hw5

Nathan Hawkins

2/16/2022

1. The survival package has a built in data set called “heart” with data on patients waiting for and receiving heart transplants. For this problem we only want the data from after the transplant was received (we will use the full data set later), so create a new data frame with the subset using code like:

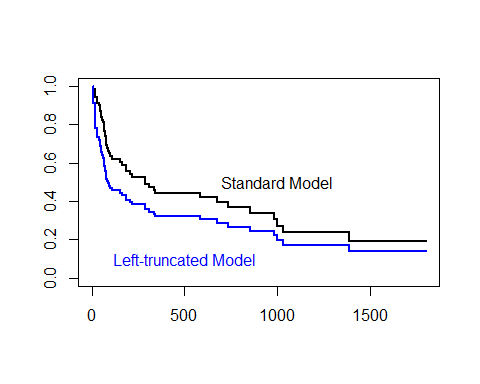
heart2 <- heart[heart$transplant==1,]

Now fit a Kaplan-Meier curve to the data ignoring truncation, i.e. use the stop variable as the time and event as the censoring indicator.

Fit a second model that uses the start variable to indicate left truncation.

Plot both fits (using the lines function on a survfit object will add the lines to the existing plot) in the same plot and compare the survival curves. Also compare the median survival times and the number of people at risk at time points 5 and 16. How does this affect the survival estimates?

library(asaur)  
library(survival)  
heart = survival::heart  
heart2 <- heart[heart$transplant==1,]  
kp\_model = survfit(Surv(stop, event) ~ 1, data = heart2)  
  
kp\_model2alt = survfit(Surv(start, stop, event) ~ 1, data = heart2)  
  
plot(kp\_model, lwd = 2, conf.int = FALSE)  
lines(kp\_model2alt, col = "blue", lwd = 2, conf.int = FALSE)  
text(1000, 0.5, "Standard Model", col = "black")  
text(500, 0.1, "Left-truncated Model", col = "blue")



# The median is 285 in the standard model and 90 in the left-truncated model.  
kp\_model

## Call: survfit(formula = Surv(stop, event) ~ 1, data = heart2)  
##   
## n events median 0.95LCL 0.95UCL   
## 69 45 285 153 852

kp\_model2alt

## Call: survfit(formula = Surv(start, stop, event) ~ 1, data = heart2)  
##   
## records n.max n.start events median 0.95LCL 0.95UCL   
## 69 45 11 45 90 53 334

# The number of people at risk at time 5 and 16 for the standard model are 69 and 69.  
# The number of people at risk at time 5 and 16 for the left-truncated model are 11 and 21.  
summary(kp\_model)

## Call: survfit(formula = Surv(stop, event) ~ 1, data = heart2)  
##   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 5 69 1 0.986 0.0144 0.9577 1.000  
## 16 68 2 0.957 0.0246 0.9096 1.000  
## 17 66 1 0.942 0.0281 0.8885 0.999  
## 28 65 1 0.928 0.0312 0.8683 0.991  
## 30 64 1 0.913 0.0339 0.8489 0.982  
## 39 63 1 0.899 0.0363 0.8301 0.973  
## 43 61 1 0.884 0.0386 0.8113 0.963  
## 45 60 1 0.869 0.0407 0.7929 0.953  
## 51 59 1 0.854 0.0426 0.7748 0.942  
## 53 58 1 0.840 0.0443 0.7571 0.931  
## 58 57 1 0.825 0.0459 0.7396 0.920  
## 61 56 1 0.810 0.0474 0.7224 0.909  
## 66 55 1 0.795 0.0488 0.7053 0.897  
## 68 54 2 0.766 0.0512 0.6719 0.873  
## 72 52 2 0.737 0.0533 0.6391 0.849  
## 77 50 1 0.722 0.0543 0.6229 0.836  
## 78 49 1 0.707 0.0551 0.6069 0.824  
## 80 48 1 0.692 0.0559 0.5910 0.811  
## 81 47 1 0.678 0.0566 0.5752 0.798  
## 90 46 1 0.663 0.0573 0.5596 0.785  
## 96 45 1 0.648 0.0579 0.5441 0.772  
## 100 44 1 0.633 0.0584 0.5287 0.759  
## 110 42 1 0.618 0.0589 0.5130 0.745  
## 153 40 1 0.603 0.0594 0.4969 0.731  
## 165 39 1 0.587 0.0599 0.4810 0.717  
## 186 37 1 0.572 0.0603 0.4647 0.703  
## 188 36 1 0.556 0.0607 0.4485 0.688  
## 207 35 1 0.540 0.0610 0.4325 0.674  
## 219 34 1 0.524 0.0613 0.4166 0.659  
## 285 32 2 0.491 0.0616 0.3840 0.628  
## 308 30 1 0.475 0.0617 0.3680 0.613  
## 334 29 1 0.458 0.0617 0.3521 0.597  
## 343 27 1 0.441 0.0617 0.3356 0.581  
## 584 20 1 0.419 0.0625 0.3132 0.562  
## 675 16 1 0.393 0.0638 0.2860 0.540  
## 733 15 1 0.367 0.0647 0.2597 0.519  
## 852 13 1 0.339 0.0656 0.2317 0.495  
## 980 10 1 0.305 0.0672 0.1979 0.470  
## 996 9 1 0.271 0.0678 0.1660 0.442  
## 1032 8 1 0.237 0.0672 0.1360 0.413  
## 1387 5 1 0.190 0.0685 0.0935 0.385

summary(kp\_model2alt)

