Survival Analysis

Lecture 8

Marta Fiocco[1,2] & Hein Putter [1]

(1) Department of Medical Statistics and Bioinformatics
Leiden University Medical Center
(2) Mathematical Institute Leiden University

Outline

The objective

- We are studying survival data
- Time to an event plus status variable
- We have discussed ways of estimating survival curves, hazards
- We have discussed the problem of testing whether two (or more) survival curves are equal (log-rank test)
- We would like to *quantify* the effect of covariates on survival
  - We would like to have an effect size, not only a P-value, when comparing two survival curves
  - We would like to study the effect of continuous covariates, like age
  - We would like to look at several covariates at the same time

Regression in general

The basic problem

$$Z_1, Z_2, \ldots, Z_p \Rightarrow Y$$

- Interest in the relation between $Z_1, Z_2, \ldots, Z_p$ and Y
- $Z_1, Z_2, \ldots, Z_p$:
  - Predictors
  - Explanatory variables
  - "Independent" variables
  - Covariates
  - Prognostic factors
- Y:
  - Response variable
  - Dependent variable
  - Outcome variable

Regression models

Statistical relationship

- The statistical relationship between $Z_1, \ldots, Z_p$ and Y can be studied by means of a regression model
- The type of regression model depends on the type of the distribution of Y given the Z's
  - Y continuous (approximately normal): linear regression
  - Y dichotomous: logistic regression model
  - Y Poisson (count): Poisson regression model

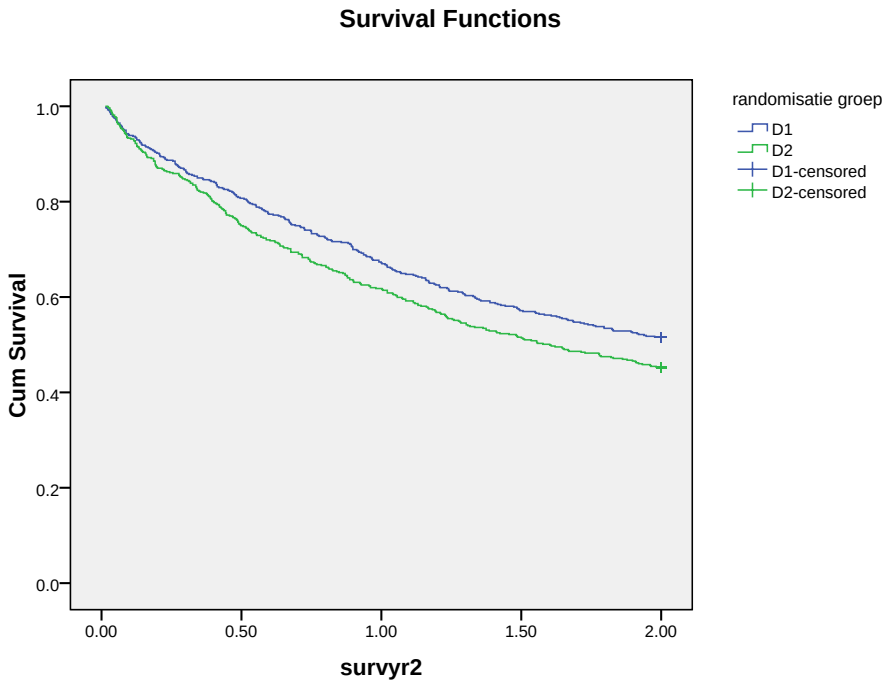
Regression models for survival data

- Y survival data: cannot use linear regression or logistic regression
- Special regression models for survival data
  - Accelerated failure time model
  - Poisson regression
  - Cox's proportional hazards model
  - The last one is by far the most popular

Motivation

- Recall the D1/D2 study

Two survival curves

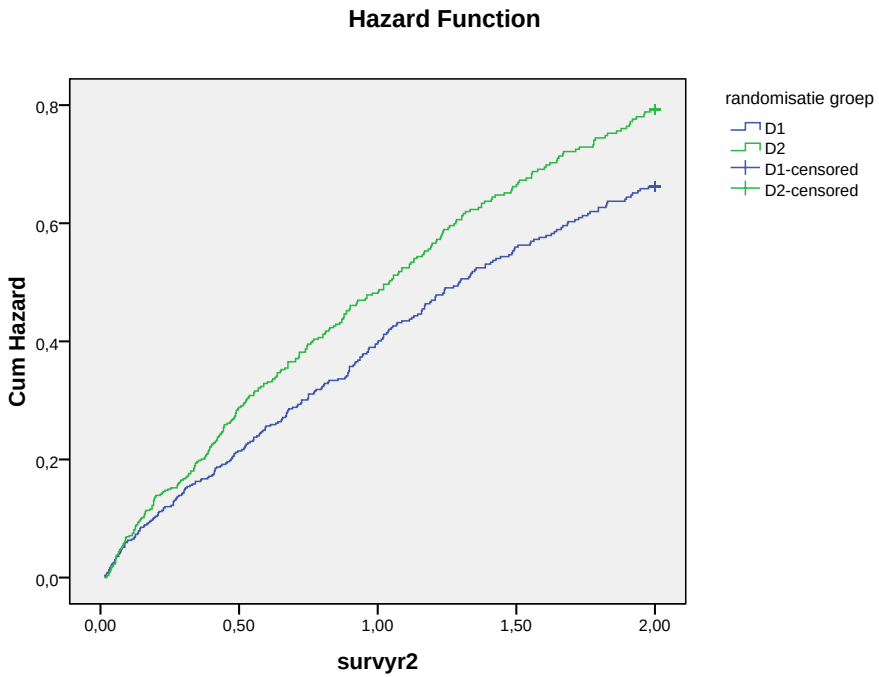


Question

- We have already tested equality of these survival curves
- We tested equality at 2 years, which gave P=0.038
- Now we would like to quantify how much better D1 is than D2

A look at the hazards

The two cumulative hazards



$$h(t) \Rightarrow H(t) = \int_0^t h(s)\,ds \Rightarrow S(t) = e^{-H(t)}$$

Proportional hazards

- $h_1(t)$: the hazard rate of D1
- $h_2(t)$: the hazard rate of D2
- $HR(t) = \frac{h_2(t)}{h_1(t)}$ is the ratio of these hazards, the hazard ratio
- Since both $h_1(t)$ and $h_2(t)$ depend on time, in principle $HR(t)$ depends on time

Proportional hazards assumption

- $HR(t)$ does not depend on time, but is a constant HR

$$\frac{h_2(t)}{h_1(t)} = HR$$

Other notation

Same story

- Z: covariate, treatment
  - $Z = 0$ corresponds to D1-dissection
  - $Z = 1$ corresponds to D2-dissection
  - D1 is called *reference category*
- $h_0(t)$: hazard rate corresponding to reference category ($Z = 0$), also called baseline hazard
- Model:

$$h(t \mid Z) = h_0(t) \exp(\beta Z)$$

What does it mean?

