Survival Models: Week 6

Example: Partial Likelihood

The times of death, or censoring (*), following a particular operation are given below. Information on each person's weight and gender is also provided.

	time	Gender	Weight	$\beta^T x$
	4	M	60	$\beta_1 + 60\beta_2$
censored ←	5*	F	80	$80\beta_2$
tie death	6	F	50	$50\beta_2$
	6	M	70	$\beta_1 + 70\beta_2$
	7*	M	100	$\beta_1 + 100\beta_2$

Note: Gender will be coded as 1 for Male and 0 for female.

Example: Partial Likelihood

Based on the information in the table, the first two terms in the PL will be:

$$\frac{\exp(\beta_1 + 60\beta_2)}{\exp(\beta_1 + 60\beta_2) + \exp(80\beta_2) + \exp(50\beta_2) + \exp(\beta_1 + 70\beta_2) + \exp(\beta_1 + 100\beta_2)}$$

and,

$$\frac{\exp((1+0)\beta_1 + (50+70)\beta_2)}{[\exp(50\beta_2) + \exp(\beta_1 + 70\beta_2) + \exp(\beta_1 + 100\beta_2)]^2}.$$

R Example - Recidivism

```
Call:
coxph(formula = Surv(week, arrest) ~ fin + age + race + wexp +
    mar + prio, data = Rossi)

  n= 432, number of events= 114

                   coef exp(coef) se(coef)      z Pr(>|z|)
finyes         -0.37352   0.68831  0.19082 -1.957 0.050295 .
age            -0.05640   0.94516  0.02184 -2.583 0.009796 **
raceother      -0.30983   0.73357  0.30780 -1.007 0.314133
wexpyes        -0.15331   0.85786  0.21218 -0.723 0.469957
marnot married  0.44339   1.55799  0.38136  1.163 0.244958
prio            0.09336   1.09785  0.02832  3.296 0.000981 ***
```

$$\lambda(t,\beta) = \lambda_0(t)\cdot e^{\overset{\downarrow}{\underset{1\times p}{\beta^T}} \overset{\downarrow}{\underset{p\times 1}{x}}}$$

e.g. $\lambda_1(t;\beta) = \lambda_0(t)\cdot e^{\beta^T x_1}$

$\lambda_2(t;\beta) = \lambda_0(t)\cdot e^{\beta^T x_2}$

Holding either variable constant let the first variable $\underbrace{x_{2,1}}_{\text{first variable, 2nd individual}} = \underbrace{x_{1,1}}_{\text{first variable, 1st individual}} + 1$

$$\frac{\lambda_1(t;\beta)}{\lambda_2(t;\beta)} = \frac{e^{\beta_1 x_{2,1}}}{e^{\beta_1 x_{1,1}}} = \boxed{e^{\beta_1}}$$

1 unit ↑ in $\underbrace{x_1}_{\text{first variable}}$, will increase $\lambda(t;\beta)$ to e^{β_1} times

For example

$e^{\beta_1} \approx 69\%$ for "fin"

compared with fin = 0, fin = 1

$\lambda(t,\beta)$ will be decreased to

$e^{\beta_1} = 69\%$

$$\frac{\lambda_2(t;\beta) \leftarrow \text{fin}=1}{\lambda_1(t;\beta) \nwarrow \text{fin}=0} = 69\%$$

alternatively, $\lambda(t,\beta)$ will be decreased by $1-e^{\beta_1} = 31\%$

b/c $\lambda_2 - \lambda_1 = -31\%\,\lambda_1$

R Example - Recidivism

The *relative risk* for a not-married individual relative to a married individual is $\exp(0.44) = 1.55$.

$\Rightarrow$ similar effect to maximize likelihood

The parameter estimates obtained from maximizing the PL have an approximately Normal sampling distribution with mean β and covariance matrix that can be found by inverting the information matrix. Tests of whether particular covariates are important can be conducted by comparing the statistic:

null: $\beta = 0$
alt: $\beta \neq 0$

$\Rightarrow z = \frac{\hat{\beta} - \beta}{se(\hat{\beta})}$

$$\frac{\hat{\beta}}{SE(\hat{\beta})}$$

$\sim$ z-statistic

to the $N(0, 1)$ distribution. For example, the test-statistic for *married* is 1.16 with a p-value of 0.24. This suggests that married is not an important explanatory variable.

