

## Contents

- Preliminaries
- 0: Introduction
- 1: Background & Basic Concepts
- 2: Single Sample Methods
  - 2.1 Non-parametric Methods
  - 2.6 Parametric Models
- 3: Two-Sample Comparisons
- 4: Regression Models
  - 4.2 Parametric Regression Models
  - 4.4 Proportional Hazards models

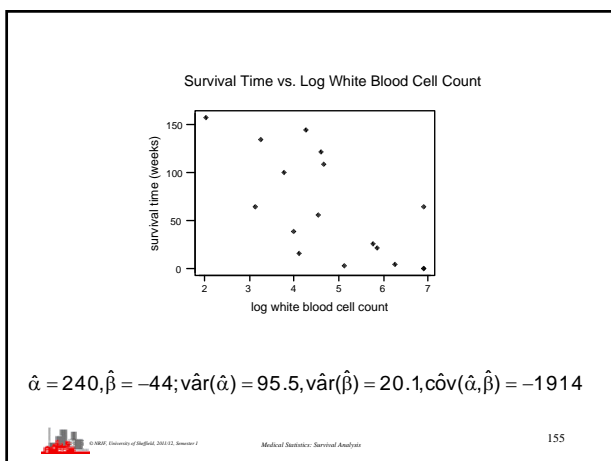
- Individual hazard rate  $h_i(t)$ 
  - e.g.  $T_i \sim \text{Ex}(\lambda_i)$  with
 
$$f_i(t) = \lambda_i e^{-\lambda_i t}$$
    - could estimate  $\lambda_i$  individually
    - Better: link  $\lambda_i$  together – fewer parameters
    - e.g.  $E[T_i] = \lambda_i^{-1} = \alpha + \beta x_i$ 
      - (exponential regression)
      - (c.f. Normal regression)

- Estimate  $\alpha$  &  $\beta$  by ML – standard theory
  - Construct likelihood by  $f(t_i)$  for uncensored and  $S(t_i)$  for censored observations
  - Get estimates and CI for  $\lambda_i$  from  $\alpha$  &  $\beta$
  - Strictly need condition  $\alpha + \beta x_i > 0$  (since  $\lambda_i > 0$ )

Example

myelogenous leukemia survival times with covariate white blood cell count

Patient	(AG positive) n=17	
	WBC $\times 10^2$	Survival Time (weeks)
1	23	65
2	7.5	156
3	43	100
4	26	134
5	60	16
6	105	108
7	100	121
8	170	4
9	54	39
10	70	143
11	94	56
12	320	26
13	350	22
14	1000	1
15	1000	1
16	520	1
17	1000	5
Median values	100	56



- Minitab implementation
- Stat>Reliability/Survival> Regression with Life Data
  - Actually models  $\log(\text{survival})$  so different answers [slightly] from above
- Similarly S-Plus and R, details in notes



### Two-Sample Example

- ◆  $h_1(t)=\lambda_1$  and  $h_2(t)=\lambda_2$
- ◆ Define dummy variable  
0 for group 1, 1 for group 2
- ◆  $h(t;x) = \lambda e^{\beta x} = \lambda$  for  $x=0$  (group 1)  
or  $= \lambda e^{\beta}$  for  $x=1$  (group 2)
- ◆ The sign of  $\beta$  determines whether  
 $\lambda_2 > \lambda_1$  ( $\beta > 0$ ) or  $\lambda_1 > \lambda_2$  ( $\beta < 0$ )



### Notes

- ◆  $\log_e h(t;x) = \log_e(\lambda) + \beta x$   
i.e. *log-linear model* for hazard function
- ◆ guarantees  $h(t;x) > 0$
- ◆ extend to several groups with  $h(t;x) = \lambda e^{\beta'x}$
- ◆ Could let  $\underline{x}$  be any vector of
  - covariates
  - prognostic factors



- ◆ allowing for covariates gives  
more sensitive tests
- ◆ can correct for imbalance  
in initial randomization
- ◆ incorporating prognostic factors for  
investigating factors in their own right
- ◆ can investigate **interactions**
  - see later re interactions
  - see later for implementation **R**  
of parametric and non-parametric models



## Contents

### Preliminaries

#### 0: Introduction

#### 1: Background & Basic Concepts

#### 2: Single Sample Methods

##### 2.1 Non-parametric Methods

##### 2.6 Parametric Models

#### 3: Two-Sample Comparisons

#### 4: Regression Models

##### 4.2 Parametric Regression Models

##### 4.4 Proportional Hazards models



#### Generally model

$$h(t) \rightarrow h(t;\underline{x}) = h_0(t) \exp\{\beta'\underline{x}\}$$

where  $h_0(t)$  is the 'underlying' hazard rate

◆  $\beta_j$  reflects the effect of  $x_j$  on survival

- if  $\beta_j > 0$  :  $x_j \nearrow \Rightarrow$  hazard  $\nearrow \Rightarrow$  survival prospect  $\searrow$
- if  $\beta_j < 0$  :  $x_j \nearrow \Rightarrow$  hazard  $\searrow \Rightarrow$  survival prospect  $\nearrow$
- if  $\beta_j = 0$  :  $x_j \Rightarrow$  no effect on survival.



