Survival Analysis HW 3

Faizan Khalid Mohsin

March 14, 2019

Contents

1 Question 2 **2**
1.1 Question 2 a. 2
1.2 Question 2 c. 3

2 Question 3 **5**

1 Question 2

1.1 Question 2 a.

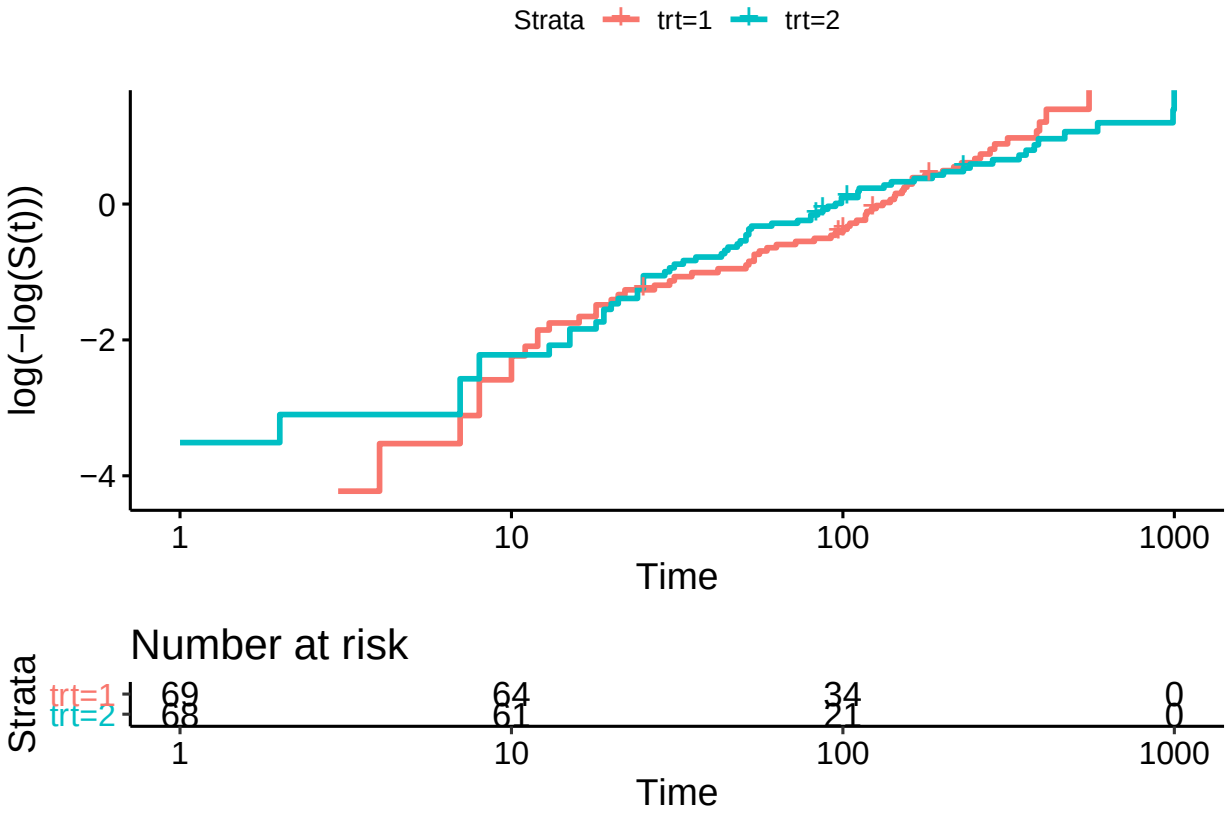
```
data(veteran)
# str(veteran)
# unique(veteran$trt)
# head(veteran)
# names(veteran)
table(veteran$trt, veteran$status)
```

```
##
##      0  1
##   1  5 64
##   2  4 64
```

