HW3 - Applied Survival Analysis

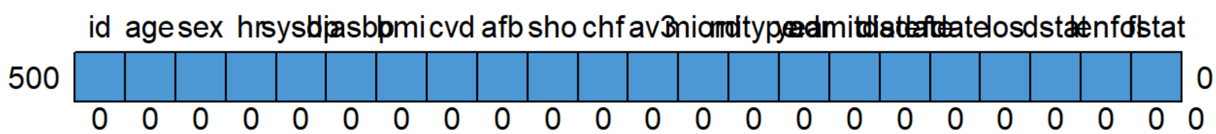
Benjamin Panny

2023-10-16

For this problem, use the WHAS500 data with non-missing values for sex (the primary covariate of interest) and bmi. Use length of follow-up as the survival time variable, and status at last follow-up as the censoring variable.

```
whas %>% md.pattern()
```

```
##  /\     /\
## {  `---'  }
## {  O   O  }
## ==>  V <==  No need for mice. This data set is completely observed.
##  \  \|/  /
##   `-----'
```




```
##     id age sex hr sysbp diasbp bmi cvd afb sho chf av3 miord mitype year
## 500  1   1   1  1     1      1   1   1   1   1   1   1     1      1    1
##      0   0   0  0     0      0   0   0   0   0   0   0     0      0    0
##     admitdate disdate fdate los dstat lenfol fstat
## 500         1       1     1   1     1      1     1 0
##             0       0     0   0     0      0     0 0
```

```
whas %>% glimpse
```

```
## Rows: 500
## Columns: 22
## $ id        <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 1…
## $ age       <dbl> 83, 49, 70, 70, 70, 70, 57, 55, 88, 54, 48, 75, 48, 54, 67, …
## $ sex       <dbl> 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, …
## $ hr        <dbl> 89, 84, 83, 65, 63, 76, 73, 91, 63, 104, 95, 154, 85, 95, 93…
## $ sysbp     <dbl> 152, 120, 147, 123, 135, 83, 191, 147, 209, 166, 160, 193, 1…
## $ diasbp    <dbl> 78, 60, 88, 76, 85, 54, 116, 95, 100, 106, 110, 123, 80, 65,…
## $ bmi       <dbl> 25.54051, 24.02398, 22.14290, 26.63187, 24.41255, 23.24236, …
## $ cvd       <dbl> 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0, 1, …
## $ afb       <dbl> 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, …
## $ sho       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ chf       <dbl> 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, …
## $ av3       <dbl> 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, …
## $ miord     <dbl> 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, …
## $ mitype    <dbl> 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, …
## $ year      <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, …
## $ admitdate <chr> "01/13/1997", "01/19/1997", "01/01/1997", "02/17/1997", "03/…
## $ disdate   <chr> "01/18/1997", "01/24/1997", "01/06/1997", "02/27/1997", "03/…
## $ fdate     <chr> "12/31/2002", "12/31/2002", "12/31/2002", "12/11/1997", "12/…
## $ los       <dbl> 5, 5, 5, 10, 6, 1, 5, 4, 4, 5, 5, 10, 7, 21, 4, 1, 13, 14, 6…
## $ dstat     <dbl> 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ lenfol    <dbl> 2178, 2172, 2190, 297, 2131, 1, 2122, 1496, 920, 2175, 2173,…
## $ fstat     <dbl> 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, …
```

a. Fit the proportional hazards model containing sex and estimate the hazard ratio (between two sex groups), pointwise and confidence interval.

```
cox.fit.1<-coxph(Surv(lenfol,fstat)~sex, data=whas)
cox.fit.1 %>% summary
```

```
## Call:
## coxph(formula = Surv(lenfol, fstat) ~ sex, data = whas)
##
##   n= 500, number of events= 215
##
##       coef exp(coef) se(coef)     z Pr(>|z|)
## sex 0.3815    1.4645   0.1376 2.773  0.00556 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##     exp(coef) exp(-coef) lower .95 upper .95
## sex     1.464     0.6828     1.118     1.918
##
## Concordance= 0.542  (se = 0.018 )
## Likelihood ratio test= 7.6  on 1 df,   p=0.006
## Wald test            = 7.69  on 1 df,   p=0.006
## Score (logrank) test = 7.78  on 1 df,   p=0.005
```

The hazard ratio between the two sex groups is 1.4645. That is, when the value of sex == 1 instead of 0, the hazard rate is ~46.5% higher on average.