Proportional hazards model

$$h(t \mid Z) = h_0(t) \exp(\beta Z)$$

- $Z = 0$: $h(t \mid Z = 0) = h_0(t) \exp(\beta \cdot 0) = h_0(t)$
- $Z = 1$: $h(t \mid Z = 1) = h_0(t) \exp(\beta \cdot 1) = h_0(t) \exp(\beta)$
- Hazard rate of D1-dissection ($Z = 0$): $h_0(t)$ (previously called $h_1(t)$
- Hazard rate of D2-dissection ($Z = 1$): $h_0(t) \exp(\beta)$ (previously called $h_2(t)$

- The ratio of these hazards, the hazard ratio is given by

$$\frac{h(t \mid Z=1)}{h(t \mid Z=0)} = \frac{h_0(t)\exp(\beta)}{h_0(t)} = \exp(\beta)$$

- $\exp(\beta)$ is the hazard ratio (does not depend on time), β is the log-hazard ratio

D1/D2 trial

Table with estimates

- Recall:
  - $Z = 0$: D1-dissection
  - $Z = 1$: D2-dissection

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
randgr	,187	,085	4,863	1	,027	1,206	1,021	1,425

The Cox model in general

- It has become the most used procedure for modeling the relationship of covariates to a survival or other censored outcome
- X: time to some event
- $\mathbf{Z}_j$: vector of covariates (risk factors) for the j^{th} individual at time t which may affect the survival distribution of X; covariates can be fixed or vary over time (ex repeated laboratory test); in the latter case the notation is $\mathbf{Z}_j(t)$
- data consist of $(T_j, \delta_j, \mathbf{Z}_j(t))$

The Cox model in general

- The Cox model specifies the hazard $h(t|\mathbf{Z})$ for individual i as

$$h(t|\mathbf{Z}) = h_0(t)\exp(\beta^\top \mathbf{Z}) = h_0(t)\exp(\sum_{k=1}^{p} \beta_k Z_k)$$

- h_0: baseline hazard rate; β: parameter vector of coefficients
- It is a semi-parametric model
  - A parametric form is assumed for the covariate effect
  - The baseline hazard is non-parametric
- Event rate $h(t|\mathbf{Z})$ must be positive $h_0(t)$ positive
  - $\exp(\beta^\top \mathbf{Z})$ ensures that $h(t|\mathbf{Z})$ is positive (as long as $h_0(t)$ is)

Proportional hazards

- The hazard ratio for two subjects with fixed covariates vectors $\mathbf{Z}_i$ and $\mathbf{Z}_j$

$$\frac{h(t|\mathbf{Z}_i)}{h(t|\mathbf{Z}_j)} = \frac{h_0(t)\exp(\sum_{k=1}^{p} \beta_k \mathbf{Z}_{ik})}{h_0(t)\exp(\sum_{k=1}^{p} \beta_k \mathbf{Z}_{jk})}$$

- Is constant over time
- The hazards are proportional

- Relation between covariate and hazard

$$h(t \mid Z) = h_0(t) \exp(\beta Z)$$

- Relation between covariate and *cumulative* hazard

$$H(t \mid Z) = H_0(t) \exp(\beta Z)$$

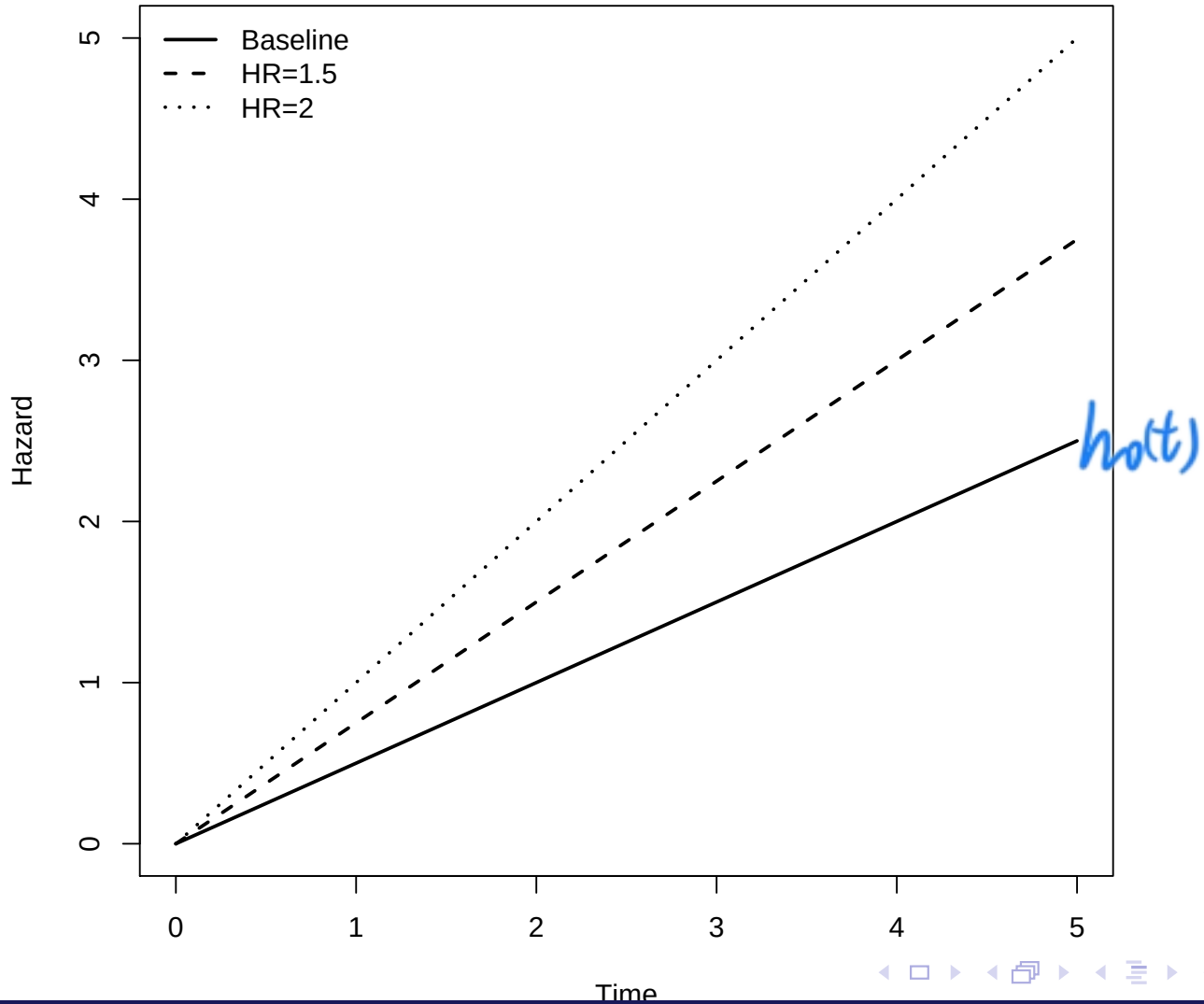
- Relation between covariate and survival function

$$\begin{aligned} S(t \mid Z) = \exp(-H(t \mid Z)) &= \exp(-H_0(t) \exp(\beta Z)) \\ &= S_0(t)^{\exp(\beta Z)} \end{aligned}$$

$$e^{-H(t)} = e^{-H_0 e^{\beta Z}} = (e^{-H_0})^{e^{\beta Z}} = S_0^{e^{\beta Z}}$$

