04\_homework2

Randy

2/4/2021

# BIOS7721 Homework2

## Introduction

* dataset contains 500 patients
* human tissue valve in aortic position
* subcoronary implantation (SI) or root replacement (RR)
* followed over time and longitudinal aortic gradient measurements
* at risk of experiencing death following their surgery

aort <- here::here("aort\_new.csv") %>%   
 read\_csv() %>%  
 janitor::clean\_names()   
  
## aort data is in longitudial form  
## each row is for one visit  
## one subject has many visit  
# View(aort)

## Question1. Survival analysis with a time-varying covariate

### a. create the start/stop time data set

* the beginning of a time interval represents a measurement time
* the end of the final time window represents the survival time
* You also need a new status indicator
* indicator value of 0 for all intervals
* the last indicator is 1 if an event is observed
* the last indicator is 0 if a patient is censored
* print the rows for Patients 1 and 2.

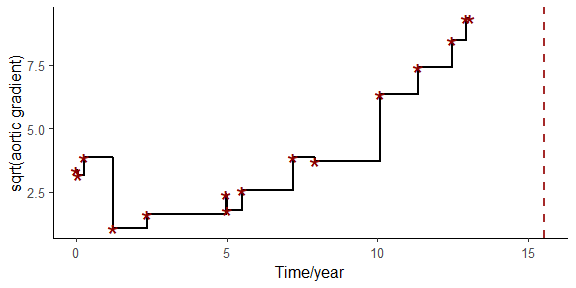
aort1 <- aort %>%  
 ## tmerge cannot bear duplicate id  
 filter(time == 0) %>%  
 ## time based merge for survial data  
 survival::tmerge(  
 data1 = .,  
 data2 = .,  
 id = id,   
 ## the tdc and event use   
 ## the final value in the data  
 ## 4 types of operational arguments:  
 ## tdc/cumtdc/event/cumevent  
 death = event(survtime, event)) %>%  
 ## start stop death added  
 survival::tmerge(  
 data1 = .,  
 data2 = aort,  
 id = id,  
 sqrt\_aort\_grad = tdc(time, sqrt(aort\_grad))) %>%  
 ## given time, sqrt\_aort\_grad added  
 select(id, tstart,   
 tstop, death,   
 sqrt\_aort\_grad,   
 oper, sex)  
  
aort1 %>%  
 filter(id %in% c(1,2)) %>%  
 knitr::kable("simple", align = "c")

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| id | tstart | tstop | death | sqrt\_aort\_grad | oper | sex |
| 1 | 0.0000000 | 0.7504862 | 0 | 2.077925 | SI | Male |
| 1 | 0.7504862 | 1.8606248 | 0 | 1.909819 | SI | Male |
| 2 | 0.0000000 | 3.2322893 | 0 | 2.872840 | SI | Male |
| 2 | 3.2322893 | 3.8214459 | 0 | 2.535935 | SI | Male |
| 2 | 3.8214459 | 4.7061867 | 0 | 5.235966 | SI | Male |
| 2 | 4.7061867 | 5.3569110 | 0 | 4.049932 | SI | Male |
| 2 | 5.3569110 | 5.6207138 | 0 | 4.250990 | SI | Male |
| 2 | 5.6207138 | 6.6390331 | 0 | 4.405298 | SI | Male |
| 2 | 6.6390331 | 6.8962116 | 0 | 4.795410 | SI | Male |
| 2 | 6.8962116 | 7.8946465 | 0 | 4.006678 | SI | Male |
| 2 | 7.8946465 | 8.7633751 | 0 | 4.948445 | SI | Male |
| 2 | 8.7633751 | 9.4291267 | 0 | 7.366542 | SI | Male |
| 2 | 9.4291267 | 9.8154872 | 0 | 5.797056 | SI | Male |
| 2 | 9.8154872 | 10.7623612 | 0 | 6.578867 | SI | Male |
| 2 | 10.7623612 | 11.0586943 | 0 | 7.638576 | SI | Male |
| 2 | 11.0586943 | 12.0011893 | 0 | 7.368440 | SI | Male |
| 2 | 12.0011893 | 12.6317556 | 0 | 9.113807 | SI | Male |

