Assignment 11

① (a) $h(t) = h_0(t)\, e^{\beta_1 a + \beta_2 d_1 + \beta_3 d_2 + \beta_4 d_3 + \beta_5 d_4}$

(b)

	d_1	d_2	d_3	d_4	$h(t)$	
Obs Only	1	0	0	0	$h_0(t)\, e^{\beta_1 a}$	e^{β_2}
Placebo	0	0	0	0	$h_0(t)\, e^{\beta_1 a}$	
100 mg	0	1	0	0	$h_0(t)\, e^{\beta_1 a}$	e^{β_3}
500 mg	0	0	1	0	$h_0(t)\, e^{\beta_1 a}$	e^{β_4}
2,000 mg	0	0	0	1	$h_0(t)\, e^{\beta_1 a}$	e^{β_5}

(c) (i) $H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

(ii)

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$L \qquad \beta \qquad = \quad 0$$

(1d) (i) $H_0: \beta_2 = 0$

(ii)

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = 0$$

(e) (i) H_0: ~~$\beta_3 = \beta_4 = \beta_5$~~ $\beta_3 = \beta_4 = \beta_5 = 0$

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$L \qquad \beta = 0$$

(f) $\binom{4}{2} = 6$

(1g) (i) $H_0: \beta_5 = 0$

(ii)

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = 0$$

(h) (i) $H_0: \beta_3 = \beta_4$

$$\begin{pmatrix} 0 & 0 & 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = 0$$

