Survival Models for the Social and Political Sciences

Week 4: Cox Regression

JEFF GILL
Professor of Political Science
Professor of Biostatistics
Professor of Surgery (Public Health Sciences)
Washington University, St. Louis

Proportional Hazards

- If $h_0(t)$ and $h_1(t)$ are hazard functions from two separate distributions, they are *proportional* if:

$$h_1(t) = \phi h_0(t), \quad \forall t \geq 0$$

  where $\phi > 0$.

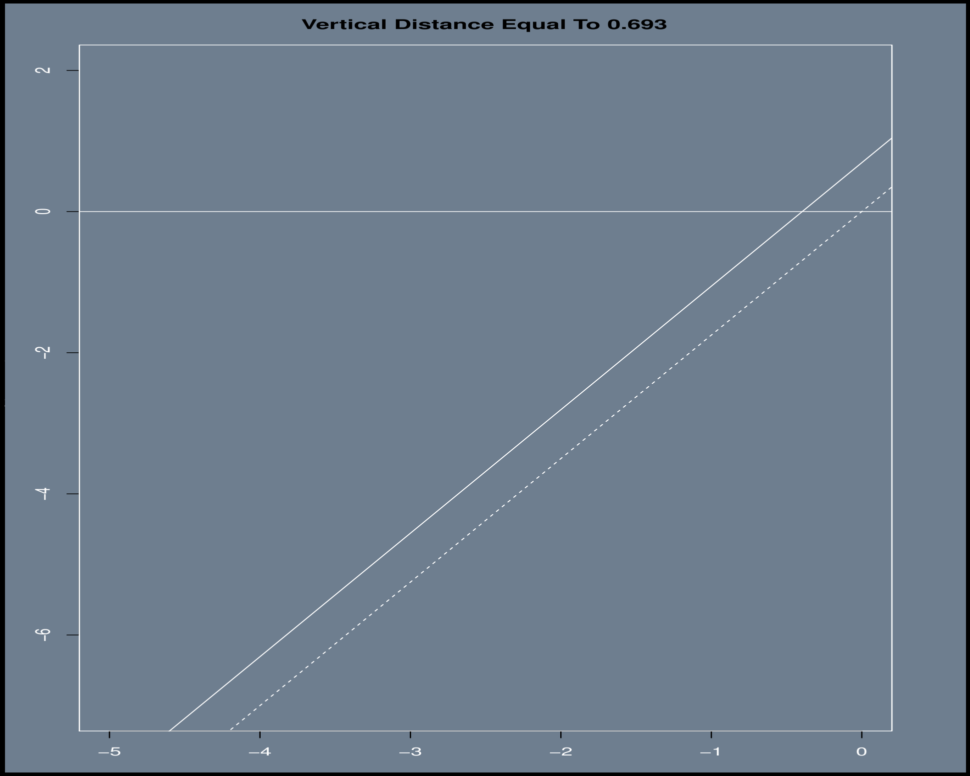
- This property caries over to the corresponding cumulative hazard functions:

$$H_1(t) = \phi H_0(t), \quad \forall t \geq 0$$

- Note that that ϕ is constant and therefore does not depend on t (eg. women have a survival advantage at all ages).

More on Proportional Hazards

- The proportional hazards assumption is important for Cox models.
- One way to check is a log-log plot: time versus hazard, also called a *Weibull plot*:
- For example with $\phi = 2$ then the vertical distance is $\log(2) = 0.693$:



A Note On Proportional Hazards

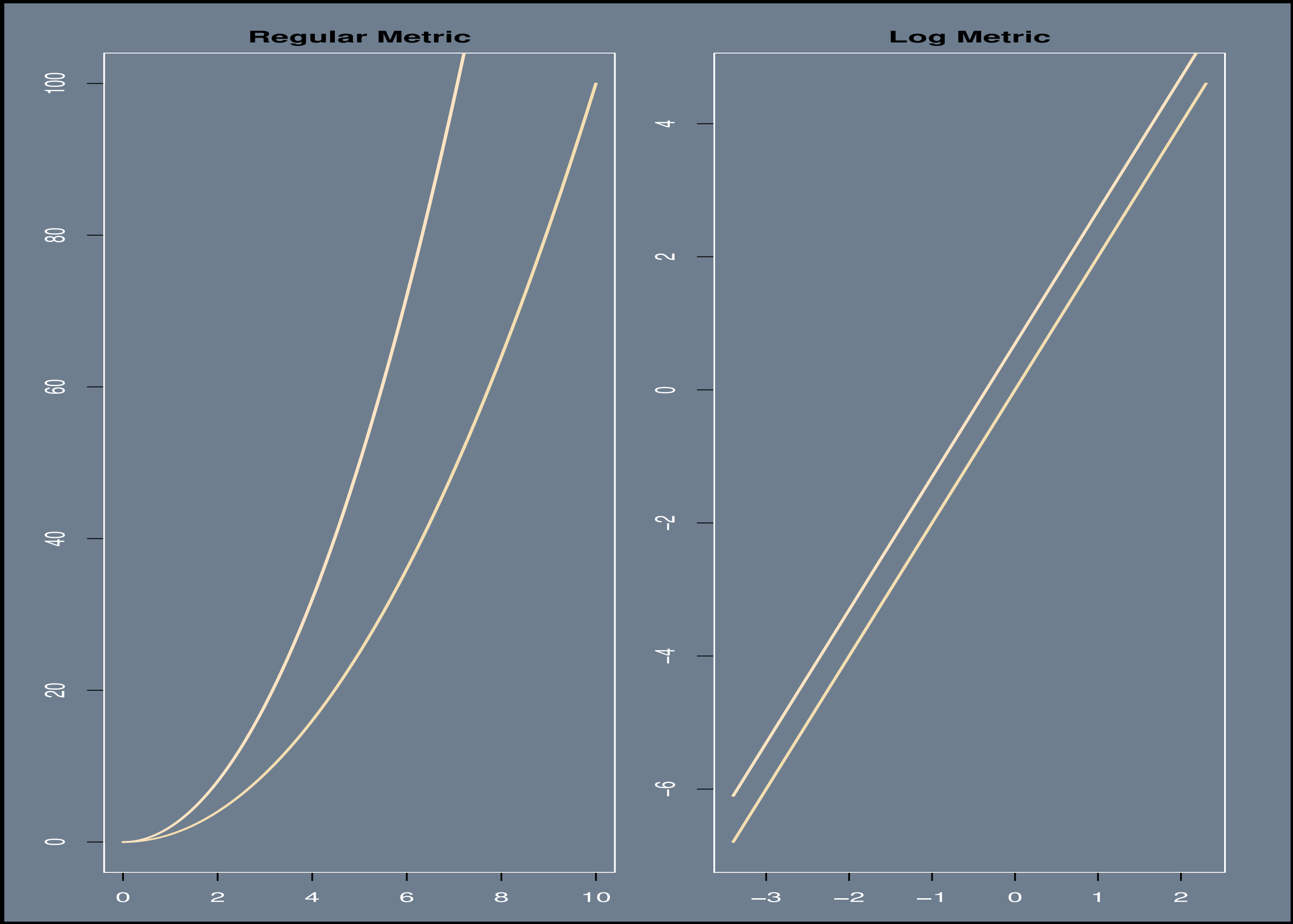
- To increase our intuition, consider the simple functions:

$$h_0(t) = t^2, \quad h_1(t) = 2h_0(t) = 2t^2, \quad t \geq 0.$$

- Let's plot these on the regular and log scale:

```
par(oma=c(1,1,1,1),mar=c(2,2,2,1),mfrow=c(1,2),col.axis="white",
    col.lab="white",col.sub="white",col="white", bg="slategray")
dur <- seq(0,10,length=300)
plot(dur,dur^2,type="l",lwd=2,col="wheat", main="Regular Metric")
lines(dur,2*dur^2,lwd=2,col="bisque")
plot(log(dur),log(dur^2),type="l",lwd=2,col="wheat", main="Log Metric")
lines(log(dur),log(2*dur^2),lwd=2,col="bisque")
```

A Note On Proportional Hazards



Log-Rank Test

- This is a comparison of hazards for two groups.
- Book example with two samples: `letters`(A,B,C,D,E) and `numbers`(1,2,3,4,5), where we are concerned about whether the groups have the same survival profile:
  - H_0 : No difference in survival between `letters` and `numbers`.
- The expected deaths is calculated under the null by assuming the marginals are fixed.
- The table with syntax is:

Group	Deaths	Survivors	Total
I	d_1	$n_1 - d_1$	n_1
II	d_2	$n_2 - d_2$	n_2
Total	d	$n - d$	n

Log-Rank Test, Fake Data

Group	Time	Event
`numbers`	4.0	1
`numbers`	2.0	0
`numbers`	6.0	1
`numbers`	1.0	1
`numbers`	3.5	0
`letters`	5.0	1
`letters`	3.0	1
`letters`	6.0	0
`letters`	1.0	1
`letters`	2.5	0

Log-Rank Test

- Observe $O_i = d_i$.
- The expected value for the first group under the null is:

$$E_1 = \frac{d \times n_1}{n}.$$

- The variance under the null is:

$$V = \frac{(n-d)dn_1n_2}{n^2(n-1)}$$

- Repeating for all sub-tables, the aggregate for the log-rank statistic:

$$T = \frac{(\sum_{i=1}^{k}(O_i - E_i))^2}{\sqrt{\sum_{i=1}^{k} V_i}}$$

  where this statistic is distributed χ^2 with one degree of freedom (for J groups it has $J-1$ degrees of freedom).

