Name Jerry

Student Number

STA 312 s2019 Quiz 11

1. Consider a proportional hazards regression in which the cases are newly married couples with both people from the same religion. The explanatory variables are total family income and religion. Religion is coded A, B, C and None (let's call "None" a religion), and the response variable is time to divorce.

   (a) (1 point) In the space below, write the hazard function for this problem, including the baseline hazard. Denote income by x. There are no interactions. You do not have to say how the dummy variables are defined. You will do that in the next part.

   $h_0(t) \exp\{\beta_1 x + \beta_2 \pi_1 + \beta_3 \pi_2 + \beta_4 \pi_3\}$

   (b) (2 points) Make a table with four rows, showing how you would set up indicator dummy variables for Religion, with None as the reference category. Add a column showing the hazard function, including the baseline hazard. Symbols for the dummy variables must not appear in the last column, because they are all zero or one, and different for each row.

Religion	π_1	π_2	π_3	$h(t)$
A	1	0	0	$h_0(t)\, e^{\beta_1 x}\, e^{\beta_2}$
B	0	1	0	$h_0(t)\, e^{\beta_1 x}\, e^{\beta_3}$
C	0	0	1	$h_0(t)\, e^{\beta_1 x}\, e^{\beta_4}$
None	0	0	0	$h_0(t)\, e^{\beta_1 x}$

   (c) (1 point) For a constant value of income, the hazard of divorce is e^{β_3} as great for a couple from religion B as for a couple with no religion. Answer in terms of the β symbols of your model.

   (d) (1 point) You want to know whether marriages from Religions A and C have the same hazard of divorce. Write the null hypothesis in matrix form as $\mathbf{L}\boldsymbol{\beta} = \mathbf{h}$.

$$\begin{pmatrix} 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix}$$

2. (1 point) For the `channing` data (survival in nursing homes), you fit a proportional hazards model with no interactions. Consider the test of gender controlling for age.

   (a) (points) Fill in the table below. On your printout, circle the test statistic and p-value, and write "Question 2" beside them.

Chi-squared or Z Statistic	p-value	Reject H_0 at $\alpha = 0.05$? (Yes or No)
$Z = -2.187$	$p = 0.0288$	Yes

   (b) (2 points) Describe this finding in plain, non-statistical language. Complete the sentence below.

   Correcting for age of nursing home resident ...

   Women tended to survive longer.

3. (1 point) For the liver disease data, you tested whether the effect of experimental treatment depends on the patient's blood platelet count. Fill in the table below. On your printout, circle the test statistic and p-value, and write "Question 3" beside them.

Chi-squared or Z Statistic	p-value	Reject H_0 at $\alpha = 0.05$? (Yes or No)
$\chi^2 = 2.0372$	$p = 0.3611$	No

4. (1 point) In plain, non-statistical language, what do you conclude from that last test?

   There is not sufficient evidence to conclude that the effect of experimental treatment depends on blood platelet level.

Please attach *both* printouts. Make sure your name is on them.

```
R version 3.3.3 (2017-03-06) -- "Another Canoe"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.69 (7328) x86_64-apple-darwin13.4.0]

[Workspace restored from /Users/brunner/.RData]
[History restored from /Users/brunner/.Rapp.history]

> # channing
> rm(list=ls()); # options(scipen=999)
> # install.packages("KMsurv",dependencies=TRUE) # Only need to do this once
> library(KMsurv); library(survival)
> data(channing) # For some reason this is necessary
>
> retired = within(channing,
+ {
+ cage = (ageentry-mean(ageentry))/12 # Centered age in years
+ gender = gender-1 # 1=F, 0=M
+ })
> summary(retired)
      obs            death          ageentry          age            time           gender
 Min.   :  1.0   Min.   :0.000   Min.   : 733.0   Min.   : 777   Min.   :  0.00   Min.   :0.00
 1st Qu.:116.2   1st Qu.:0.000   1st Qu.: 854.0   1st Qu.: 939   1st Qu.: 35.00   1st Qu.:1.00
 Median :231.5   Median :0.000   Median : 900.0   Median : 990   Median : 81.50   Median :1.00
 Mean   :231.5   Mean   :0.381   Mean   : 905.7   Mean   : 986   Mean   : 80.12   Mean   :0.79
 3rd Qu.:346.8   3rd Qu.:1.000   3rd Qu.: 955.8   3rd Qu.:1031   3rd Qu.:137.00   3rd Qu.:1.00
 Max.   :462.0   Max.   :1.000   Max.   :1140.0   Max.   :1207   Max.   :137.00   Max.   :1.00
      cage
 Min.   :-14.3885
 1st Qu.: -4.3052
 Median : -0.4719
 Mean   :  0.0000
 3rd Qu.:  4.1740
 Max.   : 19.5281
>
>
> mod1 = coxph( Surv(time,death) ~ cage + gender, data=retired); summary(mod1)
Call:
coxph(formula = Surv(time, death) ~ cage + gender, data = retired)

  n= 462, number of events= 176

           coef exp(coef) se(coef)      z Pr(>|z|)
cage    0.08549   1.08925  0.01259  6.790 1.12e-11 ***
gender -0.37591   0.68666  0.17191 -2.187   0.0288 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Question 2