Overall Significance

Testing the overall significance of the Cox model equates to testing the following:

$$\text{null:} \quad \beta_1 = \beta_2 = \ldots \beta_p = 0$$

$$\text{alt: at least one of the } \beta_i \neq 0 \;\; i = 1, \ldots, p.$$

One way to conduct this test is too look at the likelihoods correspond to the models specified under the null and alternative. Assuming the null hypothesis is correct the quantity $-2(LL(null) - LL(alt))$ has an approximate chi-square distribution with degrees of freedom equal to the difference in the number of parameters in the models specified under the null and alternative.

R Example - Recidivism

```
library(RcmdrPlugin.survival)
data(Rossi) #see J Fox notes for more details
cox.null<-coxph(Surv(week,arrest)~1,data=Rossi)
cox.alt<-coxph(Surv(week,arrest)~fin+age+race+wexp+mar+prio,data=Rossi)
cox.null$loglik
[1] -675.3806
cox.alt$loglik
[1] -675.3806 -658.8411
TS<--2*(cox.null$loglik[1]-cox.alt$loglik[2])
TS
[1] 33.07898
1-pchisq(TS,df=6)
[1] 1.012523e-05
```

R Example - Recidivism

```
> summary(cox.alt)
Call:
coxph(formula = Surv(week, arrest) ~ fin + age + race + wexp +
    mar + prio, data = Rossi)

  n= 432, number of events= 114

#I have deleted some output here##

               exp(coef) exp(-coef) lower .95 upper .95
finyes            0.6883     1.4528    0.4735    1.0005
age               0.9452     1.0580    0.9056    0.9865
raceother         0.7336     1.3632    0.4013    1.3410
wexpyes           0.8579     1.1657    0.5660    1.3003
marnot married    1.5580     0.6419    0.7378    3.2898
prio              1.0979     0.9109    1.0386    1.1605

Concordance= 0.642  (se = 0.027 )
Rsquare= 0.074   (max possible= 0.956 )
Likelihood ratio test= 33.08  on 6 df,   p=1.013e-05
Wald test            = 32.01  on 6 df,   p=1.625e-05
Score (logrank) test = 33.43  on 6 df,   p=8.68e-06
```

R Example - Recidivism

How about testing whether *race* and *married* are needed in a model that contains the other four covariates? This corresponds to the following test:

$$\text{null:} \quad \beta_{race} = \beta_{married} = 0$$

$$\text{alt: at least one of } \beta_{race} \text{ or } \beta_{married} \neq 0.$$

This test is easily conducted using a likelihood ratio test. The following slide shows how the test is conducted. Note: in this R code we need to make sure that race and married are included as the last two variables in the cox.alt model.

R Example - Recidivism

```
cox.null<-coxph(Surv(week,arrest)~fin+age+wexp+prio,data=Rossi)
cox.alt<-coxph(Surv(week,arrest)~fin+age+wexp+prio+race+mar,data=Rossi)
cox.null$loglik
[1] -675.3806 -660.2845
cox.alt$loglik
[1] -675.3806 -658.8411
TS<--2*(cox.null$loglik[2]-cox.alt$loglik[2])
TS
[1] 2.886664
1-pchisq(TS,df=2)
[1] 0.2361397
anova(cox.alt,cox.null)
Analysis of Deviance Table
 Cox model: response is  Surv(week, arrest)
 Model 1: ~ fin + age + wexp + prio + race + mar
 Model 2: ~ fin + age + wexp + prio
   loglik  Chisq Df P(>|Chi|)
1 -658.84
2 -660.28 2.8867  2    0.2361
```

Survival Curves

Estimated survival curves can also be produced based on a fitted Cox regression model. These estimates are derived using the following relationship:

$$S_0(t)^{\exp(\beta^T x)},$$

and plugging in the coefficient estimates obtained from the Cox regression model. The baseline survival function, $S_0(t)$, is also estimated based on the Cox regression (See O'Neill notes page 27 for details). Often estimated survival curves are plotted based on the covariates being set at their mean or median levels.

```
#R code for plot on next slide.
cox.full<-coxph(Surv(week,arrest)~fin+age+race+wexp+mar+prio,data=Rossi)
plot(survfit(cox.full))
```

