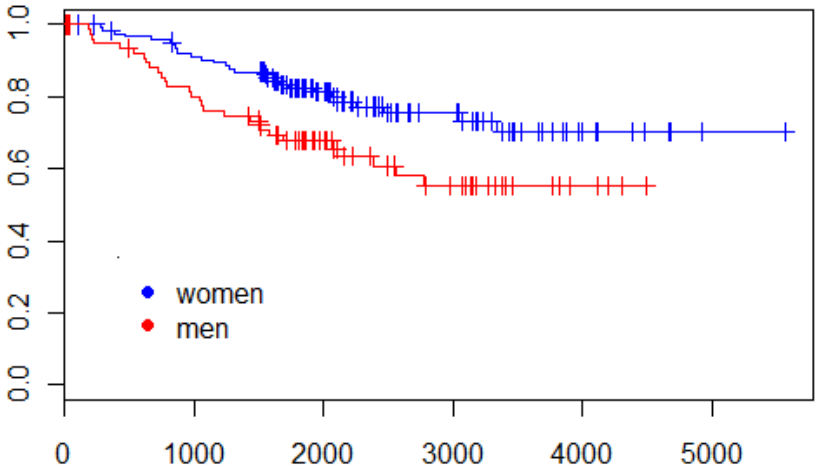
LECTURE 14

Survival analysis II: log-rank tests and Cox regression

Survival analysis

- ***Log-rank test***: tests of statistical differences between two survival curves
- ***Cox (proportional hazard) regression***: estimating effect of factors on between-group differences in survival curves



Log-rank test

- In addition to visual examination, we can test for statistical differences in survival between groups

- The *log-rank test* is nonparametric (it does not assume that the distribution of deaths by time is normally distributed)

- Based on simple logic:
  - at any given time, there is a number of individuals at risk in each group, and a number of deaths
  - we can test whether proportion of observed deaths differs from predicted deaths across groups (as in a two independent proportions or chi-square test)
  - a new interval is created every time a death happens, but procedure lumps intervals together and counts total of observed vs. predicted deaths by group

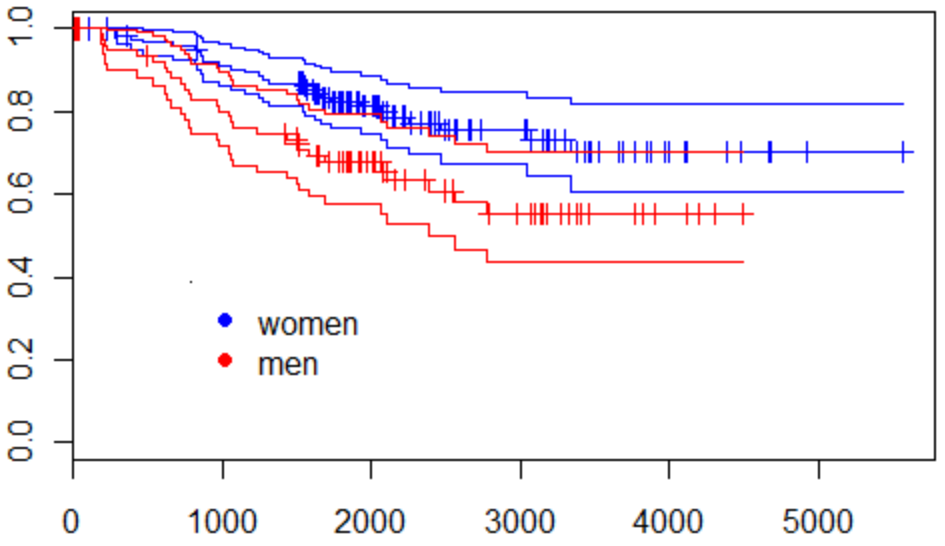
Log-rank test

- Null hypothesis: groups have same survival curve S(t) and *hazard rate*
  - *Hazard rate:* approximately our 'mortality rate'=deaths in interval i divided by people alive at start of interval i
- To run log-rank test, use function *survdiff*
  - **don't forget to load package *survival***
- Test statistic is based on total variance (V), but approximation uses expected deaths (E)
- >survdiff(Surv(melanom$days, melanom$status==1)~sex)

```
> survdiff(Surv(days,status==1)~sex)
Call:
survdiff(formula = Surv(days, status == 1) ~ sex)

        N  Observed  Expected  (O-E)^2/E  (O-E)^2/V
sex=1 126        28      37.1       2.25       6.47
sex=2  79        29      19.9       4.21       6.47

 Chisq= 6.5  on 1 degrees of freedom, p= 0.011 …
```