```
       exp(coef) exp(-coef) lower .95 upper .95
cage      1.0892     0.9181    1.0627    1.1165
gender    0.6867     1.4563    0.4902    0.9618

Concordance= 0.647  (se = 0.023 )
Rsquare= 0.102   (max possible= 0.985 )
Likelihood ratio test= 49.48  on 2 df,   p=1.804e-11
Wald test            = 51.51  on 2 df,   p=6.541e-12
Score (logrank) test = 52.6  on 2 df,   p=3.78e-12

> cox.zph(mod1)
          rho chisq     p
cage   0.0494 0.424 0.515
gender 0.0314 0.173 0.677
GLOBAL     NA 0.586 0.746
>
> mod2 = coxph( Surv(time,death) ~ cage*gender, data=retired); summary(mod2)
Call:
coxph(formula = Surv(time, death) ~ cage * gender, data = retired)

  n= 462, number of events= 176

               coef exp(coef) se(coef)      z Pr(>|z|)
cage         0.05952   1.06132  0.02739  2.173   0.0298 *
gender      -0.43168   0.64942  0.17751 -2.432   0.0150 *
cage:gender  0.03278   1.03332  0.03068  1.068   0.2854
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

            exp(coef) exp(-coef) lower .95 upper .95
cage           1.0613     0.9422    1.0059    1.1198
gender         0.6494     1.5398    0.4586    0.9196
cage:gender    1.0333     0.9678    0.9730    1.0974

Concordance= 0.652  (se = 0.023 )
Rsquare= 0.104   (max possible= 0.985 )
Likelihood ratio test= 50.62  on 3 df,   p=5.896e-11
Wald test            = 51.33  on 3 df,   p=4.152e-11
Score (logrank) test = 52.68  on 3 df,   p=2.144e-11

>
> # Uncentered
> coxph(formula = Surv(time, death) ~ ageentry + gender, data = channing)
Call:
coxph(formula = Surv(time, death) ~ ageentry + gender, data = channing)

             coef exp(coef) se(coef)     z       p
ageentry  0.00712   1.00715  0.00105  6.79 1.1e-11
gender   -0.37591   0.68666  0.17191 -2.19   0.029

Likelihood ratio test=49.5  on 2 df, p=1.8e-11
n= 462, number of events= 176
> summary(coxph( Surv(time,death) ~ ageentry + gender + ageentry:gender, data=channing) )
Call:
coxph(formula = Surv(time, death) ~ ageentry + gender + ageentry:gender,
    data = channing)

  n= 462, number of events= 176

                     coef exp(coef)  se(coef)      z Pr(>|z|)
ageentry         0.002228  1.002231  0.004704  0.474    0.636
gender          -2.905381  0.054728  2.365764 -1.228    0.219
ageentry:gender  0.002731  1.002735  0.002557  1.068    0.285
```

```
                exp(coef) exp(-coef) lower .95 upper .95
ageentry          1.00223     0.9978 0.9930335     1.012
gender            0.05473    18.2722 0.0005302     5.649
ageentry:gender   1.00274     0.9973 0.9977227     1.008

