Midterm

Nathan Hawkins

3/2/2022

1) The standard uniform distribution has a height of 1 between 0 and 1 and 0 elsewhere. For the standard uniform compute F(t), S(t), h(t), and H(t).

$f(x) = 1$

$F(x) = x$

$S(x) = 1 - x$

$h(x) = \frac{1}{1-x}$

$H(x) = -ln(1 - x)$

2)

I estimate that the mean of the women's normal curve is 66.04 with sd = 5.830 and the mean of the men's normal curve is 74 with standard deviation 5.832.

```
island = read.csv("islandheights.csv")

llik = function(x,par){
m=par[1]
s=par[2]
n=length(x[,3])
# log of the normal likelihood
# -n/2 * log(2*pi*s^2) + (-1/(2*s^2)) * sum((x-m)^2)
ll = - (n/2)*(log(2*pi*s^2)) + (-1/(2*s^2)) * sum((x[,1]-m)^2) - (n/2)*(log(2*pi*s^2)) + (-1/(2*s^2)) *

# return the negative to maximize rather than minimize
return(-ll)
}

x = cbind(island$upper[1:10], island$lower[1:10], island$count.Freq[1:10])
nisland = island$count.Freq
res0 = optim(par=c(.5,.5), llik, x=x)



llik = function(x,par){
m=par[1]
s=par[2]
n=length(x[,3])
```

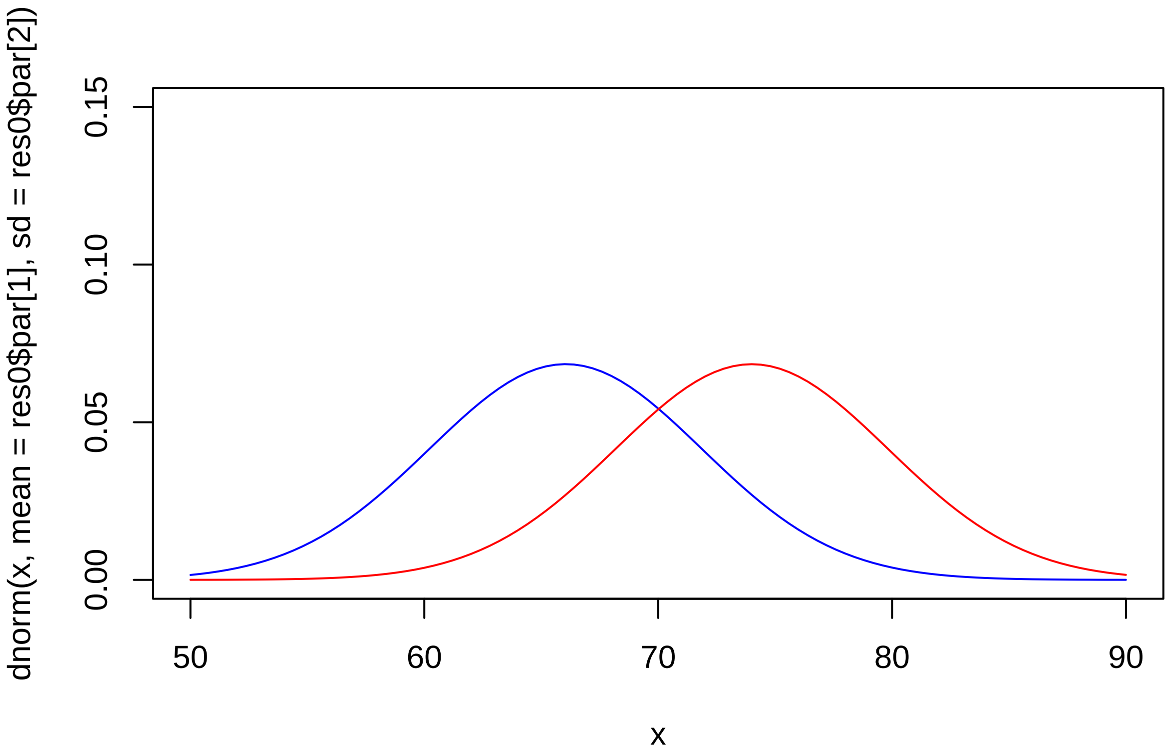
```
ll = -(n/2)*(log(2*pi*s^2)) + (-1/(2*s^2)) * sum((x[,1]-m)^2) - (n/2)*(log(2*pi*s^2)) + (-1/(2*s^2)) * s

return(-ll)
}

x = cbind(island$upper[5:14], island$lower[5:14], island$count.Freq[5:14])
res1 = optim(par=c(75,1), llik, x=x)

# Here I show the curves of the two distributions

curve(dnorm(x, mean = res0$par[1], sd = res0$par[2]), from = 50, to = 90,ylim = c(0,0.15), col = "blue")
curve(dnorm(x, mean = res1$par[1], sd = res1$par[2]), from = 50, to = 90, ylim = c(0,0.3), add = TRUE, c
```