#### Exponential Model

◆  $h(t;\underline{x}) = \lambda \exp\{\beta'\underline{x}\}$

- $f(t;\underline{x}) = \lambda \exp\{\beta'\underline{x}\} \exp\{-\lambda t \exp\{\beta'\underline{x}\}\}$
- $S(t;\underline{x}) = \exp\{-\lambda t \exp\{\beta'\underline{x}\}\}$

#### Weibull Model

◆  $h(t;\underline{x}) = \lambda \gamma t^{\gamma-1} \exp\{\beta'\underline{x}\}$ .

#### Estimation by [numerical] ML



#### Proportional Hazards Model

◆ semi-parametric model [Cox, 1972]

◆  $h(t;\underline{x}) = h_0(t) \exp\{\beta'\underline{x}\}$

◆  $h_0(t)$  — baseline hazard

- i.e. hazard of a patient with  $\underline{x}=\underline{0}$

◆ Useful for investigating prognostic factors when actual survival distribution not of immediate interest.



◆  $h(t;\underline{x}) = h_0(t) \exp\{\beta'\underline{x}\}$

◆ Dependence on factors and covariates is precisely modelled

◆ Distribution of failure not specified

- $h_0(t)$  is not specified

◆ Useful in medical situations

- important to know which prognostic variables have an effect and to what extent



#### Can give answers such as


- ◆ Treatment halves hazard rate
- ◆ Smoking trebles hazard rate

#### But not answers such as

- ◆ Red wine increases survival by 10 years
- ◆ Coffee shortens life by 6 months




- **special cases**
  - ◆ Exponential
    - $h(t; \underline{x}) = \lambda \cdot \exp\{\beta' \underline{x}\}$
    - $h_0(t) = \lambda$
  - ◆ Weibull
    - $h(t; \underline{x}) = \lambda \gamma t^{\gamma-1} \exp\{\beta' \underline{x}\}$
    - $h_0(t) = \lambda \gamma t^{\gamma-1}$

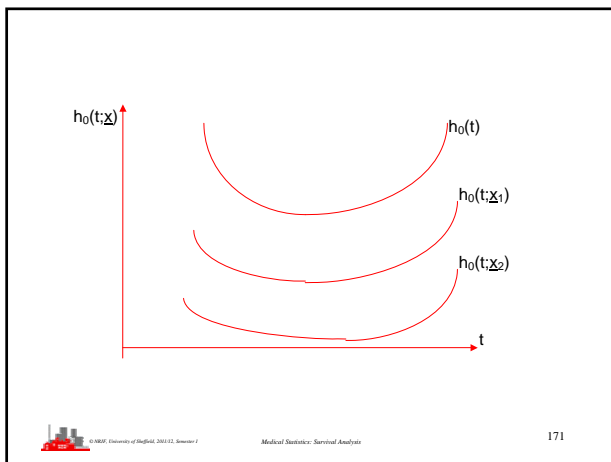


©NRJF, University of Sheffield, 2011/12, Semester 1      Medical Statistics: Survival Analysis      169


- patients with covariates  $\underline{x}_1$  and  $\underline{x}_2$ 
  - $h(t; \underline{x}_1) / h(t; \underline{x}_2) = h_0(t) \exp\{\beta' \underline{x}_1\} / h_0(t) \exp\{\beta' \underline{x}_2\}$
  - $= \exp\{\beta' (\underline{x}_1 - \underline{x}_2)\}$
- ◆ **Independent of t**
- ◆ hazard functions proportional over time
- ◆ linear component does not vary with time



©NRJF, University of Sheffield, 2011/12, Semester 1      Medical Statistics: Survival Analysis      170




- Patients have the same 'shape' of hazard function
  - but shifted multiplicatively according to  $\underline{x}$ .
- **Hazard functions can never cross**
  - ◆ Preliminary check on K-M estimates
  - ◆ Proportional hazard models are not appropriate if K-M estimates clearly cross




©NRJF, University of Sheffield, 2011/12, Semester 1      Medical Statistics: Survival Analysis      172

- **Parameter Estimation**
  - ◆ Observations
    - Survival times  $t_1, t_2, \dots, t_n$
    - Censorings  $\delta_1, \delta_2, \dots, \delta_n$
    - Covariates  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$
  - ◆ likelihood  $= \prod_{i=1}^n [h(t_i; \underline{x}_i)]^{\delta_i} S(t_i; \underline{x}_i)$
  - ◆  $S(t; \underline{x}_i) = [S_0(t; \underline{x}_i)] \exp\{\beta' \underline{x}_i\}$



