BIOS 6312: Modern Regression Analysis

Andrew J. Spieker, Ph.D.

Assistant Professor of Biostatistics
Vanderbilt University Medical Center

Set 5 supplementary slides for `R` enthusiasts

Version: 01/25/2021

Examples for Set 5

Examples for R enthusiasts:

- Kaplan-Meier by gender (Slide 517)
- Log-rank test for gender (Slide 530)
- Cox regression by gender (Slide 541)
- Time-dependent Cox for heart transplant (Slide 565)
- Cumulative incidence of cardiovascular death (Slide 577)

Examples for Set 5

Examples for R enthusiasts:

- **Kaplan-Meier by gender (Slide 517)**
- Log-rank test for gender (Slide 530)
- Cox regression by gender (Slide 541)
- Time-dependent Cox for heart transplant (Slide 565)
- Cumulative incidence of cardiovascular death (Slide 577)

Kaplan-Meier by gender

Reading in the MRI data:

- Read in data:

```
mri.data <- read.csv("mri.csv",
                     header = TRUE,
                     stringsAsFactors = FALSE)
```

Kaplan-Meier by gender

Survival library:

- Many methods to model time-to-event data rely on the `survival` package.

```
library("survival")
```

Kaplan-Meier by gender

Kaplan-Meier estimation:

- The `survfit` function allows us to perform Kaplan-Meier estimation.

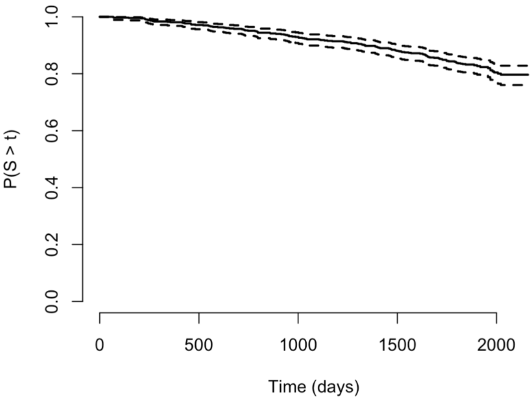
```
s.overall <- survfit(Surv(obstime, death) ~ 1,
                     conf.type = "log-log",
                     data = mri.data)

plot(s.overall,
     frame.plot = FALSE,
     col = c("black"),
     lwd = 2,
     xlab = "Time (days)",
     ylab = "P(S > t)")
```

- Confidence intervals from a log-log transformation will allow us to come close to Stata's results.

Kaplan-Meier by gender

Kaplan-Meier estimation: Overall



- Confidence intervals included if only presenting one group.

Kaplan-Meier by gender

Kaplan-Meier estimation:

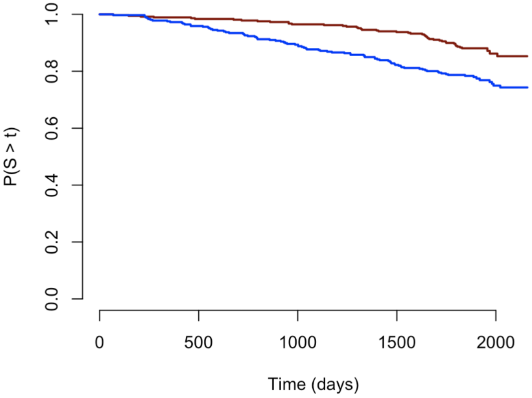
- Curves stratified by gender.

```
s.gender <- survfit(Surv(obstime, death) ~ male,
                    conf.type = "log-log",
                    data = mri.data)

plot(s.gender,
     frame.plot = FALSE,
     conf.int = FALSE,
     col = c("darkred", "blue"),
     lwd = c(2,2),
     xlab = "Time (days)",
     ylab = "P(S > t)")
```

Kaplan-Meier by gender

Kaplan-Meier estimation: By gender



- Confidence intervals not included if presenting 2+ groups.

Kaplan-Meier by gender

Kaplan-Meier estimation:

- Extracting restricted mean (overall).

```
summary(s.overall)$table[5:6]

## Output
    *rmean *se(rmean)
1974.46913   16.56625
```

- Agrees with Stata output (Slide 520).

Kaplan-Meier by gender

Kaplan-Meier estimation:

- Extracting restricted mean (by gender).

```
summary(s.gender)$table[,5:6]

## Output
          *rmean *se(rmean)
male=0 2049.954   17.55878
male=1 1899.064   27.59734
```

- Approximately agrees with Stata output (Slide 521).

Kaplan-Meier by gender

Kaplan-Meier estimation:

- Extracting quantiles (overall).

```
quantile(s.overall, 0.10)

## Output
$quantile
  10
1338

$lower
  10
1045

$upper
  10
1519
```

- Agrees with Stata output (Slide 522).

Kaplan-Meier by gender

Kaplan-Meier estimation:

- Extracting quantiles (by gender).

```
quantile(s.gender, 0.20)

## Output
$quantile
         20
male=0   NA
male=1 1707

$lower
         20
male=0   NA
male=1 1457

$upper
         20
male=0   NA
male=1 1988
```

- Approximately agrees with Stata output (Slide 523).