I couldn't get this example to work

```
# llik = function(dat,par){
# m1=par[1]
# m2=par[2]
# s1=par[3]
# s2=par[4]
#
#
# # log of the normal likelihood
# # -n/2 * log(2*pi*s^2) + (-1/(2*s^2)) * sum((x-m)^2)
#
#
```

```r
# ll_vec = NA
# for(i in 1:nrow(island)){
#   ll_vec[i] =sum(dat[i,4]*(emdbook::dmvnorm(x = c(dat[i,1], dat[i,1]), mu = c(m1,m2),
#                               Sigma = matrix(c(s1,0,0,s1),nrow = 2, ncol = 2, byrow = TRUE),
#                    log = TRUE) -
#                 dat[i,4]*emdbook::dmvnorm(x = c(dat[i,2], dat[i,2]), mu = c(m1,m2),
#                               Sigma = matrix(c(s1,0,0,s1),nrow = 2, ncol = 2, byrow = TRUE),
#                    log = TRUE)))
# }
#
# ll = sum(ll_vec)
#
#
# # return the negative to maximize rather than minimize
# return(-ll)
# }
```

3)

My final model has Hours of Sleep, ClassTime, Light, Student Sex, Instructor Sex, and Student-Instructor Sex interaction term. After comparing AIC values, this model had the lowest while including the variables that were significant at the alpha = 0.05 level. Note that Student and Instructor Sex alone aren't significant but the interaction is.

```r
sleep = read.csv("sleep.csv")

fit1 = coxph(Surv(Start, Stop, Status)~  HoursSleep + ClassTime +
               StudentSex*InstructorSex + Light + cluster(ID),
        data=sleep)

summary(fit1)
```

```
## Call:
## coxph(formula = Surv(Start, Stop, Status) ~ HoursSleep + ClassTime +
##     StudentSex + InstructorSex + Light + StudentSex:InstructorSex,
##     data = sleep, cluster = ID)
##
##   n= 1792, number of events= 383
##
##                               coef exp(coef) se(coef) robust se      z
## HoursSleep                -0.13538   0.87339  0.02698   0.02878 -4.705
## ClassTime                  0.10994   1.11621  0.02134   0.01906  5.768
## StudentSexM                0.17050   1.18590  0.15353   0.15560  1.096
## InstructorSexM             0.23251   1.26177  0.14352   0.14245  1.632
## LightLow                   0.41754   1.51821  0.10390   0.10261  4.069
## StudentSexM:InstructorSexM -0.41619   0.65955  0.20668   0.20815 -2.000
##                           Pr(>|z|)
## HoursSleep                2.54e-06 ***
## ClassTime                 8.01e-09 ***
## StudentSexM                 0.2732
## InstructorSexM              0.1026
```

```
## LightLow                    4.72e-05 ***
## StudentSexM:InstructorSexM   0.0456 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                            exp(coef) exp(-coef) lower .95 upper .95
## HoursSleep                    0.8734     1.1450    0.8255    0.9241
## ClassTime                     1.1162     0.8959    1.0753    1.1587
## StudentSexM                   1.1859     0.8432    0.8742    1.6088
## InstructorSexM                1.2618     0.7925    0.9544    1.6682
## LightLow                      1.5182     0.6587    1.2416    1.8564
## StudentSexM:InstructorSexM    0.6596     1.5162    0.4386    0.9918
##
## Concordance= 0.622  (se = 0.014 )
## Likelihood ratio test= 72.44  on 6 df,   p=1e-13
## Wald test            = 80.34  on 6 df,   p=3e-15
## Score (logrank) test = 73.17  on 6 df,   p=9e-14,   Robust = 71.18  p=2e-13
##
##   (Note: the likelihood ratio and score tests assume independence of
##      observations within a cluster, the Wald and robust score tests do not).
```

```
AIC(fit1)
```

```
## [1] 5061.263
```