R Example - Recidivism

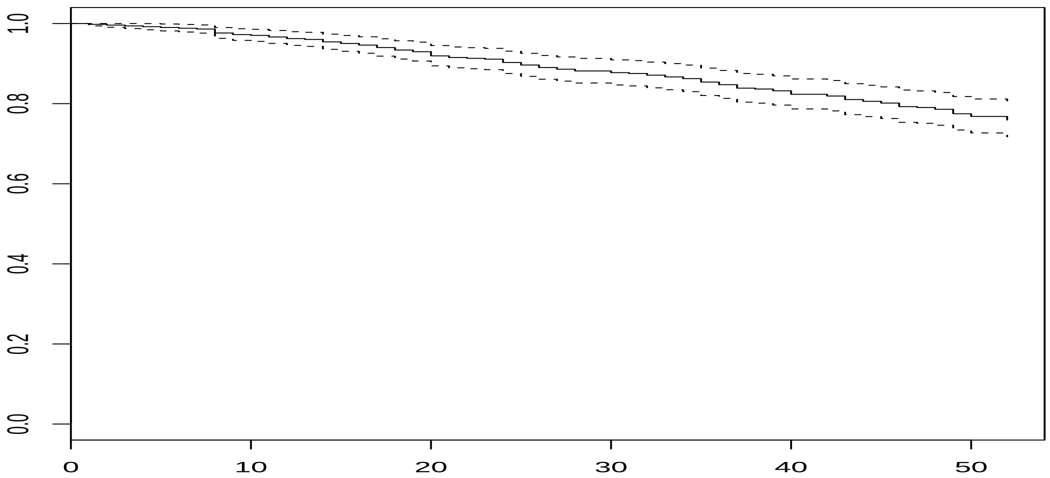

Figure 1: Estimated survival curves for Recidivism data. All covariates at their mean levels.

R Example - Recidivism

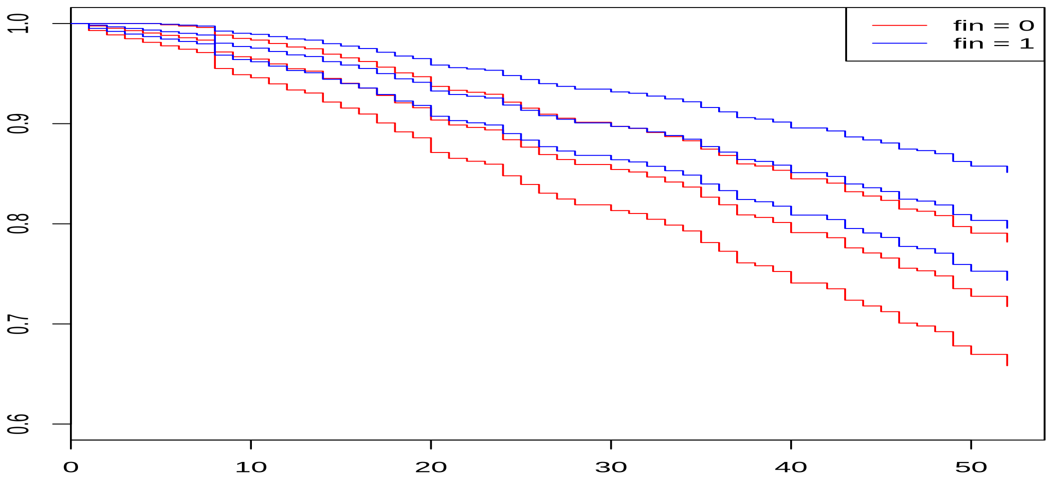

Figure 2: Estimated survival curves based on financial aid for Recidivism data. All other covariates at their mean levels.

