Midterm Exam Math 6710

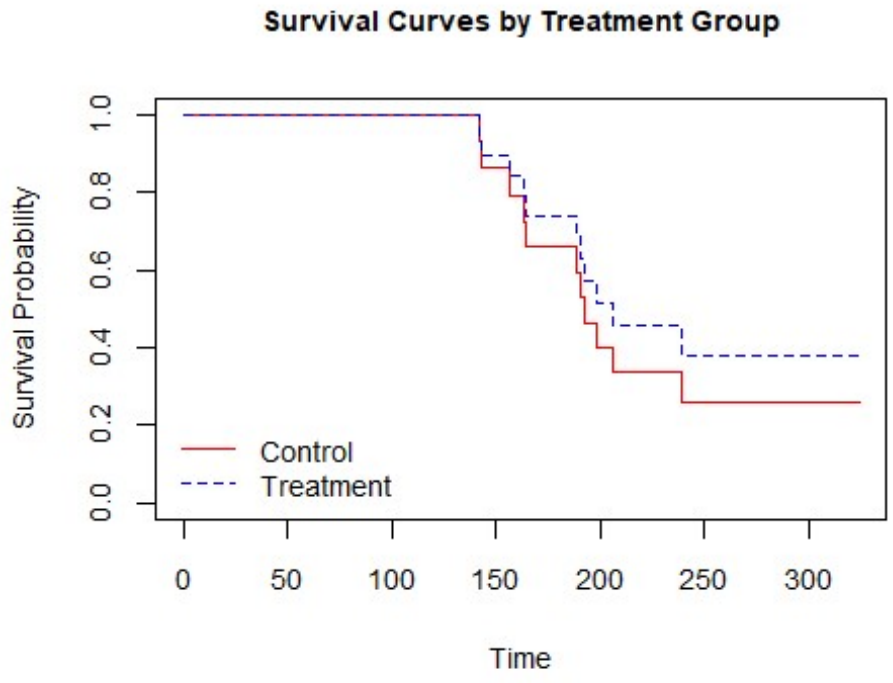
Wenjuan Bian

2023-02-25

```
#####
# 1. (30 points = 10+10+10) The time data (in days) from insult with the carcinogen
 DMBA to mortality from vaginal cancer in two groups of rats, 8 rats in each group
with a different pretreatment regimen were given as follows (the right-censored tim
es are indicated by an asterisk):
## Group 1: 143, 164, 188, 190, 192, 206, 216*, 244*
## Group 2: 142, 156, 163, 198, 239, 240*, 305*, 324*
#####
#(a) Consider Group 1 as the control group and Group 2 as the treatment group. Fit
the time data with the hazard function in relative risk Cox regression model.
#####

library(survival)
library(asaur)
library(survminer)

## Loading required package: ggplot2

## Loading required package: ggpubr

## Attaching package: 'survminer'

## The following object is masked from 'package:survival':
##
##     myeloma

library(numDeriv)
tt <- c(143, 164, 188, 190, 192, 206, 216, 244,142, 156, 163, 198, 239, 240, 305, 3
24)
delta <- c(1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0)
trt <- c(0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1)


middata <- data.frame(tt = c(143, 164, 188, 190, 192, 206, 216, 244,142, 156, 163,
198, 239, 240, 305, 324), delta = c(1, 1, 1, 1, 1, 1, 0, 0,1, 1, 1, 1, 1, 0, 0,0),
trt = c(0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1))

middata

##     tt delta trt
## 1  143     1   0
## 2  164     1   0
## 3  188     1   0
```

```
## 4  190     1   0
## 5  192     1   0
## 6  206     1   0
## 7  216     0   0
## 8  244     0   0
## 9  142     1   1
## 10 156     1   1
## 11 163     1   1
## 12 198     1   1
## 13 239     1   1
## 14 240     0   1
## 15 305     0   1
## 16 324     0   1
```

```
result.cox <- coxph(Surv(tt, delta) ~ trt, data=middata)
surv.cox <- survfit(result.cox, newdata=data.frame(trt=c(0, 1)))
plot(surv.cox, col=c("red", "blue"), lty=1:2, xlab="Time", ylab="Survival Probabili
ty", cex.axis = 0.9, cex.lab=0.9,
     main="Survival Curves by Treatment Group", cex.main = 0.9, ,lwd=1.5)
legend("bottomleft", legend=c("Control", "Treatment"), bty="n", col=c("red", "blue"
),cex = 0.9, lty=1:2)
```



```
####
# (b) Test the hypothesis in the Cox model at significance level α = 5% that the tw
o pretreatment regimens make no difference on the survival time.
#####

summary(result.cox)
```

```
## Call:
## coxph(formula = Surv(tt, delta) ~ trt, data = middata)
##
##   n= 16, number of events= 11
##
##        coef exp(coef) se(coef)      z Pr(>|z|)
## trt -0.3246    0.7228   0.6200 -0.523    0.601
##
##     exp(coef) exp(-coef) lower .95 upper .95
## trt    0.7228      1.383    0.2144     2.437
##
## Concordance= 0.523  (se = 0.087 )
## Likelihood ratio test= 0.28  on 1 df,   p=0.6
## Wald test            = 0.27  on 1 df,   p=0.6
## Score (logrank) test = 0.28  on 1 df,   p=0.6
```

