

# BIOSTAT 675 – Homework #5 Solutions

Kyle Kumbier

## Problem 1

Data on  $n = 64$  patients with severe anemia are contained in the SAS file anemia2.sas7bdat located in the "Data Sets" folder. The failure time of interest is time until graft-versus-host-disease (GVHD), measured in days. The variate CSP.MTX is a treatment indicator (0 = no, 1 = yes), while LAF is an indicator for being assigned to an airflow isolation room. AGE is recorded in years.

**(a) Fit a Cox model with CSP.MTX (Zi), LAF (Li) and AGE (Ai) as covariates. Interpret each of the hazard ratios.**

```
library(survival)
model <- coxph(Surv(obs_time, GVHD) ~ CSP_MTX + LAF + age, data = anemia2)
summary(model)
```

```
Call:
coxph(formula = Surv(obs_time, GVHD) ~ CSP_MTX + LAF + age, data = anemia2)

n= 64, number of events= 20

              coef exp(coef) se(coef)      z Pr(>|z|)
CSP_MTX -1.39645    0.24747  0.53511 -2.610  0.00906 **
LAF      -0.54295    0.58103  0.49411 -1.099  0.27183
age       0.06129    1.06320  0.02743  2.234  0.02547 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
CSP_MTX      0.2475      4.0408    0.08671    0.7063
LAF           0.5810      1.7211    0.22060    1.5303
age           1.0632      0.9406    1.00755    1.1219

Concordance= 0.697 (se = 0.067 )
Rsquare= 0.16 (max possible= 0.915 )
Likelihood ratio test= 11.16 on 3 df,  p=0.01
Wald test               = 9.63 on 3 df,  p=0.02
Score (logrank) test = 10.2 on 3 df,  p=0.02
```

CSP.MTX Hazard Ratio Interpretation: Those who receive CSP.MTX have a 75% reduction in hazard in comparison to those who do not receive CSP.MTX, with LAF and age held constant.

LAF Hazard Ratio Interpretation: Those who receive LAF have a 42% reduction in hazard in comparison to those who do not receive LAF, with CSP.MTX and age held constant.

Age Hazard Ratio Interpretation: The hazard increases 6.3% for each one-year increase in age, holding CSP.MTX and LAF constant.

**(b) Refit the model, with  $A_i$  replaced by  $A_i/5$ . Compare each parameter estimate to that from (a) and comment on their similarity or differences.**

```
library(dplyr)
anemia2b <- mutate(anemia2, age5 = age/5)
model2 <- coxph(Surv(obs_time, GVHD) ~ CSP_MTX + LAF + age5, data = anemia2b)
summary(model2)
```

Call:  
coxph(formula = Surv(obs\_time, GVHD) ~ CSP\_MTX + LAF + age5,  
data = anemia2b)

n= 64, number of events= 20

	coef	exp(coef)	se(coef)	z	Pr(> z )	
CSP_MTX	-1.3964	0.2475	0.5351	-2.610	0.00906	**
LAF	-0.5430	0.5810	0.4941	-1.099	0.27183	
age5	0.3064	1.3586	0.1372	2.234	0.02547	*

---  
Signif. codes:  
0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
CSP_MTX	0.2475	4.0408	0.08671	0.7063
LAF	0.5810	1.7211	0.22060	1.5303
age5	1.3586	0.7361	1.03834	1.7776

Concordance= 0.697 (se = 0.067 )  
Rsquare= 0.16 (max possible= 0.915 )  
Likelihood ratio test= 11.16 on 3 df, p=0.01  
Wald test = 9.63 on 3 df, p=0.02  
Score (logrank) test = 10.2 on 3 df, p=0.02

CSP.MTX Hazard Ratio Interpretation: Does not change from (a).

LAF Hazard Ratio Interpretation: Does not change from (a)

Age Hazard Ratio Interpretation: The hazard increases 35.9% for each five-year increase in age, holding CSP.MTX and LAF constant.