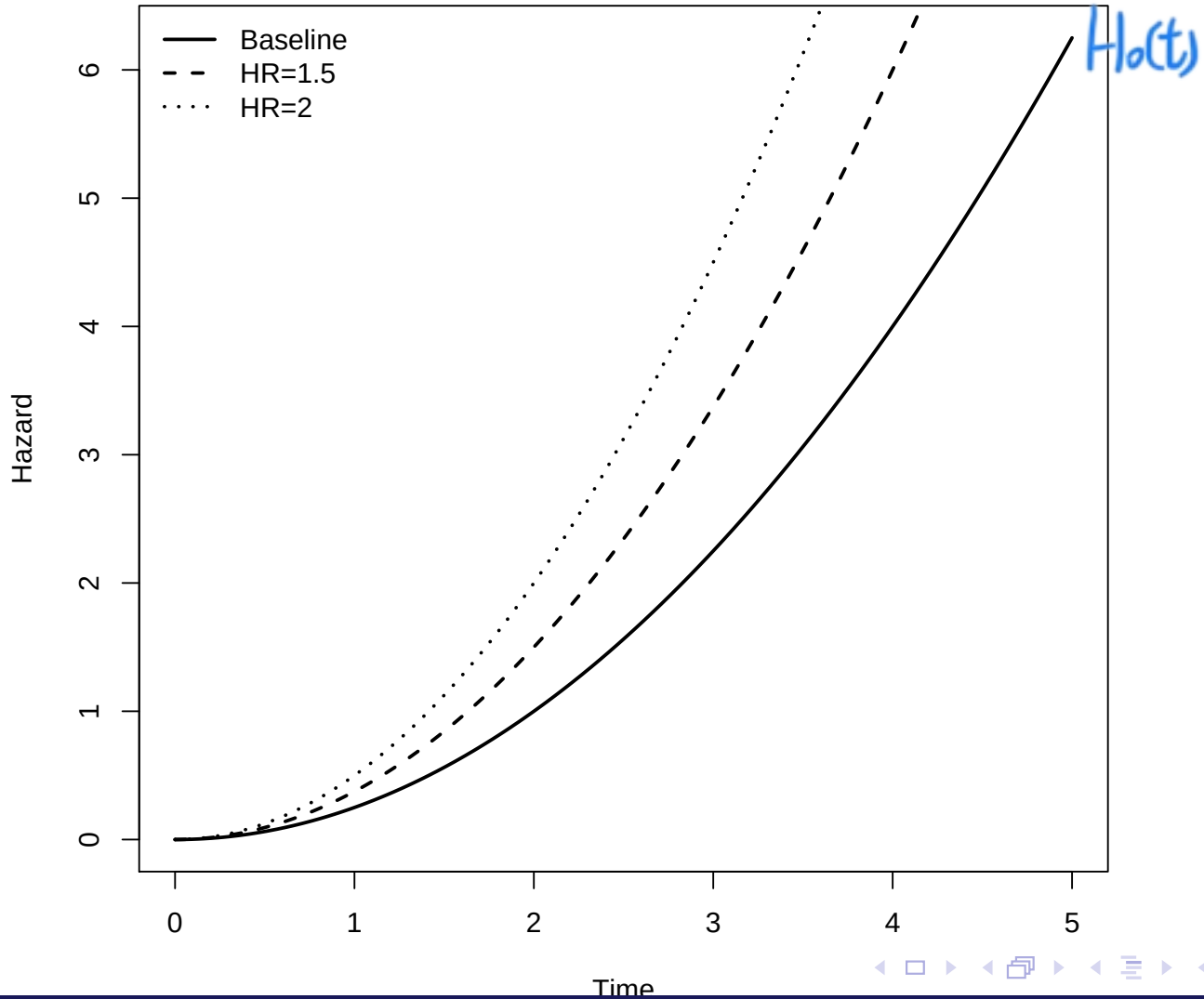
Relation illustrated (Weibull(2,2))

Hazards



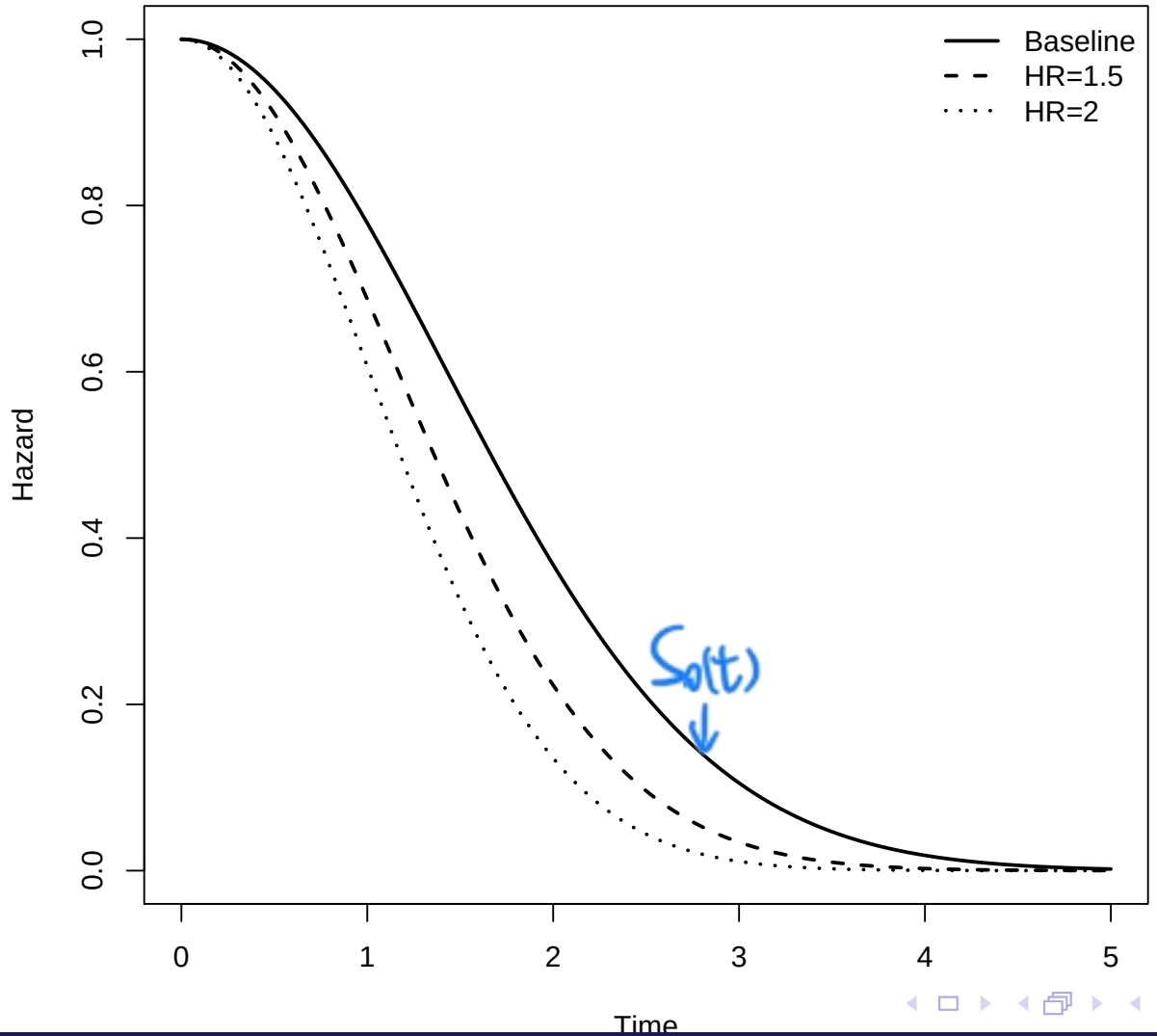
Relation illustrated (Weibull(2,2)