©NRJF, University of Sheffield, 2011/12, Semester 1      Medical Statistics: Survival Analysis      173

- Cannot maximize this without  $h_0(t)$
- Instead use **partial likelihood**
  - ◆ Consider time points at which deaths occur and  $P[\text{individual } i \text{ dies at time } t_{(i)}]$
  - ◆ Construct likelihood from this by a combinatorial argument



©NRJF, University of Sheffield, 2011/12, Semester 1      Medical Statistics: Survival Analysis      174




◆ P[it is individual i who dies at time  $t_{(i)}$ ]

$$= \frac{h_0(t_{(i)})e^{\beta'x_{(i)}}}{\sum_{j \in R(t_{(i)})} h_0(t_{(i)})e^{\beta'x_{(j)}}}$$

$$= \frac{e^{\beta'x_{(i)}}}{\sum_{j \in R(t_{(i)})} e^{\beta'x_{(j)}}$$

◆ by proportional hazards assumption




©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 175

■ And then obtain partial likelihood

$$L(\beta) = \prod_{i=1}^n \left\{ \frac{e^{\beta'x_{(i)}}}{\sum_{j \in R(t_{(i)})} e^{\beta'x_{(j)}} \right\}^{\delta_{(i)}}$$


■ maximize this w.r.t.  $\beta$



©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 176


■ Notes

- ◆ If no censored observations this is a **conditional** likelihood
  - conditional on the observed  $t_{(1)}, t_{(2)}, \dots, t_{(n)}$
- ◆ With censored observations this is known as a **partial** likelihood
- ◆ Cox (1975) showed that the usual likelihood methods apply in this case
  - Also known as **Cox Regression**



©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 177


- ◆ Asymptotic normality
- ◆ Asymptotic variance on second partial derivative of log-likelihood
  - i.e. treat partial likelihood just as if it was a full likelihood
- ◆ Various adjustments for ties etc.



©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 178

■ Example


Variable	Coefficient	Standard Error	$\chi^2$ statistic (using lrt)	coeff/s.e
treatment (0=A, 1=B)	-1.42	0.64	4.89	-2.22
age (years)	-0.004	0.034	0.01	-0.12
sex (1=M, 0=F)	0.31	0.72	0.18	0.43
volume of heart (mm)	0.0076	0.0036	4.44	2.11
Duration of symptoms (months)	-0.004	0.063	0.00	-0.06
digitalisation	-0.59	0.73	0.66	-0.81



©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 179

■ Can use  $\chi^2$  statistic or coeff/s.e.


- ◆  $\chi^2$  statistic if factor with > 2 levels
- ◆ coeff/s.e. if factor with only 2 levels or continuous covariate
  - then  $\chi^2 = (\text{coeff/s.e.})^2$  for one coefficient
    - [continuous covariate of binary factor]
  - $\chi^2 = \sum (\text{coeff/s.e.})^2$  over the k-1 dummy variables
    - k-level factor



©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 180




- Treatment:
  - ◆  $| -2.22 | > 1.96$
  - ◆ Good evidence of effect of treatment
  - ◆ Coeff  $< 0$  so treatment = 1 decreases hazard, i.e. treatment B is 'better'
- Heart volume:
  - ◆ coeff/s.e. =  $+2.11 > 1.96$
  - ◆ Increased heart volume decreases relapse time



©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 181


- No evidence that other factors affect relapse time
- **NB Not shown that other factors have no effect**
- Useful to calculate CIs:
  - ◆ 95% CI for  $\beta_3$  (M/F) is  $0.31 \pm 2 \times 0.72 = (-1.13, 1.75)$ 
    - i.e could be large difference between M & F



©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 182


## Interpretation of $\beta$

- Consider model  $h(t; \underline{x}) = h_0(t) \exp\{\beta' \underline{x}\}$   
 $= h_0(t) \exp\{\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k\}$ 
  - where  $x_1$  is a factor indicating treatment (say)
  - $x_1=1$  for treatment,  $x_1=0$  for placebo
- ◆ hazard for those on treatment is  $h_0(t) \exp\{\beta_1 + \beta_2 x_2 + \dots + \beta_k x_k\}$  &  
 $h_0(t) \exp\{\beta_2 x_2 + \dots + \beta_k x_k\}$  on placebo
- ◆ So **hazard ratio** for treatment is  $\exp\{\beta_1\}$ 
  - So of interest to estimate  $\exp\{\beta_1\}$  [&/or  $\beta_1$ ]
    - With confidence intervals etc.