- This test has high power with proportional odds, but weak otherwise.

Log-Rank Table

Time	Group	Deaths	Survivals	Total	Expected	Difference	Variance
$t_{(1)}$	`numbers`	1	4	5	1	0.0	
	`letters`	1	4	5	1	0.0	
	Total	2	8	10			0.4444
$t_{(2)}$	`numbers`	0	3	3	1/2	-1/2	
	`letters`	1	2	3	1/2	1/2	
	Total	1	5	6			0.25
$t_{(3)}$	`numbers`	1	1	2	1/2	1/2	
	`letters`	0	2	2	1/2	-1/2	
	Total	1	3	4			0.25
$t_{(4)}$	`numbers`	0	1	1	1/3	-1/3	
	`letters`	1	1	2	1/3	2/3	
	Total	1	2	3			0.2222
$t_{(5)}$	`numbers`	1	0	1	1/2	1/2	
	`letters`	0	1	1	1/2	-1/2	
	Total	1	1	2			0.25
Sum for `numbers`		3	9	12	2.8333	0.16667	1.4166
Sum for `letters`		3	10	13	2.8333	0.16667	1.4166

Log-Rank Table

- It does not matter which group we pick with this version of the test.
- The test statistic is:

$$T = \frac{\left(\sum_{i=1}^{k}(O_i - E_i)\right)^2}{\sqrt{\sum_{i=1}^{k} V_i}} = \frac{0.1667}{\sqrt{1.4166}} = 0.01961.$$

- Because:

```
pchisq(0.01961,df=1,lower.tail=FALSE)
[1] 0.88863
```

then we conclude that there is no evidence to claim a difference in the hazards of the two groups.

Log-Rank Table, in R

- Use the Cox PH regression function:

```
lapply("eha","survival")
ex.data <- data.frame("group"=rep(1:2, each = 5),
    "time"=c(4,2,6,1,3.5,5,3,6,1,2.5),"event"=c(1,0,1,1,0,1,1,0,1,0))
ex.fit <- coxph(Surv(time,event) ~ group, data=ex.data)
ex.fit$score
[1] 0.018868
```

- Note that this is slightly different than the hand calculations: this toy example is too small to meet the asymptotic assumptions of the test and the R function uses the accompanying distributions.

Log-Rank Table, in R

- So let's use the the old age mortality dataset, even though we will care about only one value:

```
data(oldmort)
oldmort.fit <- coxph(Surv(enter, exit, event) ~ sex, data = oldmort)
summary(oldmort.fit)
  n= 6495, number of events= 1971

             coef exp(coef) se(coef)     z Pr(>|z|)
sexfemale -0.1929    0.8245   0.0456 -4.23  2.3e-05

          exp(coef) exp(-coef) lower .95 upper .95
sexfemale     0.825       1.21     0.754     0.902

Concordance= 0.532  (se = 0.007 )
Rsquare= 0.003   (max possible= 0.985 )
Likelihood ratio test= 17.7  on 1 df,   p=2.54e-05
Wald test            = 17.9  on 1 df,   p=2.32e-05
Score (logrank) test = 18  on 1 df,     p=2.25e-05
```

Log-Rank Table, in R

- Since:

```
oldmort.fit$score
[1] 17.961
pchisq(oldmort.fit$score,df=1,lower.tail=FALSE)
[1] 2.2544e-05
```

  we can claim a difference by sex.

- The extent of this difference is the proportionality constant, using:

```
             coef exp(coef) se(coef)      z Pr(>|z|)
sexfemale -0.1929    0.8245   0.0456 -4.23  2.3e-05
```

  with the exponential of the coefficient to undo the log metric we see that female mortality is 0.8245 that of male mortality (a value of 1 would mean no difference).

- We can also look at this in reverse:

```
          exp(coef) exp(-coef) lower .95 upper .95
sexfemale     0.825       1.21     0.754     0.902
```

  showing that male mortality is 1.21 that of female mortality ($1/1.213 = 0.8244$).

Model Assessment from the Output

- Concordance looks at all of the individual predictions and calculates the probability that for two randomly chosen cases one with the shorter survival also has the larger risk score. So higher values are better.

```
Concordance= 0.532  (se = 0.007 )
```

- R-Square in this context is the Cox & Snell pseudo R-squared. Ignore it.

```
Rsquare= 0.003   (max possible= 0.985 )
```

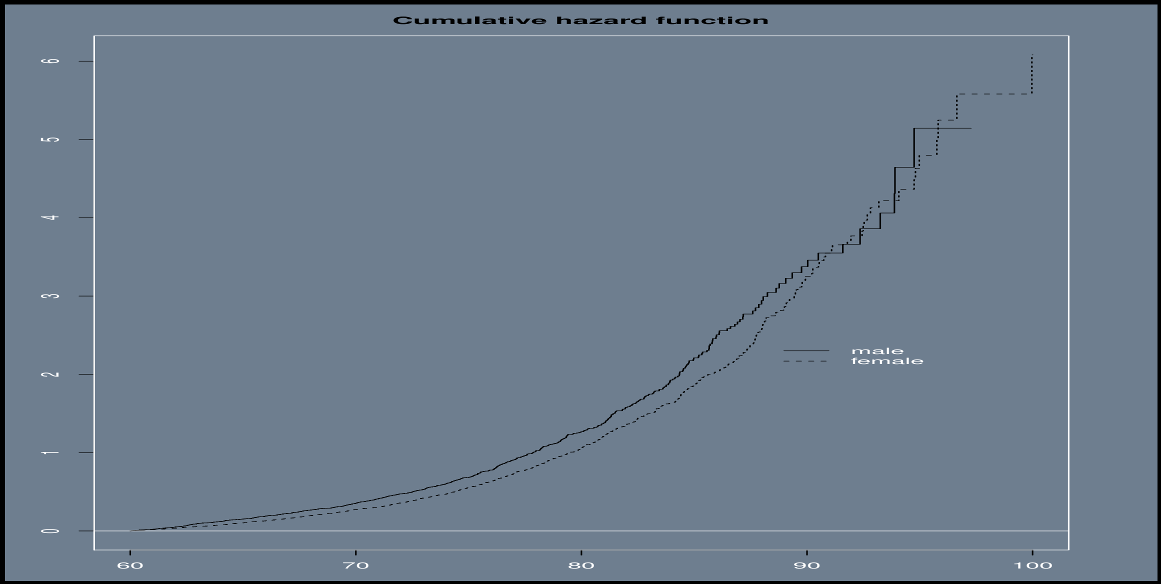
- The last three lines give the likelihood-ratio, Wald, and score chi-square statistics, and are asymptotically equivalent chi-square distributed tests of the null hypothesis that all of the β coefficients are zero. Here the test statistics are basically the same, and the null hypothesis is rejected.

```
Likelihood ratio test= 17.7  on 1 df,   p=2.54e-05
Wald test            = 17.9  on 1 df,   p=2.32e-05
Score (logrank) test = 18  on 1 df,     p=2.25e-05
```

Log-Rank Table, in R

- We can check the proportional hazards assumption graphically as done before:

```
par(mfrow=c(1,1),mar=c(3,3,3,3),col.axis="white",
    col.lab="white", col.sub="white",col="white",bg="slategray")
with(oldmort, plot(Surv(enter, exit, event), strat=sex))
```



Several Samples

- Now generalize from the 2×2 case to $k \times 2$, with $k - 1$ degrees of freedom instead of $2 - 1$:

```
oldmort.fit2 <- coxph(Surv(enter, exit, event) ~ birthplace, data=oldmort)
summary(oldmort.fit2)
  n= 6495, number of events= 1971

                   coef exp(coef) se(coef)    z Pr(>|z|)
birthplaceregion 0.0594    1.0612   0.0552 1.08     0.28
birthplaceremote 0.1008    1.1060   0.0595 1.69     0.09

                 exp(coef) exp(-coef) lower .95 upper .95
birthplaceregion      1.06      0.942     0.952      1.18
birthplaceremote      1.11      0.904     0.984      1.24

Concordance= 0.507  (se = 0.007 )
Rsquare= 0.001   (max possible= 0.985 )
Likelihood ratio test= 3.25  on 2 df,   p=0.197
Wald test            = 3.28  on 2 df,   p=0.194
Score (logrank) test = 3.28  on 2 df,   p=0.194
```

Several Samples

- Since:

```
table(oldmort$birthplace)
parish region remote
  3598   1503   1394
```

  we know that `parish` is the references category.