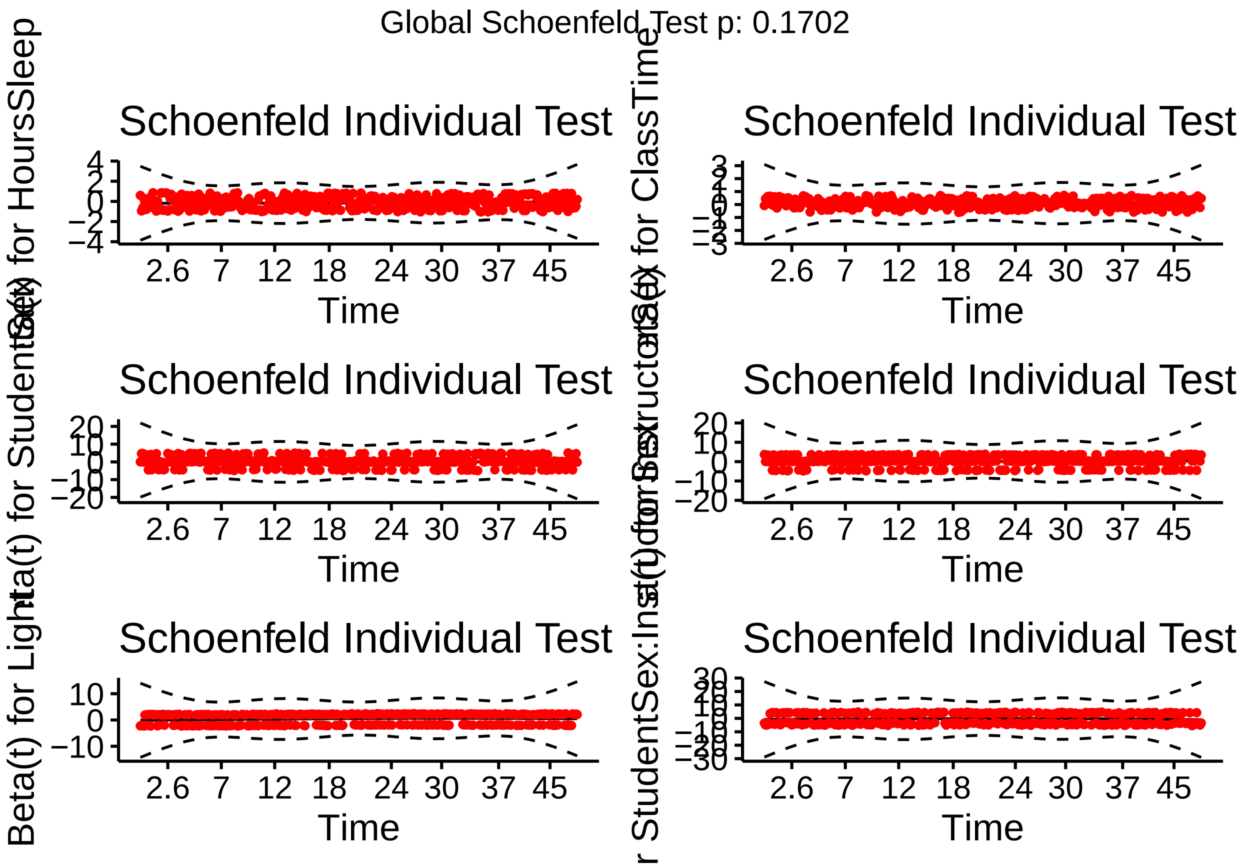
Having low lights, a earlier start time for a class, getting less sleep, or having a teacher-student be the different genders increases the risk of falling asleep.

Below I check the proportional hazards assumption. With no significant p-values and plots that show the values within acceptable ranges, I conclude that the proportional hazards assumption is met.

```
test.ph <- cox.zph(fit1)
test.ph
```