The 95% confidence interval for the hazard ratio is

```
exp(confint(cox.fit.1))
```

```
##        2.5 %   97.5 %
## sex 1.118349 1.917785
```

If we follow the same procedures, our constructed 95% confidence interval will capture the true hazard ratio for sex==1 vs. sex==0 95% of the time

b. Add bmi to the model fit. Is bmi a confounder of the effect of sex? Explain the reasons for your answer.

```
cox.fit.2<-coxph(Surv(lenfol,fstat)~sex+bmi, data=whas)
cox.fit.2 %>% summary
```

```
## Call:
## coxph(formula = Surv(lenfol, fstat) ~ sex + bmi, data = whas)
##
##   n= 500, number of events= 215
##
##         coef exp(coef) se(coef)      z Pr(>|z|)
## sex  0.21894   1.24476  0.14140  1.548    0.122
## bmi -0.09308   0.91112  0.01491 -6.241 4.34e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##     exp(coef) exp(-coef) lower .95 upper .95
## sex    1.2448     0.8034    0.9435    1.6423
## bmi    0.9111     1.0976    0.8849    0.9381
##
## Concordance= 0.647  (se = 0.019 )
## Likelihood ratio test= 50.71  on 2 df,   p=1e-11
## Wald test            = 48.32  on 2 df,   p=3e-11
## Score (logrank) test = 48.46  on 2 df,   p=3e-11
```

```
cox.fit.3<-coxph(Surv(lenfol,fstat)~bmi, data=whas)
cox.fit.3 %>% summary
```

```
## Call:
## coxph(formula = Surv(lenfol, fstat) ~ bmi, data = whas)
##
##   n= 500, number of events= 215
##
##         coef exp(coef) se(coef)      z Pr(>|z|)
## bmi -0.09847   0.90622  0.01475 -6.676 2.46e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##     exp(coef) exp(-coef) lower .95 upper .95
## bmi    0.9062      1.103    0.8804    0.9328
##
## Concordance= 0.648  (se = 0.019 )
## Likelihood ratio test= 48.33  on 1 df,   p=4e-12
## Wald test            = 44.57  on 1 df,   p=2e-11
## Score (logrank) test = 44.31  on 1 df,   p=3e-11
```

```
lm(bmi ~ sex, data = whas) %>% summary
```

```
##
## Call:
## lm(formula = bmi ~ sex, data = whas)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.5856  -3.3509  -0.6785   2.7801  19.2078
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  27.2689     0.3089  88.266  < 2e-16 ***
## sex          -1.6378     0.4885  -3.353 0.000861 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.351 on 498 degrees of freedom
## Multiple R-squared:  0.02208,    Adjusted R-squared:  0.02011
## F-statistic: 11.24 on 1 and 498 DF,  p-value: 0.0008607
```

I believe BMI is a confounder of sex because it is significantly related to both the time-to-outcome and sex, and it also alters the estimate of the hazard ratio for the sex groups when included in the same model.

```
(.381-.218)/.218*100
```

```
## [1] 74.77064
```