©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 183

- **Computer implementation:**
  - ◆ Available in R, S--PLUS and SPSS (*not Minitab*)
  - ◆ All packages produce a table of parameter estimates and standard errors for each factor
  - ◆ **R**
    - Construct `Surv( . )` object first then use `coxph( )`
  - ◆ **S-PLUS:**
    - [Statistics>Survival>Cox Proportional Hazards...](#)
  - ◆ **SPSS:**
    - [Analyze>Survival>Cox Regression](#)



©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 184

- **Methtrex data:**
  - ◆ Data on liver survival:
    - Time variable is **FOLLOWUP**
    - Censoring in **STATUS**
    - Various covariates and prognostic factors
    - **TREATMNT** (0=placebo, 1=methtrex)
    - **MAYO** is a key covariate




©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 185

```

> library(survival)
Loading required package: splines
> attach(methtrex)
> methtrex[1:4, ]
  TREATMNT STATUS FOLLOWUP   MAYO LUDWIG BILIRUBEN PROTHROM ALBUMIN AGE AMA
1         0      0        28 4.796029      2      15   12.86   4.2 63  1
2         0      1        32 5.883894      2      74   12.00   4.3 61  1
3         0      1        34 4.868391      4      33   11.76   4.7 60  1
4         0      0        37 4.531851      2      16   12.20   3.6 48  1

> meth.sv<-Surv(FOLLOWUP, STATUS)
> meth.ph<-coxph(meth.sv~TREATMNT+MAYO)
> summary(meth.ph)
    
```



©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 186



```
> summary(meth.ph)
Call:
coxph(formula = meth.sv ~ TREATMNT + MAYO)
  n= 60
      coef exp(coef) se(coef)      z Pr(>|z|)
TREATMNT1 0.4358  1.5462  0.5646 0.772  0.44
MAYO      0.7543  2.1261  0.1811 4.164 3.12e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
TREATMNT1    1.546    0.6468    0.5113    4.675
MAYO         2.126    0.4703    1.4907    3.032

Rsquare= 0.224 (max possible= 0.826 )
Likelihood ratio test= 15.25 on 2 df,  p=0.000487
Wald test              = 19.56 on 2 df,  p=5.657e-05
Score (logrank) test = 22.58 on 2 df,  p=1.250e-05
```

	coef	exp(coef)	se(coef)	z	p
TREATMNT	0.436	1.55	0.565	0.772	0.440000
MAYO	0.754	2.13	0.181	4.164	0.000031

	exp(coef)	exp(-coef)	lower .95	upper .95
TREATMNT	1.55	0.647	0.511	4.68
MAYO	2.13	0.470	1.491	3.03

Note that this gives estimates & confidence intervals for  $\beta$  and  $\exp(\beta)$

### S-PLUS output

	coef	exp(coef)	se(coef)	z	p
TREATMNT	0.436	1.55	0.565	0.772	0.440000
MAYO	0.754	2.13	0.181	4.164	0.000031

	exp(coef)	exp(-coef)	lower .95	upper .95
TREATMNT	1.55	0.647	0.511	4.68
MAYO	2.13	0.470	1.491	3.03

### SPSS output

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
TREATMNT	.446	.565	.621	1	.431	1.561
MAYO	.753	.181	17.301	1	.000	2.124

### SPSS output

Variables in the Equation


	B	SE	Wald	df	Sig.	Exp(B)
TREATMNT	.446	.565	.621	1	.431	1.561
MAYO	.753	.181	17.301	1	.000	2.124

(Wald statistic is same as chi-squared)

- More on interpretation of coefficients
  - Interactions:
    - If 2 covariates **interact** it means that the effect of one of them **depends** upon the value of the other
      - i.e effect different for different levels
    - e.g. treatment  $\times$  stage of cancer interaction
    - treatment only effective for stage 2 not stage 1




- Could be that the two **main effects** have non-significant coefficients but their [2-way] interaction is significant
- Cannot ignore main effects
  - even if main coefficients 'insignificant'
  - ◆ Need to include main effect coefficients as well as interaction term in model interpretation




©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 193

- Mathematically express [linear] interactions as a product of the two variables which interact
  - ◆ e.g. A & B 2-level factors coded by  $x_1$  &  $x_2$  taking values 0 & 1
  - ◆ Define interaction as  $x_3 = x_1 \times x_2$




©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 194

- ◆ So if model is  $h(t) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)$ 
  - $x_3 = x_1 \times x_2$
- ◆ Then
  - $h(t) = h_0(t)$  for  $x_1 = 0, x_2 = 0$
  - $h(t) = h_0(t) \exp(\beta_1)$  for  $x_1 = 1, x_2 = 0$
  - $h(t) = h_0(t) \exp(\beta_2)$  for  $x_1 = 0, x_2 = 1$
  - $h(t) = h_0(t) \exp(\beta_1 + \beta_2 + \beta_3)$  for  $x_1 = 1, x_2 = 1$ 
    - Similarly if one or both of A & B are continuous