## Call: survfit(formula = Surv(start, stop, event) ~ 1, data = heart2)  
##   
## time n.risk n.event censored survival std.err lower 95% CI upper 95% CI  
## 5 11 1 0 0.909 0.0867 0.7541 1.000  
## 16 21 2 0 0.823 0.0977 0.6517 1.000  
## 17 20 1 0 0.781 0.1011 0.6064 1.000  
## 28 33 1 0 0.758 0.1008 0.5839 0.983  
## 30 34 1 0 0.735 0.1002 0.5630 0.961  
## 39 43 1 1 0.718 0.0993 0.5478 0.942  
## 43 43 1 0 0.702 0.0984 0.5329 0.924  
## 45 42 1 0 0.685 0.0975 0.5182 0.905  
## 51 43 1 0 0.669 0.0965 0.5042 0.888  
## 53 44 1 0 0.654 0.0955 0.4910 0.871  
## 58 44 1 0 0.639 0.0945 0.4781 0.854  
## 61 45 1 0 0.625 0.0935 0.4660 0.838  
## 66 44 1 0 0.611 0.0924 0.4538 0.821  
## 68 45 2 0 0.583 0.0903 0.4308 0.790  
## 72 44 2 0 0.557 0.0881 0.4084 0.759  
## 77 42 1 0 0.544 0.0870 0.3973 0.744  
## 78 41 1 0 0.530 0.0859 0.3861 0.728  
## 80 41 1 0 0.517 0.0847 0.3753 0.713  
## 81 40 1 0 0.504 0.0836 0.3646 0.698  
## 90 41 1 0 0.492 0.0825 0.3544 0.684  
## 96 40 1 0 0.480 0.0813 0.3443 0.669  
## 100 40 1 0 0.468 0.0802 0.3344 0.655  
## 110 38 1 1 0.456 0.0790 0.3243 0.640  
## 153 37 1 1 0.443 0.0778 0.3142 0.625  
## 165 37 1 0 0.431 0.0766 0.3044 0.611  
## 186 35 1 1 0.419 0.0754 0.2944 0.596  
## 188 34 1 0 0.407 0.0742 0.2844 0.581  
## 207 33 1 0 0.394 0.0730 0.2744 0.567  
## 219 33 1 0 0.382 0.0717 0.2647 0.552  
## 285 31 2 1 0.358 0.0692 0.2448 0.523  
## 308 29 1 0 0.345 0.0679 0.2349 0.508  
## 334 29 1 0 0.333 0.0666 0.2254 0.493  
## 343 27 1 1 0.321 0.0653 0.2156 0.478  
## 584 20 1 6 0.305 0.0639 0.2023 0.460  
## 675 16 1 3 0.286 0.0627 0.1861 0.440  
## 733 15 1 0 0.267 0.0614 0.1701 0.419  
## 852 13 1 1 0.246 0.0600 0.1529 0.397  
## 980 10 1 2 0.222 0.0588 0.1318 0.373  
## 996 9 1 0 0.197 0.0572 0.1116 0.348  
## 1032 8 1 0 0.172 0.0551 0.0922 0.323  
## 1387 5 1 2 0.138 0.0538 0.0642 0.296

# Having less people in the risk set probably gives a more accurate estimate because it doesn't account for people until they are actually at risk. This means their survival estimate is lower.

1. Fit a Cox proportional Hazards model to the subset of the Heart data used above with a spline term for age, another spline term for year (when the patient was entered into the study) and surgery (whether the patient had a prior heart surgery). Use the summary output as well as computing AIC values to find a simpler model that fits the data and makes sense, interpret and justify your final model.

The model that has covariates age, surgery, and year has a slightly smaller AIC (296 < 298) but has fewer terms than the model with the splines. We will favor the simpler model that fits the data better.

library(splines)  
  
  
  
model3 = coxph(Surv(start, stop, event) ~ surgery + pspline(age, df = 3) +  
 pspline(heart2$year, df = 3),  
 data = heart2)  
  
model3.1 = coxph(Surv(start, stop, event) ~ surgery + age + year,  
 data = heart2)  
  
  
AIC(model3, model3.1)

## df AIC  
## model3 6.955234 298.9438  
## model3.1 3.000000 296.1922