The crude estimate for sex is about 75% larger than the adjusted estimate for sex when BMI is included

c. Estimate the bmi-adjusted hazard ratio (between two sex groups), pointwise and confidence interval.

```
cox.fit.2 %>% summary
```

```
## Call:
## coxph(formula = Surv(lenfol, fstat) ~ sex + bmi, data = whas)
##
##   n= 500, number of events= 215
##
##         coef exp(coef) se(coef)      z Pr(>|z|)
## sex  0.21894   1.24476  0.14140  1.548    0.122
## bmi -0.09308   0.91112  0.01491 -6.241 4.34e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##     exp(coef) exp(-coef) lower .95 upper .95
## sex    1.2448     0.8034    0.9435    1.6423
## bmi    0.9111     1.0976    0.8849    0.9381
##
## Concordance= 0.647  (se = 0.019 )
## Likelihood ratio test= 50.71  on 2 df,   p=1e-11
## Wald test            = 48.32  on 2 df,   p=3e-11
## Score (logrank) test = 48.46  on 2 df,   p=3e-11
```

```
whas %>% summarise(mean_bmi=mean(bmi))
```

```
## # A tibble: 1 × 1
##   mean_bmi
##      <dbl>
## 1     26.6
```

The BMI-adjusted hazard ratios between the two sexes is 1.248, with a 95% confidence interval of (.94, 1.64). Clearly, including BMI in the model has shifted the effect of sex towards the null hypothesis of zero effect on hazard rates.

d. Is there a significant interaction between bmi and sex? Justify your answer.

```
cox.fit.4<-coxph(Surv(lenfol,fstat)~sex*bmi, data=whas)
cox.fit.4 %>% summary
```

```
## Call:
## coxph(formula = Surv(lenfol, fstat) ~ sex * bmi, data = whas)
##
##   n= 500, number of events= 215
##
##             coef exp(coef) se(coef)      z Pr(>|z|)
## sex     -0.80152   0.44865  0.76538 -1.047    0.295
## bmi     -0.11716   0.88944  0.02346 -4.993 5.93e-07 ***
## sex:bmi  0.04094   1.04179  0.03021  1.355    0.175
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##         exp(coef) exp(-coef) lower .95 upper .95
## sex        0.4486     2.2289    0.1001    2.0109
## bmi        0.8894     1.1243    0.8495    0.9313
## sex:bmi    1.0418     0.9599    0.9819    1.1053
##
## Concordance= 0.65  (se = 0.019 )
## Likelihood ratio test= 52.56  on 3 df,   p=2e-11
## Wald test            = 46.92  on 3 df,   p=4e-10
## Score (logrank) test = 48.47  on 3 df,   p=2e-10
```

There is not a statistically significant interaction effect, given the Wald Test Statistic we observe for the interaction coefficient with a $p > 0.05$

```
anova(cox.fit.2, cox.fit.4)
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(lenfol, fstat)
##  Model 1: ~ sex + bmi
##  Model 2: ~ sex * bmi
##   loglik  Chisq Df Pr(>|Chi|)
## 1  -1202
## 2  -1201 1.8427  1     0.1746
```

This is reiterated by the partial log likelihood ratio test

e. Explain why the main effect coefficients for sex in the models fit in 1(b) and 1(d) are different.

The main effect coefficients are different between the models for 1(b) and 1(d) because in 1(d), the main effect coefficient now represents the hazard ratio for the two sexes when BMI is 0.

As can be seen here,

```
exp(-.801+0*-.117+.04*0)/exp(0*-.117)
```

```
## [1] 0.4488799
```

However, it can also be seen that the hazard ratio represented by the main effect coefficient in 1(b) can be recovered if we set BMI to a plausible value, such as its rounded average.

```
exp(-.801+26*-.117+.04*26)/exp(26*-.117)
```

```
## [1] 1.269979
```

which is very close the main effect term in 1(b)