©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 195


- Example:-
  - ◆  $x_1 = 0$  or 1 for placebo & treatment
  - ◆  $x_2 = \log_{10}$ (white blood cell count)
  - ◆  $x_3 = x_1 \times x_2 =$  interaction



©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 196

- Estimates: =

factor	estimate	st. error
$x_1$ (treatment)	-1.34	0.72
$x_2$ ( $\log_{10}$ {w.b.c})	+1.65	0.82
$x_3 = x_1 \times x_2$ (interaction)	-1.31	0.41




©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 197

- Estimates: =

Factor	estimate	st. error	p-value
$x_1$ (treatment)	-1.34	0.72	> 0.05
$x_2$ ( $\log_{10}$ {w.b.c})	+1.65	0.82	~ 0.05
$x_3 = x_1 \times x_2$ (interaction)	-1.31	0.41	< 0.05

- ◆ Not justified to say treatment has no effect
  - Treatment-by-wbc interaction **important**



©NRJF, University of Sheffield, 2011/12, Semester 1 Medical Statistics: Survival Analysis 198





- On placebo model estimated as
  - ♦  $h(t) = h_0(t)\exp\{1.65 \times \log_{10}(wbc)\}$ 
    - i.e. survival poorer with higher wbc
- On treatment model estimated as
  - ♦  $h(t) = h_0(t)\exp\{-1.34 + 1.65 \times \log_{10}(wbc) - 1.31 \times \log_{10}(wbc)\}$   
 $= h_0(t)\exp\{-1.34 + 0.34 \times \log_{10}(wbc)\}$ 
    - i.e. effect of treatment is to lessen greatly effect of increased wbc
    - $\log_{10}(wbc)$  has to be above  $1.34/0.34 \approx 4$  for increased wbc to severely affect survival of those on active treatment
      - i.e. wbc > 10,000 before treatment is overwhelmed

- **Model checking**
  - ♦ log-log plots
    - (log-minus-log)
  - ♦ e.g. two groups: plot  $\log_e[-\log_e\{\hat{S}_j(t)\}]$  vs t for each group

Proportional hazards  $\Rightarrow$  parallel curves  
 Curves cross  $\Rightarrow$  not proportional hazards

- **Log-log plots in R (& S-plus)**
  - Need to fit model separately to levels of factor
    - First change data type to factor
    - Then fit this as a 'strata'
  - Produces separate R 'survival objects'
    - One for each stratum, i.e. each factor level
  - Can then plot separate survival curves

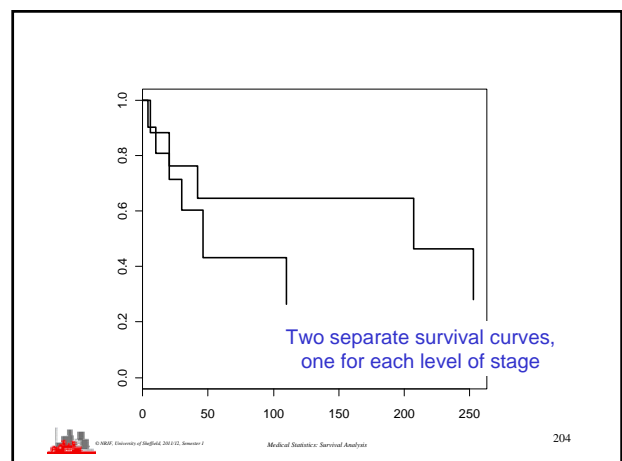
- **Example: lymphoma data**
  - Variables time, censor, stage
  - ♦ FIRST need to 'attach' data set
    - `attach(lymphoma)`
  - ♦ Change data type to factor
    - `stage <- factor(stage)`
  - ♦ Use R function for Cox model, `coxph()`

```

> attach(lymphoma)
> stage <- factor(stage)
> lymph.cox <- coxph(Surv(time, censor) ~ strata(stage))
> lymph.cox
Call: coxph(formula = Surv(time, censor) ~ strata(stage))

Null model
log likelihood = -17.77164
n = 18
> plot(survfit(lymph.cox))
    
```

Note use of `strata()` & use of `survfit()` to obtain object for plotting

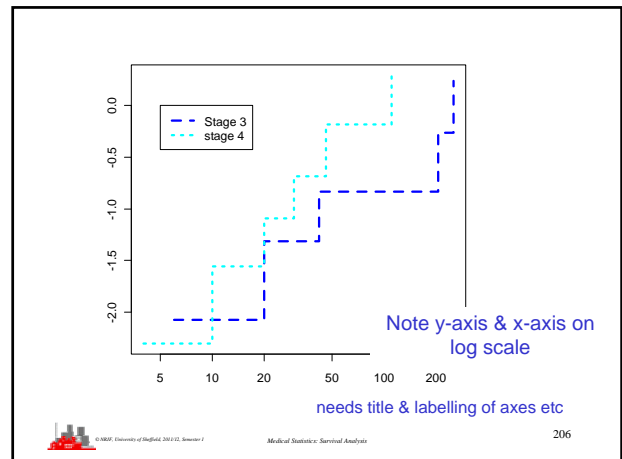


- Now need to plot on different vertical scale:
  - i.e. log-log or 'complementary log log'
    - Use argument `fun="cloglog"` in plot call
    - can also choose line styles with `lty=...`
    - and colours with `col=...`

```

> plot(survfit(lymph.cox), fun="cloglog", lty=2:3, col=4:5)
> legtext<-c("Stage 3", "stage 4")
> legend(5, 0, legtext, lty=2:3, col=4:5, lwd=3)
    
```