Concordance= 0.652  (se = 0.023 )
Rsquare= 0.104   (max possible= 0.985 )
Likelihood ratio test= 50.62  on 3 df,   p=5.896e-11
Wald test            = 51.33  on 3 df,   p=4.152e-11
Score (logrank) test = 52.68  on 3 df,   p=2.144e-11

>
>
```

```
R version 3.3.3 (2017-03-06) -- "Another Canoe"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.69 (7328) x86_64-apple-darwin13.4.0]

[Workspace restored from /Users/brunner/.RData]
[History restored from /Users/brunner/.Rapp.history]

> # Liver disease data
> rm(list=ls())
> # # install.packages("survival",dependencies=TRUE) # Only need to do this once
> library(survival)
> liver = read.table('http://www.utstat.toronto.edu/~brunner/data/legal/liver.data.txt')
> head(liver)
  id  drug age sex platelet failed start end
1  1 Drug2  33   0      239      0     0   1
2  1 Drug2  33   0      275      0     1   2
3  1 Drug2  33   0      185      0     2   3
4  1 Drug2  33   0      148      0     3   4
5  1 Drug2  33   0      289      0     4   5
6  1 Drug2  33   0      397      0     5   6
> summary(liver)
       id             drug          age              sex            platelet         failed
 Min.   :  1.0   Drug1  :954   Min.   :21.00   Min.   :0.0000   Min.   :-69.0   Min.   :0.0000
 1st Qu.: 91.0   Drug2  :999   1st Qu.:43.00   1st Qu.:0.0000   1st Qu.:196.0   1st Qu.:0.0000
 Median :191.0   Placebo:659   Median :49.00   Median :0.0000   Median :261.0   Median :0.0000
 Mean   :188.5                 Mean   :49.64   Mean   :0.4786   Mean   :260.4   Mean   :0.1436
 3rd Qu.:286.2                 3rd Qu.:56.00   3rd Qu.:1.0000   3rd Qu.:327.0   3rd Qu.:0.0000
 Max.   :375.0                 Max.   :80.00   Max.   :1.0000   Max.   :585.0   Max.   :1.0000
     start             end
 Min.   : 0.000   Min.   : 1.000
 1st Qu.: 2.000   1st Qu.: 3.000
 Median : 4.000   Median : 5.000
 Mean   : 4.722   Mean   : 5.722
 3rd Qu.: 8.000   3rd Qu.: 9.000
 Max.   :11.000   Max.   :12.000
>
> # Basic stats: One line per case
> onions = subset(liver,end==1); dim(onions)
[1] 375   8
> summary(onions)
       id             drug          age              sex            platelet         failed
 Min.   :  1.0   Drug1  :134   Min.   :21.00   Min.   :0.0000   Min.   :-39.0   Min.   :0.0000
 1st Qu.: 94.5   Drug2  :125   1st Qu.:44.00   1st Qu.:0.0000   1st Qu.:199.0   1st Qu.:0.0000
 Median :188.0   Placebo:116   Median :51.00   Median :0.0000   Median :267.0   Median :0.0000
 Mean   :188.0                 Mean   :51.13   Mean   :0.4853   Mean   :267.7   Mean   :0.3333
 3rd Qu.:281.5                 3rd Qu.:58.00   3rd Qu.:1.0000   3rd Qu.:338.0   3rd Qu.:1.0000
 Max.   :375.0                 Max.   :80.00   Max.   :1.0000   Max.   :518.0   Max.   :1.0000
```