- Also since:

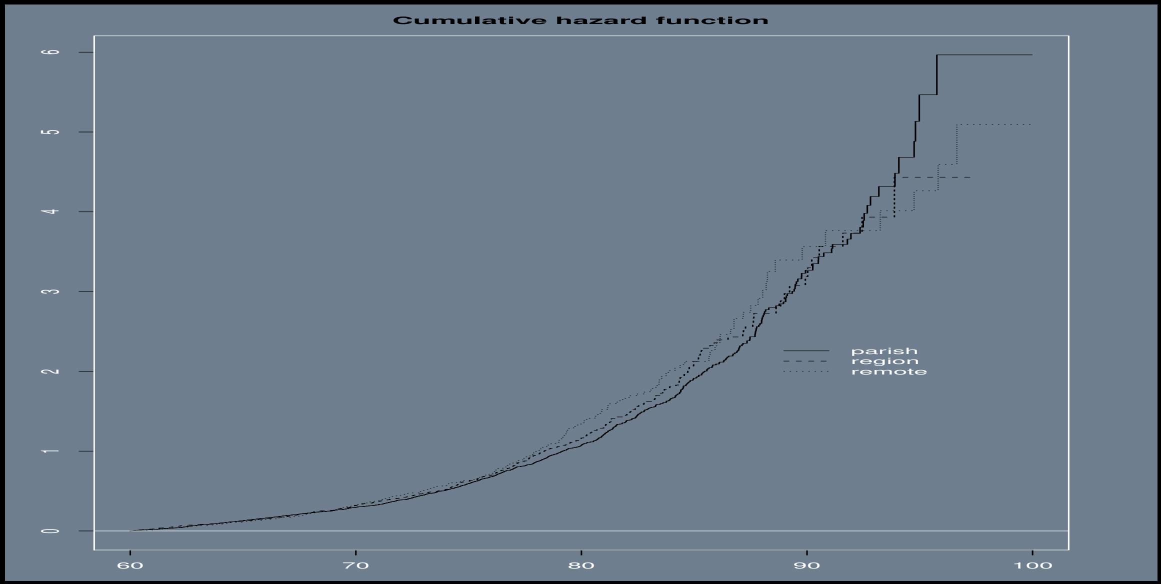
```
oldmort.fit2$score
[1] 3.2804
pchisq(oldmort.fit2$score,df=2,lower.tail=FALSE)
[1] 0.19394
```

  we cannot claim a difference by region of birth.

Several Samples

- We can check the proportional hazards assumption graphically as done before:

```
par(mfrow=c(1,1),mar=c(3,3,3,3),col.axis="white",
    col.lab="white", col.sub="white",col="white",bg="slategray")
with(oldmort, plot(Surv(enter, exit, event), strat=birthplace))
```

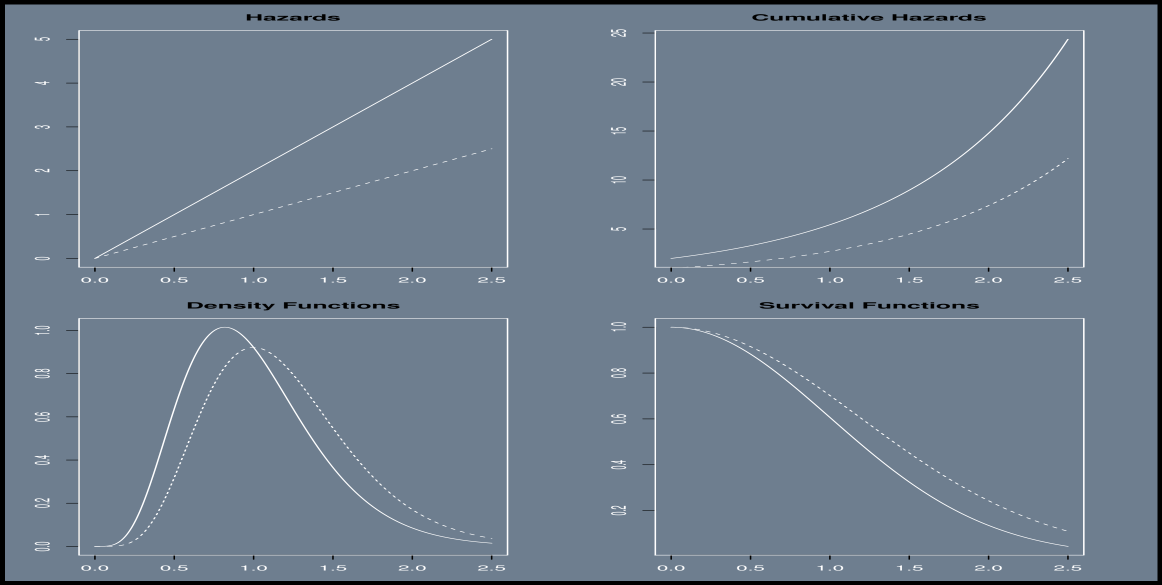


When the Proportional Hazards Assumption is Rejected

- Look at the structural form of the linear additive component. It may be that they have a more nuanced relationship with the hazard.
- Try stratifying on levels of the exposure variable. Other variables my have important strata as well.
- Fit Kaplan Meier curves for each exposure group independently and report differences.
- Subset the time of analysis such that proportional hazards assumption holds for a given time interval of interest.
- Relatedly, fit a model over two (or more) time intervals such that the proportional hazards assumption fits in each era. Two (or more) different Cox model results may even be more interesting.
- Specify a model with time-dependent explanatory variables.
- Use a parametric model instead that does not require the proportional hazards assumption.
- Specify an extended Cox model that has an interaction of exposure and time.

Continuous Time

- It is still true that $h_1(t) = \phi_1 h_0(t)$ and $h_2(t) = \phi_2 h_0(t)$, for all $t \geq 0$.
- So even though the hazards and cumulative hazards are proportional, the density and survival functions are not guaranteed to be unless $\phi = 0$.



Proportional Hazards, Two Groups, Continuous Time

- Rewrite the proportional hazards equation with an additional variable:

$$h_x(t) = \phi^x h_0(t), \quad t \geq 0,\ x = \{0, 1\},\ \phi > 0.$$

- So when $x = 0$ then the hazards are equivalent and when $x = 1$ then we get proportional hazards as seen before.

- Since $\phi > 0$ then we can rewrite it as $\beta = \log(\phi)$, meaning that $\phi = \exp(\beta)$.

- Therefore:

$$h_x(t) = (\exp(\beta))^x h_0(t) = \exp(\beta x) h_0(t), \quad t \geq 0, x = \{0, 1\}, -\infty < \beta < \infty,$$

  where the Broström book also labels the LHS $h(t; x)$.

- This is a "two sample situation" in that x indexes one subgroup or another in the data.

Proportional Hazards, $k+1$ Groups, Continuous Time

- Consider $k+1$ groups where each one has its own hazard function, starting with the reference group, $h_0(t) \sim \text{group}\, 0$, for pairwise comparisons:

$$\begin{aligned} h_1(t) &= \phi_1 h_0(t) \sim \text{group}\, 1 \\ h_2(t) &= \phi_2 h_0(t) \sim \text{group}\, 2 \\ &\vdots \\ h_{k-1}(t) &= \phi_{k-1} h_0(t) \sim \text{group}\, k-1 \\ h_k(t) &= \phi_k h_0(t) \sim \text{group}\, k. \end{aligned}$$

- Now the definition of x is expanded to be a k-length vector of indicator values, $\mathbf{x} = (x_0, x_1, x_2, \dots, x_k)$, according to:

$\mathbf{x}$	1	2	...	$k-1$	k
Group 0	0	0	...	0	0
Group 1	1	0	...	0	0
Group 2	0	1	...	0	0
$\vdots$					
Group $k-1$	0	0	...	1	0
Group k	0	0	...	0	1

General Proportional Hazards Model

- This provides the specification:

$$h(t;\mathbf{x}) = h_0(t)\exp(x_1\beta_1 + x_2\beta_2 + \cdots + x_{k-1}\beta_{k-1} + x_k\beta_k) = h_0(t)\exp(\mathbf{x}\boldsymbol{\beta}).$$

- We want to allow $\mathbf{x}$ to take on any values, not just binary indicators, as in standard regression.
- For $i = 1, \dots, n$ cases in the data, define the ith data element as $(t_{i0}, t_i, d_i, \mathbf{x}_i)$, where:
  - t_{i0} is a left truncation time: if $y_{i0} = 0,\ \forall i$, then drop this
  - t_i is end time
  - d_i is event indicator: 1 if TRUE, 0 if FALSE
  - $\mathbf{x}_i$ is a vector of explanatory variables.
- The *survival object* is then (t_{i0}, t_i, d_i).
- So now we need to estimate $h_0(t)$ and $\boldsymbol{\beta}$.

General Proportional Hazards Model

- Re-expressing the general PH model so that the log hazard ratio is on the LHS (Box-Steffensmeier and Jones, page 49):

$$\log\left[\frac{h_i(t;\mathbf{x}_i)}{h_0(t)}\right] = x_{1i}\beta_1 + x_{2i}\beta_2 + \cdots + x_{k-1i}\beta_{k-1i} + x_{ki}\beta_k$$

  for a single case i.

- Notice that the constant term, β_0 is absorbed into the baseline hazard function, $h_0(t)$.

- This last expression of the Cox PH model reveals clearly the interpretation of the estimated coefficients:

  - positive and statistically reliable coefficients imply that increasing values of the corresponding x variable increase the hazard (shorten the period)

  - negative and statistically reliable coefficients imply that increasing values of the corresponding x variable decrease the hazard (lengthen the period)

  - statistically unreliable coefficients give no evidence of an effect of the corresponding $x$variable on the hazard.