Cumulative hazards



Relation illustrated (Weibull(2,2)

Survival functions



Coding Covariates

- Numerical (quantitative) or categorical (qualitative) covariates (explanatory variables)
- Numerical
  - Blood pressure, blood glucose levels, age or waiting time until a transplant
- Categorical
  - Gender, smoking behavior, stage of disease, presence/absence of something, treatment yes/no
- Categorical variables in regression analysis: care needs to be taken in the coding and interpretation
- Different ways of coding categorical variables

Coding Covariates

- Dichotomous (for instance gender): obvious way is to code one of the genders as 0, the other as 1
- Coding is arbitrary
- Interpretation of the results will depend on the way the coding is done

Data Section 1.5

```
> data(btrial)
> head(btrial)
```

```
  time death im
1   19     1  1
2   25     1  1
3   30     1  1
4   34     1  1
5   37     1  1
6   46     1  1
```

- Cox model with immunoperoxidase status (`im`) as single covariate
- Coded as 1=negative, 2=positive

```
> table(btrial$im)
```

```
 1  2
36  9
```

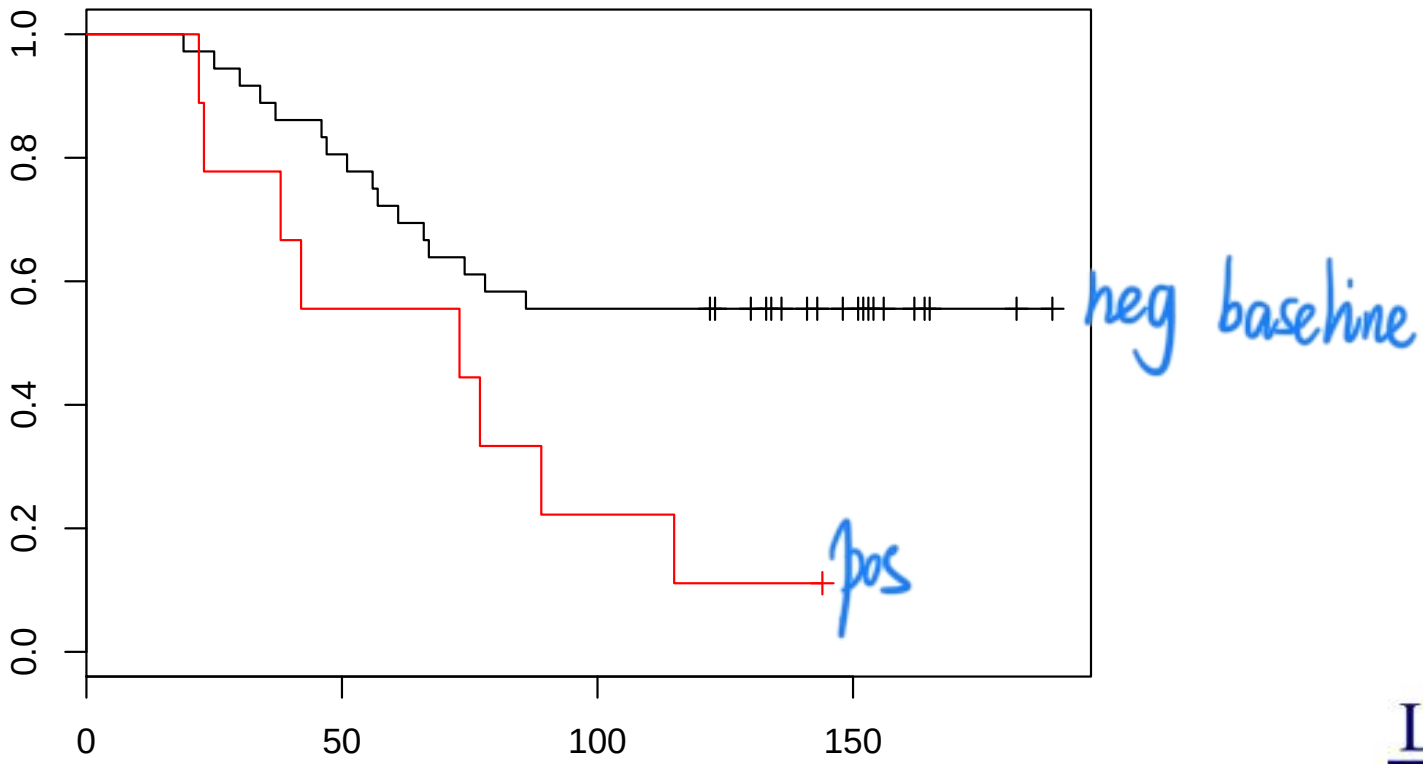
- Define

$$Z_1 = \begin{cases} 1 & \text{if immunoperoxidase positive (im+)} \\ 0 & \text{if immunoperoxidase negative (im-)} \end{cases}$$

```
> btrial$z1 <- ifelse(btrial$im==2,1,0)
> table(btrial$z1,btrial$im)

    1  2
0  36  0
1   0  9
```

A plot

```
> plot(survfit(Surv(time,death) ~ im, data=btrial), col=1:2)
```



Model

- The Cox model specifies

$$h(t \mid Z_1) = h_0(t) \exp(\beta Z_1)$$

- $\exp(\beta)$: hazard ratio of patient being im+ relative to the patient being im-

Software

Function *coxph* from the *survival* package

```
> c1 <- coxph(Surv(time,death) ~ z1, data=btrial)
> c1

Call:
coxph(formula = Surv(time, death) ~ z1, data = btrial)


      coef exp(coef) se(coef)     z     p
z1 0.98       2.66     0.435 2.25 0.024

Likelihood ratio test=4.45  on 1 df, p=0.035  n= 45, number of events=
```