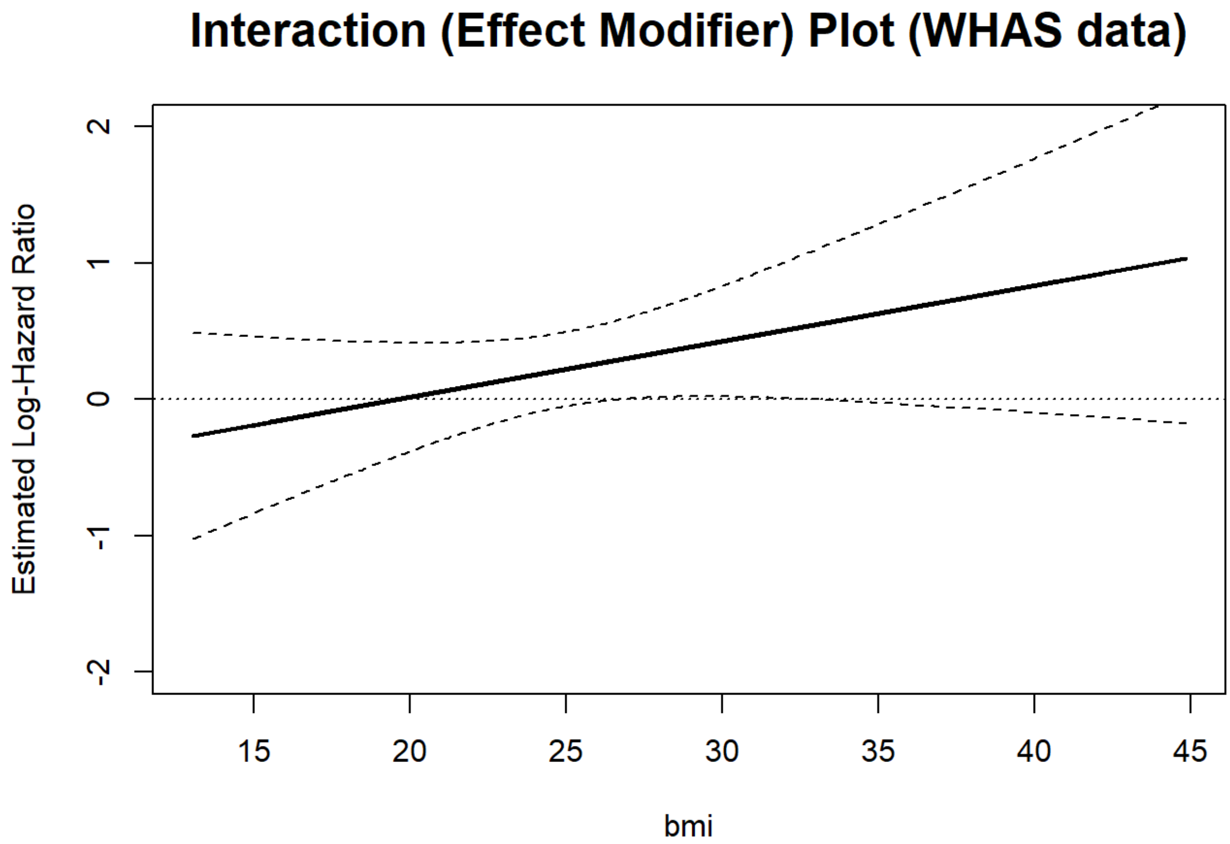
f. Using the interaction model fitted in 1(d), plot the estimated hazard ratio between two sex groups with confidence limits (upper and lower) as a function of bmi.

```
cox.fit.4$coef[1]
```