### b. for patient 3 create a stepped line plot

* square root aortic gradient
* vertical dashed line for patient’s observed survival time

aort1\_id3 <- aort1 %>%  
 filter(id == 3)   
  
plot\_id3 <- aort1\_id3 %>%  
 ggplot(aes(tstart, sqrt\_aort\_grad)) +  
 geom\_step(direction = "hv",  
 size = 1) +  
 geom\_point(color = "darkred",  
 shape = "\*",  
 size = 7) +  
 geom\_vline(xintercept = max(aort1\_id3$tstop),  
 linetype = "dashed",  
 color = "brown",  
 size = 1) +  
 theme\_classic2() +  
 xlab("Time/year") +  
 ylab("sqrt(aortic gradient)")  
  
plot\_id3



### c. fit an extended Cox survival model

* square root aortic gradient as a time-varying covariate
* operation type and sex as a time-independent baseline covariate
* interpret the coefficient estimates
* how it is the different from model with only the baseline values?

The aortic gradient level can significantly affect the subject’s survival status (p << 0.001). On average, one unit increase on the sqrt of aortic gradient can increase the risk of event to 1.5 folds (95% CI 1.39 1.62).  
After adjusted for the biomarker aortic gradient, the effect of operation type on survival becomes on significant, and effect size exp(operation) changes from 0.95 fold to 0.67; given patient’s aort gradient level, operation SI has stronger effects on survival improvement for certain patient.  
Also the adjustment of biomarker revert the effect of gender, from increasing risk of event to decreasing risk of event, even though neither of model show these gender effects significant.  
Overall the model performance gets improved in model cox1 (AIC = 2284); the AIC decreased, comparing to the baseline value model cox0 (AIC = 2400).

## add the survobj into aort1 as part of dataset  
aort1$survobj <- with(aort1, Surv(tstart, tstop, death))  
cox0 <- coxph(survobj ~ oper + sex,   
 data = aort1)  
cox1 <- coxph(survobj ~ oper + sex + sqrt\_aort\_grad,   
 data = aort1)  
tidy0 <- tidy(cox0) %>%  
 tibble() %>%  
 mutate(model = "cox-ex")  
tidy1 <- tidy(cox1) %>%  
 tibble() %>%  
 mutate(model = "cox-ex")  
glance0 <- glance(cox0)  
glance1 <- glance(cox1)

summary(cox0)

## Call:  
## coxph(formula = survobj ~ oper + sex, data = aort1)  
##   
## n= 4183, number of events= 243   
##   
## coef exp(coef) se(coef) z Pr(>|z|)  
## operSI -0.05155 0.94976 0.13087 -0.394 0.694  
## sexMale 0.02349 1.02377 0.12983 0.181 0.856  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## operSI 0.9498 1.0529 0.7349 1.227  
## sexMale 1.0238 0.9768 0.7938 1.320  
##   
## Concordance= 0.501 (se = 0.021 )  
## Likelihood ratio test= 0.2 on 2 df, p=0.9  
## Wald test = 0.2 on 2 df, p=0.9  
## Score (logrank) test = 0.2 on 2 df, p=0.9

summary(cox1)

## Call:  
## coxph(formula = survobj ~ oper + sex + sqrt\_aort\_grad, data = aort1)  
##   
## n= 4183, number of events= 243   
##   
## coef exp(coef) se(coef) z Pr(>|z|)   
## operSI -0.40736 0.66541 0.13574 -3.001 0.00269 \*\*   
## sexMale -0.14437 0.86557 0.13147 -1.098 0.27214   
## sqrt\_aort\_grad 0.40747 1.50301 0.03847 10.591 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## operSI 0.6654 1.5028 0.510 0.8682  
## sexMale 0.8656 1.1553 0.669 1.1200  
## sqrt\_aort\_grad 1.5030 0.6653 1.394 1.6207  
##   
## Concordance= 0.63 (se = 0.023 )  
## Likelihood ratio test= 117.4 on 3 df, p=<2e-16  
## Wald test = 112.3 on 3 df, p=<2e-16  
## Score (logrank) test = 115 on 3 df, p=<2e-16

rbind(glance0, glance1) %>%  
 tibble() %>%  
 rownames\_to\_column("model") %>%  
 select(model, AIC, BIC, logLik) %>%  
 knitr::kable("simple", align = "c")

|  |  |  |  |
| --- | --- | --- | --- |
| model | AIC | BIC | logLik |
| 1 | 2399.754 | 2406.741 | -1197.877 |
| 2 | 2284.523 | 2295.002 | -1139.261 |