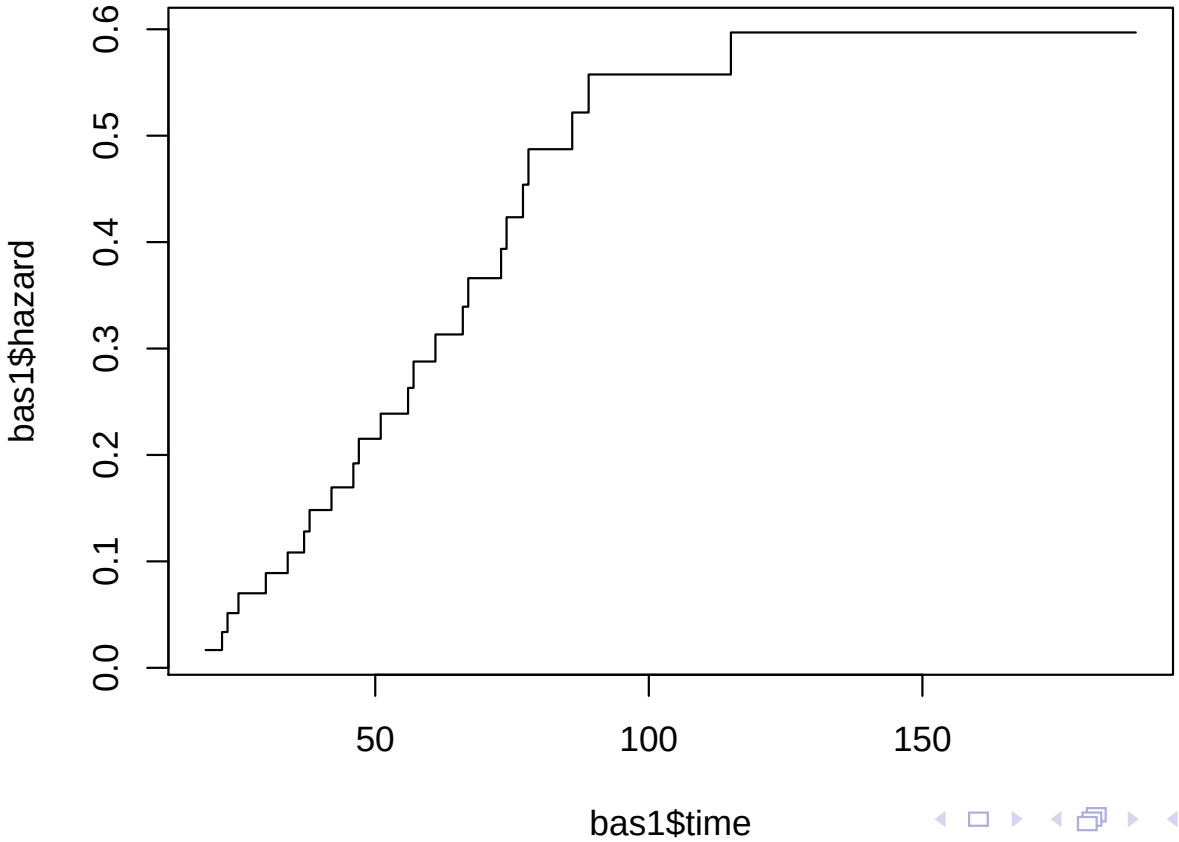
- Hazard ratio for an im+ patient relative to an im- patient is $\exp(0.98) = 2.67$
- Patient who is im+ has 2.67 times higher risk of dying than an im- patient

More detail with summary()

```
> summary(c1)

Call:
coxph(formula = Surv(time, death) ~ z1, data = btrial)

  n= 45, number of events= 24

     coef exp(coef) se(coef)     z Pr(>|z|)
z1 0.9802    2.6650   0.4349 2.254   0.0242 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

   exp(coef) exp(-coef) lower .95 upper .95
z1     2.665     0.3752     1.136      6.25

Rsquare= 0.094   (max possible= 0.976 )
Likelihood ratio test= 4.45  on 1 df,   p=0.03498
Wald test            = 5.08  on 1 df,   p=0.0242
Score (logrank) test = 5.49  on 1 df,   p=0.01908
```

The cumulative baseline hazard

```
> bas1 <- basehaz(c1, centered=FALSE)
> plot(bas1$time, bas1$hazard, type="s")
```



Reversing reference category

- Let's use a different coding of immunoperoxidase status
- $Z_2 = 0$: im+
- $Z_2 = 1$: im-

reverse

$$h(t \mid Z_2) = \tilde{h}_0(t) \exp(\tilde{\beta} Z_2)$$

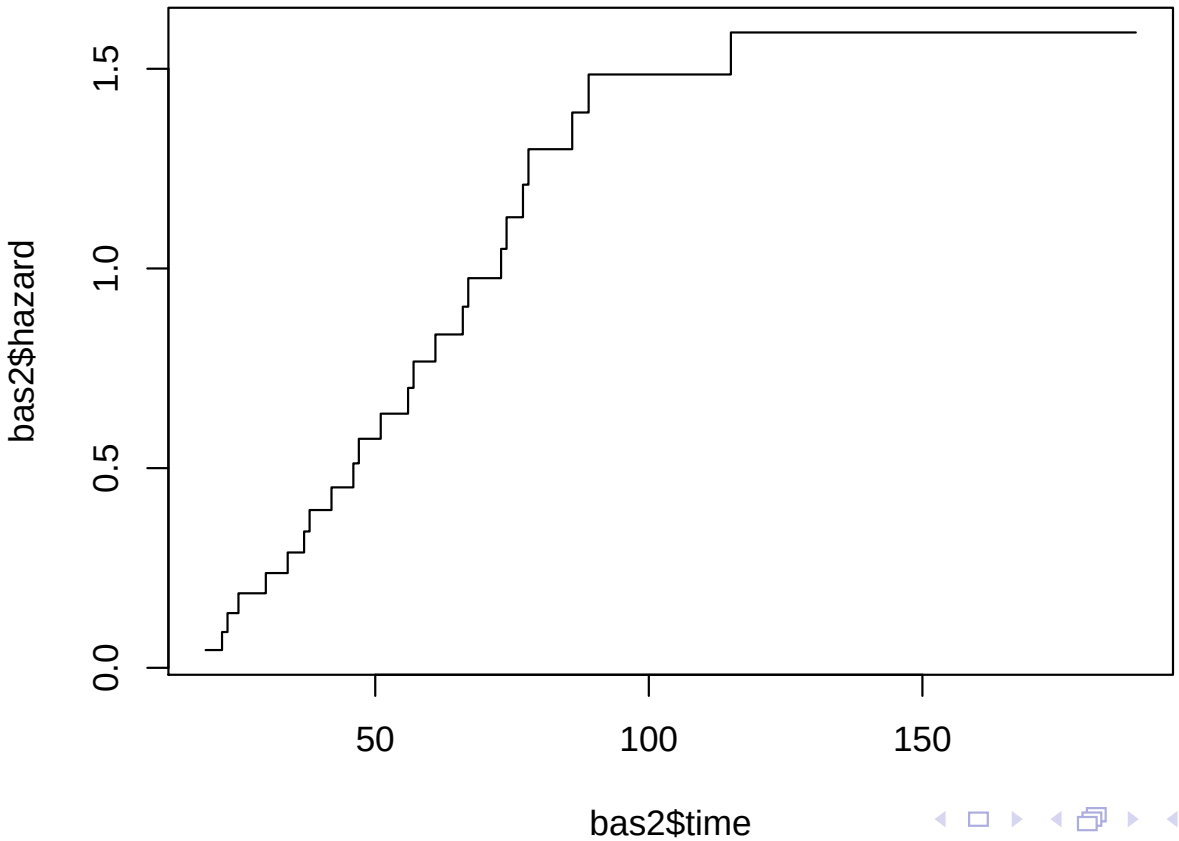
- $h(t \mid Z_2 = 0) = \tilde{h}_0(t)$
- Since $Z_2 = 0$ is the same as $Z_1 = 1$, we have $\tilde{h}_0(t) = h_0(t) \exp(\beta)$
- $h(t \mid Z_2 = 1) = \tilde{h}_0(t) \exp(\tilde{\beta})$
- Since $Z_2 = 1$ is the same as $Z_1 = 0$, we have $\tilde{h}_0(t) \exp(\tilde{\beta}) = h_0(t)$, so $\tilde{h}_0(t) = h_0(t) \exp(-\tilde{\beta})$
- $\exp(\tilde{\beta}) = \exp(-\beta) = \frac{1}{\exp(\beta)}$; $\tilde{\beta} = -\beta$