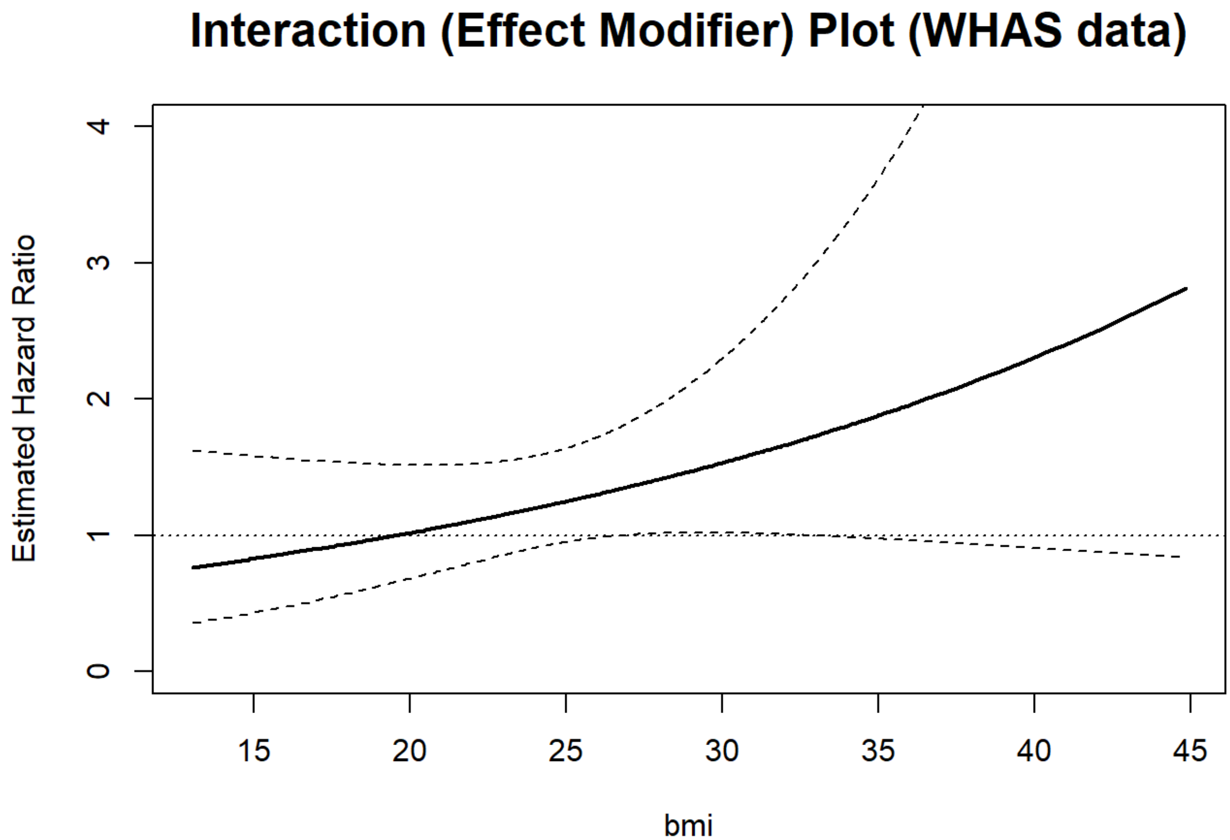
```
##        sex
## -0.8015171
```

```
bmi.num<-as.numeric(whas$bmi)

low.lim<-cox.fit.4$coef[1]+cox.fit.4$coef[3]*bmi.num-1.96*sqrt(cox.fit.4$var[1,1]+cox.fit.4$var[3,3]*bmi.num^2
+2*bmi.num*cox.fit.4$var[1,3])
high.lim<-cox.fit.4$coef[1]+cox.fit.4$coef[3]*bmi.num+1.96*sqrt(cox.fit.4$var[1,1]+cox.fit.4$var[3,3]*bmi.num^
2+2*bmi.num*cox.fit.4$var[1,3])
plot(bmi.num,cox.fit.4$coef[1]+cox.fit.4$coef[3]*bmi.num,ylim=c(-2,2),xlab="bmi",ylab="Estimated Log-Hazard Ra
tio",type="l",lty=1,lwd=2,)
lines(bmi.num[order(bmi.num)],low.lim[order(bmi.num)],lty=2)
lines(bmi.num[order(bmi.num)],high.lim[order(bmi.num)],lty=2)
abline(h=0,lty=3)
title(main="Interaction (Effect Modifier) Plot (WHAS data)",cex.main=1.5)
```




```
low.lim<-exp(cox.fit.4$coef[1]+cox.fit.4$coef[3]*bmi.num-1.96*sqrt(cox.fit.4$var[1,1]+cox.fit.4$var[3,3]*bmi.num^2+2*bmi.num*cox.fit.4$var[1,3]))
high.lim<-exp(cox.fit.4$coef[1]+cox.fit.4$coef[3]*bmi.num+1.96*sqrt(cox.fit.4$var[1,1]+cox.fit.4$var[3,3]*bmi.num^2+2*bmi.num*cox.fit.4$var[1,3]))
plot(bmi.num[order(bmi.num)],exp(cox.fit.4$coef[1]+cox.fit.4$coef[3]*bmi.num)[order(bmi.num)],ylim=c(0,4),xlab="bmi",ylab="Estimated Hazard Ratio", type = "l", lwd=2)
lines(bmi.num[order(bmi.num)],low.lim[order(bmi.num)],lty=2)
lines(bmi.num[order(bmi.num)],high.lim[order(bmi.num)],lty=2)
abline(h=1,lty=3)
title(main="Interaction (Effect Modifier) Plot (WHAS data)",cex.main=1.5)
```