### d. comment on why it is not appropriate of the extended Cox model

The extended Cox model assumptions is only valid for exogenous time dependent covariates. However in this case, the aortic gradient level is a endogenous bio-marker, the level of which closely related to failure status.  
The biomarker level cannot be predetermined or totally immune to measurement errors.  Also for extended Cox model, we would assume that the aortic gradient changes only at the measurement times and remain constant between two measurements, as step-function approximation.  
Therefore, extended Cox model is not appropriate, or at least not the optimal method.

## Question2. Two stage model

### a. fit a mixed effects model

* outcome: square root aortic gradient
* fixed effects: linear time, operation type, and sex
* random effects: intercept and linear slope for time
* interpret the coefficient estimates from this model

lme1 <- nlme::lme(sqrt\_aort\_grad ~   
 oper + sex + tstart,   
 random = (~ 1 + tstart | id),  
 data = aort1)  
summary(lme1)

## Linear mixed-effects model fit by REML  
## Data: aort1   
## AIC BIC logLik  
## 12225.41 12276.11 -6104.703  
##   
## Random effects:  
## Formula: ~1 + tstart | id  
## Structure: General positive-definite, Log-Cholesky parametrization  
## StdDev Corr   
## (Intercept) 0.9339484 (Intr)  
## tstart 0.1750445 -0.147  
## Residual 0.8455028   
##   
## Fixed effects: sqrt\_aort\_grad ~ oper + sex + tstart   
## Value Std.Error DF t-value p-value  
## (Intercept) 2.2863387 0.08294554 3682 27.56434 0.0000  
## operSI 0.7374438 0.09190882 497 8.02365 0.0000  
## sexMale 0.0397345 0.09183342 497 0.43268 0.6654  
## tstart 0.3379494 0.01052127 3682 32.12060 0.0000  
## Correlation:   
## (Intr) operSI sexMal  
## operSI -0.600   
## sexMale -0.579 0.042   
## tstart -0.165 0.001 -0.002  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -3.152709486 -0.592187596 0.006059147 0.601982291 4.206817111   
##   
## Number of Observations: 4183  
## Number of Groups: 500

As seen in model lme1, time has a very highly significant effect on the level of aortic gradient level (p << 0.001); on average, every year will increase the subject’s sqrt aortic gradient level 0.33 unit. Operation type SI can increase sqrt aortic gradient 0.737 compare to RR.

There is some variability for each patient baseline aortic gradient level sd\_(Intercept) = 0.934, and the random linear time effects sd\_slope = 0.175. The random intercept and random slope are not strongly related with each other. We can see that the variability for within individual is still pretty high time to time.

### b. the mixed effects model as subject-specific predictions

* the contributions from the random intercept and random slope
* use as a time-varying covariate in a Cox survival model
* Cox model includes operation type and sex as time independent baseline covariates
* interpret the coefficient estimates from this model

## why this is subject specific? this is marginal  
aort1$sqrt\_aort\_pred <- c(predict(lme1))  
  
cox2 <- coxph(survobj ~ oper + sex + sqrt\_aort\_pred,  
 data = aort1)  
tidy2 <- tidy(cox2) %>%  
 tibble() %>%  
 mutate(model = "two-time")  
glance2 <- glance(cox2)  
summary(cox2)

## Call:  
## coxph(formula = survobj ~ oper + sex + sqrt\_aort\_pred, data = aort1)  
##   
## n= 4183, number of events= 243   
##   
## coef exp(coef) se(coef) z Pr(>|z|)   
## operSI -0.36616 0.69340 0.13748 -2.663 0.00774 \*\*   
## sexMale -0.11116 0.89480 0.13124 -0.847 0.39701   
## sqrt\_aort\_pred 0.35229 1.42232 0.04475 7.872 3.5e-15 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## operSI 0.6934 1.4422 0.5296 0.9078  
## sexMale 0.8948 1.1176 0.6919 1.1573  
## sqrt\_aort\_pred 1.4223 0.7031 1.3029 1.5527  
##   
## Concordance= 0.597 (se = 0.02 )  
## Likelihood ratio test= 63.46 on 3 df, p=1e-13  
## Wald test = 62.09 on 3 df, p=2e-13  
## Score (logrank) test = 63.01 on 3 df, p=1e-13

### c. bootstrap

* compute the standard errors for Cox component of the two-stage model
* the differences from the standard errors estimated in the model2?
* reasons of differences between two methods for inference?