Cox with z2

```
> btrial$z2 <- ifelse(btrial$im==1,1,0)
> table(btrial$z2,btrial$im)

    1  2
0   0  9
1  36  0

> coxph(Surv(time,death) ~ z2, data=btrial)

Call:
coxph(formula = Surv(time, death) ~ z2, data = btrial)

      coef exp(coef) se(coef)     z     p
z2 -0.98      0.375    0.435 -2.25 0.024

Likelihood ratio test=4.45  on 1 df, p=0.035  n= 45, number of events=
```

The cumulative baseline hazard

```
> bas2 <- basehaz(c2, centered=FALSE)
> plot(bas2$time, bas2$hazard, type="s")
```



With factors in R

```
> btrial$im12 <- factor(btrial$im,levels=1:2,labels=c("neg","pos"))
> table(btrial$im12)

neg pos
 36   9

> btrial$im21 <- factor(btrial$im,levels=2:1,labels=c("pos","neg"))
> table(btrial$im21)

pos neg
  9  36
```

Cox with these factors

```
> coxph(Surv(time,death) ~ im12, data=btrial)

Call:
coxph(formula = Surv(time, death) ~ im12, data = btrial)


        coef exp(coef) se(coef)    z     p
im12pos 0.98      2.66    0.435 2.25 0.024

Likelihood ratio test=4.45  on 1 df, p=0.035  n= 45, number of events=
> coxph(Surv(time,death) ~ im21, data=btrial)

Call:
coxph(formula = Surv(time, death) ~ im21, data = btrial)


         coef exp(coef) se(coef)     z     p
im21neg -0.98     0.375    0.435 -2.25 0.024

Likelihood ratio test=4.45  on 1 df, p=0.035  n= 45, number of events=
```

The design matrix

```
> btrial[35:38,]
   time death im im12 im21 z1 z2
35  182     0  1  neg  neg  0  1
36  189     0  1  neg  neg  0  1
37   22     1  2  pos  pos  1  0
38   23     1  2  pos  pos  1  0
> mm1 <- model.matrix(coxph(Surv(time,death) ~ im12, data=btrial))
> mm1[35:38,,drop=FALSE]
   im12pos
35       0
36       0
37       1
38       1
> mm2 <- model.matrix(coxph(Surv(time,death) ~ im21, data=btrial))
> mm2[35:38,,drop=FALSE]
   im21neg
35       1
36       1
37       0
38       0
```

More than two categories

- Data Section 1.8 (used for the trend test last time)

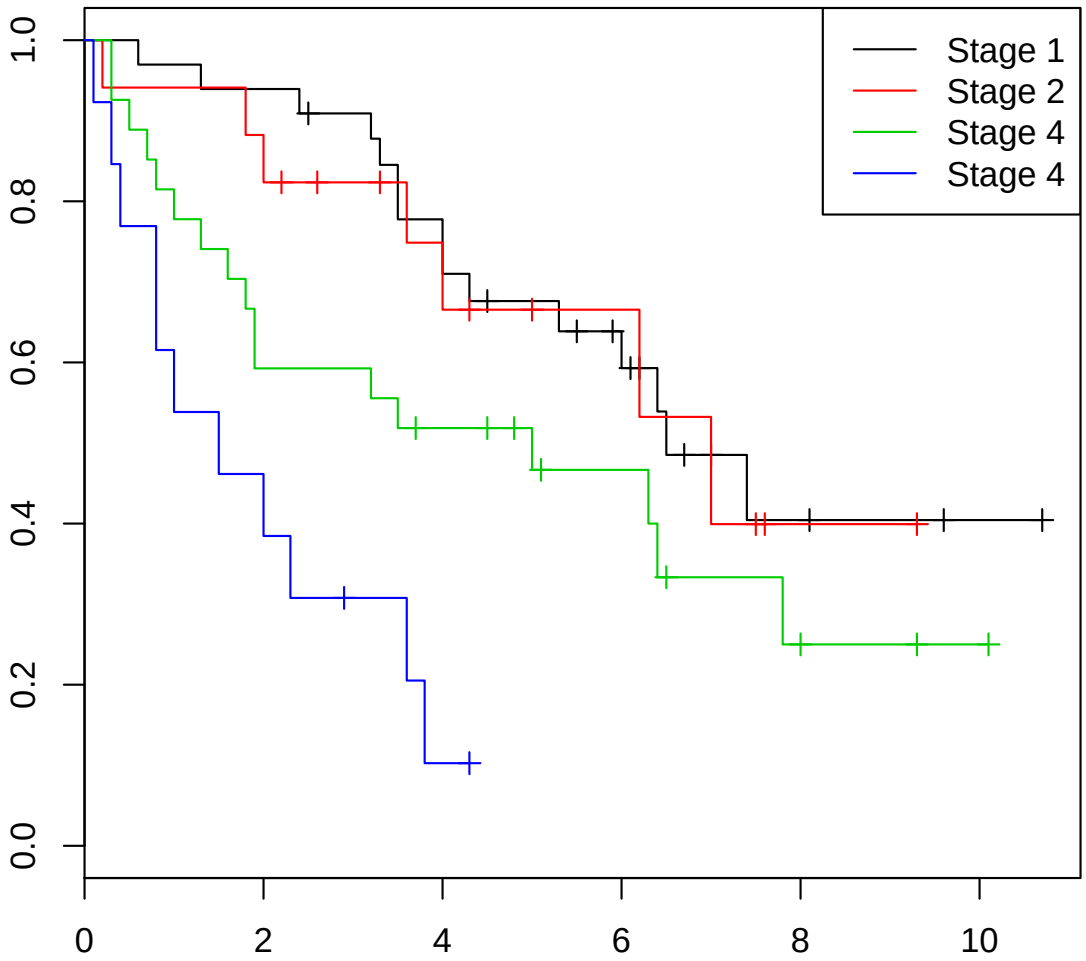
```
> data <- larynx
> head(data)
```

```
  stage time age diagyr delta
1     1  0.6  77     76     1
2     1  1.3  53     71     1
3     1  2.4  45     71     1
4     1  2.5  57     78     0
5     1  3.2  58     74     1
6     1  3.2  51     77     0
```

```
> table(data$stage)
```

```
 1  2  3  4
33 17 27 13
```

```
> plot(survfit(Surv(time,delta) ~ stage, data=larynx), col=1:4)
>  legend("topright",c("Stage 1","Stage 2","Stage 4","Stage 4"),
 + lwd=1,col=1:4)
```



