Bios 323 Lab IX

Topic

- Cox proportional hazards model 1

Data Set

bmt: bone marrow transplant

group: Disease Group. 1-ALL, 2-AML Low Risk, 3-AML High Risk
t1: Time to Death or On Study Time
d1: Death Indicator (1-Dead, 0-Alive)
z2: Donor Age in Years
z7: Waiting time to transplant in days
z10: MTX Used as a Graft-Versus-Host-Prophylactic (1-Yes, 0-No)

R

```
# import data #
> library(KMsurv)
> data(bmt)
> names(bmt)[which(names(bmt) %in% c("z2", "z7", "z10"))] <- c("age", "waiting", "mtx")
> library(survival)

# Research question: is there a difference in survival after transplant between the three #
# disease groups? We also expect survival time will be affected by donor's age, MTX use    #
# and waiting time to transplant #
# fit Cox PH model #
> f1 <- coxph(Surv(t1, d1) ~ factor(group) + mtx + age + I(age^2) + waiting, data=bmt)
> summary(f1)

# ways to handle ties #
> with(bmt, sum(duplicated(t1) & d1==1))  # number of ties
> system.time(f1 <- coxph(Surv(t1, d1) ~ factor(group) + mtx + age + I(age^2) + waiting,
+ data=bmt, method="efron"))
> system.time(f2 <- coxph(Surv(t1, d1) ~ factor(group) + mtx + age + I(age^2) + waiting,
+ data=bmt, method="exact"))
> system.time(f3 <- coxph(Surv(t1, d1) ~ factor(group) + mtx + age + I(age^2) + waiting,
+ data=bmt, method="breslow"))

# test on disease group #
# likelihood ratio test #
f1a <- update(f1, . ~ . - factor(group))
(p.lik <- 1-pchisq(2*(f1$loglik[2] - f1a$loglik[2]), df=2))

# wald test #
stat <- t(f1$coef[1:2]) %*% solve(f1$var[1:2,1:2]) %*% f1$coef[1:2]
(p.wald <- 1-pchisq(as.numeric(stat), df=2))
```

```
# score test #
f1b <- update(f1, init=c(0, 0, f1a$coef), iter=0)
stat <- f1b$score
(p.score <- 1-pchisq(stat, df=2))

# estimate of (baseline) survival function #
> bmt$age.s <- bmt$age^2
> bmt$group1 <- ifelse(bmt$group==1, 1, 0)
> bmt$group2 <- ifelse(bmt$group==2, 1, 0)
> f1c <- update(f1, .~.- factor(group) - I(age^2) + age.s + group1 + group2)

> s1 <- basehaz(f1c)
> s2 <- survfit(f1c, conf.type="none")
> s3 <- survfit(f1c, newdata=data.frame(group1=mean(bmt$group1), group2=mean(bmt$group2),
age=mean(bmt$age), age.s=mean(bmt$age.s), mtx=mean(bmt$mtx), waiting=mean(bmt$waiting)), conf.type="none")

> exp(-s1$hazard[s1$time %in% summary(s2)$time])
> summary(s2)$surv
> summary(s3)$surv

> plot(s2, mark.time=FALSE, ylab="S", xlab="days")
> plot(s2, mark.time=FALSE, fun="cumhaz", ylab="Cumulative Hazards", xlab="days")

> s4 <- survfit(f1, newdata=data.frame(age=mean(bmt$age), mtx=median(bmt$mtx), group=c(1,2,3),
+ waiting=mean(bmt$waiting)), conf.type="none")
> s5 <- survfit(Surv(t1, d1) ~ factor(group), data=bmt, conf.type="none")
> plot(s4, mark.time=FALSE, col=c("black", "red","blue"))
> legend("topright", lty=1, col=c("black", "red","blue"), legend=c("group=1", "group=2", "group=3"))
> par(new=TRUE)
> plot(s5, mark.time=FALSE, col=c("black", "red","blue"), lty=2)

# trend test #
> bmt$age2 <- cut(bmt$age, breaks=c(0,20,30,40,100))
> plot(survfit(Surv(t1,d1) ~ age2, data=bmt), lty=1:4, mark.time=FALSE, ylab="S", xlab="Days")
> legend("topright", lty=1:4, legend=c("<20", "20~30", "30~40", ">40"))
> f4 <- coxph(Surv(t1, d1) ~ as.numeric(age2), data=bmt, method="exact")
> summary(f4)$sctest
> f5 <- survdiff(Surv(t1, d1) ~ age2, data=bmt)
> zt <-  f5$obs - f5$exp
> sigma <- f5$var
> alp <- c(1,2,3,4)
> stat <- alp %*% zt /sqrt(alp %*% sigma %*% alp)
> (p.value <- 2*(1-pnorm(abs(stat))))
```

Exercise

1. Continuing with the model f1,

    (a) After accouting for donor's age, waiting time and MTX use, what is estimated relative risk of dying for patients of disease group 3 compared to disease group 2? And 95% CI?

(b) Patient A and B both used MTX, belonged to disease group 2, and had donors aged 30, but patient A waited for 500 days to transplant, and patient B waited for 100 days. What is the hazard ratio for these two patients?

(c) We also suspect MTX use may affect the survival of patients in the three disease groups differently. How would you test this?

(d) Use the method described in section 8.7 in the textbook and decide which other variables in the data set (z1, z3, z4, z5, z6, z8, z9) should also be included in the model.

2. Perform log-rank test on disease group using survdiff(). Is the p-value exactly the same as that of the score test from a univariable Cox PH model on disease group?

```
# 1(a) make inference on b2-b1 #
> x <- f1$coef[2] - f1$coef[1]
> v <- t(c(1,-1)) %*% f1$var[1:2,1:2] %*% c(1,-1)
> lower <- x - 1.96*sqrt(v)
> upper <- x + 1.96*sqrt(v)
> hr <- exp(x)
> hr.lower <- exp(lower)
> hr.upper <- exp(upper)
> stat <- x^2/v
> p.value <- 1-pchisq(stat, df=1)
# or set disease group 2 as reference level #
> f1d <- update(f1, .~.-factor(group) + factor(group, levels=c(2,1,3)))
> summary(f1d)

# 1(b) make inference on 400*b6 #
> hr <- exp(400*f1$coef[6])
> hr.lower <- summary(f1)$conf.int[6,3]*exp(400)
> hr.upper <- summary(f1)$conf.int[6,4]*exp(400)

# 1(c) include interaction terms #
> f1e <- update(f1, .~.+factor(group):mtx)
> anova(f1e, f1)

# 1(d) forward selection of variables by AIC #
> library(MASS)
> f.final <- stepAIC(f1, scope=list(upper=~factor(group)+mtx+age+I(age^2)+waiting+z1+z3+z4+z5+z6+z8+z9,
+ lower=~factor(group)+mtx+age+I(age^2)+waiting), trace=TRUE,direction="forward")
> f.final
> extractAIC(f.final)
```