By the cox model summary, the likelihood ratio test, Wald test, and Score (logrank) test all have p-values of 0.6, which indicates that the two pretreatment regimens make no significant difference. Therefore, we cannot reject the null hypothesis.

```
lgl <- logLik(result.cox)
result.cox.re <- coxph(Surv(tt, delta) ~ 1, data = middata)
lgl_re <- logLik(result.cox.re)
test_stat <- -2 * (lgl_re - lgl)
p_value <- pchisq(test_stat, df = 1, lower.tail = FALSE)

cat("Test statistic:", round(test_stat, 2), "\n")
```

```
## Test statistic: 0.28
```

```
cat("p-value:", format(p_value, scientific = FALSE), "\n")
```

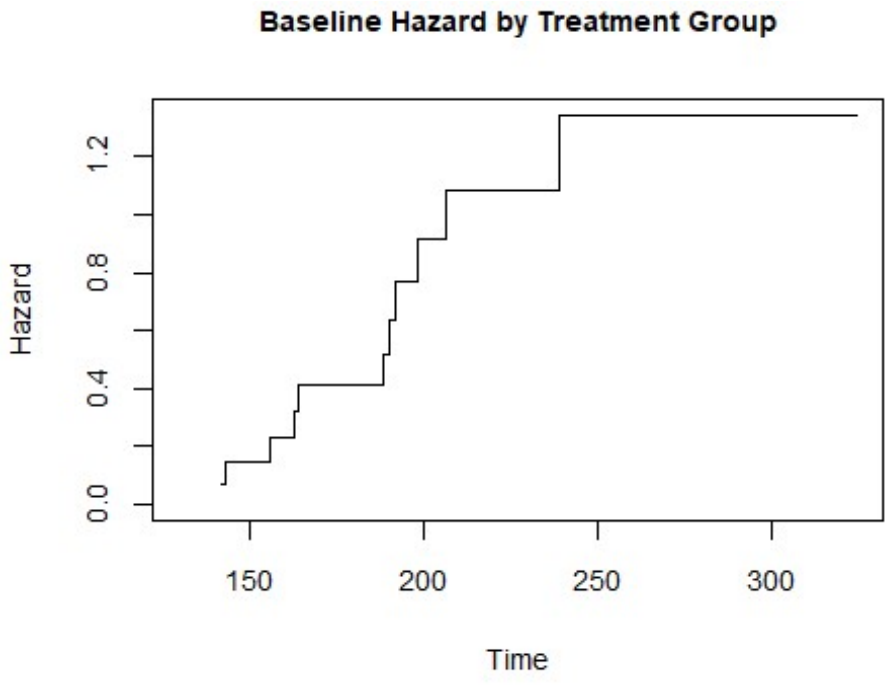
```
## p-value: 0.5993819
if (p_value < 0.05) {
  cat("The two pretreatment regimens make a significant difference on the survival
time (reject null hypothesis).\n")
} else {
  cat("The two pretreatment regimens do not make a significant difference on the su
rvival time (fail to reject null hypothesis).\n")
}
```

```
## The two pretreatment regimens do not make a significant difference on the surviv
al time (fail to reject null hypothesis).
```

```
######
# (c)Use the basehaz function in R to obtain an estimate of the baseline cumulative
 hazard function.
######

result.bh <- basehaz(result.cox, centered = F)
result.bh
```

```
##       hazard time
## 1  0.0725547  142
## 2  0.1491252  143
## 3  0.2320448  156
## 4  0.3202513  163
## 5  0.4144648  164
## 6  0.5184777  188
## 7  0.6345653  190
## 8  0.7658990  192
## 9  0.9170890  198
## 10 1.0868293  206
## 11 1.0868293  216
## 12 1.3438094  239
## 13 1.3438094  240
## 14 1.3438094  244
## 15 1.3438094  305
## 16 1.3438094  324
```

```
plot(result.bh$hazard ~ result.bh$time, type="s", ylim=c(0,max(result.bh$hazard)),
xlim=c(130,max(result.bh$time)), xlab="Time", ylab="Hazard",cex.lab=0.9, cex.axis =
 0.9, main="Baseline Hazard by Treatment Group",  cex.main = 0.9, lwd=1.5)
```



```
######
# 2. (20 points = 10+10) Now consider the 16 rats as a random sample by combining t
he two groups.
# (a) Obtain the Kaplan-Meier estimate for the survival function of the failure tim
e. Report the results in a table similar to Table 3.1 together with a plot for the
KM estimate for the survival function.
######
```