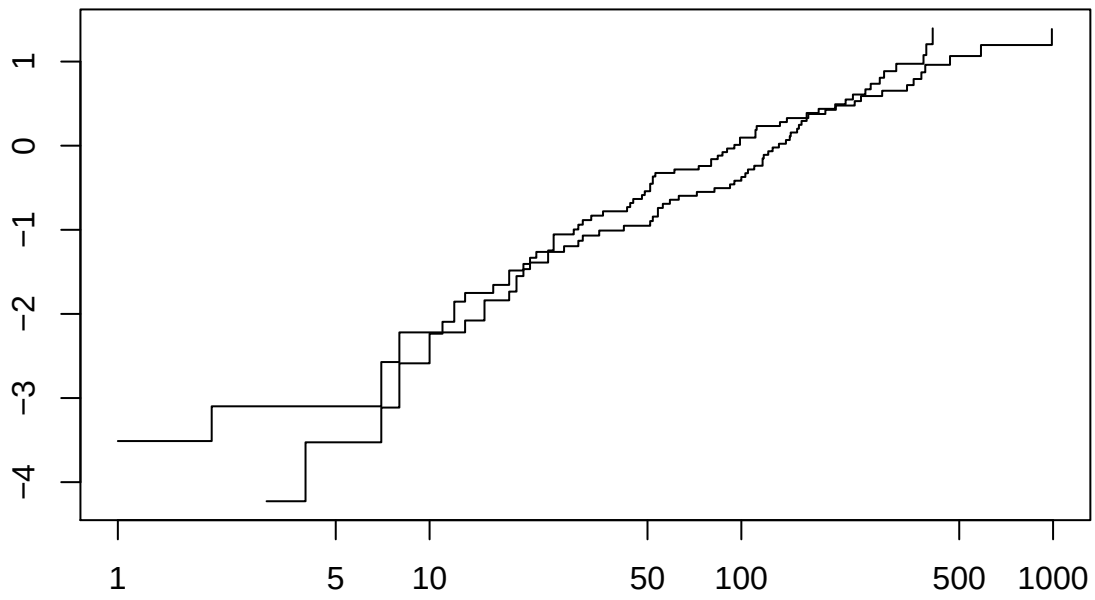
```
# No missing data.
all(!is.na(veteran)==TRUE)
```

```
## [1] TRUE
```

```
fit <- survfit(Surv(time, status) ~ trt , data=veteran )
ggsurvplot(fit, data = veteran, risk.table = TRUE, fun="cloglog")
```




```
#summary(fit)
plot(fit, fun="cloglog")
```



From the above plot we see that the two survival curves for the two treatments are practically parallel, hence, the proportionality of hazard functions in the context of Cox models is not violated.