R Example - Recidivism

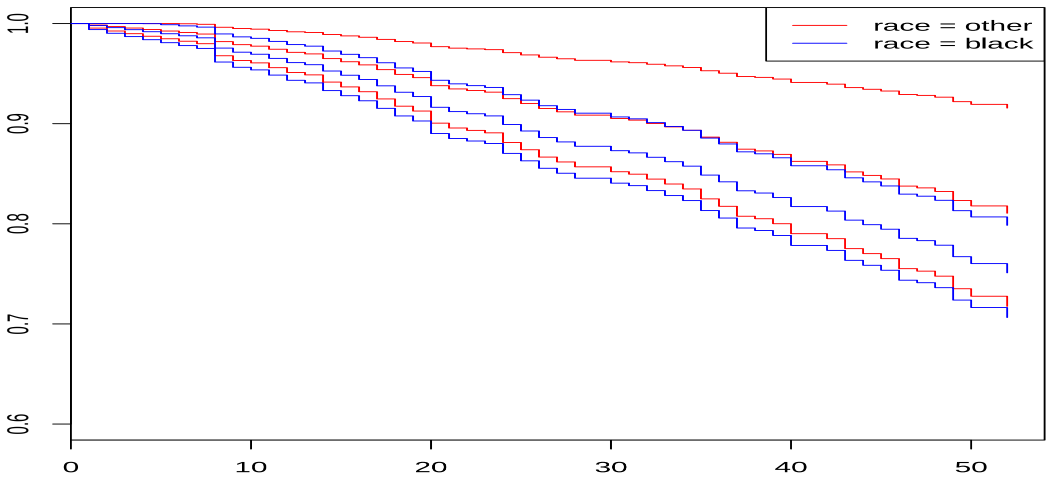

Figure 3: Estimated survival curves based on race for Recidivism data. All other covariates at their mean levels.

R Example - Addict

In week 5 we produced the following survival curves using KM.

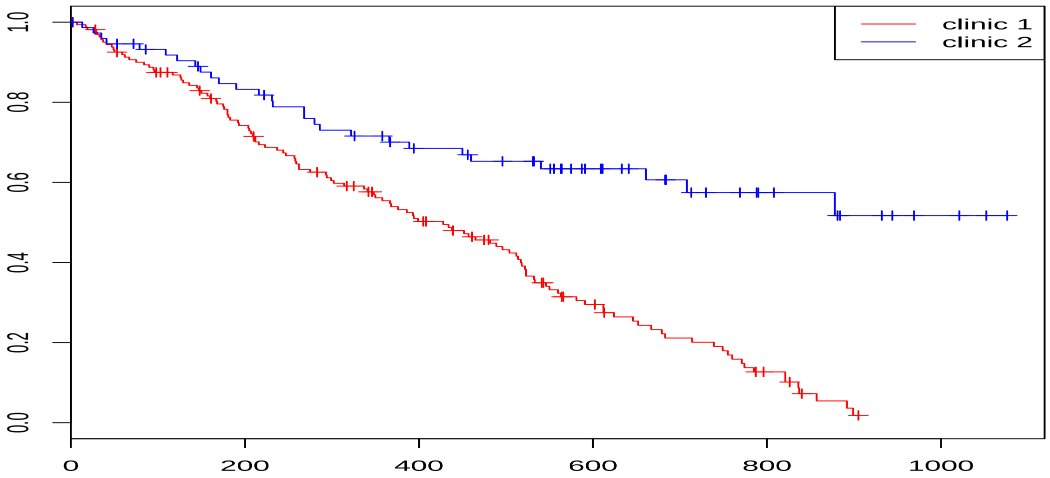

Figure 4: Estimated KM survival curves for Addict data

R Example - Addict

Fitting a Cox regression model to this data gives the following output:

```
attach(addict)
cox.mod<-coxph(Surv(time,status)~as.factor(clinic)+prison+dose,data=addict)
summary(cox.mod)
Call:
coxph(formula = Surv(time, status) ~ as.factor(clinic) + prison +
    dose, data = addict)

  n= 238, number of events= 150

                         coef exp(coef)  se(coef)      z Pr(>|z|)
as.factor(clinic)2 -1.009896  0.364257  0.214889 -4.700 2.61e-06 ***
prison              0.326555  1.386184  0.167225  1.953   0.0508 .
dose               -0.035369  0.965249  0.006379 -5.545 2.94e-08 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

                   exp(coef) exp(-coef) lower .95 upper .95
as.factor(clinic)2    0.3643     2.7453    0.2391    0.5550
prison                1.3862     0.7214    0.9988    1.9238
dose                  0.9652     1.0360    0.9533    0.9774
```

```
Concordance= 0.665  (se = 0.026 )
Rsquare= 0.238   (max possible= 0.997 )
Likelihood ratio test= 64.56  on 3 df,   p=6.228e-14
Wald test            = 54.12  on 3 df,   p=1.056e-11
Score (logrank) test = 56.32  on 3 df,   p=3.598e-12
```

R Example - Addict

Is there an interaction between prison and dose?

```
cox.mod<-coxph(Surv(time,status)~as.factor(clinic)+prison+dose+I(prison*dose),data=addict)
summary(cox.mod)
Call:
coxph(formula = Surv(time, status) ~ as.factor(clinic) + prison +
    dose + I(prison * dose), data = addict)

  n= 238, number of events= 150

                        coef exp(coef)  se(coef)      z Pr(>|z|)
as.factor(clinic)2 -0.995839  0.369413  0.215784 -4.615 3.93e-06 ***
prison             -0.206669  0.813289  0.766871 -0.269    0.788
dose               -0.038905  0.961842  0.008086 -4.811 1.50e-06 ***
I(prison * dose)    0.009084  1.009126  0.012727  0.714    0.475
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

R Example - Addict

Estimated effect of dose is -0.038905 if prison=0 and -0.038905+0.009084 if prison=1. However, the interaction term is not significant (p-value=0.475).

We could also include a squared term to model the effect of dose. Perhaps the impact of dose increases up to a certain point and then "flattens out".

