Cox ED Demonstration, for R

Jonathan Kropko and Jeffrey J. Harden

May 22, 2017

In this example, we will demonstrate the use of the `coxed()` function after fitting a Cox proportional hazards model for a lung study from the Veteran's Administration, an example originally used by Kalbfleisch and Prentice (1980). The data are publicly available as part of the `survival` package.

```
library(survival)
```

Data

The data are contained in a data frame named `veteran`.

```
summary(veteran)
```

```
##       trt             celltype       time           status
##  Min.   :1.000   squamous :35   Min.   :  1.0   Min.   :0.0000
##  1st Qu.:1.000   smallcell:48   1st Qu.: 25.0   1st Qu.:1.0000
##  Median :1.000   adeno    :27   Median : 80.0   Median :1.0000
##  Mean   :1.496   large    :27   Mean   :121.6   Mean   :0.9343
##  3rd Qu.:2.000                  3rd Qu.:144.0   3rd Qu.:1.0000
##  Max.   :2.000                  Max.   :999.0   Max.   :1.0000
##      karno          diagtime           age             prior
##  Min.   :10.00   Min.   : 1.000   Min.   :34.00   Min.   : 0.00
##  1st Qu.:40.00   1st Qu.: 3.000   1st Qu.:51.00   1st Qu.: 0.00
##  Median :60.00   Median : 5.000   Median :62.00   Median : 0.00
##  Mean   :58.57   Mean   : 8.774   Mean   :58.31   Mean   : 2.92
##  3rd Qu.:75.00   3rd Qu.:11.000   3rd Qu.:66.00   3rd Qu.:10.00
##  Max.   :99.00   Max.   :87.000   Max.   :81.00   Max.   :10.00
```

In the data, the outcome variable is `time` and the censoring variable is `status`. There is a randomly-assigned, binary treatment named `trt`. Although randomization implies that no control variables are necessary, we use the following control variables for the purposes of this example (these descriptions are quoted from the documentation for the data, accessible by typing `?veterans`):

- `celltype`: 1=squamous, 2=smallcell, 3=adeno, 4=large
- `karno`: Karnofsky performance score (100=good)
- `diagtime`: months from diagnosis to randomisation
- `age`: in years
- `prior`: prior therapy 0=no, 1=yes

Cox Proportional Hazards Model

We use the Cox proportional hazards model to assess the efficacy of the treatment and controls on the prognosis of the subjects in the study. The summary of the output from the `coxph()` function displays both coefficients (the `coef` column) and hazard ratios (the `exp(coef)` column).

```
coxmodel <- coxph(Surv(time, status) ~ trt + celltype + karno + diagtime +
                        age + prior, data=veteran)
summary(coxmodel)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ trt + celltype + karno +
##     diagtime + age + prior, data = veteran)
##
##   n= 137, number of events= 128
##
##                          coef  exp(coef)   se(coef)      z Pr(>|z|)
## trt                 2.946e-01  1.343e+00  2.075e-01  1.419  0.15577
## celltypesmallcell   8.616e-01  2.367e+00  2.753e-01  3.130  0.00175 **
## celltypeadeno       1.196e+00  3.307e+00  3.009e-01  3.975 7.05e-05 ***
## celltypelarge       4.013e-01  1.494e+00  2.827e-01  1.420  0.15574
## karno              -3.282e-02  9.677e-01  5.508e-03 -5.958 2.55e-09 ***
## diagtime            8.132e-05  1.000e+00  9.136e-03  0.009  0.99290
## age                -8.706e-03  9.913e-01  9.300e-03 -0.936  0.34920
## prior               7.159e-03  1.007e+00  2.323e-02  0.308  0.75794
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                    exp(coef) exp(-coef) lower .95 upper .95
## trt                   1.3426     0.7448    0.8939    2.0166
## celltypesmallcell     2.3669     0.4225    1.3799    4.0597
## celltypeadeno         3.3071     0.3024    1.8336    5.9647
## celltypelarge         1.4938     0.6695    0.8583    2.5996
## karno                 0.9677     1.0334    0.9573    0.9782
## diagtime              1.0001     0.9999    0.9823    1.0182
## age                   0.9913     1.0087    0.9734    1.0096
## prior                 1.0072     0.9929    0.9624    1.0541
##
## Concordance= 0.736  (se = 0.03 )
## Rsquare= 0.364   (max possible= 0.999 )
## Likelihood ratio test= 62.1  on 8 df,   p=1.799e-10
## Wald test            = 62.37  on 8 df,   p=1.596e-10
## Score (logrank) test = 66.74  on 8 df,   p=2.186e-11
```