Cox Model of UN Peacekeeping Missions

- Box-Steffensmeier and Jones, page 49:

Variable	Estimate	(s.e.)
Civil War	0.73	(0.38)
Interstate Conflict	-0.86	(0.50)
N	54	
Log-Likelihood	-127.16	

- Neither of these coefficients has a t-score bigger than 1.96 in absolute value.
- Recall that the reference category is Internationalized Civil War, which the value 0 in a treatment contrast: $(1, 0), (0, 1), (0, 0)$.
- We get the hazard ratio between Civil War and Interstate Conflict by: $h(t)/h_0(t) = \exp(\hat{\beta}_{CW}(X_{CW} - X_{IC})) = \exp(0.73(1-0)) = 2.075$.
- We get the hazard ratio between Interstate Conflict and Civil War by: $h(t)/h_0(t) = \exp(\hat{\beta}_{IC}(X_{IC} - X_{CW})) = \exp(-0.86(1-0)) = 0.423$.
- The hazard ratio for Internationalized Civil War to either of the other categories must be $h(t)/h_0(t) = \exp((X_{ICW} - X_{other})) = \exp((0-0)) = 1$.

Estimation of the Continuous Baseline Cumulative Hazard

- First consider the interval to be summed: $j : t_j < t$, which is all of the time periods that come before j,
- Define R_j to be the number at risk at time period j.
- We say that $\ell \in R_j$ if the ℓth case (out of the i possible) is in the risk group at time j.
- $\hat{\boldsymbol{\beta}}$ is the estimated coefficient vector from MLE.
- $\mathbf{x}_i$ is the ith case's vector of explanatory variables, continuous or discrete.
- The standard estimate is:

$$\hat{H}_0(t) = \sum_{j:t_j<t} \frac{d_i}{\sum_{\ell \in R_j} \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}})}$$

Estimation of the Continuous Baseline Cumulative Hazard

- If $\hat{\boldsymbol{\beta}}$ is a vector of all zeros (no regression inferences), then

$$\hat{H}_0(t) = \sum_{j:t_j<t} \frac{d_i}{\sum_{\ell \in R_j} \exp(\mathbf{x}_i \boldsymbol{\beta})}$$

  reduced to:

$$\hat{H}_0(t) = \sum_{j:t_j<t} \frac{d_i}{n_j}$$

  since:

$$\sum_{\ell \in R_j} \exp(\mathbf{x}_i \mathbf{0}) = \sum_{\ell \in R_j} (1) = n_j,$$

  which is the Nelson-Aalen estimator

- The `eha` package from the Broström book centers the explanatory variables at their means during this estimation:

$$\hat{H}_0(t, \text{eha}) = \sum_{j:t_j<t} \frac{d_i}{\sum_{\ell \in R_j} \exp((\mathbf{x}_i - \bar{\mathbf{x}})\boldsymbol{\beta})}$$

  (presumably for numerical stability in the `R` calculations).

Continuous Time, Factor Covariate Interpretation

- For individual i at time t with explanatory variable vector $\mathbf{x}_i$ the hazard function is:

$$\hat{h}(t;\mathbf{x}_i) = \hat{h}_0(t)\exp(\mathbf{x}_i\boldsymbol{\beta})$$

which shows the porportionality of the regression information relative to the baseline:

$$\frac{\hat{h}(t;\mathbf{x}_i)}{\hat{h}_0(t)} = \exp(\mathbf{x}_i\boldsymbol{\beta})$$

(relative risk to the baseline), and also

$$\log\left(\frac{\hat{h}(t;\mathbf{x}_i)}{\hat{h}_0(t)}\right) = \mathbf{x}_i\boldsymbol{\beta}$$

(log relative risk to the baseline).

- $\hat{h}(t;\mathbf{x}_i)$ allows calculation of the survival function for the ith case:

$$\hat{S}(t,\mathbf{x}_i) = \exp(-\hat{h}(t;\mathbf{x}_i)).$$

Continuous Time, Continuous Coefficient Interpretation

- Assume for the moment a single explanatory variable just for simplicity:

$$h(t;x) = h_0(t)\exp(x\boldsymbol{\beta})$$

- Consider the effect of adding 1 to this explanatory variable the way we often discussion linear models:

$$\frac{h(t;x+1)}{h(t;x)} = \frac{h(t)\exp(\beta(x+1)}{h(t)\exp(\beta x)} = \frac{\exp(\beta(x+1)}{\exp(\beta x)} = \frac{\exp(\beta x)\exp(\beta)}{\exp(\beta x)} = \exp(\beta).$$

- So incrementing x by one increases the relative risk (hazard ratio) by $\exp(\beta)$, which easy to interprete.

- This is why `R` (and other packages) routinely provide the exponent of the coefficient as well in the output.

- Skip Broström Section 3.9.1.

Proportional Hazards In Discrete Time

- Proportional hazards in discrete time is a set of conditional probabilities, which are by definition bounded by $[0:1]$.
- So ratios can give awkard numbers that exceed one.
- Fix: assume continuous time that is segmented into a set of k period by the mechanism of measurement: $0 = t_0 < t_1 < t_2 < \cdots < t_k = \infty$.
- Then for the random variable T we have:

$$p(t_i \leq T < t_i | T > t_{i-1}, \mathbf{x}) = \frac{S(t_{i-1}|\mathbf{x}) - S(t_i|\mathbf{x})}{S(t_{i-1}|\mathbf{x})}$$

Since survival functions work "backwards" relative to cumulative hazard functions: $S(t) = p(T \geq t) = 1 - H(t)$, and continuing

$$\cdots = 1 - \frac{S(t_i|\mathbf{x})}{S(t_{i-1}|\mathbf{x})} = 1 - \left(\frac{S_0(t_i)}{S_0(t_{j-1})}\right)^{\exp(\mathbf{x}\boldsymbol{\beta})} = 1 - (1 - h_i)^{\exp(\mathbf{x}\boldsymbol{\beta})}$$

where:

$$h_i = p(t_{i-1} \leq T < t_i | T \geq t_{i-1}, \mathbf{x}_0)$$

to give the definition of proportional hazards in discrete time.

Cox Likelihood

- Suppose there are n cases, with r events (deaths, etc.), and $n - r$ right censorings.
- Assume initially that only one individual dies at each at each death time: no ties, so that we can order the death times: $t_{(1)} < t_{(2)} < \cdots < t_{(r-1)} < t_{(r)}$.
- The subjects/cases at risk at time $t_{(j)}$ are denoted $R(t_{(j)})$ (alive and uncensored at the beginning of time $t_{(j)}$).
- Denote $\mathbf{x}_{(j)}$ as the vector of explanatory variables for the subject that dies at time $t_{(j)}$ (note that we are not indexing cases by $i = 1, \ldots, n$).
- Sir David Cox (1972) derived the following PH likelihood function:

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{r} \frac{\exp(\mathbf{x}_{(j)}\boldsymbol{\beta})}{\sum_{\ell \in R(t_{(j)})} \exp(\mathbf{x}_{\ell}\boldsymbol{\beta})}$$

- So each of the r terminations contributes one element to the numerator, and the denominator for each contribution is the sum of risk set at unique termination time $t_{(j)}$.

Cox Likelihood

- Onservations about...

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{r} \frac{\exp(\mathbf{x}_{(j)}\boldsymbol{\beta})}{\sum_{\ell \in R(t_{(j)})} \exp(\mathbf{x}_{\ell}\boldsymbol{\beta})}$$

- All of the individual contributions, terminations and non-terminations are positive.
- So terminations raise the likelihood function.
- Non-terminations do not contribute at all to the numerator of the likelihood function, but they are in the sum in the denominator, meaning survivals lower the likelihood function.
- Observe that the initial ordering of the terminations, $t_{(1)} < t_{(2)} < \cdots < t_{(r-1)} < t_{(r)}$, means that the likelihood function depends only the rank of the termination times, giving the termination in the numerator and the risk set in the denominator.

Cox Likelihood, Computational Considerations

- To setup Newton-Raphson style estimation of the unknown coefficients, we need to re-express the likelihood function.
- For n cases denote $t_1, t_2, \ldots, t_{n-1}, t_1$ survival times.
- Define ξ_i as an event indicator according to:

$$\xi_i = \begin{cases} 0 & \text{if } t_i \text{ is right censored} \\ 1 & \text{otherwise.} \end{cases}$$

- Now the likelihood function is:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\frac{\exp(\mathbf{x}_i \beta)}{\sum_{\ell \in R(t_i)} \exp(\mathbf{x}_i \boldsymbol{\beta})} \right]^{\xi_i}$$

where we now have a slightly different definition of the risk set: $R(t_i)$.

- In log terms this is:

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \xi_i \left[\mathbf{x}_i \beta - \log \sum_{\ell \in R(t_i)} \exp(\mathbf{x}_i \boldsymbol{\beta}) \right]$$

Partial Likelihood

- Because Cox's likelihood function is based on ranks and not actual survival times it does not produce a true likelihood in the conventional sense: *the actual time intervals between events are assumed to carry no information between the explanatory variables and the ordered events.*

- However, it has since been proven that all of the important properties carry over to this partial likelihood: the parameter estimates are *asymptotically normal*, *asymptotically efficient*, *consistent*, and *invariant to reparameterization.*

- Note that the denominator changes for each contribution to the likelihood function.