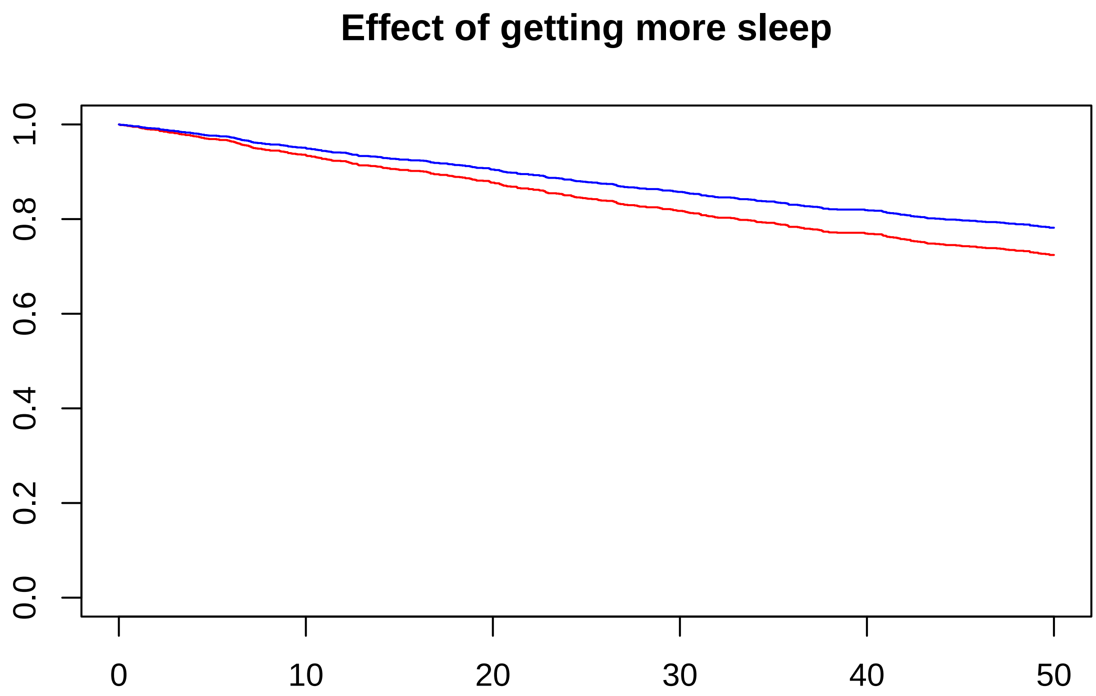
```
##                               chisq df     p
## HoursSleep                 3.213605  1 0.073
## ClassTime                  0.000235  1 0.988
## StudentSex                 2.113971  1 0.146
## InstructorSex              0.053548  1 0.817
## Light                      3.386436  1 0.066
## StudentSex:InstructorSex 1.180852  1 0.277
## GLOBAL                     9.060903  6 0.170
```

```
ggcoxzph(test.ph)
```



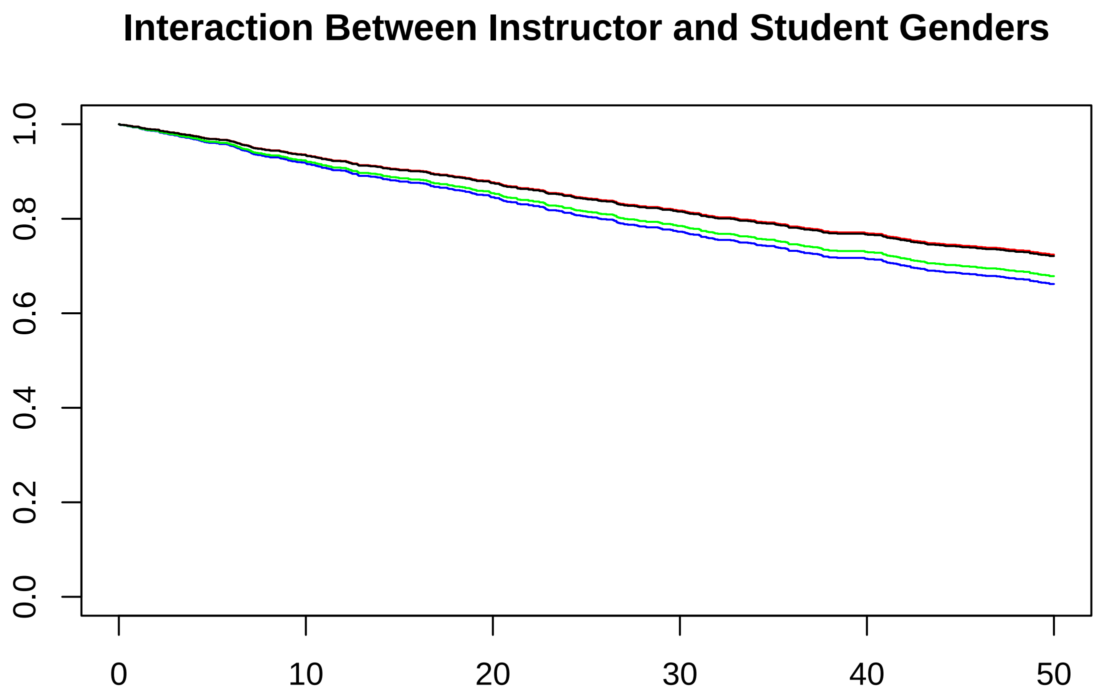
This plot shows the difference in survival curves between a student with 7 hours of sleep (the median, in blue) and a student with 9 hours of sleep (3rd quartile, in red) holding all else equal.

```
new.dat1 <- data.frame(Light=c("High","High"), HoursSleep = c(7,9),
                       ClassTime = c(12,12), StudentSex = c("M", "M"),
                       InstructorSex = c("M", "M"))
plot(survfit(fit1, newdata=new.dat1),
     col=c('red','blue'), main = "Effect of getting more sleep")
```