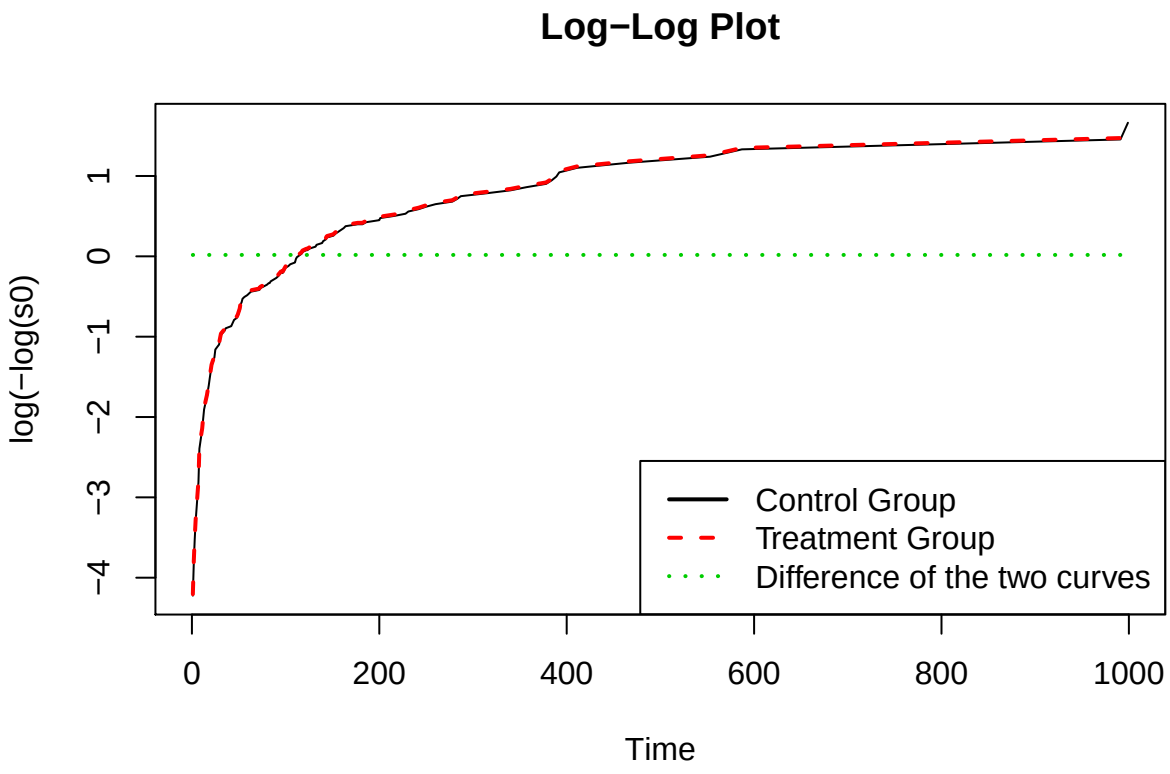
Is it problamatic if the curves cross?

1.2 Question 2 c.

```
model <- coxph(Surv(time, status) ~ trt , data=veteran)
summary(model)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ trt, data = veteran)
##
##   n= 137, number of events= 128
##
##        coef exp(coef) se(coef)     z Pr(>|z|)
## trt 0.01774   1.01790  0.18066 0.098    0.922
##
##     exp(coef) exp(-coef) lower .95 upper .95
## trt     1.018     0.9824    0.7144      1.45
##
## Concordance= 0.525  (se = 0.026 )
## Rsquare= 0   (max possible= 0.999 )
## Likelihood ratio test= 0.01  on 1 df,   p=0.9218
## Wald test            = 0.01  on 1 df,   p=0.9218
## Score (logrank) test = 0.01  on 1 df,   p=0.9218
```

```r
cox<-coxph(Surv(time, status) ~ trt , data=veteran)
b1<-cox$coefficients[1]
bh <- basehaz(model)
breslow <- bh
s0<-exp(-breslow[,1])
s1<-exp(-breslow[,1]*exp(b1))
difference = log(-log(s1)) - log(-log(s0))
plot(bh[,2],log(-log(s0)),
     type="l",lty=1, col=1,
     xlab="Time",
     ylab="log(-log(s0)",
     main="Log-Log Plot")
points(bh[,2], log(-log(s1)),col=2, type="l", lty=2, lwd=2)
points(bh[,2],difference, type="l", col=3, lty=3, lwd=2)
legend("bottomright", col=1:3,lty = 1:3 ,
       legend=c("Control Group", "Treatment Group", "Difference of the two curves")
       , lwd=2)
```



This is not useful for testing proportionallity of hazard functions because that is an assumption of the model. So once we fit the model which assumes proportionality, the estimates of the model will already have proportionality assumption reflected in them. Hence, there is no point in testing it this way.