Partial Likelihood Example

- Box Steffensmeir and Jones (page 53) give the following contrived example in the order of events or censoring (modified from Collett 2003, page 66).

$t_{(j)}$	Case	Duration	Right Censored
1	7	7	No
2	4	15	No
3	5	21	No
4	2	28	Yes
5	9	30	Yes
6	3	36	No
7	8	45	Yes
8	1	46	No
9	6	51	No

- Consider how to manually create the Cox PH likelihood,

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{r} \frac{\exp(\mathbf{x}_{(j)}\boldsymbol{\beta})}{\sum_{\ell \in R(t_{(j)})} \exp(\mathbf{x}_{\ell}\boldsymbol{\beta})},$$

for this example.

Partial Likelihood Example

- Now define $\psi(k) = \exp(\mathbf{x}_k\boldsymbol{\beta})$ for non-censored cases and $\zeta(k) = \exp(\mathbf{x}_k\boldsymbol{\beta})$ fo the censored cases, such that the partial likelihood is given by:

$$L(\boldsymbol{\beta}) = \frac{\psi(7)}{\psi(1)+\zeta(2)+\psi(3)+\psi(4)+\psi(5)+\psi(6)+\psi(7)+\zeta(8)+\zeta(9)}$$
$$\times \frac{\psi(4)}{\psi(1)+\zeta(2)+\psi(3)+\psi(4)+\psi(5)+\psi(6)+\zeta(8)+\zeta(9)}$$
$$\times \frac{\psi(5)}{\psi(1)+\zeta(2)+\psi(3)+\psi(5)+\psi(6)+\zeta(8)+\zeta(9)}$$
$$\times \frac{\psi(3)}{\psi(1)+\psi(3)+\psi(6)+\zeta(8)}$$
$$\times \frac{\psi(1)}{\psi(1)+\psi(6)}$$
$$\times \frac{\psi(6)}{\psi(6)}$$

- Note also that censored cases never contribute to the numerator, only to the denominator..

The Trouble With Ties

- This last exercise neatly illustrates that the likelihood function is based on ordering not actual time.
- Since the likelihood is built on an ordering of unique durations, there is no facility for ties.
- Ties occur for various reasons, but mostly because of the non-granularity of how time is measured.
- Therefore we cannot tell which of two tied cases occur in the measurement interval.
- So there needs to be some modification of the partial likelihood to accomodate ties.
- Four major approaches in the literature: *Breslow, Efron, Average Likelihood, Exact Discrete* (in order of popularity).

The Breslow Method For Ties

- **Basic Idea:** keep the risk set the same for ties and consecutively product them into the likelihood function, so that they belong to the same risk set at the tied time of events.

- BSJ example: four cases with event times: $5, 5, 8, 14$ (no censoring), so the first two contributions to the likelihood function are:

$$\ell_1 = \frac{\psi(1)}{\psi(1) + \psi(2) + \psi(3) + \psi(4)} \qquad \ell_2 = \frac{\psi(2)}{\psi(1) + \psi(2) + \psi(3) + \psi(4)}$$

- Non-tied contributions remain the same and the approximate likelihood function contributions for the k tied cases is:

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{k} \frac{\exp(\boldsymbol{s}_j \boldsymbol{\beta})}{\left[\sum_{\ell \in R(t_{(j)})} \exp(\mathbf{x}_\ell \boldsymbol{\beta})\right]^{d_j}}$$

  where $\boldsymbol{s}_j$ is the sum of the explanatory variable vectors for the tied cases and d_j are the number of deaths at time t_i ($d_j = 2$ for one tie*).

- This approximation works well as long as the number of tied cases is modest.

Efron's Method For Ties

- **Basic Idea:** assuming that each of the tied cases is equally likely to be *actually* first, so average the risk set between them.

- BSJ example again, the truth is one of:

$$\ell_{1,2} = \frac{\psi(1)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)}\frac{\psi(2)}{\psi(2)+\psi(3)+\psi(4)} \qquad \ell_{2,1} = \frac{\psi(2)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)}\frac{\psi(1)}{\psi(1)+\psi(3)+\psi(4)}$$

- So include include the full risk set and the average second risk set in one combined contribution to the likelihood function that removes double counting the average:

$$\ell_{12} = \frac{\psi(1)+\psi(2)}{(\psi(1)+\psi(2)+\psi(3)+\psi(4)) \times [\psi(1)+\psi(2)+\psi(3)+\psi(4)-(\psi(1)+\psi(2))/2]}$$

- The complete statement of the likelihood function is given by:

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{r} \frac{\exp(\boldsymbol{s}_j\boldsymbol{\beta})}{\prod_{k=1}^{d_j}\left[\sum_{\ell\in R(t_j)}\exp(\mathbf{x}_\ell) - (k-1)\frac{1}{d_j}\sum_{\ell\in D(t_j)}\exp(\mathbf{x}_\ell\boldsymbol{\beta})\right]}$$

  where the terms are the same except a new term, $D(t_j)$, for the set of cases that die at time t_j.

- B. and E. methods give similar results, particularly with large data and/or small numbers of ties.

Averaged Likelihood Method

- **Basic Idea:** average the likelihoods produced my combinatorically producing all possible likelihood contributions for the tied period (Therneau and Grambsch 2000).

- This can be computationally challenging for cases where there are many ties in a given period.

- While this method is sometimes called the *exact method*, it is still given as an approximation by software solutions.

- Suppose we modify the BSJ example to have four cases with event times: $5, 5, 5, 14$ (no censoring), giving one three-way tie.

Averaged Likelihood Method

- There are now $3! = 6$ possible orderings within the tied period, giving six possible contributions to the partial likelihood function:

$$\ell_{1,2,3} = \frac{\psi(1)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)} \times \frac{\psi(2)}{\psi(2)+\psi(3)+\psi(4)} \times \frac{\psi(3)}{\psi(3)+\psi(4)}$$

$$\ell_{1,3,2} = \frac{\psi(1)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)} \times \frac{\psi(3)}{\psi(2)+\psi(3)+\psi(4)} \times \frac{\psi(2)}{\psi(2)+\psi(4)}$$

$$\ell_{2,1,3} = \frac{\psi(2)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)} \times \frac{\psi(1)}{\psi(1)+\psi(3)+\psi(4)} \times \frac{\psi(3)}{\psi(3)+\psi(4)}$$

$$\ell_{2,3,1} = \frac{\psi(2)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)} \times \frac{\psi(3)}{\psi(1)+\psi(3)+\psi(4)} \times \frac{\psi(1)}{\psi(1)+\psi(4)}$$

$$\ell_{3,1,2} = \frac{\psi(3)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)} \times \frac{\psi(1)}{\psi(1)+\psi(2)+\psi(4)} \times \frac{\psi(2)}{\psi(2)+\psi(4)}$$

$$\ell_{3,2,1} = \frac{\psi(3)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)} \times \frac{\psi(2)}{\psi(1)+\psi(2)+\psi(4)} \times \frac{\psi(1)}{\psi(1)+\psi(4)}$$

- The likelihood contribution of these three tied cases is $L^* = \sum \ell()/((k-1)!) = \sum \ell_{1:3}/2$.

Exact Discrete Method

- **Basic Idea:** the exact discrete method is an approximation that uses the probability that an event occurs at time t_i conditional on the risk set.
- The previous three methods assumed that the underlying time metric was continuous but that ties occur due to data recording practices that produce discrete periods.
- In the truly discrete case, ties can be actual ties rather than a consequence of recording.
- BSJ use another contrived example as an example with *matched observations* according to:

Observation Number	Risk Period	Event Occurence	Duration
1	1	0	5
2	1	0	5
3	1	0	5
4	1	1	5
5	1	1	5
1	2	0	11
2	2	0	11
3	2	1	11

- So for the first period there are 5 matched cases with 2 cases (event) and 3 controls (no event) for a $3:2$ ratio.

Exact Discrete Method

- Meaning…Period One: 2 tied events, 3 non-events; 5 at risk, Period Two: 1 event, 2 non-events, 3 at risk.
- The idea is to use the censoring indicator as an outcome.
- The exact discrete approximation of the partial likelihood comes from an estimation of a response pattern of 0's and 1's in the risk set conditional on the cases in the risk set.
- In the first risk period there are $\binom{5}{2} = 10$ possible ways to get 2 events from 5 cases.
- So what is the probability that the observed 2/5 occur?

Exact Discrete Method

- Suppose there are $k = 1, 2, \ldots, K$ risk periods and J_k observations at risk in the kth period.
- The response pattern of the outcome variable at the kth period is: $\mathbf{y}_k = \sum_{y \in k}$.
- The total number of events (cases) observed in the kth risk period is:

$$n_{1k} = \sum_{i=1}^{J} y_{ki}.$$

- The total numer of non-events (controls) observed in the kth risk period is:

$$n_{0k} = J_k - n_{1k}.$$

Exact Discrete Method

- The probability of observing the response pattern $\mathbf{y}_k$ is then:

$$p(\mathbf{y}_k | \sum_{i=1}^{J} y_{ki} = n_{1k}) = \frac{\exp(\sum_{i=1}^{J} \mathbf{x}_{ki}\mathbf{y}_{ki}\boldsymbol{\beta})}{\sum_{\boldsymbol{d}_k \in R_k} \exp(\sum_{1=1}^{J} \mathbf{x}_{ki} d_{ki})}$$

where R_k is the set of all possible combinations of 0's and 1's in the kth risk period, $d_k = (d_{k1}, d_{k2}, \ldots, d_{kJ},\ d_{ki} = 0$ or 1.