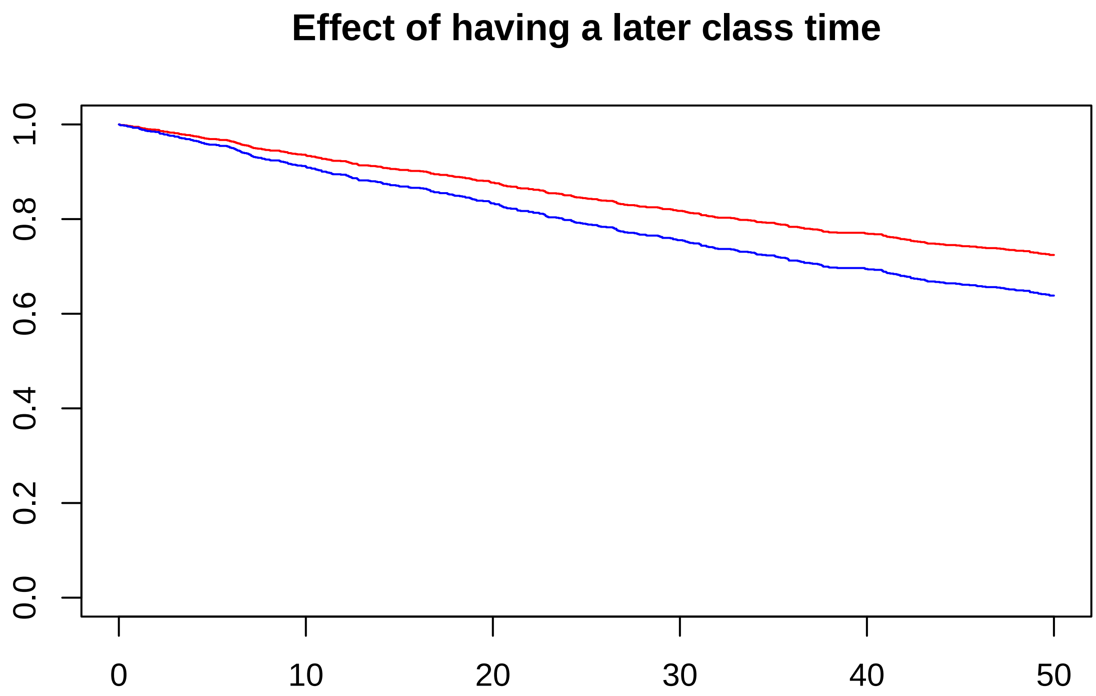
This plot shows the importance of the interaction between the sexes of the student and teacher. Black and red lines are survival curves for students that are the same gender as their instructor Also green and blue lines show the survival curve for students that are the opposite gender of their instructor, all else being held equal.

```
new.dat1 <- data.frame(Light=c("High","High", "High", "High"), HoursSleep = c(7,7,7,7),
                       ClassTime = c(12,12,12,12), StudentSex = c("M", "F", "M", "F"),
                       InstructorSex = c("M", "M", "F", "F"))
plot(survfit(fit1, newdata=new.dat1),
     col=c('red','blue', "green", "black"), main = "Interaction Between Instructor and Student Genders")
```



Having a later class time is shown in the blue line compared to red time holding all else equal in the following survival plot. Notice that survival curve is lower for the blue line, indicating that students are more likely to fall asleep in classes that start later in the day, holding all else equal.

```
new.dat1 <- data.frame(Light=c("High","High"), HoursSleep = c(7,7),
                       ClassTime = c(12,15), StudentSex = c("M", "M"),
                       InstructorSex = c("M", "M"))
plot(survfit(fit1, newdata=new.dat1),
     col=c('red','blue'), main = "Effect of having a later class time")
```


Effect of having a later class time
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50