**(c) Is the treatment effect (i.e., effect of  $Z_i$ ) different for subjects of different ages? Carry out an appropriate Wald test.**

```
model3 <- coxph(Surv(obs_time, GVHD) ~ CSP_MTX + LAF + age + CSP_MTX*age,  
data = anemia2)  
summary(model3)
```

Call:  
coxph(formula = Surv(obs\_time, GVHD) ~ CSP\_MTX + LAF + age +  
CSP\_MTX \* age, data = anemia2)

n= 64, number of events= 20

	coef	exp(coef)	se(coef)	z	Pr(> z )	
CSP_MTX	1.70562	5.50479	1.39744	1.221	0.22226	
LAF	-0.68505	0.50406	0.49861	-1.374	0.16946	
age	0.10085	1.10611	0.03368	2.995	0.00275	**

```

CSP_MTX:age -0.13380    0.87477    0.06230 -2.148    0.03175 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
CSP_MTX      5.5048      0.1817     0.3558    85.1621
LAF           0.5041      1.9839     0.1897     1.3394
age           1.1061      0.9041     1.0355     1.1816
CSP_MTX:age   0.8748      1.1432     0.7742     0.9884

Concordance= 0.727 (se = 0.067 )
Rsquare= 0.225 (max possible= 0.915 )
Likelihood ratio test= 16.32 on 4 df, p=0.003
Wald test               = 16.09 on 4 df, p=0.003
Score (logrank) test = 19.93 on 4 df, p=5e-04

```

From the table of parameter estimates, we see that the interaction term is significant based on the Wald test with a p-value = 0.03175.

**(d) Re-evaluate the hypothesis from (c), but this time use a likelihood ratio test. Compare your results to that obtained through the Wald test and comment.**

```

like_full <- model3$loglik[2]
like_rdc <- model$loglik[2]
LRT <- 2*(like_full - like_rdc)
pval_LRT <- 1 - pchisq(LRT, 1)

> LRT
[1] 5.162259
> pval_LRT
[1] 0.02308286

```

Since  $5.16 > 3.84$ , we also reject the null hypothesis with the likelihood ratio test (consistent with the Wald test done in part (c)). The two test statistics are not identical, but we wouldn't expect them to be as they are only asymptotically equivalent.

**(e) Give interpretations for each HR for the model fitted in (c).**

CSP.MTX Hazard Ratio Interpretation: For those age 0, those who receive CSP.MTX have a 550% increase in hazard in comparison to those who do not receive CSP.MTX, with LAF held constant.

LAF Hazard Ratio Interpretation: Those who receive LAF have a 50% reduction in hazard in comparison to those who do not receive LAF, with CSP.MTX and age held constant.

Age Hazard Ratio Interpretation: For those not treated with CSP.MTX, the hazard increases 10.6% for each one-year increase in age, holding LAF constant.

CSP.MTX-Age Hazard Ratio Interpretation: The increase in hazard rate for each one-year increase in age for those treated with CSP.MTX is 12.5% lower in comparison to those not treated, holding LAF and age constant.