- So we estimate this probability of the response parttern for each risk period, for all possible combination of events and non-events amonth the J observations at risk.

- Note that this is equivalent to a conditional logit model.

Broström Example of the Four Methods

```
library(eha)
data(fert)
first <- fert[fert$parity == 1,]
# Default method is Efron, so specifying is redundant
fit.E <- coxreg(Surv(next.ivl, event) ~ year + age, data = first, method="efron")
# Breslow
fit.B <- coxreg(Surv(next.ivl, event) ~ year + age, data = first, method="breslow")
# Maximum Partial Likelihood
fit.X <- coxreg(Surv(next.ivl, event) ~ year + age, data = first, method="mppl")
# Maximum Likelihood
fit.D <- coxreg(Surv(next.ivl, event) ~ year + age, data = first, method="ml")
```

Broström Example of the Four Methods

- This gives the following Cox PH regression results:

	year	age
Efron	0.0017	-0.0390
Breslow	0.0016	-0.0390
MPPL	0.0016	-0.0390
ML	0.0016	-0.0390

- So there are basically no differences here.
- Note that for the combinatorical method there are approximately 7×10^{192} permutations, which is computationally impractical.
- It is totally unclear what Broström means for the second two models, even in the `cran` documentation.

Broström's Advice On Model Selection (page 49)

- All models are wrong, some are useful. TRUE.
- More than one model may be useful. TRUE.
- Keep important covariates in the model. TRUE, with caveats.
- Avoid automatic stepwise procedures. TRUE.
- In using interactions, the main effects must be present. FALSE.

My *Additional* Advice On Model Selection

- Think about model choice in terms of Leamer's "inside the horizon" and "outside the horizon" analogy.
- It's okay to keep explanatory variables in the model if they aren't reliable predictors of the outcome variable if they help in some other way.
- All models with non-identity link functions imply interactions.
- Deviance comparisons are generally the best way to compare nested models.
- R^2 does not mean much.

Male Mortality In Ages 40-60, Nineteenth Century

- Males born in the years 1800-1820 and surving at least 40 years in the parish Skellefte in northern Sweden are followed from their fortieth birthday until death or the sixtieth birthday, whichever comes first.
- 2058 observations with 6 variables.
- `id`: personal identification number.
- `enter`: start of duration in years since the fortieth birthday.
- `exit`: end of duration in years since the fortieth birthday.
- `event`:a logical vector indicating death at end of interval.
- `birthdate`: birthdate in decimal form.
- `ses`: socio-economic status, a factor with levels lower (565), upper (643).

Male Mortality In Ages 40-60, Nineteenth Century

```
data(mort)
dim(mort)
[1] 1208    6
mort[1:10,]
   id  enter   exit event birthdate   ses
1   1  0.000 20.000     0  1800.010 upper
2   2  3.478 17.562     1  1800.015 lower
3   3  0.000 13.463     0  1800.031 upper
4   3 13.463 20.000     0  1800.031 lower
5   4  0.000 20.000     0  1800.064 lower
6   5  0.000  0.089     0  1800.084 lower
7   5  0.089 20.000     0  1800.084 upper
8   6  0.000 20.000     0  1800.094 upper
9   7  0.000  3.388     0  1800.105 upper
10  7  3.388 14.063     1  1800.105 lower
```

Cox Proportional Hazards Regression Model

- We have already run limited versions this model with `coxph` and `coxreg`.
- Note that `coxreg` is a "wrapper" for `coxph` in the `survival` package, but gives different printing of the results.
- The `survfit` gives summaries from a fit `codeph` object, including estimates of $S(t)$ at the mean values of the covariates and the plot method for objects estimated surivival function, along with 95-percent confidence bands.
- A requirement with these functions is that we stipulate a survival object on the LHS of the model specification with the function `Surv`.
- Run a model with only two explanatory variables. . .

Cox Proportional Hazards Regression Model

```
res <- coxph(Surv(enter, exit, event) ~ ses * birthdate, data=mort)
summary(res) # HERE IT MATTERS
  n= 1208, number of events= 276
                coef exp(coef) se(coef)     z Pr(>|z|)
sesupper           -5.35e+01  5.74e-24  3.98e+01 -1.35     0.18
birthdate           1.62e-02  1.02e+00  1.42e-02  1.15     0.25
sesupper:birthdate  2.93e-02  1.03e+00  2.19e-02  1.33     0.18

                    exp(coef) exp(-coef) lower .95 upper .95
sesupper            5.74e-24   1.74e+23  8.12e-58  4.06e+10
birthdate           1.02e+00   9.84e-01  9.89e-01  1.05e+00
sesupper:birthdate  1.03e+00   9.71e-01  9.86e-01  1.07e+00

Concordance= 0.585  (se = 0.017 )
Rsquare= 0.02   (max possible= 0.953 )
Likelihood ratio test= 24.9  on 3 df,   p=1.64e-05
Wald test            = 22.8  on 3 df,   p=4.4e-05
Score (logrank) test = 23.7  on 3 df,   p=2.86e-05
```

Cox Proportional Hazards Regression Model

```
res <- coxreg(Surv(enter, exit, event) ~ ses + birthdate, data=mort)
summary(res) # OR JUST res (IT DOESN'T MATTER HERE)
Covariate           Mean       Coef     Rel.Risk   S.E.    Wald p
ses
           lower    0.416     0         1 (reference)
           upper    0.584    -0.484     0.616     0.121     0.000
birthdate        1811.640     0.029     1.029     0.011     0.008

Events                    276
Total time at risk         17038
Max. log. likelihood      -1841.7
LR test statistic         23.08
Degrees of freedom        2
Overall p-value           9.72167e-06
```

Determing Model Quality with Likelihood Ratio Tests

- With one statement we can LRT both explanatory variables for inclusion:

```
drop1(res, test = "Chisq")
Single term deletions

Model:
Surv(enter, exit, event) ~ ses + birthdate
          Df    AIC     LRT  Pr(>Chi)
<none>        3687.3
ses        1 3701.4 16.1095 5.978e-05
birthdate  1 3692.6  7.2752  0.006991
```

- `<none> 3687.3` is the AIC for the full model for comparison.
- The Broström book provides further testing of an interaction model, `ses*birthdate`, that doesn't work at all.

Determing Model Quality with Likelihood Ratio Tests

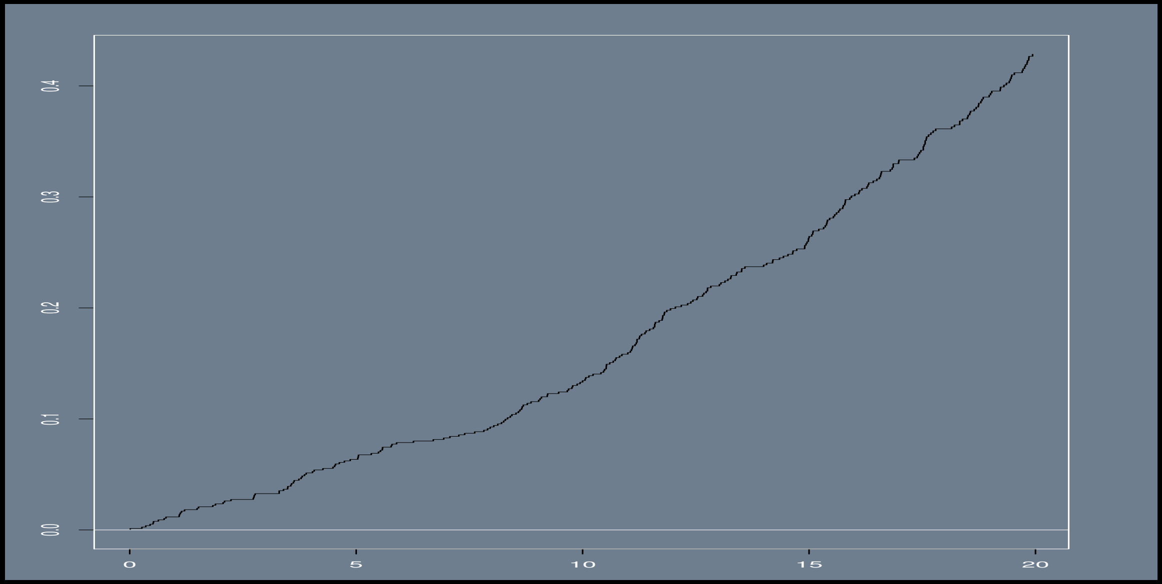
- Equivalently:

```
2*(
coxph(Surv(enter, exit, event) ~ ses + birthdate, data=mort)$loglik[2] -
coxph(Surv(enter, exit, event) ~       birthdate, data=mort)$loglik[2] )
[1] 16.10947        # TEST WITH 1 DF

2* (
coxph(Surv(enter, exit, event) ~ ses + birthdate, data=mort)$loglik[2] -
coxph(Surv(enter, exit, event) ~ ses            , data=mort)$loglik[2] )
[1] 7.275183       # TEST WITH 1 DF

2*(
coxph(Surv(enter, exit, event) ~ ses + birthdate, data=mort)$loglik[2] -
coxph(Surv(enter, exit, event) ~ 1              , data=mort)$loglik )
[1] 23.08231        # TEST WITH 2 DF
```

Plotting the Estimated Baseline Cumulative Hazard Function

- You must use `coxreg` not `coxph` to do this:

```
res <- coxreg(Surv(enter, exit, event) ~ ses + birthdate, data=mort)
par(mfrow=c(1,1),mar=c(3,3,3,3),col.axis="white",
    col.lab="white", col.sub="white",col="white",bg="slategray")
plot(res)
```



More on `coxph`

- To begin:

```
library(survival)
args(coxph)
function (formula, data, weights, subset, na.action, init, control,
    method = c("efron", "breslow", "exact"),
    singular.ok = TRUE, robust = FALSE, model = FALSE, x = FALSE, y = TRUE, ...)
```

- The right-hand side of the formula is the same as for a `lm()` or `glm()` model.
- The left-hand side is a survival object, created by the `Surv` function.
- In the basic case of just right-censored data, the call is `Surv(time, event)`, where `time` is either the event time or the censoring time, and `event` is a 0/1 variable coded 1 if the event is observed or 0 if the observation is censored.

