Proportional Hazards Regression with R: Part One*

```
> rm(list=ls())
> # install.packages("survival",dependencies=TRUE) # Only need to do this once
> library(survival) # Do this every time
> mux = 50; sdx = 12; n = 500
> beta0 = 2.3; beta1 = 0.05
>
> # Simulate one data set
> set.seed(9999); options(scipen=999)
> x = rnorm(n,mean=mux,sd=sdx)
> epsilon = rexp(n)
> tstar = exp(beta0 + beta1*x) * epsilon
> U = abs(rnorm(n, mean=0, sd=100)) # Censoring times
> t = pmin(tstar,U) # Vector of minima, n values
> delta = as.numeric(tstar<U); table(delta)
delta
  0   1
281 219

> # Try exponential regression and proportional hazards regression
> stime = Surv(t,delta)
>
> expo = survreg(stime ~ x , dist='exponential')
> summary(expo)

Call:
survreg(formula = stime ~ x, dist = "exponential")
              Value Std. Error    z                    p
(Intercept) 2.28167    0.27697 8.24 <0.0000000000000002
x           0.04968    0.00586 8.47 <0.0000000000000002

Scale fixed at 1

Exponential distribution
Loglik(model)= -1217   Loglik(intercept only)= -1253.3
      Chisq= 72.56 on 1 degrees of freedom, p= 0.000000000000000016
Number of Newton-Raphson Iterations: 5
n= 500

>
> phaz = coxph(stime ~ x)
> summary(phaz)
Call:
coxph(formula = stime ~ x)

  n= 500, number of events= 219

       coef exp(coef)  se(coef)      z            Pr(>|z|)
x -0.049942  0.951284  0.006072 -8.225 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

  exp(coef) exp(-coef) lower .95 upper .95
x    0.9513      1.051      0.94    0.9627

Concordance= 0.666  (se = 0.02 )
Likelihood ratio test= 69.57  on 1 df,   p=<0.0000000000000002
Wald test            = 67.65  on 1 df,   p=<0.0000000000000002
Score (logrank) test = 68.19  on 1 df,   p=<0.0000000000000002
```

* Copyright information is on the last page.

```
> # Pharmaco-smoking
> rm(list=ls());  options(scipen=999)
> # install.packages("survival",dependencies=TRUE) # Only need to do this once
> library(survival) # Do this every time
> # install.packages("asaur",dependencies=TRUE) # Only need to do this once
> library(asaur)
> # Make fixed-up data frame called quit
> quit = within(pharmacoSmoking,{
+ DayOfRelapse = Surv(ttr+1,relapse)
+ contrasts(grp) = contr.treatment(2,base=2) # Patch only is reference category
+ colnames(contrasts(grp)) = c('Combo') # Names of dummy vars -- just one
+ # Collapse race categories
+ Race = as.character(race) # Small r race is a factor. This is easier to modify.
+ Race[Race!='white'] = 'blackOther'; Race=factor(Race)
+ }) # Finished making data frame quit
>
> w_All = survreg(DayOfRelapse ~ grp + age + gender + Race + employment +
yearsSmoking + levelSmoking + priorAttempts, dist='weibull', data=quit);
summary(w_All)

Call:
survreg(formula = DayOfRelapse ~ grp + age + gender + Race +
    employment + yearsSmoking + levelSmoking + priorAttempts,
    data = quit, dist = "weibull")
                     Value Std. Error     z            p
(Intercept)        1.12177    0.97726  1.15       0.2510
grpCombo           1.09225    0.38186  2.86       0.0042
age                0.08432    0.03411  2.47       0.0134
genderMale         0.03631    0.41417  0.09       0.9301
Racewhite          0.25145    0.39143  0.64       0.5206
employmentother   -1.28799    0.46719 -2.76       0.0058
employmentpt      -1.28482    0.58631 -2.19       0.0284
yearsSmoking      -0.02351    0.03250 -0.72       0.4696
levelSmokinglight -0.07347    0.43151 -0.17       0.8648
priorAttempts     -0.00105    0.00200 -0.52       0.6000
Log(scale)         0.54194    0.08917  6.08 0.0000000012

Scale= 1.72

Weibull distribution
Loglik(model)= -463.8   Loglik(intercept only)= -476.5
      Chisq= 25.41 on 9 degrees of freedom, p= 0.0025
Number of Newton-Raphson Iterations: 5
n= 125

> ph_All = coxph(DayOfRelapse ~ grp + age + gender + Race + employment +
yearsSmoking + levelSmoking + priorAttempts, data=quit); summary(ph_All)
Call:
coxph(formula = DayOfRelapse ~ grp + age + gender + Race + employment +
    yearsSmoking + levelSmoking + priorAttempts, data = quit)

  n= 125, number of events= 89

                       coef  exp(coef)   se(coef)      z Pr(>|z|)
grpCombo          -0.5994057  0.5491379  0.2203690 -2.720  0.00653 **
age               -0.0479631  0.9531689  0.0198258 -2.419  0.01555 *
genderMale         0.0069130  1.0069369  0.2409092  0.029  0.97711