```
cox.mod<-coxph(Surv(time,status)~as.factor(clinic)+prison+dose+I(dose^2),data=addict)
summary(cox.mod)
```

R Example - Addict

Compare KM estimate Cox estimate

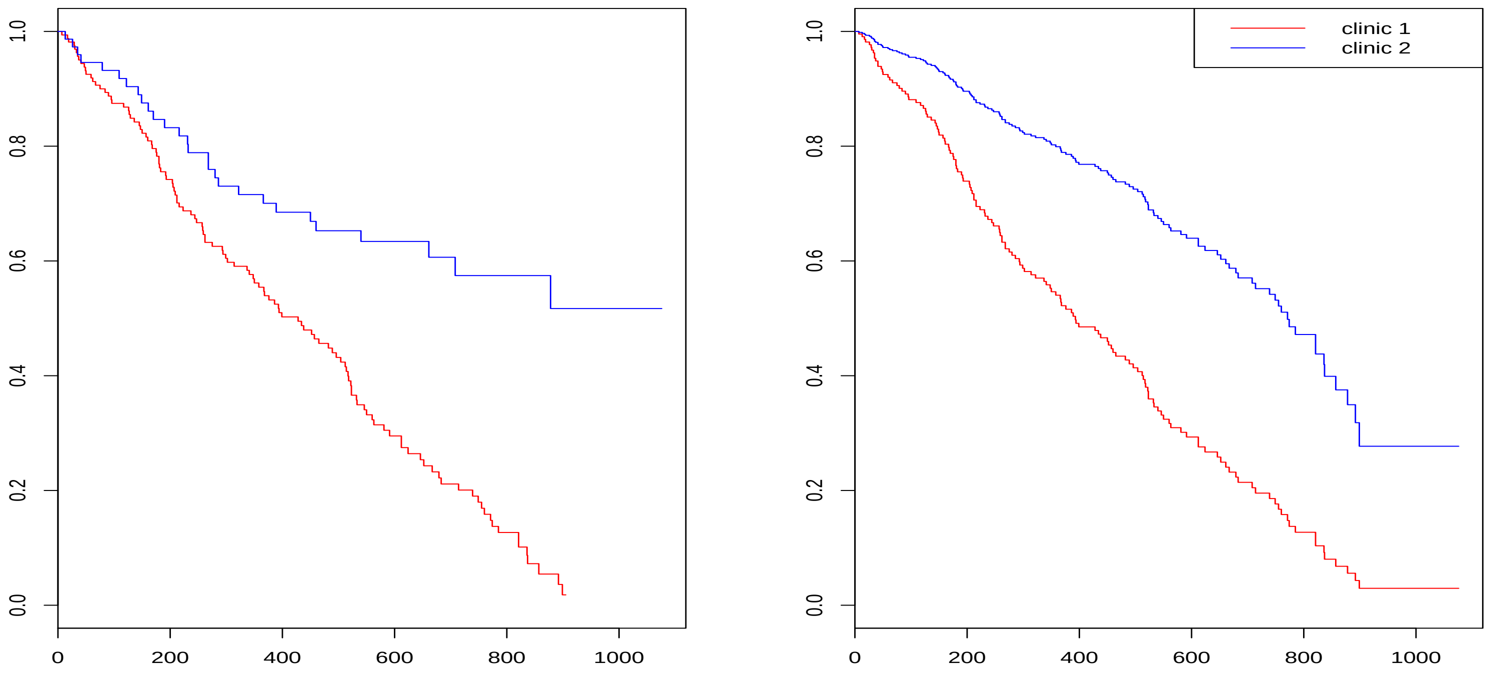

Figure 5: Estimated KM (left) and Cox (right) survival curves for Addict data

Diagnostic Checks

It is possible to conduct diagnostic checks to see whether the following two assumptions hold:

- Proportional Hazards assumption
- Linearity.
- Influential observations