Results

- In men (**sex=2**), number of deaths is larger than in women
  - P=0.011: this indicates significant difference between curves
- Comparing to K-M plot obtained by *survfit*
  - seems correct, although 95% CI bands partially overlap



Stratified test

- More than one variable can be used in a stratified analysis
  - one option is to compare survival between groups controlling for levels of a second variable (a confounder)
  - *note: in logistic regression and mixed models we 'control' for variables in a different way; but stratified tests are popular in epidemiology*

Example:

- We can test for difference in survival curve by sex, controlling for ulceration stage (1=present; 2=absent)

- > survdiff(Surv(melanom$days,melanom$status==1) ~ melanom$sex+strata(melanom$ulc))

- This will test whether sex affects survival, after controlling for differences in level of ulceration between the groups
  - i.e. we want to test whether ulceration is a confounder

Stratified test

Results:

- Controlling for stage of melanoma, difference between sexes is no longer significant!
  - why? Men may see a doctor when disease is at a more advanced stage (when ulceration is present)

Evidence for interpretation:

- if we run the log-rank test by ulceration stage, and control by sex, difference is highly significant (P~0)

```
>survdiff(Surv(melanom$days,melanom$status==1)~
melanom$ulc+strata(melanom$sex)
```

- the same is true for long-rank test by ulceration stage only (try it)
- presence of ulceration (present=1) prevails among men, initial stage prevails among women (sex=1)

```
> table(melanom$ulc,melanom$sex)
     sex
ulc   1   2
  1  47  43
  2  79  36
```

```
> survdiff(Surv(melanom$days,melanom$status==1) ~
sex+strata(melanom$ulc))

Call:
survdiff(formula = Surv(days, status == 1) ~ sex + strata(ulc))

        N Observed Expected  (O-E)^2/E  (O-E)^2/V
sex=1 126       28     34.7       1.28       3.31
sex=2  79       29     22.3       1.99       3.31

 Chisq= 3.3  on 1 degrees of freedom, p= 0.0687
```

Cox regression

- Log-rank test shows differences between groups, but does not quantify effect of covariates

- Linear methods (comparison of curves, estimation of group effect) can be applied to survival data more easily under the assumption of *proportional hazards* (PH)
  - when this is done, we can define a measure of mortality (log of hazard ratio) suitable to linearisation, and then compare lines between groups

Hazard rate

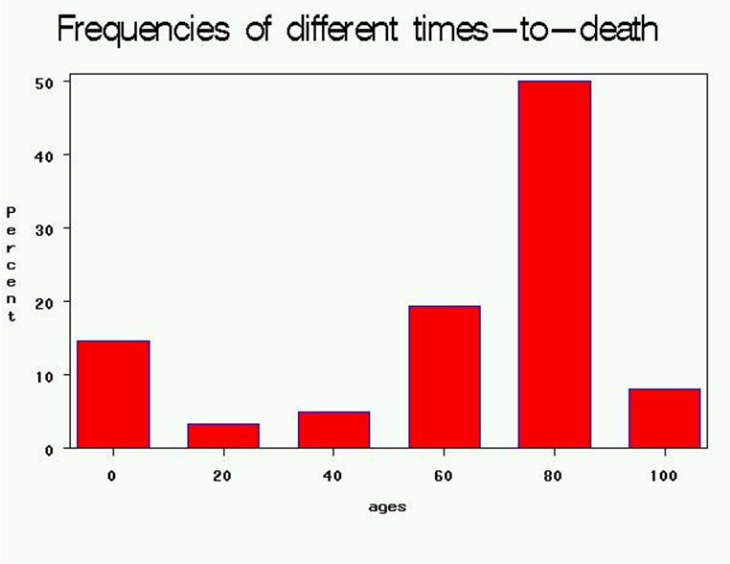
- Survival S(t) or l_x is a decreasing function; it decreases more quickly when deaths occur more frequently
  - i.e. when mortality rate is higher

