LIFE707 Survival analysis answers

Liam Dougherty

2024-09-10

Survival analysis answers

Load useful packages.

```
library(survival)
library(survminer)
```

```
## Warning: package 'survminer' was built under R version 4.4.1
```

```
## Warning: package 'ggpubr' was built under R version 4.4.1
```

View the colon data set (you may have to click on the `Chemotherapy for Stage B/C colon cancer` link in RStudio.).

```
?colon
```

Status is now coded as 0/1, rather than 1/2 in the previous data set. We have to assume that 0= censored and 1= event, because this is what the `survival` package expects. But ideally we would have better annotated data!

We are only interested in the 'death' data, rather than recurrence, so we first need to subset the data for etype==2.

```
colondeath <- subset(colon, colon$etype=="2")
```

We can then fit a survival curve separately by sex.

```
sfit_sex <- survfit(Surv(time, status)~sex, data=colondeath)
sfit_sex
```

```
## Call: survfit(formula = Surv(time, status) ~ sex, data = colondeath)
##
##         n events median 0.95LCL 0.95UCL
## sex=0 445    215     NA    2077      NA
## sex=1 484    237   2542    2052    2910
```

I asked you to then have a look at specific time points for both males and females. You will need to create a sequence for this.

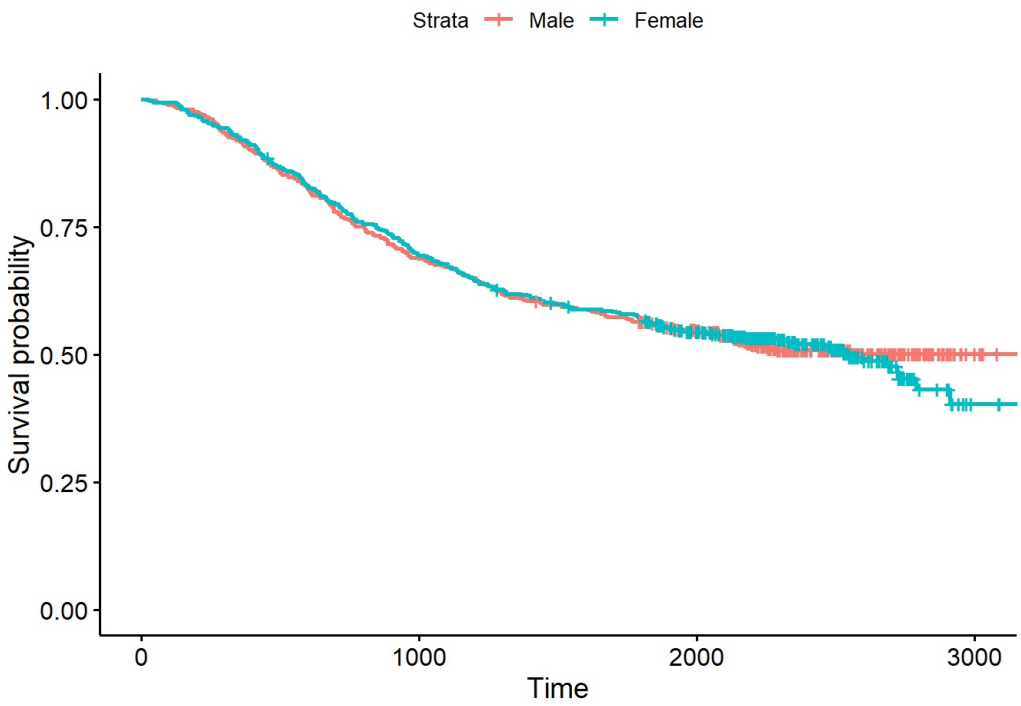
```
summary(sfit_sex, times=seq(0, 3000, 500))
```

```
## Call: survfit(formula = Surv(time, status) ~ sex, data = colondeath)
##
##                 sex=0
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     0    445       0    1.000  0.0000        1.000        1.000
##   500    381      64    0.856  0.0166        0.824        0.889
##  1000    306      75    0.688  0.0220        0.646        0.732
##  1500    265      40    0.598  0.0232        0.554        0.645
##  2000    218      22    0.547  0.0236        0.503        0.596
##  2500     89      13    0.508  0.0244        0.463        0.558
##  3000     10       1    0.501  0.0249        0.455        0.553
##
##                 sex=1
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     0    484       0    1.000  0.0000        1.000        1.000
##   500    418      65    0.866  0.0155        0.836        0.897
##  1000    335      83    0.694  0.0210        0.654        0.736
##  1500    287      46    0.598  0.0223        0.556        0.644
##  2000    238      25    0.545  0.0227        0.503        0.592
##  2500     84      10    0.510  0.0241        0.465        0.559
##  3000      7       8    0.403  0.0438        0.326        0.499
```

Males and females look very similar in terms of survival for the entire duration.

We can then plot a Kaplan-Meier plot:

```
library(survminer)
ggsurvplot(sfit_sex,legend.labs=c("Male","Female"))
```



There is no obvious difference between the two lines for most of the study. There does seem to be a drop in survival probability at the end- is this enough to result in a significant difference?

We can test for a significant difference in survival between the sexes using a Cox proportional hazards model.

```
coxph_sex <- coxph(Surv(time, status)~sex, data=colondeath)
summary(coxph_sex) # P= 0.888
```

```
## Call:
## coxph(formula = Surv(time, status) ~ sex, data = colondeath)
##
##   n= 929, number of events= 452
##
##        coef exp(coef) se(coef)     z Pr(>|z|)
## sex 0.01332   1.01341  0.09420 0.141    0.888
##
##     exp(coef) exp(-coef) lower .95 upper .95
## sex     1.013     0.9868    0.8426     1.219
##
## Concordance= 0.497  (se = 0.012 )
## Likelihood ratio test= 0.02  on 1 df,   p=0.9
## Wald test            = 0.02  on 1 df,   p=0.9
## Score (logrank) test = 0.02  on 1 df,   p=0.9
```

There is no significant difference between the sexes. We can see from the `exp(coef)` part of the output that males are only 1.01 times as likely to die as females- this is a very similar risk.