**(f) Is the effect of age linear? Support your response empirically by fitting an appropriate main effects model and providing the appropriate plot.**

```
summary(anemia2$age)
anemia2c <- mutate(anemia2b,
  age_cat1 = 1*(age > 0)*(age <= 14),
  age_cat2 = 1*(age > 14)*(age <= 21),
  age_cat3 = 1*(age > 21)*(age <= 27.25),
  age_cat4 = 1*(age > 27.25)*(age <= 42))
model5 <- coxph(Surv(obs_time, GVHD) ~ CSP_MTX + LAF + age_cat2 + age_cat3 +
  age_cat4, data = anemia2c)
summary(model5)
betas <- c(0,model5$coefficients[3:5])
names(betas)[1] <- "age_cat1"
mean1 <- mean(anemia2c$age[anemia2c$age_cat1 == 1])
mean2 <- mean(anemia2c$age[anemia2c$age_cat2 == 1])
mean3 <- mean(anemia2c$age[anemia2c$age_cat3 == 1])
mean4 <- mean(anemia2c$age[anemia2c$age_cat4 == 1])
means <- c(mean1,mean2,mean3,mean4)
plot_data <- data.frame(betas,means)
library(ggplot2)
plot01 <- ggplot(plot_data, aes(x = means, y = betas)) +
  geom_point() +
  geom_line() +
  xlab("Age") +
  ylab("Age Effect Estimate (Betas)")
plot01
```

```
Call:
coxph(formula = Surv(obs_time, GVHD) ~ CSP_MTX + LAF + age_cat2 +
  age_cat3 + age_cat4, data = anemia2c)

n= 64, number of events= 20

      coef exp(coef) se(coef)      z Pr(>|z|)
CSP_MTX -1.3130    0.2690  0.5389 -2.437  0.0148 *
LAF      -0.3951    0.6736  0.4698 -0.841  0.4004
age_cat2  1.6262    5.0844  0.7936  2.049  0.0404 *
age_cat3  1.1766    3.2434  0.8690  1.354  0.1757
age_cat4  1.9149    6.7860  0.8452  2.266  0.0235 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

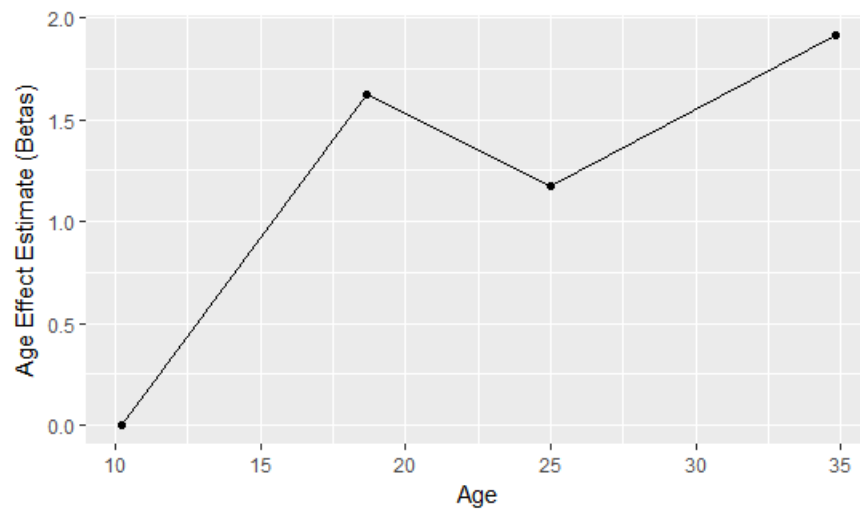
      exp(coef) exp(-coef) lower .95 upper .95
CSP_MTX    0.2690    3.7172   0.09356    0.7735
LAF         0.6736    1.4845   0.26821    1.6918
age_cat2    5.0844    0.1967   1.07336   24.0839
age_cat3    3.2434    0.3083   0.59066   17.8104
age_cat4    6.7860    0.1474   1.29479   35.5653

Concordance= 0.704 (se = 0.067 )
Rsquare= 0.19 (max possible= 0.915 )
Likelihood ratio test= 13.52 on 5 df, p=0.02
Wald test              = 10.64 on 5 df, p=0.06
Score (logrank) test = 11.81 on 5 df, p=0.04
```

```

> summary(anemia2$age)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.00   14.00   21.00   21.61   27.25   42.00
> betas
age_cat1 age_cat2 age_cat3 age_cat4
0.000000 1.626169 1.176633 1.914857
> means
[1] 10.22222 18.64706 25.00000 34.81250

```



Based on the plot above, it does NOT appear that the effect of age is linear.

## Problem 2

An investigator wishes to combine her prior knowledge of anemia with the *first* Cox model (i.e., the *main effects model*) you fitted in Problem 1.

