- 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

- No evidence that other factors affect relapse time
- **NB Not shewn that other factors have no effect**
- Useful to calculate CIs:
  - 95% CI for β_3 (M/F) is $0.31 \pm 2 \times 0.72$ $=(-1.13, 1.75)$
    - i.e could be large difference between M & F

Interpretation of β

- Consider model $h(t;\underline{x}) = h_0(t)\exp\{\underline{\beta}'\underline{x}\}$ $= h_0(t)\exp\{\beta_1x_1+\beta_2x_2+\ldots+\beta_kx_k\}$
  - 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 $h_0(t)\exp\{\boldsymbol{\beta_1}+\beta_2x_2+\ldots+\beta_kx_k\}$ & $h_0(t)\exp\{\ \ \beta_2x_2+\ldots+\beta_kx_k\}$ on placebo
  - So ***hazard ratio*** for treatment is $\boldsymbol{\exp\{\beta_1\}}$
    - So of interest to estimate $\exp\{\beta_1\}$ [&/or β_1]
      - With confidence intervals etc.

- **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**

- **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


```
> library(survival)
Loading required package: splines
> attach(methtrex)
> methtrex[1:4,]
  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
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)
```

```
> 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 β and exp(β)

SPLUS 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)
TREATMN	.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)
TREATMN	**.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 × 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

- 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 $x_3 = x_1 \times x_2$

- ♦ So if model is
  $h(t)=h_0(t)\exp(\beta_1x_1+\beta_2x_2+\beta_3x_3)$
  - $x_3 = x_1 \times x_2$
- ♦ 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**

- 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

- 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

- Estimates:=

Factor	estimate	st. error	pvalue
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**

- 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

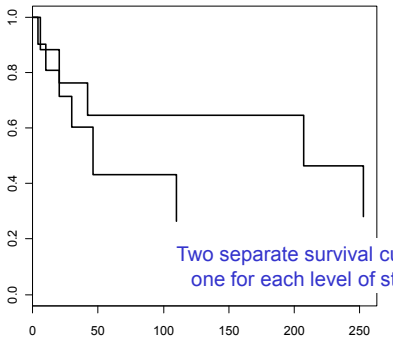
  Curves cross $\Rightarrow$ not proportional hazards

- 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))
```

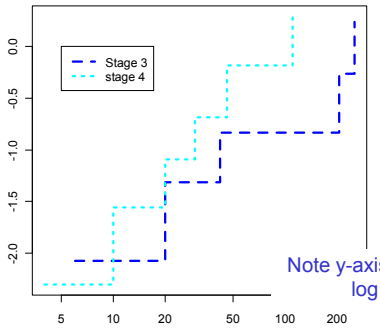
Note use of `strata()` & use of `survfit()` to obtain object for plotting



Two separate survival curves, one for each level of stage

- 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)
```

Note y-axis & x-axis on log scale

needs title & labelling of axes etc

- 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()`

- 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.

Example: Methtrex 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)
```

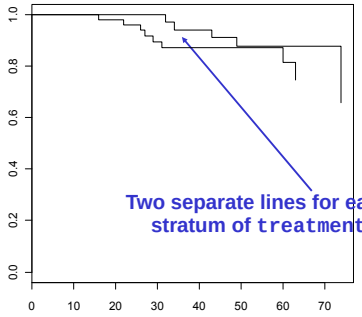
Create survival object for plotting and dependent object in regression

Example: Methtrex 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))
```

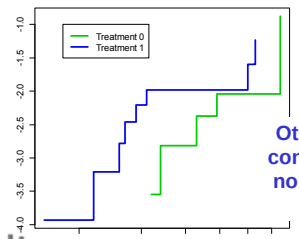
Re-fit with `treatment` as a stratum indicator for diagnostic plotting



Two separate lines for each stratum of `treatment`


```
> plot(survfit(meth.trph),fun="cloglog",
+ lwd=3,col=c(3,4))
> legtext=c("Treatment 0","Treatment 1")
> legend(18,-1, legtext,lwd=3,col=c(3,4))
```

Create log-log plot for treatment with thicker lines and a legend

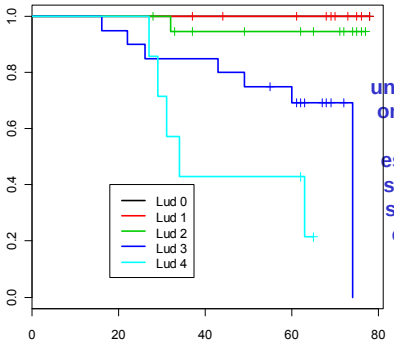


Lines look "reasonably parallel"

Otherwise would need to consider fitting parametric non-proportional hazards model

- 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))
>
```

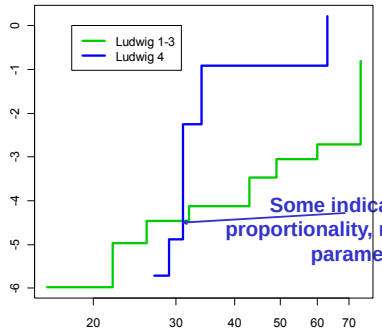
Not enough uncensored events on Ludwig stages 0, 1 and 2 to estimate survival so combine with stage 3 which is clearly distinct from stage 4.


```
>
> 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))
```

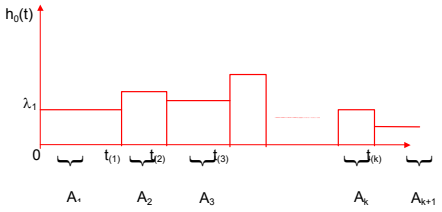
Re-fit with LUD2 as a stratum indicator for diagnostic plotting



Some indication of lack of proportionality, maybe worth fitting parametric model

- 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

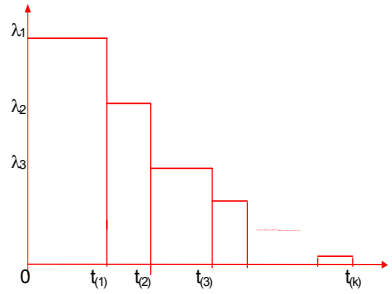
- **Estimation of h(t)**
  - No information about h(t) in intervals with no failures –
    - $h_0(t)$ is arbitrary (possibly even 0 in interval)
  - Assume that the hazard constant (at λ_j) between adjacent death times
    - (since no information that it is **not** constant)



- Estimate $\underline{\beta}$ by max. partial likelihood
- Define $\lambda_j = h_0(t)$ if $t \in A_j = (t_{(j1)}, t_{j)}]$
- Estimate λ_j by MLE, replacing $\underline{\beta}$ by $\hat{\underline{\beta}}$

$$\Rightarrow \hat{\lambda}_j = \frac{d_j}{\sum_{i \in R(t_{(j)})} e^{\hat{\underline{\beta}}' \underline{x}} (t_{(j)} - t_{(j-1)}) + \sum_{i \in A_j} e^{\hat{\underline{\beta}}' \underline{x}} (t_i - t_{(j-1)})}$$

- pattern in estimates of λ_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 loglog 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 σ_i^2 for i^{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