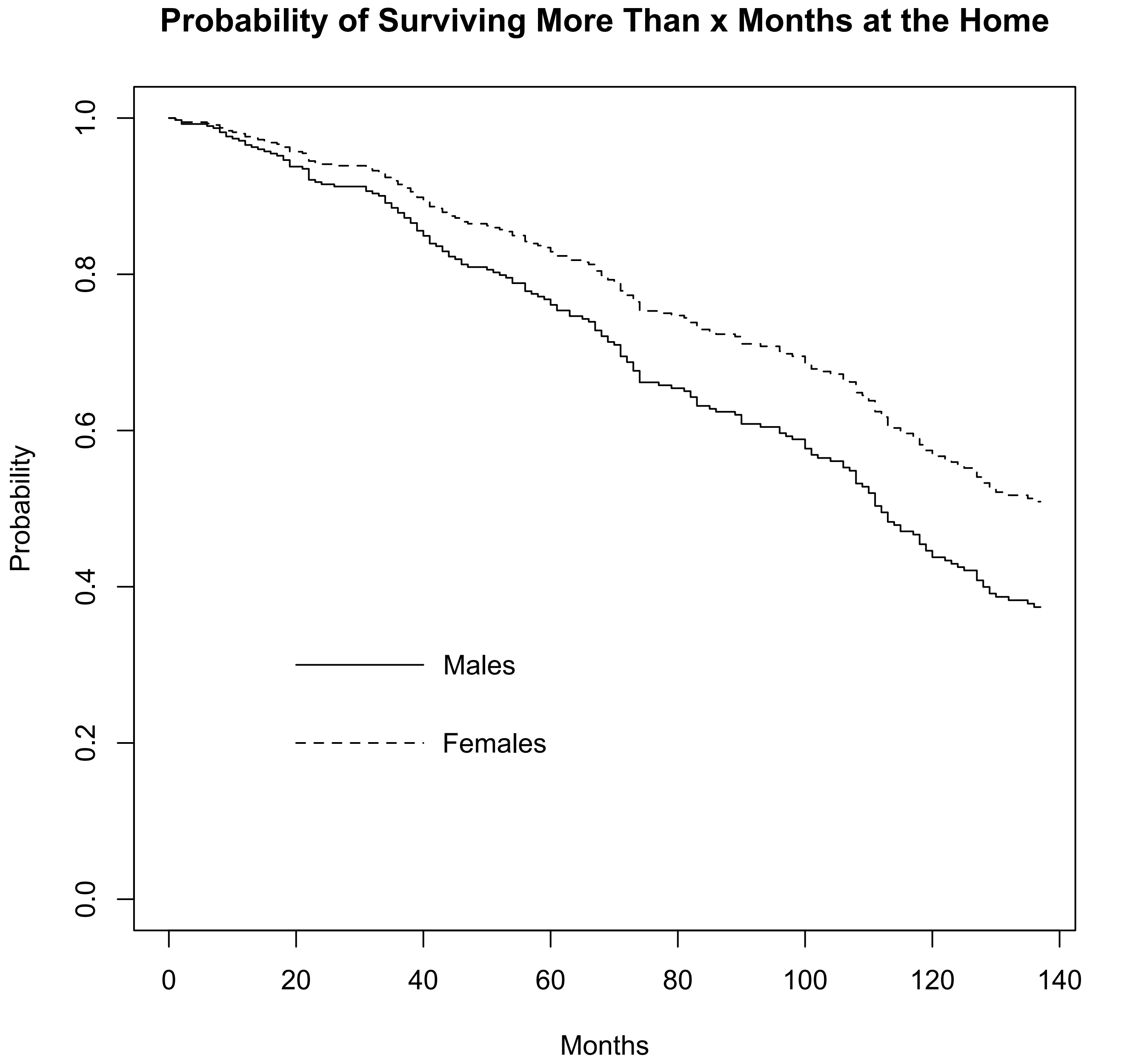
$$L \quad \beta = 0$$

Q2

```
R version 4.2.3 (2023-03-15) -- "Shortstop Beagle"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (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 1.79 (8198) x86_64-apple-darwin17.0]

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

> # Q2: 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
>
> # We want age at entry. Age at censoring or death is the response variable
>
> retired = within(channing,
+ {
+ ageentry = ageentry/12 # Age in years
+ cageentry = (ageentry-mean(ageentry)) # Centered ageentry in years
+ gender = gender-1 # 1=F, 0=M
+ })
> summary(retired) # Q2a
      obs            death          ageentry          age            time       
 Min.   :  1.0   Min.   :0.000   Min.   :61.08   Min.   : 777   Min.   :  0.00  
 1st Qu.:116.2   1st Qu.:0.000   1st Qu.:71.17   1st Qu.: 939   1st Qu.: 35.00  
 Median :231.5   Median :0.000   Median :75.00   Median : 990   Median : 81.50  
 Mean   :231.5   Mean   :0.381   Mean   :75.47   Mean   : 986   Mean   : 80.12  
 3rd Qu.:346.8   3rd Qu.:1.000   3rd Qu.:79.65   3rd Qu.:1031   3rd Qu.:137.00  
 Max.   :462.0   Max.   :1.000   Max.   :95.00   Max.   :1207   Max.   :137.00  
     gender       cageentry       
 Min.   :0.00   Min.   :-14.3885  
 1st Qu.:1.00   1st Qu.: -4.3052  
 Median :1.00   Median : -0.4719  
 Mean   :0.79   Mean   :  0.0000  
 3rd Qu.:1.00   3rd Qu.:  4.1740  
 Max.   :1.00   Max.   : 19.5281  
>
> # 2b: False. It would be a male of average age.
>
```

```
> # 2b: (i) - (vi)
> mod = coxph( Surv(time,death) ~ cageentry + gender, data=retired); summary(mod)
Call:
coxph(formula = Surv(time, death) ~ cageentry + gender, data = retired)

  n= 462, number of events= 176 

              coef exp(coef) se(coef)      z Pr(>|z|)    
cageentry  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

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

Concordance= 0.647  (se = 0.023 )
Likelihood ratio test= 49.48  on 2 df,   p=2e-11
Wald test            = 51.51  on 2 df,   p=7e-12
Score (logrank) test = 52.6  on 2 df,   p=4e-12

> summary(mod)
Call:
coxph(formula = Surv(time, death) ~ cageentry + gender, data = retired)

  n= 462, number of events= 176 

              coef exp(coef) se(coef)      z Pr(>|z|)    
cageentry  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

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

Concordance= 0.647  (se = 0.023 )
Likelihood ratio test= 49.48  on 2 df,   p=2e-11
Wald test            = 51.51  on 2 df,   p=7e-12
Score (logrank) test = 52.6  on 2 df,   p=4e-12

> 
> # 2b(vii)
> summary(coxph( Surv(time,death) ~ ageentry + gender, data=channing) )
Call:
coxph(formula = Surv(time, death) ~ ageentry + gender, data = channing)

  n= 462, number of events= 176 

              coef exp(coef)  se(coef)      z Pr(>|z|)    
ageentry  0.007124  1.007149  0.001049  6.790 1.12e-11 ***
gender   -0.375915  0.686661  0.171907 -2.187   0.0288 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
         exp(coef) exp(-coef) lower .95 upper .95
ageentry    1.0071     0.9929    1.0051    1.0092
gender      0.6867     1.4563    0.4902    0.9618

Concordance= 0.647  (se = 0.023 )
Likelihood ratio test= 49.48  on 2 df,   p=2e-11
Wald test            = 51.51  on 2 df,   p=7e-12
Score (logrank) test = 52.6  on 2 df,   p=4e-12

>
>
> # 2b(viii)
> # Just like in PH2 with R
> guy = data.frame(cageentry=0, gender=0); gal = data.frame(cageentry=0, gender=1)
> sexcomp = rbind(guy,gal); rownames(sexcomp) = c("M","F"); sexcomp
  cageentry gender
M         0      0
F         0      1
> s = survfit(mod,newdata=sexcomp); s
Call: survfit(formula = mod, newdata = sexcomp)

    n events median 0.95LCL 0.95UCL
M 462    176    112      98      NA
F 462    176     NA     124      NA
>
>
> # Make a nice plot
> plot(s,lty = c(1,2),xlab="Months", ylab="Probability")
> title('Probability of Surviving More Than x Months at the Home')
> xm = c(20,40); ym = c(0.3,0.3); lines(xm,ym,lty=1)
> text(50,0.3,"Males  ")
> xf = c(20,40); yf = c(0.2,0.2); lines(xf,yf,lty=2)
> text(50,0.2,"  Females")
>
>
```