```
     start        end
 Min.   :0   Min.   :1
 1st Qu.:0   1st Qu.:1
 Median :0   Median :1
 Mean   :0   Mean   :1
 3rd Qu.:0   3rd Qu.:1
 Max.   :0   Max.   :1
> mod1 = coxph(Surv(start,end,failed) ~ age + sex + drug + platelet, data=liver); summary(mod1)
Call:
coxph(formula = Surv(start, end, failed) ~ age + sex + drug +
    platelet, data = liver)

  n= 2612, number of events= 375

                coef  exp(coef)   se(coef)      z Pr(>|z|)
age          0.0159747  1.0161030  0.0048012  3.327 0.000877 ***
sex         -0.0555665  0.9459491  0.1041639 -0.533 0.593720
drugDrug2   -0.1103705  0.8955023  0.1247668 -0.885 0.376365
drugPlacebo  0.2785939  1.3212707  0.1277677  2.180 0.029223 *
platelet    -0.0017445  0.9982571  0.0005462 -3.194 0.001405 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

            exp(coef) exp(-coef) lower .95 upper .95
age            1.0161     0.9842    1.0066    1.0257
sex            0.9459     1.0571    0.7713    1.1602
drugDrug2      0.8955     1.1167    0.7012    1.1436
drugPlacebo    1.3213     0.7568    1.0286    1.6973
platelet       0.9983     1.0017    0.9972    0.9993

Concordance= 0.634  (se = 0.028 )
Rsquare= 0.011   (max possible= 0.758 )
Likelihood ratio test= 28.89  on 5 df,   p=2.435e-05
Wald test            = 29.14  on 5 df,   p=2.18e-05
Score (logrank) test = 29.25  on 5 df,   p=2.068e-05

> source("http://www.utstat.toronto.edu/~brunner/Rfunctions/Wtest.txt")
> betahat = mod1$coefficients; kovbetahat = vcov(mod1)
> drugL = rbind(c(0,0,1,0,0),
+               c(0,0,0,1,0))
> Wtest(L=drugL, Tn=betahat, Vn=kovbetahat)
          W          df     p-value
9.492974870 2.000000000 0.008682138
>
> LplaceboVSdrug2 = cbind(0,0,1,-1,0)
> Wtest(L=LplaceboVSdrug2, Tn=betahat, Vn=kovbetahat)
          W          df     p-value
9.009513398 1.000000000 0.002685779
>
> # Use Placebo as the reference category.
> # Comparing W = 9.009513398, p = 0.002685779
> liver2 = liver
> contrasts(liver2$drug) = contr.treatment(3,base=3)
> coxph(Surv(start,end,failed) ~ age + sex + drug + platelet, data=liver2)
Call:
coxph(formula = Surv(start, end, failed) ~ age + sex + drug +
    platelet, data = liver2)

              coef exp(coef)  se(coef)     z       p
age       0.015975  1.016103  0.004801  3.33 0.00088
sex      -0.055566  0.945949  0.104164 -0.53 0.59372
drug1    -0.278594  0.756847  0.127768 -2.18 0.02922
drug2    -0.388964  0.677758  0.129586 -3.00 0.00269
platelet -0.001744  0.998257  0.000546 -3.19 0.00140
```

```
Likelihood ratio test=28.9  on 5 df, p=2.43e-05
n= 2612, number of events= 375
>
>
> mod2 = coxph(Surv(start,end,failed) ~ age + sex + drug*platelet, data=liver); mod2
Call:
coxph(formula = Surv(start, end, failed) ~ age + sex + drug *
    platelet, data = liver)

                          coef exp(coef)  se(coef)     z       p
age                   0.015898  1.016025  0.004807  3.31 0.00094
sex                  -0.055448  0.946061  0.104173 -0.53 0.59454
drugDrug2            -0.395456  0.673373  0.348413 -1.14 0.25637
drugPlacebo           0.477830  1.612572  0.355174  1.35 0.17852
platelet             -0.001910  0.998092  0.000923 -2.07 0.03862
drugDrug2:platelet    0.001124  1.001124  0.001303  0.86 0.38853
drugPlacebo:platelet -0.000800  0.999200  0.001370 -0.58 0.55936

Likelihood ratio test=30.9  on 7 df, p=6.41e-05
n= 2612, number of events= 375
> anova(mod1,mod2)
Analysis of Deviance Table
 Cox model: response is  Surv(start, end, failed)
 Model 1: ~ age + sex + drug + platelet
 Model 2: ~ age + sex + drug * platelet
  loglik  Chisq Df P(>|Chi|)
1  -1837
2  -1836 2.0372  2     0.3611
>
```

Question 3