- Coding of the variable stage of disease
- Fit a proportional hazards regression with only stage as covariate in the model
- Stage has four levels
- Construct the dummy (or indicator) variables
- $Z_1 = 1$: if the patient is in stage II, 0 otherwise
- $Z_2 = 1$: if the patient is in stage III, 0 otherwise
- $Z_3 = 1$: if the patient is in stage IV, 0 otherwise
- Patient with stage I cancer is the referent group ($Z_1 = Z_2 = Z_3 = 0$)
- Model

$$h(t \mid Z) = h_0(t)\exp\{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3\}$$

- We have to declare covariate `stage` as categorical

```
> larynx$stage <- factor(larynx$stage)
> table(larynx$stage)

 1  2  3  4
33 17 27 13
```

- R-code `coxph` to fit the model

```
> res <- coxph(Surv(time, delta) ~ stage, data = larynx)
> res

Call:
coxph(formula = Surv(time, delta) ~ stage, data = larynx)


          coef exp(coef) se(coef)     z        p
stage2 0.0648      1.07    0.458 0.141 8.9e-01
stage3 0.6148      1.85    0.355 1.731 8.3e-02
stage4 1.7349      5.67    0.419 4.137 3.5e-05

Likelihood ratio test=16.5  on 3 df, p=0.000902  n= 90, number of even
```

Interpretation

- Estimated HR of death for Stage II disease with respect to Stage I disease: $\exp(0.0648) = 1.07$
- Estimated HR of death for Stage III disease with respect to Stage I disease: $\exp(0.6148) = 1.85$
- Estimated HR of death for Stage IV disease with respect to Stage I disease: $\exp(1.7349) = 5.67$
- HR of death for Stage IV with respect to stage III $\exp(1.7349 - 0.6148) = 3.065$

Continuous covariates

- Code the variable as a single covariate: $Z_4 = age$ (in years)
- Hazard ratio of an event for an individual of age x years compared to an individual of age $x - 1$ years
- Hazard ratio of the event for an individual 10 years older than another individual: $\exp(10 \cdot \beta)$
- Model for larynx data with risk factors *stage* of the disease and *age*

$$h(t \mid \mathbf{Z}) = h_0(t)\exp(\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4)$$

- The model fitted with *coxph*

```
> res <- coxph(Surv(time, delta) ~ stage + age, data = larynx)
> res

Call:
coxph(formula = Surv(time, delta) ~ stage + age, data = larynx)


         coef exp(coef) se(coef)     z        p
stage2 0.1400      1.15   0.4625 0.303 7.6e-01
stage3 0.6424      1.90   0.3561 1.804 7.1e-02
stage4 1.7060      5.51   0.4219 4.043 5.3e-05
age    0.0190      1.02   0.0143 1.335 1.8e-01

Likelihood ratio test=18.3  on 4 df, p=0.00107  n= 90, number of event
```

Interpretation

- the relative risk for a 50-year-old patient compared to a 40-year-old (with the same disease stage) is $\exp(10 \cdot \beta_4) = \exp(10 \cdot 0.0190) = 1.21$
- Or: a 50-year-old patient has a 1.21 times greater risk of dying than a 40-year-old patient with the same disease stage

The design matrix

```
> larynx[c(1,34,51,78),]

   stage time age diagyr delta
1      1  0.6  77     76     1
34     2  0.2  86     74     1
51     3  0.3  49     72     1
78     4  0.1  65     72     1

> model.matrix(res)[c(1,34,51,78),]

   stage2 stage3 stage4 age
1       0      0      0  77
34      1      0      0  86
51      0      1      0  49
78      0      0      1  65
```

Interactions

- When there are interactions, the coding of the covariates becomes even more important
- Computer exercise this afternoon

Partial Likelihoods for Distinct-Event Time Data

- $t_1 < t_2 < \ldots < t_D$: ordered event times
- $Z_{(i)k}$: *k-th* covariate associated with the individual whose failure time is t_i
- $R(t_i)$: risk set at time t_i, (set of all individuals who are still under study at a time just prior to t_i
- Partial likelihood based on the hazard function is given by

$$L(\beta) = \prod_{i=1}^{D} \frac{\exp(\sum_{k=1}^{p} \beta_k Z_{(i)k})}{\sum_{j \in R(t_i)} \exp(\sum_{k=1}^{p} \beta_k Z_{jk})}$$

- (Make computations how to derive $L(\beta)$ on the blackboard)

- Log partial likelihood

$$LL(\beta) = \log\left(\prod_{i=1}^{D} \frac{\exp(\sum_{k=1}^{p} \beta_k Z_{(i)k})}{\sum_{j \in R(t_i)} \exp(\sum_{k=1}^{p} \beta_k Z_{jk})}\right)$$

- We can also write it as

$$LL(\beta) = \sum_{i=1}^{D}\sum_{k=1}^{p} \beta_k Z_{(i)k} - \sum_{i=1}^{D} \log\left[\sum_{j \in R(t_i)} \exp\left(\sum_{k=1}^{p} \beta_k Z_{jk}\right)\right]$$

- Estimate the parameters β by maximizing the partial likelihood or the log-likelihood
- Score equations: $U_h(\beta) = \partial LL(\beta)/\partial\beta_h$, $h = 1, \ldots, p$
- Information matrix $\mathcal{I}(\beta) = [\mathcal{I}_{gh}(\beta)]_{p \times p}$
- Show how to derive $U_h(\beta)$ and $\mathcal{I}(\beta)$

- The (partial) maximum likelihood estimates $\hat{\beta}_1, \ldots \hat{\beta}_p$ are found by solving the set of p nonlinear equations $U_h(\beta) = 0$
- The log-likelihood does **not** depend on the baseline hazard rate $h_0(t)$, inference may be made on the effects of the explanatory variables without knowing $h_0(t)$

Tests for the regression parameters

- $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots \hat{\beta}_p)$: partial MLE of $\boldsymbol{\beta}$;
- $\mathcal{I}(\boldsymbol{\beta})$: $p \times p$ information matrix evaluated at $\boldsymbol{\beta}$
- test of the global hypothesis: $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$

Wald's test

- Based on the asymptotic normality of the (partial) maximum likelihood estimates: $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \mathcal{I}^{-1}(\hat{\boldsymbol{\beta}}))$

$$X^2_W = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \mathcal{I}(\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \sim \chi^2_p$$

Tests for the regression parameters

Likelihood ratio test (LRT) test

$$X^2_{LR} = 2[LL(\hat{\beta}) - LL(\beta_0)] \sim \chi^2_p$$

Score test

- Based on the efficient scores (first derivative of the log partial likelihood) $\mathbf{U}(\beta) = (U_1(\beta), \ldots, U_p(\beta))^\top$

$$X^2_{SC} = \mathbf{U}(\beta_0)^\top \mathcal{I}^{-1}(\beta_0)\mathbf{U}(\beta_0) \sim \chi^2_p$$

Breast cancer data Section 1.5 (recall)

```
> data(btrial)
> head(btrial)
  time death im
1   19     1  1
2   25     1  1
3   30     1  1
4   34     1  1
5   37     1  1
6   46     1  1
> res <- coxph(Surv(time,death)~im, data=btrial)
> res

Call:
coxph(formula = Surv(time, death) ~ im, data = btrial)


   coef exp(coef) se(coef)    z     p
im 0.98      2.66    0.435 2.25 0.024

Likelihood ratio test=4.45  on 1 df, p=0.035  n= 45
```