Racewhite         -0.1394286  0.8698551  0.2279991 -0.612  0.54085
employmentother    0.7086315  2.0312096  0.2727885  2.598  0.00938 **
employmentpt       0.7005798  2.0149207  0.3418680  2.049  0.04044 *
yearsSmoking       0.0144207  1.0145252  0.0188155  0.766  0.44342
levelSmokinglight  0.0329273  1.0334754  0.2495691  0.132  0.89503
priorAttempts      0.0004572  1.0004573  0.0011500  0.398  0.69095
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
                  exp(coef) exp(-coef) lower .95 upper .95
grpCombo             0.5491     1.8210    0.3565    0.8458
age                  0.9532     1.0491    0.9168    0.9909
genderMale           1.0069     0.9931    0.6280    1.6146
Racewhite            0.8699     1.1496    0.5564    1.3599
employmentother      2.0312     0.4923    1.1900    3.4670
employmentpt         2.0149     0.4963    1.0310    3.9378
yearsSmoking         1.0145     0.9857    0.9778    1.0526
levelSmokinglight    1.0335     0.9676    0.6337    1.6855
priorAttempts        1.0005     0.9995    0.9982    1.0027

Concordance= 0.647  (se = 0.03 )
Likelihood ratio test= 23.04  on 9 df,   p=0.006
Wald test            = 22.73  on 9 df,   p=0.007
Score (logrank) test = 23.23  on 9 df,   p=0.006


> # Checking the re-parameterization, a Weibull -beta_j/sigma should be close to a
proportional hazards beta_j. Look at Racewhite.
>
> -0.25145/1.72  # Compare  -0.1394286 from PH
[1] -0.1461919
> # Not bad. This is n=125
>
```

```
> wfull =  survreg(DayOfRelapse ~ grp + age + employment , dist='weibull',
data=quit)
> summary(wfull)
Call:
survreg(formula = DayOfRelapse ~ grp + age + employment, data = quit,
    dist = "weibull")
                 Value Std. Error     z           p
(Intercept)      1.4957     0.8414  1.78     0.07545
grpCombo         1.1023     0.3793  2.91     0.00366
age              0.0643     0.0186  3.45     0.00055
employmentother -1.2880     0.4617 -2.79     0.00528
employmentpt    -1.2123     0.5616 -2.16     0.03088
Log(scale)       0.5454     0.0894  6.10 0.000000001

Scale= 1.73

Weibull distribution
Loglik(model)= -464.3   Loglik(intercept only)= -476.5
      Chisq= 24.31 on 4 degrees of freedom, p= 0.000069
Number of Newton-Raphson Iterations: 5
n= 125

> phfull =  coxph(DayOfRelapse ~ grp + age + employment, data=quit)
> summary(phfull)
Call:
coxph(formula = DayOfRelapse ~ grp + age + employment, data = quit)

  n= 125, number of events= 89

                    coef exp(coef) se(coef)      z Pr(>|z|)
grpCombo        -0.60788   0.54450  0.21837 -2.784  0.00537 **
age             -0.03529   0.96533  0.01075 -3.282  0.00103 **
employmentother  0.70348   2.02077  0.26929  2.612  0.00899 **
employmentpt     0.65369   1.92262  0.32732  1.997  0.04581 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                exp(coef) exp(-coef) lower .95 upper .95
grpCombo           0.5445     1.8365    0.3549    0.8354
age                0.9653     1.0359    0.9452    0.9859
employmentother    2.0208     0.4949    1.1920    3.4256
employmentpt       1.9226     0.5201    1.0122    3.6518

Concordance= 0.638  (se = 0.03 )
Likelihood ratio test= 22.03  on 4 df,   p=0.0002
Wald test            = 21.91  on 4 df,   p=0.0002
Score (logrank) test = 22.48  on 4 df,   p=0.0002

> # How are they getting the confidence intervals for those hazard ratios?
> L = -0.60788 -1.96*0.21837; L
[1] -1.035885

> exp(L)
[1] 0.3549121
```

```
> # Try Partial Likelihood and Wald tests for employment, controlling for age and
> # experimental treatment.
>
> # Partial Likelihood Ratio test
> nojob = coxph(DayOfRelapse ~ grp + age, data=quit)
> anova(nojob,phfull) # LR test

Analysis of Deviance Table
 Cox model: response is  DayOfRelapse
 Model 1: ~ grp + age
 Model 2: ~ grp + age + employment
   loglik  Chisq Df Pr(>|Chi|)
1 -379.24
2 -375.14 8.2037  2    0.01654 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


> # Wald test: function(L,Tn,Vn,h=0) # H0: L theta = h
> source("http://www.utstat.toronto.edu/brunner/Rfunctions/Wtest.txt")
> betahat = phfull$coefficients; Vn_hat = vcov(phfull); betahat
       grpCombo             age employmentother    employmentpt
    -0.60788405     -0.03528934      0.70347664      0.65369019


> LL = rbind(c(0,0,1,0),
+            c(0,0,0,1) )
> Wtest(LL,betahat,Vn_hat)
         W         df    p-value
8.38888814 2.00000000 0.01507912
```