205



- Note:
  - using a factor as a strata indicator fits separate models to each level so we do not have a coefficient for that factor

```

> lymph.cox2<-coxph(Surv(time,censor)~stage)
> lymph.cox2
Call:
coxph(formula = Surv(time, censor) ~ stage)
      coef exp(coef) se(coef)      z      p
stage 0.29      1.34      0.33 0.877 0.38

Likelihood ratio test=0.79 on 1 df,
p=0.374 n= 18
    
```

Note absence of `strata()`

207

- Note: to obtain any survival estimate we will have to average over values of any covariate not declared as a stratum indicator
  - e.g. for Prostatic data to produce a log-log plot for the treatment groups declare `treatment` as a strata factor variable and then estimate survivor curves averaged over age, serum, size and Gleason index
    - For log-log plots of Gleason index declare `treatment` as ordinary factor and `Gleason` as strata.

208

### Example: Methrex data

```

> library(survival)
Loading required package: splines
> attach(methtrex)
> methtrex[1:3,]
  TREATMNT STATUS FOLLOWUP  MAYO  LUDWIG  BILIRUBN  PROTHROM  ALBUMIN  AGE  AMA
1         0         0      28 4.796029      2      15  12.86      4.2  63  1
2         0         1     32 5.883894      2      74  12.00      4.3  61  1
3         0         1     34 4.868391      4      33  11.76      4.7  60  1
    
```

```

> methsurv<-Surv(FOLLOWUP, STATUS)
    
```

Create survival object for plotting and dependent object in regression

209

### Example: Methrex data

Consider first treatment factor

```

> meth.treat<-coxph(methsurv~ TREATMNT+ MAYO)
> summary(meth.treat)
Call:
coxph(formula = methsurv ~ TREATMNT + MAYO)
      n= 60
      coef exp(coef) se(coef)      z Pr(>|z|)
TREATMNT1 0.4358      1.5462      0.5646 0.772      0.44
MAYO      0.7543      2.1261      0.1811 4.164 3.12e-05
    
```

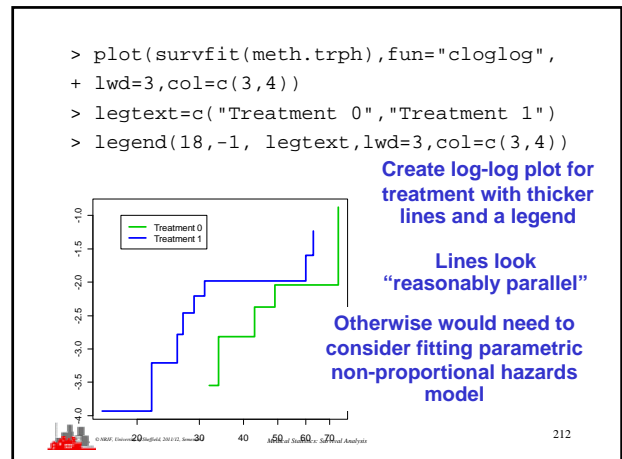
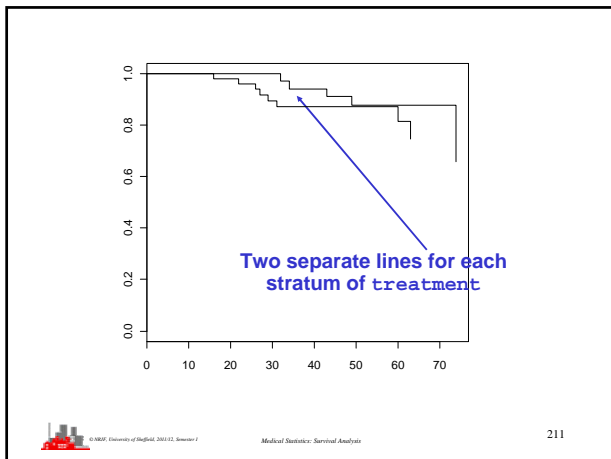
```

> meth.trph<-coxph(methsurv~strata(TREATMNT)+ MAYO)
> plot(survfit(meth.trph))
    
```