Software

- Wald's test and score test can be found with *summary*

```
> summary(res)

Call:
coxph(formula = Surv(time, death) ~ im, data = btrial)

  n= 45, number of events= 24

     coef exp(coef) se(coef)     z Pr(>|z|)
im 0.9802    2.6650   0.4349 2.254   0.0242 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

   exp(coef) exp(-coef) lower .95 upper .95
im     2.665     0.3752     1.136      6.25

Rsquare= 0.094   (max possible= 0.976 )
Likelihood ratio test= 4.45  on 1 df,   p=0.03498
Wald test            = 5.08  on 1 df,   p=0.0242
Score (logrank) test = 5.49  on 1 df,   p=0.01908
```

Software

- Test the hypothesis $H_0 : \beta = 0$ by using the three tests seen

```
> res1 <- summary(res)
> res1$logtest

      test          df     pvalue
4.44629201 1.00000000 0.03497711

> res1$waldtest

      test          df     pvalue
5.08000000 1.00000000 0.02420219

> res1$sctest

      test          df     pvalue
5.49427024 1.00000000 0.01907889
```

- Check the computations by using the `anova` function

Software

```
> anova(res)

Analysis of Deviance Table
 Cox model: response is Surv(time, death)
Terms added sequentially (first to last)

      loglik  Chisq Df Pr(>|Chi|)
NULL -83.744
im   -81.521 4.4463  1    0.03498 *
```

- From *anova* we obtain $LL(\hat{\beta})$ and $LL(\beta_0)$
- LRT:

$$X^2_{LR} = 2[LL(\hat{\beta}) - LL(\beta_0)] = 2(-81.521 - (-83.744)) = 4.446$$

```
> 1-pchisq(4.446,1)
[1] 0.0349831
```

Partial likelihoods when ties are present

- The partial likelihood for the Cox model is developed under the assumption of continuous data
- Real data sets often contain tied event times
- $t_1, \ldots, t_D$: D distinct, ordered, event times
- d_i: number of deaths at t_i;
- D_i: set of all individuals who die at time t_i
- $\mathbf{s}_i = \sum_{j \in D_i} \mathbf{Z}_j$ (sum of the vectors $\mathbf{Z}_j$ over all individuals who die at t_i)
- R_i: set of all individuals at risk just prior to t_i
- Three different algorithms are commonly used to address this problem

Partial Likelihoods When Ties Are Present

Breslow approximation

- Simplest to write down, easiest to program
  - Default method in most packages (but not in *survival* package!!)
  - Solution is the least accurate but the method is fast (see (8.4.1) in your book)

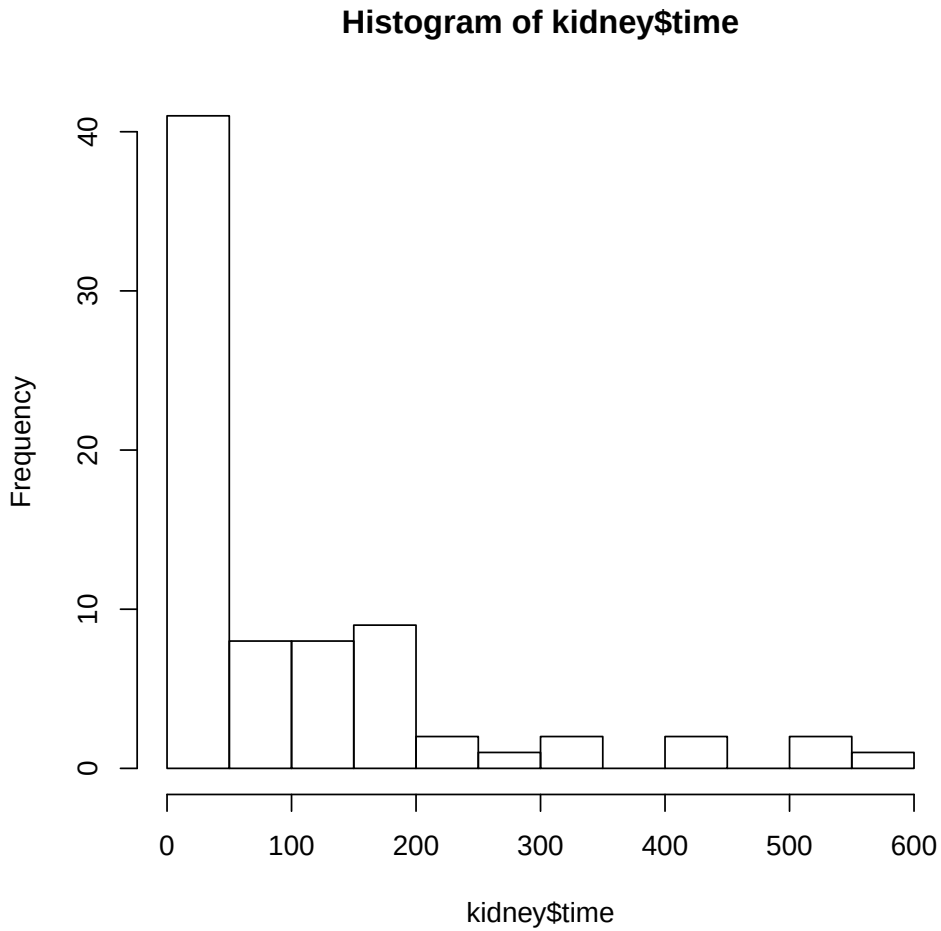
Efron approximation

- Quite accurate unless the proportion of ties relative to the size of the risk set is extremely large;
- As fast as the Breslow'method; default option in *survival* package (see (8.4.2))

Exact partial likelihood

- Exact method involves exhaustive enumeration of all possible risk sets at each tied death time
  - Can require a prohibitive amount of computation time
  - It computes the exact partial likelihood (`method="exact"` in `coxph`)
- Use data Section 1.4 to compare estimates obtained with different approximation

```
> data(kidney)
> names(kidney)
[1] "time"   "delta" "type"
> table(kidney$type)

 1  2
43 76
> hist(kidney$time)
```


Histogram of kidney$time
Frequency
0
10
20
30
40
0
100
200
300
400
500
600
kidney$time

Example

```
> coxph(formula = Surv(time, delta) ~ type, data = kidney,
method = "efron")

Call:
coxph(formula = Surv(time, delta) ~ type, data = kidney,
 method = "efron")

       coef exp(coef) se(coef)     z    p
type -0.613     0.542    0.398 -1.54 0.12

Likelihood ratio test=2.41  on 1 df, p=0.121  n= 119

> coxph(formula = Surv(time, delta) ~ type, data = kidney,
method = "breslow")

Call:
coxph(formula = Surv(time, delta) ~ type, data = kidney, method = "bre
       coef exp(coef) se(coef)     z    p
type -0.618     0.539    0.398 -1.55 0.12

Likelihood ratio test=2.45  on 1 df, p=0.118  n= 119
```

Example

```
> coxph(formula = Surv(time, delta) ~ type, data = kidney,
method = "exact")
```

- Could not estimate the model, the software is stuck!!!