difference between the KM failure estimate for CVD death and the estimated CIF for CVD death? What does the comparison of these curves tell you about the risk of the competing event? (**10 points**)
5. In a competing risks analysis, **briefly** (1-2 sentences each) define in words the following terms **(10 points)**:
   - **Cause-specific hazard**
   - **Subdistribution hazard**

(see next page)

6. Complete the following table for the cause-specific hazard ratios (csHRs) and subdistribution HRs (sHRs) you estimated **(10 points)**:

BMI	CVD csHR (95% CI)	Other csHR (95% CI)	CVD sHR (95% CI)	Other sHR (95% CI)
18.5-<25	1	1	1	1
25-<30				
≥ 30				

7. From the above table, explain how the pattern you observe in the csHRs is consistent with the pattern in the sHRs. (Consider the relationship between the csHR and sHRs.) **(10 points)**