Using hazard ratios, the effect of the treatment is interpreted as follows:

- Members of the treatment group, relative to members of the control group, have a 34.3% greater risk of death at time t conditional on survival until time t, and conditional on the controls. This change in risk is not significantly different from 0 at the 95% level.

The interpretation of the effect of the Karnofsky performance score is:

- A one-point increase in the Karnofsky performance score is associated with a 3.23% decrease in the risk of death at time t conditional on survival until time t, and conditional on the controls. This change in risk is significantly different from 0 at the 95% level.

While these results clearly convey the direction and significance of each effect, they do not clearly show us the direct impact of each variable on the survival time for subjects. For that, we turn to the `cox.ed()` function.

Cox With Expected Durations

Here we reinterpret the results for the treatment, the small cell group, and the Karnofsky performance score in terms of expected duration and marginal changes in duration.

The treatment

We employ the `cox.ed()` function to estimate the expected duration for the control group (`trt=1`) and the treatment group (`trt=2`), and the difference between the expected durations for the two groups. We use the default GAM method for generating these quantities.

```
expdur.trt <- cox.ed(cox.model = coxmodel, var.name="trt", val1=1, val2=2,
                     k=8, m=1000, ci=.95)
```

The point estimates for the expected durations for the two groups are

```
expdur.trt$ed.pe
```

```
##         1         2
## 137.48834  85.26051
```

The 95% confidence interval lower and upper bounds are

```
expdur.trt$ed.lo
```

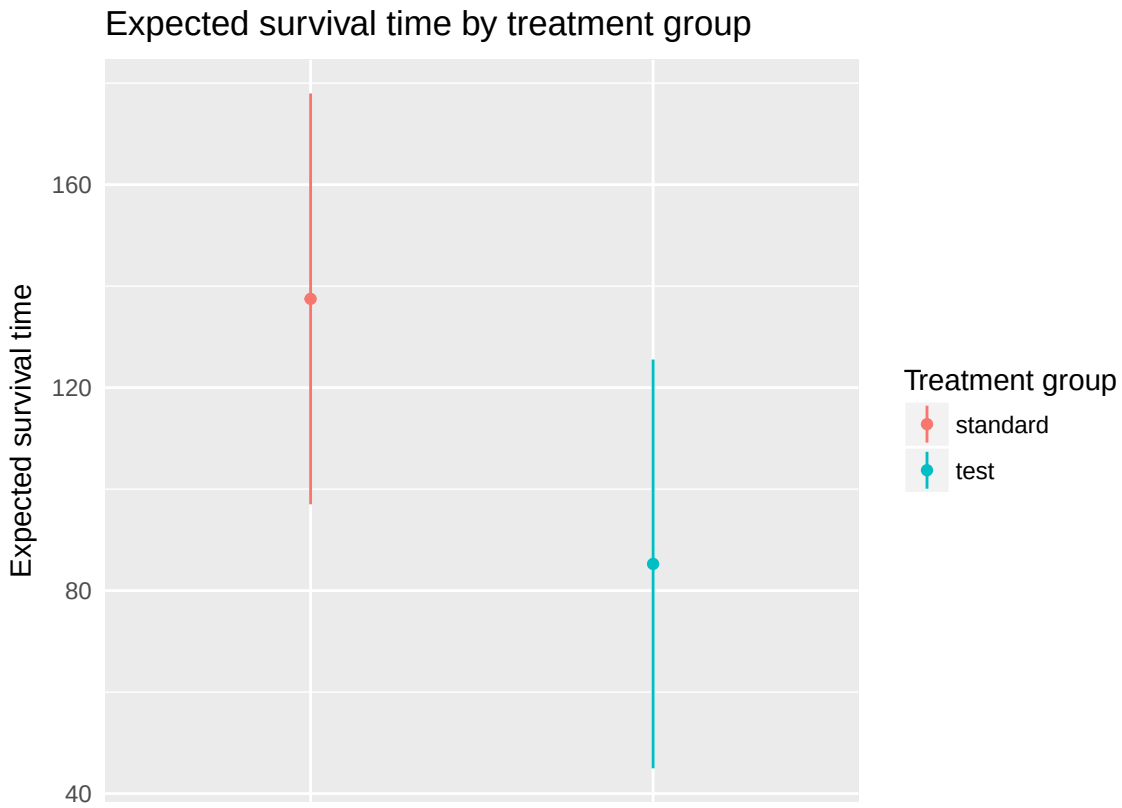
```
##        1        2
## 97.01167 44.98134
```

```
expdur.trt$ed.hi
```

```
##        1        2
## 177.9650 125.5397
```