Re-fit with `treatment` as a stratum indicator for diagnostic plotting

210

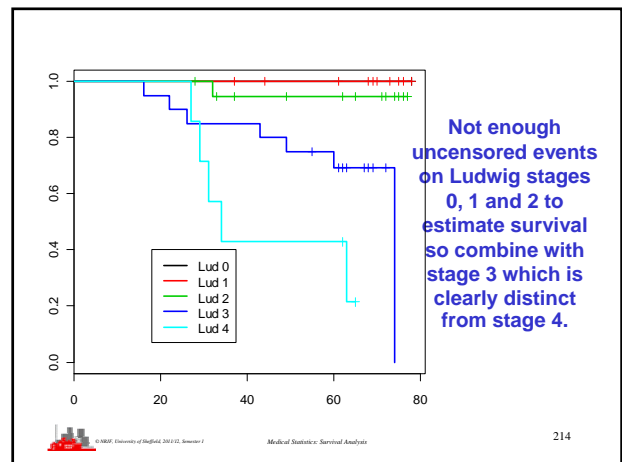




```

◆ Now consider factor LUDWIG
> summary(LUDWIG)
 0  1  2  3  4
1 12 20 20  7
# small frequencies at some levels
# exploratory K-M plot useful
> ludfit<-survfit(methsurv~LUDWIG)
> plot(ludfit,lwd=2,col=c(1:5))
> legtext<-c("Lud 0","Lud 1","Lud 2",
+ "Lud 3","Lud 4")
> legend(18,0.4, legtext,lwd=2,col=c(1:5))
>

```



```

>
> LUD2<- LUDWIG
> levels(LUD2)<- list(A=c(0,1,2,3),B=4)
> summary(LUD2)
  A  B
53  7
> meth2.anal<-coxph(methsurv~LUD2+TREATMNT+MAYO+BILIRUBN)
> summary(meth2.anal)
Call:
coxph(formula = methsurv ~ LUD2 + TREATMNT + MAYO +
      BILIRUBN)

n= 60

      coef exp(coef) se(coef)      z Pr(>|z|)
LUD2B    1.49837   4.47441  0.73770  2.031 0.042240 *
TREATMNT1 0.89523   2.44789  0.65551  1.366 0.172033
MAYO      0.67999   1.97386  0.29439  2.310 0.020897 *
BILIRUBN  0.05567   1.05725  0.01685  3.304 0.000955 ***

```

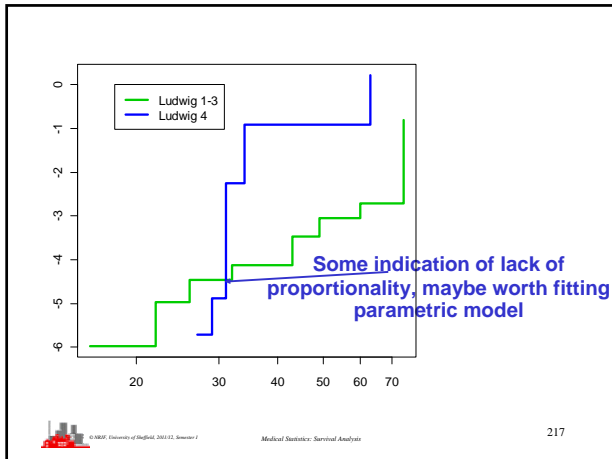
```

> meth2.ph<-coxph(methsurv~strata(LUD2)+
+ TREATMNT+ MAYO+BILIRUBN)
> plot(survfit(meth2.ph))
> plot(survfit(meth2.ph), fun="cloglog",
+ lwd=3,col=c(3,4))
> legtext=c("Ludwig 1-3","Ludwig 4")
> legend(18,0, legtext,lwd=3,col=c(3,4))

```

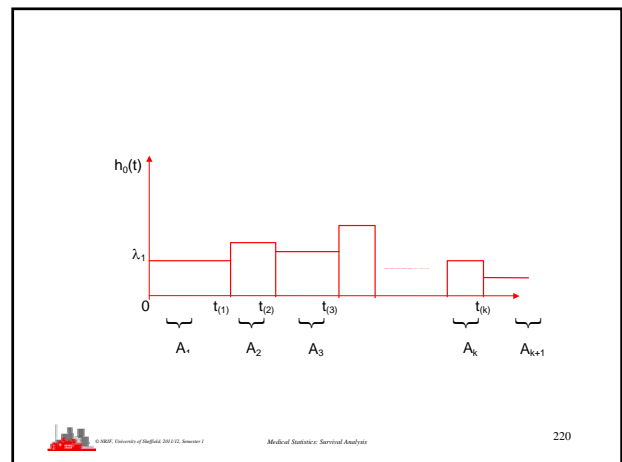
Re-fit with LUD2 as a stratum indicator for diagnostic plotting