As can be seen from the plots, the hazard ratio between the sexes increases as bmi increases. However, after about bmi = 25, the confidence interval expands rapidly to cover a very large interval.

g. Using the interaction model fit in 1(d), estimate the hazard ratio, pointwise, and confidence interval for a 5kg/m2 increase in bmi for each sex group.

The hazard ratio for a 5kg/m2 increase in bmi for each sex group is the change in hazard rate with a change in bmi given male or female.

```
coefs_4 <- cox.fit.4$coef
exp(coefs_4[1]+5*coefs_4[2]+5*coefs_4[3])/exp(coefs_4[1]+0*coefs_4[2]+0*coefs_4[3]) %>% as.vector
```

```
##       sex
## 0.6830943
```

The hazard ratio for a 5kg/m2 increase in BMI for sex == 1 is 0.683 with a 95% confidence interval of (0.646, 0.721), calculated below

```
covs <- vcov(cox.fit.4)
se_full <- sqrt(covs[2,2] + covs[3,3] + 2*covs[2,3])
as.vector(c(exp((coefs_4[2]+coefs_4[3])*5)-1.96*se_full, exp((coefs_4[2]+coefs_4[3])*5)+1.96*se_full))
```

```
## [1] 0.6456666 0.7205220
```

```
exp(5*coefs_4[2])/exp(0*coefs_4[2])
```

```
##       bmi
## 0.5566471
```

The hazard ratio for a 5kg/m2 increase in BMI for sex == 0 is 0.556, with a 95% confidence interval of (0.511, 0.602), calculated below.

```
as.vector(c(exp((coefs_4[2])*5)-1.96*sqrt(covs[2,2]), exp((coefs_4[2])*5)+1.96*sqrt(covs[2,2])))
```

```
## [1] 0.5106573 0.6026370
```

Both of these ratios indicate, given a fixed sex category, a lower hazard rate on average with each 5 unit increase in bmi.

```
exp(cox.fit.4$coef[1]+cox.fit.4$coef[3]*bmi.num-1.96*sqrt(cox.fit.4$var[1,1]+cox.fit.4$var[3,3]*bmi.num^2+2*bm
i.num*cox.fit.4$var[1,3])) %>% head
```

```
## [1] 0.9696190 0.9088964 0.8080610 0.9980134 0.9266607 0.8694313
```

h. Using the interaction model fitted in 1(d), compute and then graph the estimated survival functions for males and females with a bmi of 25 kg/m2.