For two-stage model, the estimates are obtained by fitting the corresponding mixed model using the observed responses up to given time from all subjects still at risk. Hence, there will be a bias for the implementation that remove all the measurement errors for unbiased predictors, which produce biased results by empirical Bayesian estimates from random effects. The partial likelihood asymptotic feature will no hold anymore. Hence the bootstrapping results have smaller standard errors and confidence interval.

set.seed(555)  
  
#' get\_coef() to extract the coef from one bootstrap  
#'  
#' @param data the dataset for bootstrap  
#' @param indices a placeholder for the map function  
#' @return the coefs from the model fitting  
#' @examples  
#' get\_coef(aort1, 1)  
get\_coef <- function(data, indices) {  
 data1 <- data %>%  
 group\_by(id) %>%  
 nest() %>%  
 as.data.frame()   
 index <- sample(1:nrow(data1),   
 size = nrow(data1),   
 replace =TRUE)  
 data2 <- data1[index, ] %>%  
 unnest()  
 lmm <- lme(sqrt\_aort\_grad ~   
 oper + sex + tstart,   
 random = (~ 1 + tstart | id),  
 data = data2)  
 data2$sqrt\_aort\_pred <- c(predict(lmm))  
 fit1 <- coxph(survobj ~   
 oper + sex + sqrt\_aort\_pred,   
 data = data2)  
 return(coef(fit1))  
}  
  
## bootstrapping takes too long time  
## load .Rdata file to see the result directly  
# cox\_boot <-   
# ## repeat 1000 times coxph  
# map\_df(.x = 1:1000,   
# .f = ~get\_coef(  
# data = aort1,  
# indices = .x))  
#   
# save(cox\_boot, file = "cox\_boot\_20210207.Rdata")  
  
## upload the .Rdata for convinence  
load("cox\_boot\_20210207.Rdata")

tidy(cox2) %>%  
 mutate(boot.estimate = as.numeric(map(cox\_boot, mean)),  
 boot.std.error = as.numeric(map(cox\_boot, sd))) %>%  
 select(term, estimate, boot.estimate,  
 std.error, boot.std.error) %>%  
 knitr::kable()

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | boot.estimate | std.error | boot.std.error |
| operSI | -0.3661550 | -0.3585640 | 0.1374833 | 0.1292254 |
| sexMale | -0.1111559 | -0.1050206 | 0.1312380 | 0.1254121 |
| sqrt\_aort\_pred | 0.3522863 | 0.3387036 | 0.0447545 | 0.0442396 |

### d. comparision

* the two-stage model and the time-varying covariate model
* comment on any differences
* why these differences may exist

Overall these two models are pretty similar to each other. Comparatively, the two-stage model has smaller estimate value and larger standard error.

The Cox extended model use step function approximation for the time dependent covariate, which measured without error. This may introduce bias to the estimates and standard error.  
These difference demonstrate the attenuation in the regression coefficients of the cox analysis due to the meansurement error.

rbind(tidy1, tidy2) %>%  
 select(model, everything()) %>%  
 arrange(term) %>%  
 knitr::kable("simple", align = "c")

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| model | term | estimate | std.error | statistic | p.value |
| cox-ex | operSI | -0.4073591 | 0.1357360 | -3.0011132 | 0.0026899 |
| two-time | operSI | -0.3661550 | 0.1374833 | -2.6632697 | 0.0077385 |
| cox-ex | sexMale | -0.1443701 | 0.1314674 | -1.0981437 | 0.2721417 |
| two-time | sexMale | -0.1111559 | 0.1312380 | -0.8469797 | 0.3970064 |
| cox-ex | sqrt\_aort\_grad | 0.4074694 | 0.0384742 | 10.5907213 | 0.0000000 |
| two-time | sqrt\_aort\_pred | 0.3522863 | 0.0447545 | 7.8715308 | 0.0000000 |