- Log-log plots in SPSS:
  - Obtained from menu *Analyze>Survival>Cox Regression*
  - Ensure factor variables are declared as *categorical*
  - Check *log minus log* box under *Plots...* dialogue
  - Ensure factor *separate lines for:* box is completed

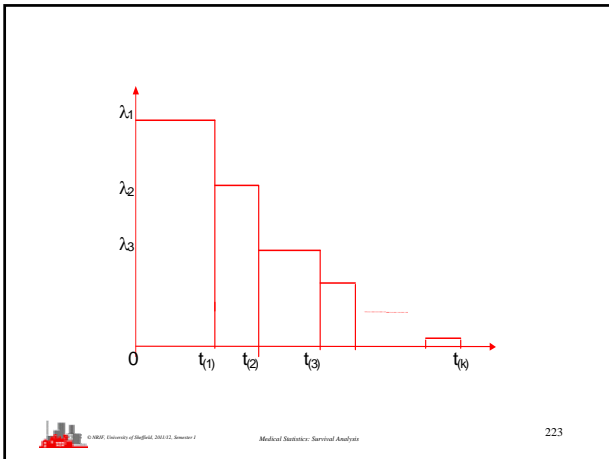
- Estimation of  $h(t)$ 
  - ◆ No information about  $h(t)$  in intervals with no failures –
    - $h_0(t)$  is arbitrary (possibly even 0 in interval)
  - ◆ Assume that the hazard constant (at  $\lambda_i$ ) between adjacent death times
    - (since no information that it is **not** constant)



- Estimate  $\underline{\beta}$  by max. partial likelihood
  - Define  $\lambda_j = h_0(t)$  if  $t \in A_j = (t_{(j-1)}, t_{(j)})]$
  - Estimate  $\lambda_j$  by MLE, replacing  $\underline{\beta}$  by  $\hat{\underline{\beta}}$
- $$\Rightarrow \hat{\lambda}_j = \frac{d_j}{\sum_{i \in R(t_{(j)})} e^{\hat{\underline{\beta}}x} (t_{(j)} - t_{(j-1)}) + \sum_{i \in A_j} e^{\hat{\underline{\beta}}x} (t_i - t_{(j-1)})}$$

- pattern in estimates of  $\lambda_j$  may identify  $\Rightarrow$  parametric form for  $h_0(t)$ .
  - ◆ roughly constant  $\Rightarrow$  exponential survival
  - ◆ Linear increasing  $\Rightarrow$  Weibull shape =2
  - ◆ Exponentially decreasing  $\Rightarrow$  Gumbel





- **Proportional Hazards Models**
  - ◆ Only models dependence on covariates
  - ◆ No statements about survival times
  - ◆ Only effect of covariates on hazards
  - ◆ Estimation by maximum partial likelihood
    - Check proportional hazards by log-log plots
  - ◆ Estimation of hazard may suggest parametric model

- Key diagnostic is log-log plots
  - ◆ Other techniques use residuals (see Schoenfeld residuals etc)
    - Try `help(residuals.coxph)` & `help(residuals.survreg)`
- No allowance for individual variability
  - ◆ i.e. no term in  $\sigma_i^2$  for  $i^{\text{th}}$  individual
  - ◆ **Frailty models** do allow for this
  - ◆ Some facilities in R



- That was the end of the course (almost)
- What other important topics are there in Survival Analysis?
  - ◆ **Accelerated failure time models**
  - ◆ (Time-dependent covariates)
  - ◆ **Competing risks models**

- ◆ **Accelerated failure time models**
  - Survivor function has the form  $S(t;\mathbf{x}) = S_0(t.e^{\beta\mathbf{x}})$
  - Effect of covariates is to 'accelerate time'
  - Weibull models have the **accelerated failure time property**
    - proportional hazards model with the acft property
      - » Others are Gompertz, Extreme Value, Loglogistic
    - name from accelerated-life testing,
      - e.g. to test electronic components raise voltage
      - need to ensure model is appropriate for this situation & not think use of model will accelerate experiment
  - ◆ Needs R library `eha` & function `aftreg(.)`
    - use `surv(.)` first to create survival object



- **Competing risks models**
  - ◆ Subject at risk from several different causes
    - Event caused by 'whichever gets them first'
      - (i.e. risks compete)
  - ◆ Hazards are 'cause-specific' & may depend differently on covariates
  - ◆ Example is death after transplant: causes could be:-
    - Rejection
    - Infection
    - Heart failure
    - etc
  - ◆ Or infection from a virus/bacterium/.....



- Key complication is censored observations are not associated with a specific cause
  - ◆ Death from one cause removes subject from risk set of death from other causes
    - (assuming you can only die once)
  - ◆ Infection from one bug may alter risks for infection from other causes
- **Competing risks models in R**
  - ◆ Many different facilities in different specialist libraries
  - ◆ Library `cmprsk`, use `surv()` first as always



**The End**