- The *hazard rate h(t) is similar to our 'mortality rate'*
  - =number of people dying at interval i divided by people surviving to interval i
  - also know as the *force of mortality* or *instantaneous death rate*
  - it is the risk of death at a given time *t* relative to the chance of survival to age *t*
    - i.e. the probability of dying at the interval t divided by S(t)

Hazard rate

- Hazard rate = f(t)/ S(t) is the same as our mortality rate
  - =the probability of dying at a given time interval f(t), divided by probability of surviving to that interval S(t)

- f(t) = probability density function
  - =the distribution of total death events by time =
  - =number of deaths per time interval
  - =di/N
  - for example, most human deaths in a given cohort (i.e. of all people born) occur in our 70s-80s; much fewer in our 90s-100s
    - but this does not mean we are less 'frail' in our 90s-100s; there are fewer deaths in the 90s-100s because fewer people reach that age!



- Survival rate S(t) is the number of people still alive at age divided by all people born = Ai/N

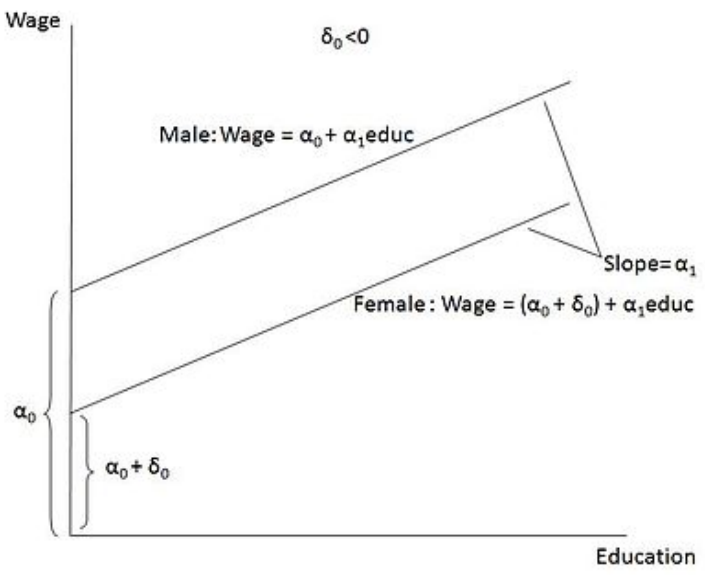
- So what is f(t)/S(t) = instantaneous dath rate, hazard rate, force of mortality?
- =(di/N)/(ai/N) = di/ai
- Di/ai is our mortality rate at interval i

- The hazard rate (mortality rate, instantaneous death rate, force of mortality) is abetter measure of 'frailty'' or vulnerability to death

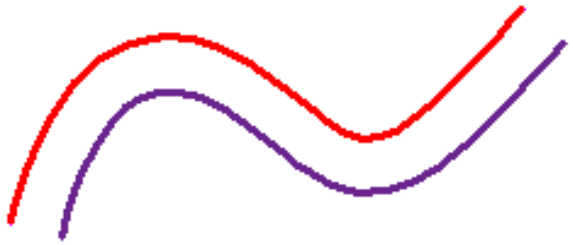
Proportional hazards assumption

- PH models simplify comparisons by assuming that compared groups maintain a constant ratio of hazard rates:

Hazard ratio = $\mathbf{h_{exposure}}$ / $\mathbf{h_{baseline}}$ = **constant**



- We are attributing differences in group survival to a constant difference in the force of mortality differs between the groups
  - this is similar to a multiple regression with continuous variable and dummy variable: two lines are parallel, and vertical distance (difference between intercepts) measures 'group' effect
  - there are other models that do not make the PH assumption



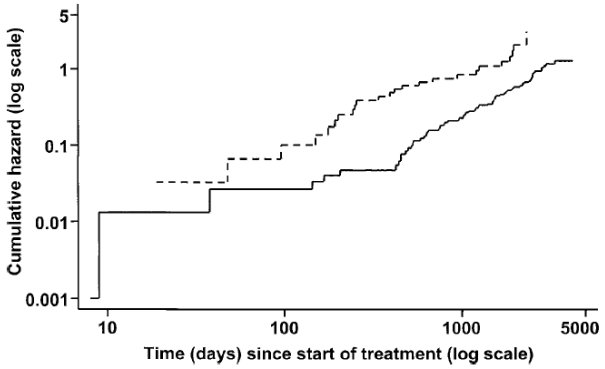
- note: the assumption of PH does not require that hazard rate varies linearly with time –only that difference (distance is constant)