2 Question 3

```
names(veteran)
```

```
## [1] "trt"      "celltype" "time"     "status"   "karno"    "diagtime"
## [7] "age"      "prior"
```

```
# "celltype" "time"     "status"   "karno"    "diagtime" "age"
# "prior"

model <- coxph(Surv(time, status) ~ strata(trt) + karno + age
               + as.factor(prior)
               + as.factor(celltype) + diagtime
               , data=veteran)


summary(model)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ strata(trt) + karno + age +
##     as.factor(prior) + as.factor(celltype) + diagtime, data = veteran)
##
##   n= 137, number of events= 128
##
##                                   coef  exp(coef)   se(coef)      z
## karno                         -0.0335972  0.9669609  0.0057067 -5.887
## age                           -0.0089083  0.9911313  0.0093653 -0.951
## as.factor(prior)10             0.0859863  1.0897914  0.2334348  0.368
## as.factor(celltype)smallcell  0.8619502  2.3677739  0.2838520  3.037
## as.factor(celltype)adeno       1.1895520  3.2856090  0.3137986  3.791
## as.factor(celltype)large       0.3657135  1.4415422  0.2946741  1.241
## diagtime                      -0.0007095  0.9992908  0.0092281 -0.077
##                               Pr(>|z|)
## karno                         3.93e-09 ***
## age                            0.34150
## as.factor(prior)10             0.71261
## as.factor(celltype)smallcell  0.00239 **
## as.factor(celltype)adeno       0.00015 ***
## as.factor(celltype)large       0.21458
## diagtime                       0.93872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                               exp(coef) exp(-coef) lower .95 upper .95
## karno                            0.9670     1.0342    0.9562    0.9778
## age                              0.9911     1.0089    0.9731    1.0095
## as.factor(prior)10               1.0898     0.9176    0.6897    1.7220
## as.factor(celltype)smallcell     2.3678     0.4223    1.3575    4.1301
## as.factor(celltype)adeno         3.2856     0.3044    1.7763    6.0775
## as.factor(celltype)large         1.4415     0.6937    0.8091    2.5684
## diagtime                         0.9993     1.0007    0.9814    1.0175
##
## Concordance= 0.728  (se = 0.043 )
## Rsquare= 0.346   (max possible= 0.998 )
```

```
## Likelihood ratio test= 58.09  on 7 df,   p=3.634e-10
## Wald test            = 56.11  on 7 df,   p=8.963e-10
## Score (logrank) test = 60.4  on 7 df,   p=1.259e-10
```

```
#basehaz(model)
```

```
martingaleres <- residuals(model, type=c('martingale'))
devianceres <- residuals(model, type=c('deviance'))
dfbeta <- residuals(model, type=c('dfbeta'))
dfbetas <- residuals(model, type=c('dfbetas'))

outpath <- '~/Dropbox/work/CHL5209H_2018/slides'

# Check martingale and deviance residuals for continuous covariates:

pdf(file.path(outpath, paste('martingale_age.pdf', sep='')), height=7, width=7, paper='special')
op <- par(mar=c(4.5,4.5,1,1))
plot(brain$age, martingaleres, xlab='Age', ylab='Martingale residual')
lines(lowess(brain$age, martingaleres), lwd=2, col='blue')
abline(h=0, lty='dotted')
par(op)
dev.off()

pdf(file.path(outpath, paste('martingale_interval.pdf', sep='')), height=7, width=7, paper='special')
op <- par(mar=c(4.5,4.5,1,1))
plot(brain$interval, martingaleres, xlab='Interval', ylab='Martingale residual')