**(a) Suppose that it is known that  $S(14) = 0.95$  for a subject age 30. If possible, under the assumed model, determine what  $S(14)$  would equal if that same subject was actually age 40? If not possible, state what information you are missing.**

Based on the main effects model,  $S(14) = 0.95^{\exp\{(40-30)*(0.06129)\}} = 0.91$

**(b) Suppose that it is known that  $S(28) = 0.90$  for a subject age 25. If possible, under the assumed model, determine  $S(56)$  if that same subject was actually age 50. If not possible, state what information you are missing.**

It is not possible to determine  $S(56)$  if that same subject was actually age 50. We would need to know  $S(56)$  for the subject at age 25 since hazard ratios only allow us to compare subjects at the same time.

## Problem 3

End-stage renal disease (ESRD; also referred to as ‘renal failure’) is increasing in many countries worldwide, including the United States and Canada. Due to the shortfall in the available donor organs, donor kidneys are now being transplanted which would in the past have been discarded; the so-called Expanded Criteria Donor (ECD) kidneys. By definition, ECD kidneys are more likely to suffer graft failure (GF), the condition wherein the transplanted kidney stops functioning sufficiently.

A random sample of U.S. transplant recipients was assembled, in order to study the effects on the mortality hazard of ECD (vs non-ECD) kidneys and graft failure (GF).

Data are contained in the file “kidney-ECD-1.sas7bdat”, with fields:

IDNUM: patient ID number

ECD: equals 1 for an ECD kidney, and 0 for non-ECD

time-to-GF: time until graft failure (missing, if GF did not occur)

time-to-death: time until death (missing, if death not observed)

time-to-censor: potential time until censoring (non-missing for all patients)

AGE: age at transplant (years)

SEX

DIABETES: indicator that diabetes was the cause of renal failure

COMORBID: number of comorbid conditions (illnesses, not counting ESRD, existing at the time of transplant)

For each of the following parts, submit your code and output as an appendix.

**(a) Fit a model which contains only factors known at the time of transplant ( $t = 0$ ). List the factors that significantly predict death.**

See appendix. Factors significant in predicting death, according to our model:

(All of them) ECD, AGE, SEX, DIABETES, and COMORBID

**(b) Interpret the ECD effect from the model from (a).**

ECD Hazard Ratio Interpretation: Those who have an ECD kidney have a 13.4% increase in death hazard in comparison to those who have a non-ECD kidney, with all other covariates being held constant.

**Appendix**

```
kidney <- read_sas("~/WORKING_DIRECTORIES/biostat.675/kidney_ecd_1.sas7bdat")

kidney2 <- mutate(kidney,
                  death = 1-is.na(time_to_death),
                  time_to_event = time_to_death)
for(i in 1:nrow(kidney2)){
  if(is.na(kidney2$time_to_event[i])){
    kidney2$time_to_event[i] <- kidney2$time_to_censor[i]
  }
}
model4 <- coxph(data = kidney2,
                formula = Surv(time_to_event, death) ~ age +
                  male + diabetes + comorbid + ECD)
summary(model4)
```

```
Call:
coxph(formula = Surv(time_to_event, death) ~ age + male + diabetes +
      comorbid + ECD, data = kidney2)

n= 6384, number of events= 5675

      coef exp(coef) se(coef)      z Pr(>|z|)
age      0.022999  1.023265 0.001345 17.095 < 2e-16 ***
male      0.175051  1.191307 0.026772  6.539 6.21e-11 ***
diabetes   0.549840  1.732975 0.031974 17.197 < 2e-16 ***
comorbid   0.147817  1.159301 0.013260 11.148 < 2e-16 ***
ECD        0.125513  1.133730 0.029124  4.310 1.64e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
age           1.023      0.9773      1.021      1.026
male           1.191      0.8394      1.130      1.255
diabetes       1.733      0.5770      1.628      1.845
comorbid       1.159      0.8626      1.130      1.190
ECD            1.134      0.8820      1.071      1.200

Concordance= 0.602 (se = 0.004 )
Rsquare= 0.101 (max possible= 1 )
Likelihood ratio test= 681.8 on 5 df, p=<2e-16
Wald test              = 703.4 on 5 df, p=<2e-16
Score (logrank) test = 711.7 on 5 df, p=<2e-16
```