We can plot the expected durations and their confidence intervals by placing these results in a separate data frame and using `ggplot`:

```
expdur.trt.df <- data.frame(trt = c("standard", "test"),
                            expdur = expdur.trt$ed.pe,
                            lb = expdur.trt$ed.lo,
                            ub = expdur.trt$ed.hi)
library(ggplot2)
g <- ggplot(expdur.trt.df, aes(x = trt, y = expdur, color=trt)) +
      geom_point() +
      geom_linerange(aes(ymin=lb, ymax=ub)) +
      xlab("") +
      ylab("Expected survival time") +
      ggtitle("Expected survival time by treatment group") +
      labs(colour = "Treatment group") +
      theme(axis.text.x = element_blank(),
            axis.ticks = element_blank())
g
```



Notice that the treated group has shorter expected survival times than the control group ("standard"), which accords with the hazard ratio for this variable being grater than 1. Also, the two expected durations have considerable overlap, which corresponds to the non-significant hazard ratio in the `coxph()` function output. The `cox.ed()` function also calculated the average difference between these two expected durations,

```
expdur.trt$diff.pe
```

```
##        1
## 52.22783
```

as well as the confidence interval of this difference

```
expdur.trt$diff.lo
```

```
##        1
## 13.62377
```

```
expdur.trt$diff.hi
```

```
##        1
## 90.83189
```

The Karnofsky performance score

We also use the `cox.ed()` function to estimate the expected duration for specific values of the Karnofsky performance score. In particular, we compare the expected survival time for the 25th percentile of the Karnofsky performance score, 40, to the 75th percentile of the Karnofsky performance score, 75:

```
expdur.kar <- cox.ed(cox.model = coxmodel, var.name="karno", val1=75, val2=40,
                     m=1000, ci=.95)
```

The point estimates for the expected durations for the two percentiles are

```
expdur.kar$ed.pe
```

```
##         1         2
## 184.23080  59.56301
```

According to the Cox proportional hazards model and the `cox.ed()` function, subjects with a Karnofsky performance score of 40 survive for an average of 59.6 time units, and subjects with a Karnofsky performance score of 75 survive for an average of 184.2. The 95% confidence interval lower and upper bounds for these point estimates are

```
expdur.kar$ed.lo
```