lines(lowess(brain$interval, martingaleres), lwd=2, col='blue')
abline(h=0, lty='dotted')
par(op)
dev.off()

pdf(file.path(outpath, paste('deviance_age.pdf', sep='')), height=7, width=7, paper='special')
op <- par(mar=c(4.5,4.5,1,1))
plot(brain$age, devianceres, xlab='Age', ylab='Deviance residual')
lines(lowess(brain$age, devianceres), lwd=2, col='blue')
abline(h=0, lty='dotted')
par(op)
dev.off()

pdf(file.path(outpath, paste('deviance_interval.pdf', sep='')), height=7, width=7, paper='special')
op <- par(mar=c(4.5,4.5,1,1))
plot(brain$interval, devianceres, xlab='Interval', ylab='Deviance residual')
lines(lowess(brain$interval, devianceres), lwd=2, col='blue')
abline(h=0, lty='dotted')
par(op)
dev.off()

# Unscaled dfbeta influence measures:

pdf(file.path(outpath, paste('dfbeta_age.pdf', sep='')), height=7, width=7, paper='special')
op <- par(mar=c(4.5,4.5,1,1))
plot(brain$age, dfbeta[,3], xlab='Age', ylab='dfbeta', ylim=c(-2/sqrt(nrow(dfbeta)), 2/sqrt(nrow(dfbeta
lines(lowess(brain$age, dfbeta[,3]), lwd=2, col='blue')
abline(h=0, lty='dotted')
par(op)
```

```
dev.off()

pdf(file.path(outpath, paste('dfbeta_interval.pdf', sep='')), height=7, width=7, paper='special')
op <- par(mar=c(4.5,4.5,1,1))
plot(brain$interval, dfbeta[,4], xlab='Interval', ylab='dfbeta')
lines(lowess(brain$interval, dfbeta[,4]), lwd=2, col='blue')
abline(h=0, lty='dotted')
par(op)
dev.off()

# Scaled dfbeta influence measures (compare to the threshold of 2/sqrt(n)):

pdf(file.path(outpath, paste('dfbetas_age.pdf', sep='')), height=7, width=7, paper='special')
op <- par(mar=c(4.5,4.5,1,1))
plot(brain$age, dfbetas[,3], xlab='Age', ylab='dfbetas')
lines(lowess(brain$age, dfbetas[,3]), lwd=2, col='blue')
abline(h=c(-2/sqrt(nrow(dfbetas)), 0, 2/sqrt(nrow(dfbetas))), lty='dotted')
par(op)
dev.off()

pdf(file.path(outpath, paste('dfbetas_interval.pdf', sep='')), height=7, width=7, paper='special')
op <- par(mar=c(4.5,4.5,1,1))
plot(brain$interval, dfbetas[,4], xlab='Interval', ylab='dfbeta')
lines(lowess(brain$interval, dfbetas[,4]), lwd=2, col='blue')
abline(h=0, lty='dotted')
abline(h=c(-2/sqrt(nrow(dfbetas)), 0, 2/sqrt(nrow(dfbetas))), lty='dotted')
par(op)
dev.off()

# Checks for proportionality:

cox.zph(model, global=FALSE)

pdf(file.path(outpath, paste('schoenfeld_age.pdf', sep='')), height=7, width=7, paper='special')
op <- par(mar=c(4.5,4.5,1,1))
plot(cox.zph(model, global=FALSE), var=3)
abline(h=0, lty='dotted')
par(op)
dev.off()

pdf(file.path(outpath, paste('schoenfeld_interval.pdf', sep='')), height=7, width=7, paper='special')
op <- par(mar=c(4.5,4.5,1,1))
plot(cox.zph(model, global=FALSE), var=4)
abline(h=0, lty='dotted')
par(op)
dev.off()

pdf(file.path(outpath, paste('schoenfeld_male.pdf', sep='')), height=7, width=7, paper='special')
op <- par(mar=c(4.5,4.5,1,1))
plot(cox.zph(model, global=FALSE), var=8)
abline(h=0, lty='dotted')
par(op)
dev.off()
```

```r
# Examples of tests for covariate-time interactions:

model <- coxph(Surv(time, status) ~ trt + tt(trt) + resect75 + age + interval + karn + race + local + ma
               tt=function(x,t, ...) x * t, data=brain)
summary(model)

model <- coxph(Surv(time, status) ~ trt + resect75 + age + interval + karn + race + local + male + tt(ma
               tt=function(x,t, ...) x * t, data=brain)
summary(model)
```