Proportional hazards assumption

- How to calculate the hazard ratio?
- The *cumulative hazard function*, H(t), is the sum of hazard rates h(t) up to time t
- It can be shown that if hazard ratio is constant, so is the ratio of cumulative hazards
  - **$H_{exp} / H_{base} = h_{exp} / h_{base}$ = constant**
- H(t) is frequently linear when logged!
- Applying log transformation:
  - **$Ln(H_{exposure}) - Ln(H_{baseline}) = ln(constant)$**

and

  - **$Ln(h_{exposure}) - Ln(h_{baseline}) = ln(constant)$**

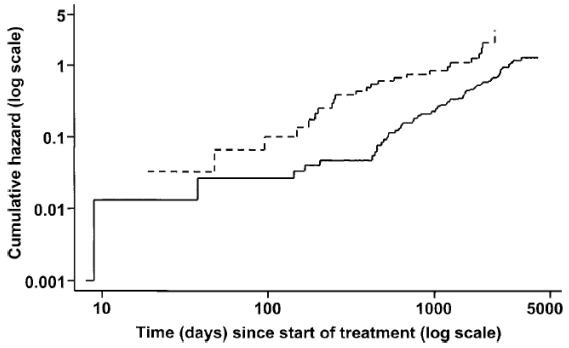


- PH assumes that the H(t) lines are parallel (same slope), and the only difference is the **intercept** (height of curve)
- If lines were not parallel, effect of group (e.g. sex) varies with time and this complicates interpretation; PH assumption eliminates time from the equation

Cox regression

- The difference in intercept is the ln(cumulative hazards ratio) = ln(hazard ratio)
  - $=\ln(h_{exp} / h_{base})$

- In Cox regression, ***ln(hazard ratio) = coefficient***
  - because groups are X=0 (baseline) and X=1 (exposure), the coefficient measures the contribution of exposure to hazard ratio
  - when exponentiated, it gives ratio of hazards between exposure and baseline
  - notice that in logistic regression coefficient is ln(odds ratio); here it is ln(hazard ratio)



- But Cox regression does not measure baseline hazard
  - we don't get an intercept

Example: sex

- Let's run Cox regression to test for effect of sex on hazard rate in melanoma patients
  - Syntax: function *coxph*

```
>summary(coxph(Surv(melanom$days,
melanom$status==1)~melanom$sex))
```

Results:

- log(hazards ratio) = sex coefficient = 0.66
  - different from 0 (i.e. there is an effect (z distribution used for logs)
  - exponentiated coefficient is the hazards ratio = 1.939
  - So force of mortality by melanoma in sex=2 (men) is almost double the value in sex =1 (women, baseline)

```
>summary(coxph(Surv(melanom$days,
melanom$status==1)~melanom$sex))

Call:
coxph(formula = Surv(days, status == 1) ~ sex)
  n= 205, number of events= 57
       coef  exp(coef)  se(coef)     z      Pr(>|z|)
sex  0.6622   1.9390     0.2651    2.498   0.0125
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
'.' 0.1 ' ' 1
      exp(coef)  exp(-coef)  lower.95  upper.95
sex     1.939      0.5157      1.153      3.26

Concordance= 0.59  (se = 0.033 )
Rsquare= 0.03   (max possible= 0.937 )
Likelihood ratio test= 6.15  on 1 df,   p=0.01314
Wald test            = 6.24  on 1 df,   p=0.01251
Score (logrank) test = 6.47  on 1 df,   p=0.01098
```