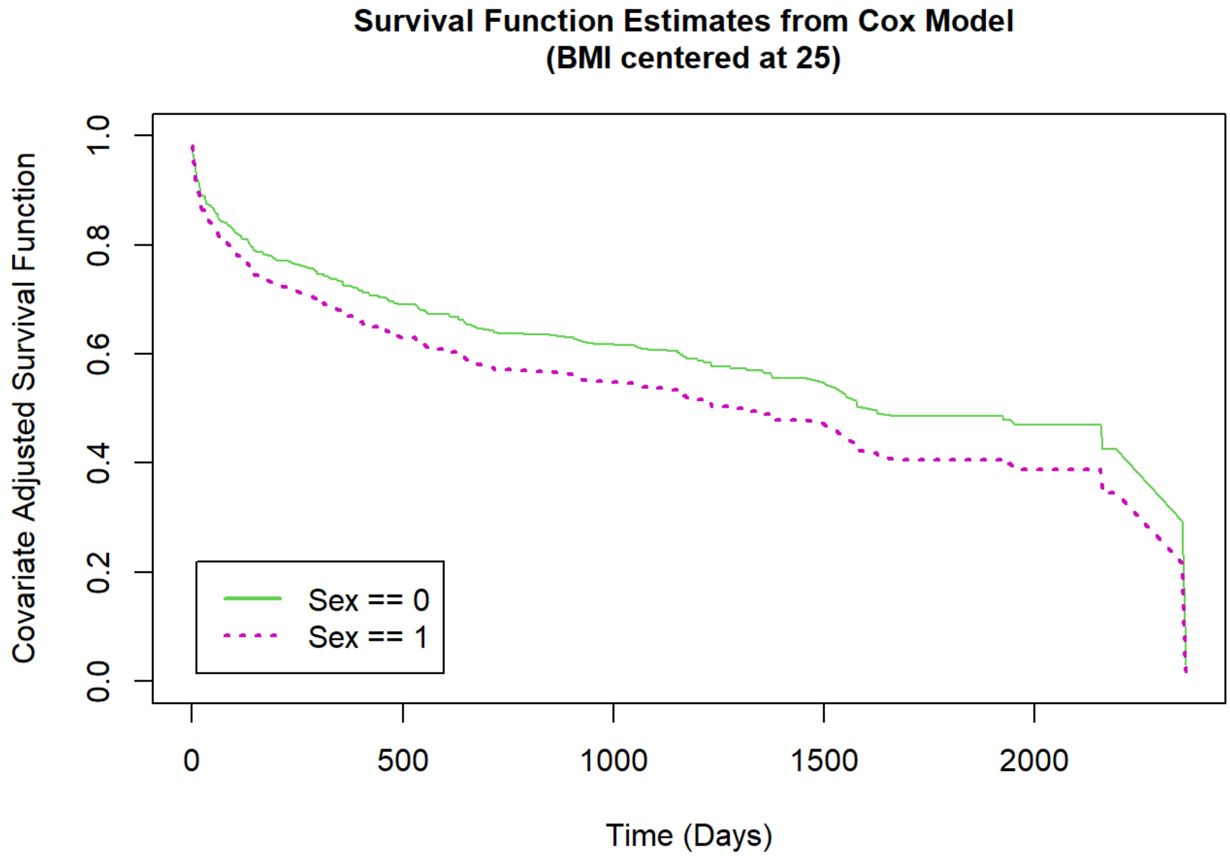
```
whas_for_st <- whas %>% mutate(bmi_25 = bmi - 25)
cox.fit.5<-coxph(Surv(lenfol,fstat)~sex*bmi_25, data=whas_for_st)
coefs_5 <- coef(cox.fit.5)
```

```
test.c<-basehaz(cox.fit.5,centered=F)
```

```
## Warning in survfit.coxph(fit, se.fit = FALSE): the model contains interactions;
## the default curve based on columm means of the X matrix is almost certainly not
## useful. Consider adding a newdata argument.
```

```
base.surv.fit<-exp(-test.c$hazard)
sex.surv.0.fit<-base.surv.fit
sex.surv.1.fit<-base.surv.fit^exp(coefs_5[1])

plot(test.c$time,sex.surv.0.fit,type="l",lty=1,col=3,ylim=c(0,1),ylab=c("Covariate Adjusted Survival Functio
n"),xlab=c("Time (Days)"))
lines(test.c$time,sex.surv.1.fit,type="l",lty=3,col=6,lwd=2)
title(main=list("Survival Function Estimates from Cox Model \n (BMI centered at 25)",cex=1))
legend(x=10,y=0.22,c("Sex == 0","Sex == 1"),lty=c(1,3),col=c(3,6),lwd=2)
```



The survival function for Sex == 1 has a similar shape to the function for when Sex == 0. However, the function for Sex == 1 seems to decrease more quickly at the beginning, a “lead” which it holds until the latest follow-up timepoint.