More on `coxph`

- The parameter `method` specifies the algorithm to handle observations that have tied survival times: `"efron"` generally preferred to `"breslow"` method, and the `"exact"` method is more computationally intensive.

- Specifying `robust is TRUE` calculates robust coefficient-variance estimates, and is unecessary unless the specification includes non-independent observations, requiring the `cluster` function in the model formula.

Application From Week 3

- Back to the `mort` dataset:

```
library(eha,survival)
data(mort)
mort[1:10,]
   id  enter   exit event birthdate   ses
1   1  0.000 20.000     0  1800.010 upper
2   2  3.478 17.562     1  1800.015 lower
3   3  0.000 13.463     0  1800.031 upper
4   3 13.463 20.000     0  1800.031 lower
5   4  0.000 20.000     0  1800.064 lower
6   5  0.000  0.089     0  1800.084 lower
7   5  0.089 20.000     0  1800.084 upper
8   6  0.000 20.000     0  1800.094 upper
9   7  0.000  3.388     0  1800.105 upper
10  7  3.388 14.063     1  1800.105 lower
```

- Here `event` means: 1 for died during the period of study, and 0 means right censored.

Application From Week 3

- Run the model and summarize:

```
fit <- coxph(Surv(enter, exit, event) ~ ses, data=mort)
summary(fit)

  n= 1208, number of events= 276

            coef exp(coef) se(coef)      z Pr(>|z|)
sesupper -0.4795    0.6191   0.1207 -3.972 7.13e-05

         exp(coef) exp(-coef) lower .95 upper .95
sesupper    0.6191      1.615    0.4886    0.7844

Concordance= 0.56  (se = 0.015 )
Rsquare= 0.013   (max possible= 0.953 )
Likelihood ratio test= 15.81  on 1 df,   p=7.014e-05
Wald test            = 15.77  on 1 df,   p=7.134e-05
Score (logrank) test = 16.08  on 1 df,   p=6.075e-05
```

Recidivism Example from John Fox

- Start with the file `Rossi.txt` that contains data on 432 male prisoners observed for a year after release from prison (Rossi et al., 1980, Allison 1995).
- Available at `https://socserv.socsci.mcmaster.ca/jfox/Books/Companion-1E/Rossi.txt`.
- These data contain the following variables:
  - `week`: time of first arrest after release or censoring time.
  - `arrest`: event indicator, 1 for those arrested during the study time and 0 otherwise.
  - `fin`: a factor, `yes` if received financial aid after release and `no` otherwise (randomly assigned!).
  - `age`: age in years at time of release.
  - `race`: a factor, `black` and `other`.
  - `wexp`: a factor, `yes` if full-time work experience prior to incarceration, and `no` otherwise.

Recidivism Example from John Fox

- Continued...
  - `mar`: a factor, `married` if married at the time of release and `not married` otherwise.
  - `paro`: a factor, `yes` if on parole, and `no` otherwise.
  - `prio`: the number of prior convictions.
  - `educ`: a categorical variable with 2 (grade 6 or less), 3 (grades 6 through 9), 4 (grades 10 and 11), 5 (grade 12), or 6 (some post-secondary).
  - `emp1emp52`:a factor with `yes` if employed in the corresponding week of the study, `no` otherwise.

Recidivism Example from John Fox

- Read in the data from John's webpage:

```
url <- "http://socserv.mcmaster.ca/jfox/Books/Companion/data/Rossi.txt"
rossi <- read.table(url, header=TRUE)
head(rossi)
    week arrest fin age  race wexp         mar paro prio educ
  1 20        1  no  27 black   no not married  yes    3    3
  2 17        1  no  18 black   no not married  yes    8    4
  3 25        1  no  19 other  yes not married  yes   13    3
  4 52        0 yes  23 black  yes     married  yes    1    5
  5 52        0  no  19 other  yes not married  yes    3    3
```

- Now run a basic model:

```
recid1 <- coxph(Surv(week, arrest) ~ fin + age + race + wexp + mar + paro
    + prio, data=rossi)
```

Recidivism Example from John Fox

- Look at results:

```
summary(recid1)
                  coef  exp(coef) se(coef)      z Pr(>|z|)
finyes         -0.3794      0.684   0.1914 -1.983   0.0470
age            -0.0574      0.944   0.0220 -2.611   0.0090
raceother      -0.3139      0.731   0.3080 -1.019   0.3100
wexpyes        -0.1498      0.861   0.2122 -0.706   0.4800
marnot married  0.4337      1.543   0.3819  1.136   0.2600
paroyes        -0.0849      0.919   0.1958 -0.434   0.6600
prio            0.0915      1.096   0.0286  3.194   0.0014

               exp(coef) exp(-coef) lower .95 upper .95
finyes             0.684      1.461     0.470     0.996
age                0.944      1.059     0.904     0.986
raceother          0.731      1.369     0.399     1.336
wexpyes            0.861      1.162     0.568     1.305
marnot married     1.543      0.648     0.730     3.261
paroyes            0.919      1.089     0.626     1.348
prio               1.096      0.913     1.036     1.159
```

Recidivism Example from John Fox

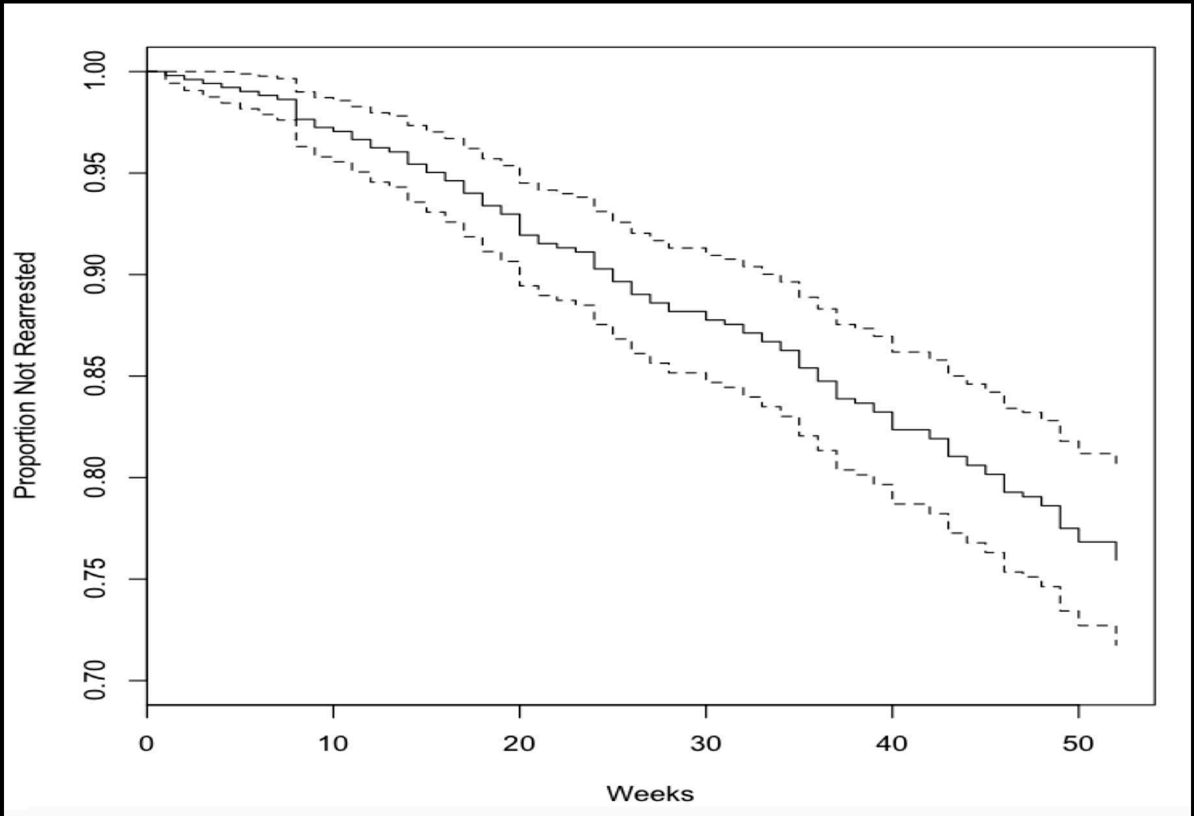
- Model quality:

```
Rsquare= 0.074 (max possible= 0.956 )
Likelihood ratio test= 33.3  on 7 df, p=0.0000236
Wald test            = 32.1  on 7 df, p=0.0000387
Score (logrank) test = 33.5  on 7 df,
p=0.0000211
```

