BIOS522: Survival Analysis Methods

Week 10: R Practice Session

In this session, you will practice implementing analyses in `R`. You will use `R markdown` to document your code and results. `R` chunks are provided for your code, and space is provided for your interpretations.

For this `R` practice session, we will use the data set `whas500`. This data set includes data extracted from the Worcester Heart Attack Study (WHAS). The study is described in Hosmer, Lemeshow and May (2013). The purpose of the study is to study factors and time trends associated with long-term survival following acute myocardial infarction. The main data set has information on more than 11,000 admissions. The data in `whas500` were sampled by taking an approximately 23% random sample from cohorts collected in 1997, 1999, and 2001, yielding 500 subjects.

The file `whas500.Rdata` includes the data. The file `whas500.txt` summarizes the variables.

The aim of this session is to practice fitting accelerated failure time models in `R`.

1. Fit a Weibull accelerated failure time model with the single predictor of atrial fibrillation. Report the acceleration factor. Write an accompanying sentence to interpret the acceleration factor, being clear about the direction of the effect (which group has longer survival if a difference in observed).

```
# ADD YOUR CODE HERE

# start by reading in the survival package and the data
library(survival)
load("whas500.RData")
colnames(whas500)
```

```
##  [1] "id"        "age"       "gender"    "hr"        "sysbp"     "diasbp"
##  [7] "bmi"       "cvd"       "afb"       "sho"       "chf"       "av3"
## [13] "miord"     "mitype"    "year"      "admitdate" "disdate"   "fdate"
## [19] "los"       "dstat"     "lenfol"    "fstat"
```

```
weibfit <- survreg(Surv(lenfol,fstat) ~ afb, data=whas500, dist="weibull")
summary(weibfit)
```

```
##
## Call:
## survreg(formula = Surv(lenfol, fstat) ~ afb, data = whas500,
##     dist = "weibull")
##               Value Std. Error     z      p
## (Intercept)  8.5212     0.1943 43.85 <2e-16
## afb         -1.1614     0.3583 -3.24 0.0012
## Log(scale)   0.7596     0.0622 12.22 <2e-16
##
## Scale= 2.14
##
## Weibull distribution
## Loglik(model)= -1747.6   Loglik(intercept only)= -1752.5
##  Chisq= 9.74 on 1 degrees of freedom, p= 0.0018
## Number of Newton-Raphson Iterations: 5
## n= 500
```

```
exp(weibfit$coefficients[1])
```

```
## (Intercept)
##    5019.983
```

The acceleration factor is exp(-1.1614) = 0.313. Patients with myocardial infarction who had atrial fibrillation survived on average 0.313 times as long as patients without atrial fibrillation. Alternatively, patients without atrial fibrillation survived on average 1/0.313 = 3.19 times longer on average than patients with atrial fibrillation. (This is the acceleration factor we would calculate if we reversed the coding of the afb covariate.)

2. Fit a Cox proportional hazards model with the single predictor of atrial fibrillation. Compare the hazard ratio from this model with the acceleration factor from the previous model. Do they provide the same conclusions?

```
# ADD YOUR CODE HERE
summary(coxph(Surv(lenfol,fstat) ~ afb, data=whas500))
```

```
## Call:
## coxph(formula = Surv(lenfol, fstat) ~ afb, data = whas500)
##
##   n= 500, number of events= 215
##
##       coef exp(coef) se(coef)     z Pr(>|z|)
## afb 0.5397    1.7156   0.1654 3.263   0.0011 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##     exp(coef) exp(-coef) lower .95 upper .95
## afb     1.716     0.5829      1.24     2.373
##
## Concordance= 0.537  (se = 0.014 )
## Likelihood ratio test= 9.58  on 1 df,   p=0.002
## Wald test            = 10.64  on 1 df,   p=0.001
## Score (logrank) test = 10.9  on 1 df,   p=0.001
```

The hazard ratio for atrial fibrillation is 1.716. The basic conclusion is the same in that the group with atrial fibrillation has poorer survival (here, higher hazard), and the p-value is the same. But note that even after inverting the acceleration factor (1/0.313 = 3.19), the magnitude of the values are not equal. As noted previously, the hazard ratio and the acceleration factor are not simply inverses of each other.

3. Examining the fit of the Weibull accelerated failure time model, report the Weibull shape parameter. Describe the shape of the hazard function (decreasing, constant, concave increasing, or convex increasing).

```
# ADD YOUR CODE HERE
weibfit <- survreg(Surv(lenfol,fstat) ~ afb, data=whas500, dist="weibull")
summary(weibfit)
```

```
##
## Call:
## survreg(formula = Surv(lenfol, fstat) ~ afb, data = whas500,
##     dist = "weibull")
##               Value Std. Error     z      p
## (Intercept)  8.5212     0.1943 43.85 <2e-16
## afb         -1.1614     0.3583 -3.24 0.0012
## Log(scale)   0.7596     0.0622 12.22 <2e-16
##
## Scale= 2.14
##
## Weibull distribution
## Loglik(model)= -1747.6   Loglik(intercept only)= -1752.5
##  Chisq= 9.74 on 1 degrees of freedom, p= 0.0018
## Number of Newton-Raphson Iterations: 5
## n= 500
```

```
1/weibfit$scale
```

```
## [1] 0.4678595
```

The Weibull shape paramter is 1/"scale" as reported by R. Because this is below 1, the hazard function is decreasing over time. This makes sense if one remembers that there is an elevated period of risk immediately following the myocardial infarction.

4. Report the Weibull rate parameter for the reference group. Include units. Recall that failure times are measured in days in this data set.

```
# ADD YOUR CODE HERE
summary(weibfit)
```

```
##
## Call:
## survreg(formula = Surv(lenfol, fstat) ~ afb, data = whas500,
##     dist = "weibull")
##               Value Std. Error     z      p
## (Intercept)  8.5212     0.1943 43.85 <2e-16
## afb         -1.1614     0.3583 -3.24 0.0012
## Log(scale)   0.7596     0.0622 12.22 <2e-16
##
## Scale= 2.14
##
## Weibull distribution
## Loglik(model)= -1747.6   Loglik(intercept only)= -1752.5
##  Chisq= 9.74 on 1 degrees of freedom, p= 0.0018
## Number of Newton-Raphson Iterations: 5
## n= 500
```

```
# can extract the coefficients from the model
exp(-coefficients(weibfit)[1])
```

```
##  (Intercept)
## 0.0001992039
```

```
# or just use R as a calculator
exp(-8.521182)
```

```
## [1] 0.0001992038
```

The rate parameter in the reference group is 0.0001992 days^-1.

5. Estimate median survival for patients with and without atrial fibrillation.

```
# ADD YOUR CODE HERE
predict(weibfit,type="quantile",p=0.5,newdata=data.frame(afb=c(1,0)))
```

```
##         1         2
##  717.9695 2293.4200
```

Median survival is 718 days for those with atrial fibrillation, and 2293 for those without.

End of practice session