Examples for Set 5

Examples for R enthusiasts:

- *Kaplan-Meier by gender (Slide 517)*
- **Log-rank test for gender (Slide 530)**
- Cox regression by gender (Slide 541)
- Time-dependent Cox for heart transplant (Slide 565)
- Cumulative incidence of cardiovascular death (Slide 577)

Log-rank test for gender

Log-rank test:

- Function for log-rank test: `survdiff`.

```
logrank.gender

## Output
Call:
survdiff(formula = Surv(obstime, death) ~ male, data = mri.data)

         N Observed Expected (O-E)^2/E (O-E)^2/V
male=0 369       47     68.8      6.89      14.3
male=1 366       86     64.2      7.38      14.3

 Chisq= 14.3  on 1 degrees of freedom, p= 2e-04
```

- Agrees with Stata output (Slide 530).

Examples for Set 5

Examples for R enthusiasts:

- *Kaplan-Meier by gender (Slide 517)*
- *Log-rank test for gender (Slide 530)*
- **Cox regression by gender (Slide 541)**
- Time-dependent Cox for heart transplant (Slide 565)
- Cumulative incidence of cardiovascular death (Slide 577)

Cox regression by gender

Proportional hazards regression:

- Function for Cox model: `coxph`.

```
model.gen <- coxph(Surv(obstime, death) ~ male,
                   ties = "breslow",
                   data = mri.data)
N <- length(model.gen$residuals)
robust.var <- sandwich(model.gen) * N/(N - 1)
```

Cox regression by gender

Proportional hazards regression:

- Results:

```
exp(c(HR = model.gen$coef,
      CI.Low = model.gen$coef - qnorm(0.975) * sqrt(robust.var),
      CI.Hi = model.gen$coef + qnorm(0.975) * sqrt(robust.var)))

## Output
 HR.male   CI.Low    CI.Hi
1.961765 1.378102 2.792624
```

- Agrees with Stata output (Slide 543).

Examples for Set 5

Examples for R enthusiasts:

- *Kaplan-Meier by gender (Slide 517)*
- *Log-rank test for gender (Slide 530)*
- *Cox regression by gender (Slide 541)*
- **Time-dependent Cox for heart transplant (Slide 565)**
- Cumulative incidence of cardiovascular death (Slide 577)

Time-dependent Cox for heart transplant

Reading in the transplant data:

- Read in data:

```
heart.data <- read.csv("transplant.csv",
                       header = TRUE,
                       stringsAsFactors = FALSE)
```

Time-dependent Cox for heart transplant

Structuring data:

- Create variable for initial time windows.

```
N <- dim(heart.data)[1]
heart.data$ptime <- c(0, heart.data$time[1:(N - 1)])
heart.data$ptime[heart.data$time == 1] <- 0
```

Time-dependent Cox for heart transplant

Time-dependent covariates:

- Account for clustering.

```
model.heart <- coxph(Surv(ptime, time, death) ~ transplant + cluster(id),
                     method = "breslow",
                     data = heart.data)
robust.var <- summary(model.heart)$coef[4]^2 * 21/20
```

Time-dependent Cox for heart transplant

Time-dependent covariates:

- Results:

```
exp(c(HR = model.heart$coef,
      CI.Low = model.heart$coef - qnorm(0.975) * sqrt(robust.var),
      CI.High = model.heart$coef + qnorm(0.975) * sqrt(robust.var)))

## Output
     HR.transplant  CI.Low.transplant CI.High.transplant
          0.261495           0.086369           0.791713
```

- Agrees with Stata output (Slide 566).

Examples for Set 5

Examples for R enthusiasts:

- *Kaplan-Meier by gender (Slide 517)*
- *Log-rank test for gender (Slide 530)*
- *Cox regression by gender (Slide 541)*
- *Time-dependent Cox for heart transplant (Slide 565)*
- **Cumulative incidence of cardiovascular death (Slide 577)**

Cumulative incidence of cardiovascular death

Subdistribution hazard in R:

- Competing risks regression can be performed using the `crr` function in the library `cmprsk`.
- We again use the MRI data.

```
library("cmprsk")

mri.data <- read.csv("mri.csv",
                     header = TRUE,
                     stringsAsFactors = FALSE)
```

Cumulative incidence of cardiovascular death

Structuring data:

- Create variable indicating censoring, event of interest, and competing event(s).

```
mri.data$event <- 0
mri.data$event[mri.data$death == 1 & mri.data$cvd == 1] <- 1
mri.data$event[mri.data$death == 1 & mri.data$cvd == 0] <- 2
```

Cumulative incidence of cardiovascular death

Subdistribution hazard regression:

- Account for competing risks.

```
model.diab <- crr(mri.data$obstime,
                  mri.data$event,
                  mri.data$diabetes)
N <- model.diab$n
robust.var <- summary(model.diab)$coef[3]^2 * N/(N - 1)
```

Cumulative incidence of cardiovascular death

Subdistribution hazard regression:

- Results:

```
exp(c(summary(model.diab)$coef[1],
      summary(model.diab)$coef[1] - qnorm(0.975) * sqrt(robust.var),
      summary(model.diab)$coef[1] + qnorm(0.975) * sqrt(robust.var)))

## Output
[1] 2.44409 1.34119 4.45391
```

- Agrees with Stata output (Slide 582).

Cumulative incidence of cardiovascular death

Plot cumulative incidence:

- Create figures:

```
plot(predict(model.diab, cov1 = 0),
     frame.plot = FALSE,
     ylim = c(0, 0.2),
     col = "darkblue",
     xlab = "Analysis time",
     ylab = "Cumulative incidence")

lines(predict(model.diab, cov1 = 1),
      col = "darkred")

legend(0, 0.20, col = c("darkblue", "darkred"),
       lwd = c(1,1), lty = c(1,1),
       c("diabetes = 0", "diabetes = 1"))
```

Cumulative incidence of cardiovascular death

Cumulative incidence: By diabetes status