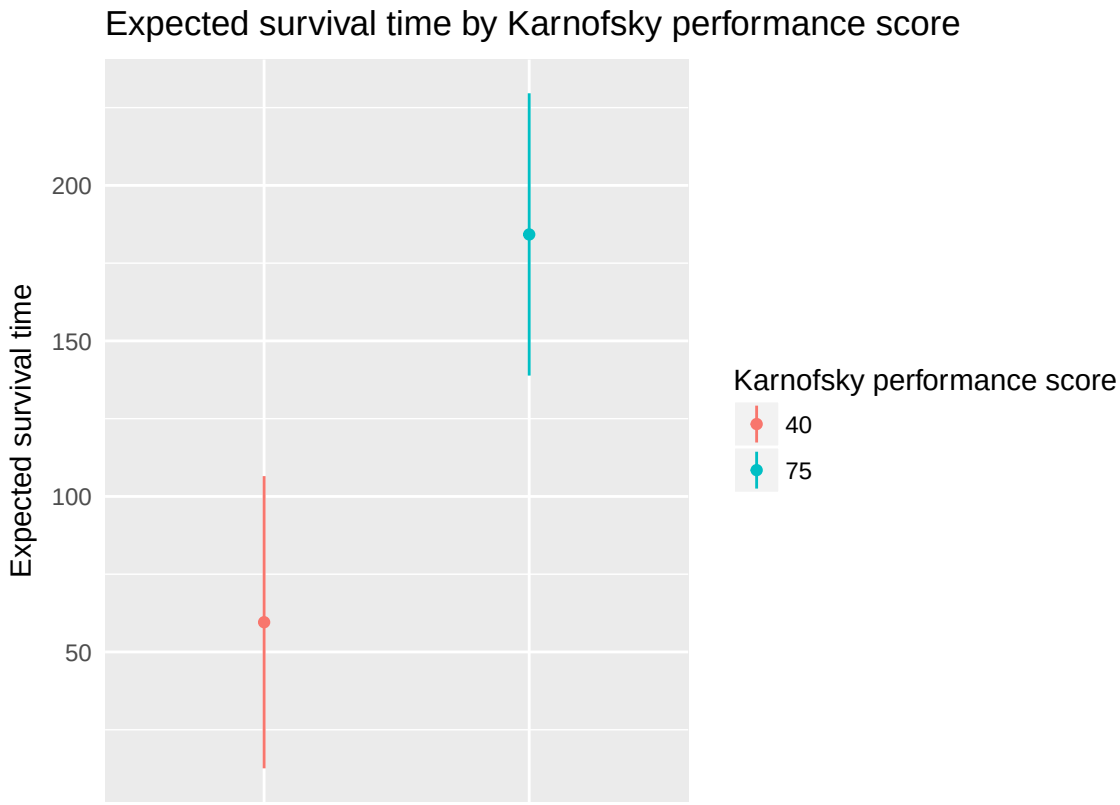
```
##         1         2
## 138.85518  12.59379
```

```
expdur.kar$ed.hi
```

```
##        1        2
## 229.6064 106.5322
```

To plot the expected durations and their confidence intervals, we place these results in a separate data frame and use `ggplot`:

```
expdur.kar.df <- data.frame(karno = c("75", "40"),
                            expdur = expdur.kar$ed.pe,
                            lb = expdur.kar$ed.lo,
                            ub = expdur.kar$ed.hi)
library(ggplot2)
g <- ggplot(expdur.kar.df, aes(x = karno, y = expdur, color=karno)) +
      geom_point() +
      geom_linerange(aes(ymin=lb, ymax=ub)) +
      xlab("") +
      ylab("Expected survival time") +
      ggtitle("Expected survival time by Karnofsky performance score") +
      labs(colour = "Karnofsky performance score") +
      theme(axis.text.x = element_blank(),
            axis.ticks = element_blank())
g
```



The hazard ratio result for the Karnofsky performance score suggested that every one-unit increase in the score is associated with a 3.4% reduction in the risk of death at time t, conditional on survival until time t. That result tells us that higher scores are associated with longer lifespans. But this result does not clearly suggest how much longer individuals live, on average, at different scores.

When the Karnofsky performance score is 75, individuals live roughly three times longer than when the Karnofsky performance score is 40. This result both confirms the hazard ratio finding and sheds more light on the dynamic that this result indicates. The `cox.ed()` function also calculated the average difference between these two expected durations,

```
expdur.kar$diff.pe
```

```
##        1
## 124.6678
```

as well as the confidence interval of this difference

```
expdur.kar$diff.lo
```

```
##        1
## 63.32789
```

```
expdur.kar$diff.hi
```

```
##        1
## 186.0077
```