Example: sex

- Inverted hazards ratio is the change from the perspective of women
  - Force of mortality is 48.4% lower in women
- 95% CI are returned unlogged
- Concordance is a version of correlation (between predicted and observed survival times)
  - notice that squared concordance does not equal R squared
- R squared measures goodness of fit of model; as seen, it is low indicating poor fit (remember that model assumes PH and linearity)
  - both concordance and R squared should be used with care; they are versions of the estimates we obtain from least square techniques

```
> summary(coxph(Surv(melanom$days, melanom$status==1)~melanom$sex))

Call:
coxph(formula = Surv(days, status == 1) ~ sex)
  n= 205, number of events= 57
       coef    exp(coef)  se(coef)     z       Pr(>|z|)
sex  0.6622   1.9390     0.2651    2.498   0.0125
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
      exp(coef)   exp(-coef)  lower.95  upper.95
sex     1.939       0.5157       1.153      3.26

Concordance= 0.59  (se = 0.033 )
Rsquare= 0.03   (max possible= 0.937 )
Likelihood ratio test= 6.15  on 1 df,   p=0.01314
Wald test            = 6.24  on 1 df,   p=0.01251
Score (logrank) test = 6.47  on 1 df,   p=0.01098
```

Significance tests

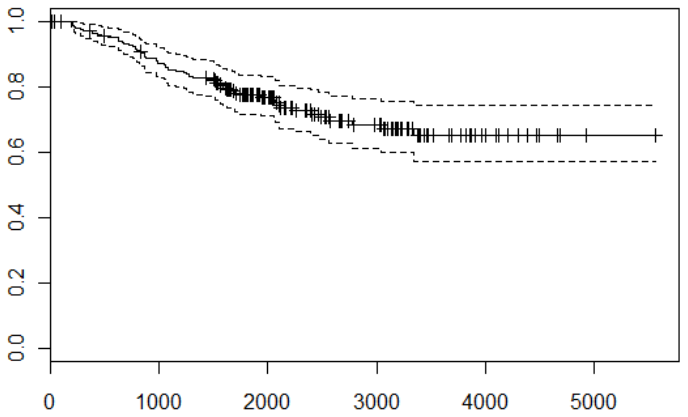
- Three tests are provided
  - *Likelihood ratio test*: based on partial likelihood estimation of null model vs. alternative model (with sex)
  - *Wald test* is based on *z* statistic that approximates a 95% CI for the group difference
  - *Score (log rank)* is the log-rank test
    - remember: the log-rank test assumes PH as null hypothesis
    - test result was P=0.011
- In large samples the three tests should give very similar results
- Cox regression works for continuous variables too; it can include more than one variable and interactions

```
> summary(coxph(Surv(melanom$days, melanom$status==1)~melanom$sex))

Call:
coxph(formula = Surv(days, status == 1) ~ sex)
  n= 205, number of events= 57
       coef      exp(coef)  se(coef)    z       Pr(>|z|)
sex  0.6622   1.9390      0.2651    2.498   0.0125
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
       exp(coef)    exp(-coef)   lower.95  upper.95
sex     1.939        0.5157        1.153       3.26

Concordance= 0.59  (se = 0.033 )
Rsquare= 0.03   (max possible= 0.937 )
Likelihood ratio test = 6.15  on 1 df,   p=0.01314
Wald test             = 6.24  on 1 df,   p=0.01251
Score (logrank) test  = 6.47  on 1 df,   p=0.01098
```

Plotting

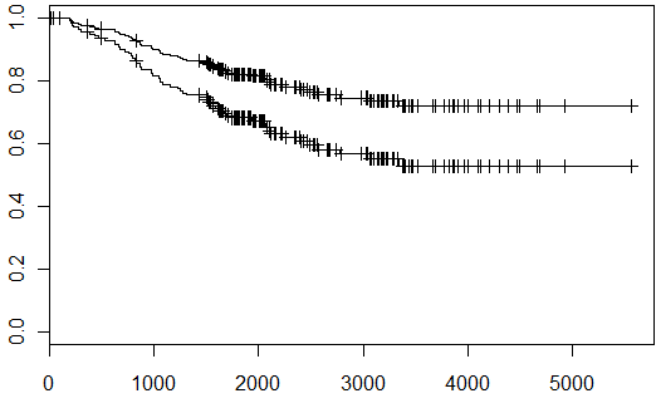
- If you plot a Cox regression, you get a plot for a pseudo-individual with average of covariates (in this case, an ‘average sex’)

```
>plot(survfit(coxph(Surv(melanom$days, melanom$status==1)~melanom$sex)))
```



- In order to plot a curve for the two sexes separately, create a new dataframe

```
> curves <- data.frame(sex=c(1,2))
> svfit <- survfit(coxph(Surv(melanom$days, melanom$status==1) ~ melanom$sex), newdata=curves)
plot(svfit)
```



s