Recidivism Example from John Fox

- Graphing the basic model:

```
plot(survfit(recid1), ylim=c(0.7, 1), xlab="Weeks",
    ylab="Proportion Not Rearrested")
```



Case Study

- **Reduction in Mortality Following Pediatric Rapid Response Team Implementation.** Nikoleta S. Kolovos, MD; Jeff Gill, PhD, MBA; Peter H. Michelson, MD, MS; Allan Doctor, MD; Mary E. Hartman, MD, MPH.
- **Objective:** To evaluate the effectiveness of a physician-led rapid response team (RRT) program on morbidity and mortality following unplanned admission to the pediatric intensive care unit (PICU).
- **Design:** Before-after study.
- **Setting:** Single center quarternary referral PICU.
- **Patients:** All unplanned PICU admissions from the ward from 2005-2011 at SLCH.
- **Interventions:** The dataset was divided into pre- and post-RRT groups for comparison. A therapeutic intensity scale (TIS) was employed to quantify resource use at the time of PICU admission.

Case Study

- **Measurements and Main Results:** A Cox proportional hazards model was used to identify the patient characteristics associated with mortality following unplanned PICU admission. Following RRT implementation, illness severity was reduced 28.1%, PICU length of stay (LOS) was less 19.8%, and mortality declined 22%. TIS ranged from 0 (1,110 cases) to 36 (1 case) with no differences in the post-RRT versus pre-RRT period. Relative risk of death following unplanned admission to the PICU after RRT implementation was 0.685.

- **Conclusions:** For children requiring unplanned admission to the PICU, RRT implementation reduced mortality, admission severity of illness and length of stay. With demonstrated increased resource use at the time of PICU admission, RRT implementation led to more proximal capture and aggressive intervention in the trajectory of a decompensating pediatric ward patient.

- Acronyms: PRISM III=Pediatric Risk of Mortality, version 3; PIM-2=Pediatric Index of Mortality, version 2; TIS=therapeutic intensity score.

Case Study, Patient Characteristics Before and After RRT Implementation

	Pre RRT (n= 1,097)	Post RRT (n=1,055)	p-value
Years	2005-2008	2008-2011	
Age in years, mean (SD)	7.1	7.4	NS
Female, %	45.7	43.6	NS
Ethnicity, %			
African American	33.3	30.3	NS
Caucasian	63.6	65.1	NS
Hispanic	0.9	1.6	NS
Other	2.2	2.9	NS
Admission day of the week, %			
Weekend	27.8	28.2	NS
Weekday	72.2	71.8	NS
PRISM III score, mean (SD)	3.17 (4.9)	2.26 (4.1)	0.001
PIM-2 score, mean (SD)	-4.59 (1.1)	-4.65 (1.1)	NS
TIS, mean (SD)	3.72 (5.1)	3.61 (5.1)	NS
PICU length of stay in days, median (IQR)	2 (1-5)	2 (1-4)	0.02
PICU mortality, %	4.9	3.8	0.001
Standardized mortality ratio	1.1	0.8	NS

Case Study

```
library(eha,survival)
tiss <- data.frame(tiss,"enter"=rep(0,nrow(tiss)),"exit" = 24*tiss$ICU_LOS_days)

surv.model <- coxph(Surv(enter,exit,dead) ~ PIM2 + strata(trachYN) + age_mo
                      + PrePostRRT + interven.scale + ContNeb
                      * venous_catheter + arterial_line + intubatedYN
                      + chest_tube + dialysis_catheter, data=tiss)
print(summary(surv.model))
surv.null <- coxreg(Surv(enter,exit,dead) ~ 1, data=tiss)
```

Stratification

- I stratified on the variable `trachYN` meaning that separate fits are created.
- Stratification involves segmenting a dataset along the categories of an explanatory variables such that there are separate baseline hazard functions are fit for each strata.
- Sometimes this is done because the strata are theoretically important and hypothesis tests are performed concerning whether the group distinctions matter.
- Sometimes this is done because the proportional hazards assumption does not hold otherwise for a variable in un-stratified form.
- Stratifying complicates likelihood estimation since separate partial likelihoods are required for each strata with a final product thereafter.

Case Study

Cox Regression Proportional Hazards Model Results for Time to Death

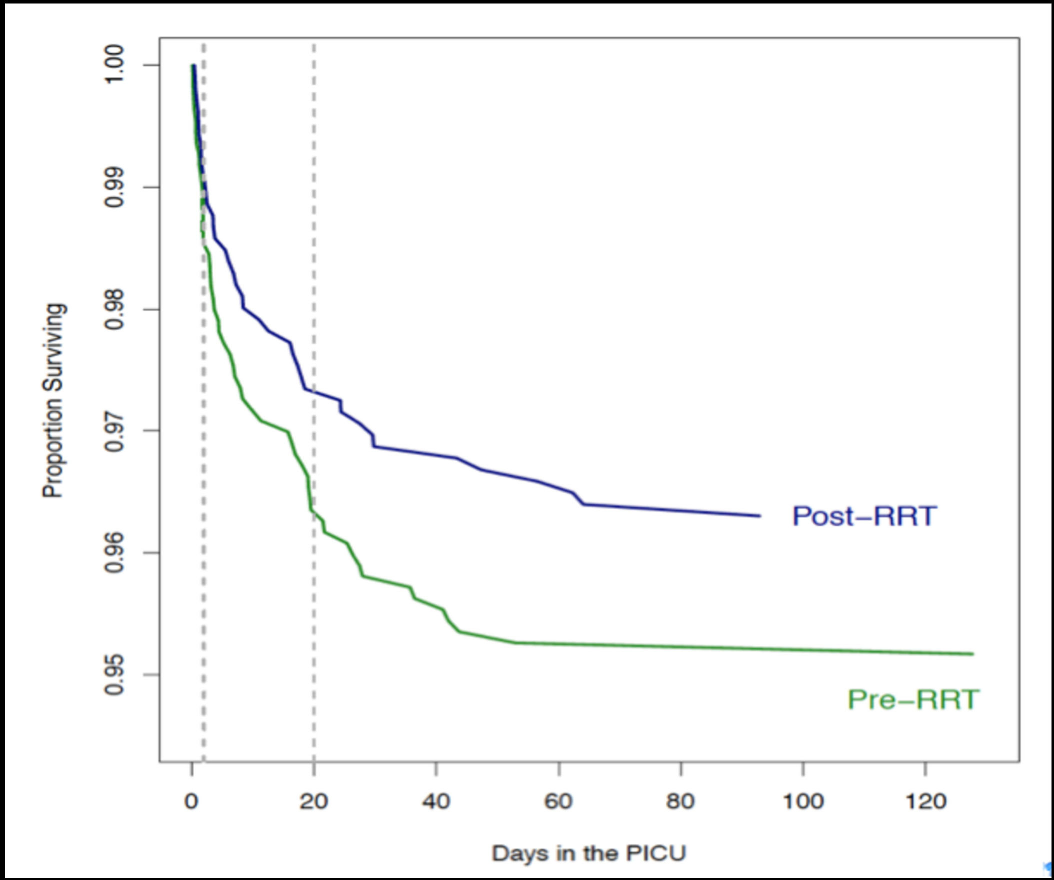
Covariate	Mean	Coefficient	Relative Risk	S.E.	1-tail p-value
PIM 2	-4.238	0.411	1.508	0.096	0.000
Age (months)	88.920	0.004	1.004	0.001	0.001
Pre Post RRT	1.442	-0.378	0.685	0.227	0.048
TIS scale	6.518	0.094	1.099	0.025	0.000
Continuous nebulization	0.034	0.301	1.351	0.740	0.342
Venous catheter	0.132	-0.199	0.819	0.342	0.280
Arterial line	0.195	-0.173	0.841	0.338	0.304
Intubated (Yes/No)	0.283	-0.508	0.601	0.337	0.066
Chest tube	0.014	-1.384	0.251	1.029	0.080
Dialysis/apheresis catheter	0.027	-1.090	0.336	0.628	0.041

Events 94, Total time at risk 271459, Max. log. Likelihood -455.6

LR test statistic 64.76, Degrees of freedom 11, Overall p-value 1.195e-09

Case Study, Survival for Unplanned PICU Admissions

The overall mortality rate for the population of unplanned PICU admissions fell significantly following RRT implementation (4.9% to 3.8%, $p < 0.001$). This difference in mortality was evident by PICU admission day 2, and was statistically reliable by day 20.



Assignment

- Problem Set 1 from Brad Jones webpage for the BSJ book.
- Rerun the the U.N. peacekeeping missions analysis from the chapter in `R`. Brad Jones has `Stata` code and data to get started.
- Note the homework requires you to use a link test. This re-estimates a Cox model where the predicted values of the tested model and their squared values are used as covariates along with constant. The (relatively weak) test is a Wald test of the significance of the squared predicted value, where being in the tail is a failure.