Probability of Surviving More Than x Months at the Home
Probability
Months
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
120
140
Males
Females

vii. Suppose you had left gender as 1-2 and you left age in months, uncentered. Would this affect the test statistics and the conclusions? Run a quick example to find out. I used the original `channing` data set.

viii. Estimate the median survival times for men of average age and women of average age. Oops! I can't get an estimate for women.

ix. Make a well-labelled plot showing the estimated survival functions for a man and a woman of average age. There should be two different line types, like maybe solid and dashed. Print the graph and bring it to the quiz. From the picture, can you see why there was no estimated median for women?

x. The following is just a comment, not a question. A crucial assumption of the censored data model is that the censoring mechanism is independent of the outcome. However, lots of nursing home residents leave the home to go to hospital, and they die there rather than at the home. In fact, they are censored *because* the staff thought they were probably about to die. To me, this helps explain the high estimated probabilities of lasting a long time that we see in these data.

3. It is very natural and tempting to use age as a time-varying covariate. This question tries to illustrate why it won't work. Consider a model with an indicator for gender, and also age. Age $= x + t$, where x is age at entry to the study. The partial likelihood can be written

$$
\begin{aligned}
\mathrm{PL}(\beta) &= \prod_{i=1}^{D} \left(\frac{e^{\mathbf{x}(t_{(i)})^\top \beta}}{\sum_{j \in R_{(i)}} e^{\mathbf{x}(t_{(i),j})^\top \beta}} \right) \\
&= \prod_{i=1}^{D} \left(\frac{e^{\beta_1 s_{(i)} + \beta_2 (x_{(i)} + t_{(i)})}}{\sum_{j \in R_{(i)}} e^{\beta_1 s_{(i),j} + \beta_2 (x_{(i),j} + t_{(i)})}} \right).
\end{aligned}
$$

Note that the event happens to individual (i) at time $t_{(i)}$, and the explanatory variables in the risk set are all assessed at exactly that same instant $t_{(i)}$. What happens?

$$= \prod_{i=1}^{n} \frac{e^{\beta_1 s_{(i)} + \beta_2 x_{(i)} + \beta_2 t_{(i)}}}{\sum_{j \in R_{(i)}} e^{\beta_1 s_{(i),j} + \beta_2 x_{(i),j} + \beta_2 t_{(i)}}}$$

$$= \prod_{i=1}^{n} \frac{e^{\beta_2 t_{(i)}} e^{\beta_1 s_{(i)} + \beta_2 x_{(i)}}}{e^{\beta_2 t_{(i)}} \sum_{j \in R_{(i)}} e^{\beta_1 s_{(i),j} + \beta_2 x_{(i),j}}}$$

Q4

```
R version 4.2.3 (2023-03-15) -- "Shortstop Beagle"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (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 1.79 (8198) x86_64-apple-darwin17.0]

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

> #Q4: Area 51
>
> library(survival)
> area51 = read.table("https://www.utstat.toronto.edu/brunner/data/legal/area51.data.txt")
> alien = coxph(Surv(time1,time2,taken) ~ age + sex + hat, data=area51)
> summary(alien)
Call:
coxph(formula = Surv(time1, time2, taken) ~ age + sex + hat,
    data = area51)

  n= 4244, number of events= 103

         coef exp(coef) se(coef)      z Pr(>|z|)
age  -0.00068   0.99932  0.01020 -0.067  0.94683
sexM  0.15126   1.16330  0.19811  0.764  0.44515
hat  -0.56542   0.56812  0.20478 -2.761  0.00576 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     exp(coef) exp(-coef) lower .95 upper .95
age     0.9993     1.0007    0.9795    1.0195
sexM    1.1633     0.8596    0.7890    1.7152
hat     0.5681     1.7602    0.3803    0.8487

Concordance= 0.591  (se = 0.028 )
Likelihood ratio test= 8.48  on 3 df,   p=0.04
Wald test            = 8.2  on 3 df,   p=0.04
Score (logrank) test = 8.4  on 3 df,   p=0.04

>
>
```