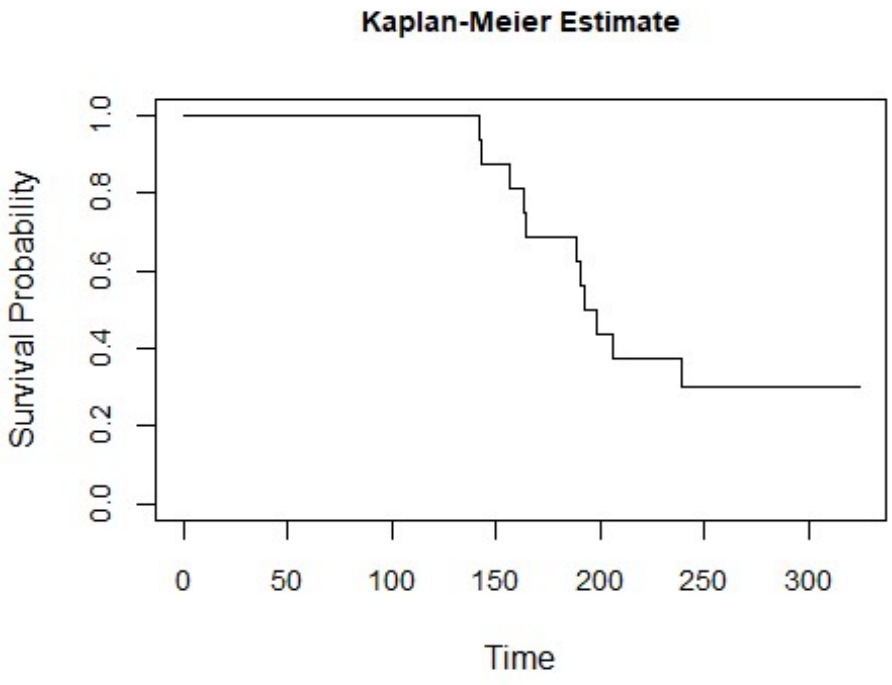
In Table 3.1, the value on column $S_i = \prod(1 - q_i)$ is computed by $S_i = \exp(-\sum q_i)$.

Rat#	t_i	n_i	d_i	$q_i = d_i/n_i$	$1 - q_i$	$S_i = (\prod(1 - q_i))$	$S_{NAA} = \exp(-\sum q_i)$
1	142	16	1	0.0625	0.9375	0.9375	0.9394
2	143	15	1	0.0667	0.9333	0.8750	0.8788
3	156	14	1	0.0714	0.9286	0.8125	0.8182
4	163	13	1	0.0769	0.9231	0.7500	0.7577
5	164	12	1	0.0833	0.9167	0.6875	0.6971
6	188	11	1	0.0909	0.9091	0.6250	0.6365
9	190	10	1	0.1000	0.9000	0.5625	0.5759
10	192	9	1	0.1111	0.8889	0.5000	0.5154
11	198	8	1	0.1250	0.8750	0.4375	0.4548
12	206	7	1	0.1429	0.8571	0.3750	0.3943
13	239	5	1	0.2000	0.8000	0.3000	0.3228

```
result.km <- survfit(Surv(tt, delta)~1, data=middata)


summary(result.km)
```

```
## Call: survfit(formula = Surv(tt, delta) ~ 1, data = middata)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   142     16       1    0.938  0.0605        0.826        1.000
##   143     15       1    0.875  0.0827        0.727        1.000
##   156     14       1    0.812  0.0976        0.642        1.000
##   163     13       1    0.750  0.1083        0.565        0.995
##   164     12       1    0.688  0.1159        0.494        0.957
##   188     11       1    0.625  0.1210        0.428        0.914
##   190     10       1    0.562  0.1240        0.365        0.867
##   192      9       1    0.500  0.1250        0.306        0.816
##   198      8       1    0.438  0.1240        0.251        0.763
##   206      7       1    0.375  0.1210        0.199        0.706
##   239      5       1    0.300  0.1178        0.139        0.648
```

```
plot(result.km, main = "Kaplan-Meier Estimate", xlab = "Time", ylab = "Survival Pro
bability",conf.int=F, cex.axis = 0.9, cex.main = 0.9,lwd=1.5)
```




```
######
# (b) Construct 95% confidence intervals by the preferred method via log-log transf
ormation for the survival functions, based on the KM estimate.
######

result.km.lg <- survfit(Surv(tt, delta) ~ 1, conf.type="log-log",
data=middata)
summary(result.km.lg)

## Call: survfit(formula = Surv(tt, delta) ~ 1, data = middata, conf.type = "log-lo
g")
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   142     16       1    0.938  0.0605        0.632        0.991
##   143     15       1    0.875  0.0827        0.586        0.967
##   156     14       1    0.812  0.0976        0.525        0.935
##   163     13       1    0.750  0.1083        0.463        0.898
##   164     12       1    0.688  0.1159        0.405        0.856
##   188     11       1    0.625  0.1210        0.349        0.811
##   190     10       1    0.562  0.1240        0.295        0.762
##   192      9       1    0.500  0.1250        0.245        0.710
##   198      8       1    0.438  0.1240        0.198        0.656
##   206      7       1    0.375  0.1210        0.154        0.598
##   239      5       1    0.300  0.1178        0.102        0.530
```

```
plot(result.km.lg, mark="|", ylab="Survival probability", xlab="Time in days",cex.l
ab=0.9,cex.axis = 0.9, lwd=1.5)
```